



HAL
open science

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

Raphael Carapito, Richard Li, Julie Helms, Christine Carapito, Sharvari Gujja, Véronique Rolli, Raony Guimaraes, Jose Malagon-Lopez, Perrine Spinnhirny, Alexandre Lederle, et al.

► To cite this version:

Raphael Carapito, Richard Li, Julie Helms, Christine Carapito, Sharvari Gujja, et al.. Identification of driver genes for critical forms of COVID-19 in a deeply phenotyped young patient cohort. *Science Translational Medicine*, 2021, pp.1-59. 10.1126/scitranslmed.abj7521 . hal-03436053

HAL Id: hal-03436053

<https://hal.science/hal-03436053>

Submitted on 19 Nov 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

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

Raphael Carapito^{1,2,3,†}, Richard Li^{4†}, Julie Helms^{1,3,5}, Christine Carapito^{3,6}, Sharvari Gujja⁴, Véronique Rolli^{1,2,3}, Raony Guimaraes⁴, Jose Malagon-Lopez⁴, Perrine Spinnhirny^{1,3}, Alexandre Lederle^{1,3}, Razieh Mohseninia⁷, Aurélie Hirschler^{3,6}, Leslie Muller^{3,6}, Paul Bastard^{8,9,10}, Adrian Gervais^{9,10}, Qian Zhang^{8,9,10}, François Danion^{1,3,11}, Yvon Ruch^{3,11}, Maleka Schenck^{3,12}, Olivier Collange^{3,13}, Thiên-Nga Chamaraux-Tran^{3,14}, Anne Molitor^{1,3}, Angélique Pichot^{1,3}, Alice Bernard^{1,3}, Ouria Tahar^{2,3}, Sabrina Bibi-Triki^{1,3}, Haiguo Wu⁴, Nicodème Paul^{1,3}, Sylvain Mayeur^{1,3}, Annabel Larnicol^{1,3}, Géraldine Laumond^{1,3}, Julia Frappier^{1,3}, Sylvie Schmidt^{1,3}, Antoine Hanauer^{1,3}, Cécile Macquin^{1,3}, Tristan Stemmelen^{1,2,3}, Michael Simons¹⁵, Xavier Mariette^{16,17}, Olivier Hermine^{10,18}, Samira Fafi-Kremer^{1,3,19}, Bernard Goichot^{3,20}, Bernard Drenou²¹, Khaldoun Kuteifan²², Julien Pottecher^{3,14}, Paul-Michel Mertes^{3,13}, Shweta Kailasan²³, M. Javad Aman²³, Elisa Pin²⁴, Peter Nilsson²⁴, Anne Thomas²⁵, Alain Viari²⁵, Damien Sanlaville²⁵, Francis Schneider^{3,12}, Jean Sibilia^{1,3,26}, Pierre-Louis Tharaux²⁷, Jean-Laurent Casanova^{8,9,10,28}, Yves Hansmann^{3,11}, Daniel Lidar^{7,29}, Mirjana Radosavljevic^{1,2,3}, Jeffrey R. Gulcher⁴, Ferhat Meziani^{3,5}, Christiane Moog^{1,3}, Thomas W. Chittenden^{4,30*}, Seiamak Bahram^{1,2,3*}

¹Laboratoire d'ImmunoRhumatologie Moléculaire, plateforme GENOMAX, INSERM UMR_S 1109, Faculté de Médecine, Institut Thématique Interdisciplinaire (ITI) de Médecine de Précision de Strasbourg, Transplantex NG, Université de Strasbourg; 67085 Strasbourg, France.

²Service d'Immunologie Biologique, Plateau Technique de Biologie, Pôle de Biologie, Nouvel Hôpital Civil; 67091 Strasbourg, France.

³Fédération Hospitalo-Universitaire (FHU) OMICARE, Fédération de Médecine Translationnelle de Strasbourg (FMTS), Centre de Recherche d'Immunologie et d'Hématologie; 67085, Strasbourg, France.

⁴Genuity AI Research Institute, Genuity Science; Boston, MA 02114, USA.

⁵Service de Médecine Intensive-Réanimation, Nouvel Hôpital Civil, Hôpitaux Universitaires de Strasbourg; 67091 Strasbourg, France.

⁶Laboratoire de Spectrométrie de Masse BioOrganique, Université de Strasbourg, CNRS, IPHC, UMR 7178; 67000, Strasbourg, France.

⁷Center for Quantum Information Science and Technology, University of Southern California; Los Angeles, 90089-0484 CA, USA.

⁸St Giles laboratory of human genetics of infectious diseases, Rockefeller Branch, The Rockefeller University; New York, NY 10065, USA.

⁹Laboratory of human genetics of infectious diseases, Necker Branch, INSERM, Necker Hospital for Sick Children; 75015 Paris, France.

¹⁰University of Paris, Imagine Institute; 75015 Paris, France.

¹¹Department of Infectious and Tropical Diseases, Hôpitaux Universitaires de Strasbourg; 67091 Strasbourg, France.

¹²Service de Médecine Intensive-Réanimation, Hôpital de Hautepierre, Hôpitaux Universitaires de Strasbourg; Avenue Molière, 67200 Strasbourg, France.

¹³ Service d'Anesthésie-Réanimation et Médecine Péri-Opératoire, Nouvel Hôpital Civil, Hôpitaux Universitaires de Strasbourg; Strasbourg, France.

¹⁴Service d'Anesthésie-Réanimation et Médecine Péri-Opératoire, Hôpital de Hautepierre, Hôpitaux Universitaires de Strasbourg; 67200 Strasbourg Cedex, France.

¹⁵Yale Cardiovascular Research Center, Departments of Medicine and Cell Biology, Yale University School of Medicine; New Haven, CT 06511, USA.

¹⁶Department of Rheumatology, Hôpital Bicêtre, Assistance Publique-Hôpitaux de Paris; 94270 Paris, France.

¹⁷Université Paris-Saclay, INSERM UMR_S 1184; 94270 Le Kremlin Bicêtre, France.

¹⁸Department of Hematology, INSERM UMR_S 1153, Imagine Institute, Necker Hospital, University of Paris, Assistance Publique-Hôpitaux de Paris, 75015 Paris, France.

¹⁹Department of Virology, Hôpitaux Universitaires de Strasbourg; 67091 Strasbourg, France.

²⁰Service de Médecine Interne, Endocrinologie et Nutrition, Hôpital de Hautepierre, Hôpitaux Universitaires de Strasbourg; 67200 Strasbourg, France.

²¹Département d'Hématologie, Groupe Hospitalier de la région Mulhouse Sud Alsace; 68100 Mulhouse, France.

²²Service de Réanimation Médicale, Groupe Hospitalier de la région Mulhouse Sud Alsace; 68100 Mulhouse, France.

²³Integrated BioTherapeutics, Inc.; Rockville, MD 20850, USA.

²⁴Division of Affinity Proteomics, Department of Protein Science, KTH Royal Institute of Technology, SciLifeLab; Stockholm, SE-171 21, Sweden.

²⁵Plateforme Auragen; 69003 Lyon, France.

²⁶Service de Rhumatologie, Centre National de Référence des Maladies Auto-immunes Systémiques Rares Est Sud-Ouest, Hôpitaux Universitaires de Strasbourg; 67200 Strasbourg, France.

²⁷INSERM (Institut de la Santé et de la Recherche Médicale), Université de Paris, Paris Cardiovascular Center-PARCC; 75015 Paris, France.

²⁸Howard Hughes Medical Institute; New York, NY 10065, USA.

²⁹Department of Electrical and Computer Engineering, Department of Chemistry, Department of Physics and Astronomy, University of Southern California; Los Angeles, CA 90089, USA.

³⁰Division of Genetics and Genomics, Boston Children's Hospital, Harvard Medical School; Boston, MA 02115, USA.

† these authors contributed equally

* correspondence: tom.chittenden@genuitysci.com (T.W.C) siamak@unistra.fr (S.B.)

Keywords: SARS-CoV-2, COVID-19, acute respiratory distress syndrome, multiomics, artificial intelligence, machine learning, disease severity signatures, ADAM9.

One sentence summary: Etiopathogenesis of critical COVID19 in a young patient population devoid of comorbidities.

Abstract

The etiopathogenesis of critical COVID-19 remains unknown. Indeed given major confounding factors (age and comorbidities), true drivers of this condition have remained elusive. Here, we employ an unprecedented multi-omics analysis, combined with artificial intelligence, in a young patient cohort where major comorbidities have been excluded at the onset. Here, we established a three-tier cohort of individuals younger than 50 years without major comorbidities. These included 47 “critical” (in the ICU under mechanical ventilation) and 25 “non-critical” (in a non-critical care ward) COVID-19 patients as well as 22 healthy individuals. The analyses included whole-genome sequencing, whole-blood RNA sequencing, plasma and blood mononuclear cells proteomics, cytokine profiling and high-throughput immunophenotyping. An ensemble of machine learning, deep learning, quantum annealing and structural causal modeling led to key findings. Critical patients were characterized by exacerbated inflammation, perturbed lymphoid/myeloid compartments, coagulation and viral cell biology. Within a unique gene signature that differentiated critical from non-critical patients, several driver genes promoted critical COVID-19 among which the upregulated metalloprotease *ADAM9* was key. This gene signature was supported in a second independent cohort of 81 critical and 73 recovered COVID-19 patients, as were *ADAM9* transcripts, soluble form and proteolytic activity. *Ex vivo* *ADAM9* inhibition affected SARS-CoV-2 uptake and replication in human lung epithelial cells. In conclusion, within a young, otherwise healthy, COVID-19 cohort, we provide the landscape of biological perturbations *in vivo* where a unique gene signature differentiated critical from non-critical patients. The key driver, *ADAM9*, interfered with SARS-CoV-2 biology. A repositioning strategy for anti-*ADAM9* therapeutic is feasible.

INTRODUCTION

Unlike many viral infections and most respiratory virus infections, COVID-19 is characterized by an extraordinarily complex and diversified spectrum of clinical manifestations, which results in the use of “syndemic” within, or *in lieu* of, pandemic (1). Indeed, upon infection with SARS-CoV-2, age-, sex-, and phenotype-matched individuals can be classified within four distinct groups, i.e., (1) asymptomatic individuals, (2) patients displaying influenza-like illnesses, (3) patients affected by respiratory dysfunction who eventually need an external oxygen supply, and (4) patients suffering from acute respiratory distress syndrome (ARDS) who need invasive mechanical ventilation in an intensive care unit (ICU). Even though the last group represents only a small fraction of COVID-19 patients, 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 despite key findings in discrete subpopulations (3-6), largely unanswered i.e. the exact pathophysiological mechanism governing the disease severity within a demographically and clinically homogeneous group of patients remains, for the majority of such patients, mostly unclear. 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 include notably older age as well as a number of comorbidities – e.g., cerebrovascular disease, types 1 and 2 diabetes, chronic kidney disease, chronic obstructive pulmonary disease, heart conditions, etc. (7) – present in COVID-19 patients. This is the case for the three-tier cohort studied in this work.

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

response is impaired at the RNA (12) and plasma (13) concentrations. Severity was also correlated with profound immune dysregulations, 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 a B-cell response activation (17) and an impaired T-cell response characterized by skewing towards a Th17 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 COVID-19 patients and controls that was established by discarding 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 to 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)/machine learning (ML)-based multiomics study of 47 young (aged <50 years) COVID-19 patients without comorbidities admitted to the ICU and under mechanical ventilation (“critical” patients), versus matched COVID-19 patients (i.e., aged <50 years with no comorbidities) needing “only” hospitalization at 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 multiomics approach included whole-genome sequencing (WGS), whole-blood RNA sequencing (RNA-seq), quantitative plasma and peripheral blood mononuclear cells (PBMCs) 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 IgG serology as well as the measurement of neutralizing anti-type I IFN auto-antibodies in the entire cohort.

RESULTS

Patients' characteristics and study design

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-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 [IQR 33; 46] years. The patients were divided into two groups: (i) a “critical” group consisting of 47 (65%) patients hospitalized at 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, nineteen (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 PaO₂/FiO₂ 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% (6 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.

Based on these two patient groups and an additional group of 22 healthy (SARS-Cov-2 negative) sex- and aged-matched controls, we applied a global multiomics analysis strategy to identify pathways and drivers of ARDS (Figure 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), detection of anti-SARS-CoV-2 neutralizing antibodies and anti-type I IFN neutralizing autoantibodies. Finally, 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 via classification of patient cohort 1 were evaluated in a second, independent cohort of 81 critical patients and 73 recovered critical patients (Table S1).

Cytokines, antibodies and immune cell hallmarks of critical COVID-19

The global proinflammatory cytokine profile showed significantly increased concentrations of IFN γ ($P=0.034$), TNF α ($P=0.022$), IL-1 β ($P=0.0002$), 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 (Figure 2A). This “cytokine storm” (25) was more pronounced in critical patients, as only IFN γ , TNF α and IL-10 are 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 a significant increase in the plasma concentration of IFN α in critical versus non-critical patients, the decrease in the IFN α concentration during the ICU stay and the reduction in the number of plasmacytoid dendritic cells, which are the main source of IFN α , suggest that the IFN response is indeed impaired in critical patients (Figure S1) (12).

At a systemic level, lymphopenia is correlated with disease severity (25-27) (Figure 2B). To further characterize the immune cells, 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 (Figure 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 cells and Th17 cells were negatively correlated with disease severity (Figure 2D). The latter observation is in line with the absence of a clear association between the plasma concentration of IL-17 and disease severity (Figure 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

(Figure 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 (Figure 2F and G). The remaining cell populations are presented in Figure S2.

Altogether, 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 *per se*, as both the global anti-SARS-CoV-2 antibody concentrations and their neutralizing activity were not significantly different in critical versus non-critical patients (Figure S3A and B).

To complete the immunologic profile, based on findings suggesting that at least 10% of critical patients have preexisting anti-type I IFN autoantibodies (30, 31), we measured anti-IFN α 2 and anti-IFN ω neutralizing autoantibodies in patients and controls. Autoantibodies against type I IFNs were identified in two critical patients (Figure S3C) but none of the non-critical patients nor the healthy controls. Interestingly, in these two patients, the presence of autoantibodies was associated with an absence of SARS-CoV-2 neutralizing activity (Figure S3D).

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

Quantitative nanoLC-MS/MS analysis of whole unfractionated plasma samples identified a total of 336 proteins, of which an average of 178 ± 7 , 189 ± 11 and 195 ± 8 proteins in healthy individuals, non-critical and critical patients, respectively were used for differential analysis (Figure 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 (Figure 3B and C). In line with previous studies

(8, 32), the antimicrobial calprotectin (heterodimer of S100A8 and S100A9) was among the top differentially expressed proteins (DEPs) in critical versus non-critical patients, which confirms that calprotectin is a robust marker for disease severity (Figure 3D). Our data also showed dysregulation of multiple apolipoproteins including APOA1, APOA2, APOA4, APOM, APOD, APOC1 and APOL1 (Figure 3C and E). Most of these proteins were associated with macrophage functions and were downregulated in critical patients. Acute-phase proteins (CRP, CPN1, CPN2, C6, CFB, ORM1, ORM2, SERPINA3, and SAA1) were strongly upregulated in critical patients (Figure 3C 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. A total of 2196 proteins were identified, of which an average of 801 ± 213 , 1050 ± 309 and 1052 ± 286 proteins in healthy individuals, non-critical and critical patients, respectively were used for differential analysis (Figure 3F). Although the human proteome coverage was relatively low after exclusion of contaminating fetal calf serum 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 (Figure 3H-I). The latter observation involving the CA2, AHSP, SLC4A1, TFRC, DMTN, FASN, and 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 HBB, HBD, HBE1, SLC4A1, PRDX2, SRI, ARF4, MANF, ITGA2, ORM1, and 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 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 (Figure 4A 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. Figure 4C shows the nGOSeq terms that were found to be statistically enriched in at least two omics datasets in critical compared with non-critical patients. In line with the cytokine profiling results (Figure 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 (Figure 2C and D) and previous studies, the B-cell response was activated, whereas the T-cell response was impaired (17, 37). As previously witnessed (8, 14, 15, 38), the activation of neutrophils and monocytes was confirmed by the enrichment of nine different nGOSeq terms (Figure 4C). The nGOSeq enrichment analysis also indicated that dysfunction of blood coagulation involves a fibrinolytic response, an observation that could, however, 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 in 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 ensemble AI/ML and probabilistic programming discovers a robust gene expression signature and driver genes that differentiate critical from non-critical patients

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

pipeline depicted in Figure 1B. Briefly, we partitioned the 69 patient 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/ML algorithms, including a quantum support vector machine (qSVM) to differentiate between non-critical and critical COVID-19 patients. 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 Figure 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/ML algorithms are generalizable.

After successfully classifying non-critical versus critical patients based on whole-transcriptome RNA-seq profiling, we assessed feature scores across the six distinct ML algorithms (see Methods section) and all partitions to determine an ensemble feature ranking, ignoring features from the partitions of patient data where the test 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 represents hundreds of genes that are differentially expressed and by itself does 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 disease. 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) in Figure 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. That is, perturbing these genes is most disruptive to the state of the network (Figure 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 upregulated in critical patients with false discovery rates (FDR) of 1.6×10^{-11} , 3.1×10^{-12} , 1.6×10^{-11} , 1.0×10^{-9} and 5.3×10^{-13} , respectively (Figure 5C).

To further assess the informativeness of this COVID-19 gene expression signature, a second, independent patient cohort consisting of critical COVID-19 patients sampled at the time of entry into the ICU and recovered critical patients sampled at three months after discharge from the ICU was used. Patients in this second cohort were a more “typical” COVID-19 ICU-population as no exclusion criteria (unlike the primary cohort) based on age (<50) or absence of comorbidities were applied (Table S1). Although non-critical COVID-19 patients cannot be assumed to be the same as recovered critical COVID-19 patients, and thus the machine learning (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 (Figure 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 (Figure S5B 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 major driver of ARDS in critical COVID-19 patients

Among the five driver genes identified by structural causal modeling, 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, which encode membrane-bound and secreted proteins. Although neither isoform could be detected using our proteomics approach, *ADAM9* was upregulated at the RNA level, and the secreted form was found at a higher concentration in the serum of critical versus non-critical patients (Figure 6A and B). The transcriptional upregulation of *ADAM9* was also found to be associated with disease severity in a previously published bulk RNA-seq dataset (Figure S6) (45). To assess a potentially increased metalloprotease activity in the critical group, we quantified the soluble form of the MICA protein (46), which is known to be cleaved by *ADAM9* (47) by ELISA. The concentration of soluble MICA was indeed significantly higher in the plasma of critical patients as compared to non-critical patients ($P=0.016$) and healthy controls ($P=0.0001$; Figure 6C). A global eQTL analysis using WGS and RNA-seq data identified eight SNPs associated with three of the top five putative driver genes with genome-wide significance (Table S6). Among these SNPs, rs7840270 is localized just 0.3 kb upstream of the *ADAM9* gene and an eQTL for blood expression reported in GTEX. In the present cohort, including all 3 groups together, the C allele was associated with a higher amount of *ADAM9* transcripts (Figure 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% vs. 50%, OR=2.48, 95% CI: [1.14-5.36], $P=0.017$) (this was not due to any difference in ethnicities between critical and non-critical groups; Figure 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 was contributing to higher

ADAM9 RNA expression most strongly in critical patients although the other phenotype subsets are smaller datasets (Figure S8).

To assess the role of ADAM9 in viral infection, we set up an *ex vivo* assay in which *ADAM9* was silenced by siRNA in Vero 76 or A549-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 RT-PCR of the culture supernatant (Figure 6E). The average silencing efficiency reached 66% in Vero 76 cells and 93% in A549-ACE2 cells (Figure 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 (Figure 6F and G). Our results collectively demonstrate that *ADAM9* is an *in vivo* upregulated driver in critical patients. We also show a higher global proteolytic activity in the sera of these 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 and/or standard bioclinical parameters such as CRP, D-dimers or SOFA scores (e.g. (55-57)). A comprehensive understanding of the immune responses to SARS-CoV-2 infection is fundamental to an explanation as of 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 multiomics 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 (Figure 5A and Table S3). Moreover, one can infer generalizability 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 patients and 73 recovered critical patients (Figure S5).

Using the top 600 gene expression features of the signature as the input for structural causal modeling, 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 GTPase 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 upregulation of *ADAM9* at the RNA and protein levels in critical patients (which might partly be linked to pre-stored *ADAM9*'s release by neutrophils; see (69)),

the increased metalloprotease activity in these same patients and our *ex vivo* validation of its effect on viral uptake/replication are indeed strong arguments supporting the targeting of this protein as a potential therapeutic strategy for the treatment or prevention of critical COVID-19. ADAM9 appears to dramatically affect viral uptake. The inhibition of this presumed mechanism of action of ADAM9 might represent a novel means for the treatment SARS-CoV-2 and/or other viral illnesses. Moreover, therapies that block viral uptake rather than host receptor binding are more likely to be variant-independent, a known virological behavior which might - at least partially - derail current vaccination efforts (70).

Due to 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 COVID-19 patients could therefore be envisioned. Alternatively, other therapeutic agents to reduce the ADAM9 concentration or activity could be pursued.

Our study has several limitations. Based on the present experimental results, we cannot conclude yet as to the molecular mechanism linking ADAM9 and viral uptake/replication. The predictive performance of ADAM9 as diagnostic marker for disease severity, as well as therapeutic target has to be evaluated in further studies. Finally, it would be interesting to test the silencing of ADAM9 on various SARS-CoV-2 variants.

In conclusion, this study presents a detailed multiomics 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 significantly minimized possible annotation biases, to gain novel insights into COVID-19 pathogenesis and to provide potential candidate diagnostic, prognostic and especially much needed therapeutic targets that might be helpful in the ongoing battle against the COVID-19 pandemic.

MATERIALS AND METHODS

Patients and 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 a quantitative real-time reverse transcriptase 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: (1) the “critical group” which included 47 patients admitted to the ICU, (2) the “non-critical group”, which was composed of 25 hospitalized patients at the non-critical care ward, and (3) 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. Additional details of the study design are presented in the first paragraph of the results section and in Figure 1.

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 4-8 days posthospitalization until discharge or death. Patient blood was collected into BD Vacutainer tubes with heparin (for plasma and PBMCs), EDTA (for DNA) or without additive (for serum) and into PAXgene® Blood RNA tubes (Becton, Dickinson and Company, Franklin Lakes, NJ, USA). 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 900 x g at room temperature for 10 min, aliquoted, and stored at -80°C until use. PBMCs were prepared within 24 h by Ficoll density gradient centrifugation. Aliquots of 1×10^6 dry cell pellets were frozen at -80°C until use for proteomics. Aliquots of a minimum of 5×10^6 cells were frozen at -80°C in 90% fetal calf serum (FCS)/10% dimethyl sulfoxide (DMSO). The EDTA and PAXgene[®] tubes were stored at -80°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- α , IL-12p70, IL-1 β , IL-2, IL-4 and IFN- γ) and the S-PLEX Human IFN- α 2a Kit following the manufacturer's instructions (Mesoscale Discovery, Gaithersburg, MD, USA). The plasma was used undiluted for the S-PLEX Human IFN- α 2a Kit and diluted 2-fold for use with the V-PLEX Proinflammatory Panel 1. The MSD plates were analyzed with an MS2400 imager (Mesoscale Discovery, Gaithersburg, MD, USA). Soluble IL-17 in undiluted serum was quantified by Quantikine[®] HS ELISA (Human IL-17 Immunoassay) following the manufacturer's instructions (R&D Systems, Minneapolis, MN, USA). All standards and samples were measured in duplicate.

Immune phenotyping by mass cytometry

PBMCs were thawed rapidly, washed twice with 10 volumes of RPMI (Roswell Park Memorial Institute) medium (Thermo Fisher Scientific, Waltham, MA, USA) and centrifuged for 7 min at 300 x g at room temperature between each washing step. Cells were then treated with 250 U of DNase (Thermo Fisher Scientific, Waltham, MA, USA) in 10 volumes of RPMI medium for 30 min at 37°C in the presence of 5% CO₂. During this step, the viability and the number of the cells were determined with Trypan Blue (Thermo Fisher Scientific, Waltham, MA, USA) and Türk's solution (Merck Millipore, Burlington, MA, USA), respectively. After the elimination of DNase by centrifugation for 7 min at 300 x g at room temperature, a total of 3×10^6 cells were used for immunostaining with the Maxpar[®] Direct Immune Profiling Assay kit (Fluidigm, San

Francisco, CA, USA), following the manufacturer's instructions, except that we used 32% paraformaldehyde (Electron Microscopy Sciences, Hatfield, PA, USA). A red blood cell lysis step was included after the immunostaining following the manufacturer's instructions. The prepared cells were stored at -80°C until use for acquisition with a Helios mass cytometer system (Fluidigm, San Francisco, CA, USA). An average of 600000 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 FCS files from each group (healthy, critical, non-critical) were then concatenated using CyTOF® software v.7.0.8493.0 for viSNE analysis (Cytobank Inc, Mountain View, CA, USA). A total of 300000 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

Sample preparation

Samples were prepared using the PreOmics iST Kit (PreOmics GmbH, Martinsried, Germany) according to the manufacturer's protocol. Two microliters of plasma was mixed with 50 µl of Lyse buffer and heated at 95°C for 10 min at 1000 rpm. The protein concentration was determined using the Bradford assay (Bio-Rad, Hercules, CA, USA) according to the manufacturer's instructions. The samples were transferred to 96-well-plate cartridges, and 50 µl of resuspended Digest solution was added. The samples were then heated at 37°C for 2 h, and 100 µl of Stop buffer was added. The samples were centrifuged to retain the peptides on the cartridge and washed twice with "Wash 1" and "Wash 2" buffers. The peptides were then eluted twice with Elute buffer before evaporation under vacuum. The peptides were then resuspended using the "LC-load" solution containing iRT peptides (Biognosys, Schlieren, Switzerland), and the samples were rapidly sonicated before being injected into the nanoLC-MS/MS system.

NanoLC-MS/MS analysis

The NanoLC-MS/MS analyses were performed with a nanoAcquity Ultra-Performance LC[®] (UPLC[®]) device (Waters Corporation, Milford, MA, USA) coupled to a Q-Exactive[™] Plus mass spectrometer (Thermo Fisher Scientific, Waltham, MA, USA). Peptide separation was performed on an ACQUITY UPLC BEH130 C18 column (250 mm × 75 μm with 1.7-μm-diameter particles) and a Symmetry C18 precolumn (20 mm × 180 μm with 5-μm-diameter particles, Waters). The solvent system consisted of 0.1% FA in water (solvent A) and 0.1% FA in ACN (solvent B). The samples (equivalent to 500 ng of proteins) were loaded into the enrichment column over 3 min at 5 μl/min with 99% solvent A and 1% solvent B. The peptides were eluted at 400 nl/min using the following gradient of solvent B: from 1 to 35% over 60 min and from 35 to 90% over 1 min. The 90 samples were injected in randomized order. The MS capillary voltage was set to 2.1 kV at 250°C. The system was operated in the data dependent acquisition mode with automatic switching between MS (mass range 300–1800 m/z with R = 70000, automatic gain control (AGC) fixed to 3×10^6 ions, and maximum injection time of 50 ms) and MS/MS (mass range of 200–2000 m/z with R = 17500, AGC fixed at 1×10^5 and maximal injection time of 100 ms) modes. The ten most abundant ions were selected on each MS spectrum for further isolation and higher energy collision dissociation fragmentation, excluding unassigned and monocharged ions. The dynamic exclusion time was set to 60s. 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.

Data analysis

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-SwissProt database (May 11 2020, 20410 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 false discovery rate 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).

Differential protein expression analysis

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 justvs function from the vsn R package (74). Missing values were imputed using the random forest 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 based on an adjusted *P*-value cutoff of 0.05 using the Benjamini-Hochberg method.

PBMC proteomics analysis

Samples were prepared using the PreOmics' iST Kit (PreOmics GmbH, Martinsried, Germany) according to the manufacturer's protocol. Briefly, PBMC pellets were resuspended in 50 µl of Lyse buffer and heated at 95°C for 10 min at 1000 rpm before being sonicated for 10 min on ice. The protein concentration of the extract was determined using the Pierce™ BCA Protein Assay Kit (Thermo Fisher, Waltham, MA, USA). The samples were transferred to 96-well plate cartridges and 50 µl of resuspended Digest solution was added. The samples were then

heated at 37°C for 2 h and 100 µl of Stop buffer was added. The samples were centrifuged to retain the peptides on the cartridge and washed twice with “Wash 1” and “Wash 2” buffers. The peptides were then eluted twice with Elute buffer before evaporation under vacuum. Finally, the peptides were resuspended using the “LC-load” solution containing iRT peptides (Biognosys, Schlieren, Switzerland), and the samples were rapidly sonicated before being injected into the nanoLC-MS/MS system.

NanoLC-MS/MS analysis

NanoLC-MS/MS analyses were performed with a nanoAcquity UPLC device (Waters Corporation, Milford, MA, USA) coupled to a Q-Exactive HF-X mass spectrometer (Thermo Fisher Scientific, Waltham, MA, USA). Peptide separation was performed on an Acquity UPLC BEH130 C18 column (250 mm × 75 µm with 1.7-µm-diameter particles) and a Symmetry C18 precolumn (20 mm × 180 µm with 5-µm-diameter particles, Waters). The solvent system consisted of 0.1% formic acid (FA) in water (solvent A) and 0.1% FA in acetonitrile (ACN) (solvent B). The samples (equivalent to 414 ng of proteins) were loaded into the enrichment column over 3 min at 5 µl/min with 99% solvent A and 1% solvent B. The peptides were eluted at 400 nl/min using the following gradient of solvent B: from 2 to 25% over 53 min, from 25 to 40% over 10 min and 40 to 90% over 2 min. The 77 samples were injected using a randomized injection sequence. The MS capillary voltage was set to 1.9 kV at 250°C. The system was operated in data dependent acquisition mode with automatic switching between MS (mass range of 300-1800 m/z with R = 60000, automatic gain control (AGC) fixed to 3×10^6 ions and maximum injection time of 50ms) and MS/MS (mass range of 200–2000 m/z with R = 15000, AGC fixed at 1×10^5 and maximal injection time of 100 ms) modes. The ten most abundant ions were selected on each MS spectrum for further isolation and higher energy collision dissociation fragmentation with the exclusion of unassigned and monocharged ions. The dynamic exclusion time was set to 60 s.

Data analysis

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 potential traces of fetal calf serum in the samples) was extracted from UniProtKB-SwissProt (8 September 2020, 26413 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 false discovery rate 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).

Differential protein expression analysis

Summed peptides normalized label-free quantification (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. Variance stabilization normalization (Vsn) was performed using the justvsfn function from the vsn R package (74). Missing values were imputed using the random forest 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 based on an adjusted *P*-value cutoff of 0.05 using the Benjamini-Hochberg method.

Serology analysis

The serum IgG reactivity towards three SARS-CoV-2 viral antigens was measured as previously described (24) by means of a multiplex antigen bead array. The three viral antigens included in our multiplex bead array comprise two representations of the Spike protein (a soluble trimeric form of the spike glycoprotein stabilized in the pre-fusion conformation expressed in HEK, and the Spike S1 domain expressed in CHO cells), and the C-terminal domain of the Nucleocapsid protein (expressed in *E.coli*). Briefly, each antigen was immobilized on the surface of uniquely color-coded magnetic beads (Bead ID; MagPlex®, Luminex Corporation, Austin, TX, USA), and the bead IDs pooled together to generate the bead array used to test the serum samples. Sera were incubated with the multiplex antigen array in a 384-well plate format, and the IgG reactive to the viral antigen was detected by means of a phycoerythrin-conjugated anti-hIgG (H10104, Invitrogen, Thermo Fisher Scientific, Waltham, MA, USA) in a FlexMap3D instrument (Luminex Corporation, Austin, TX, USA). The cut-off for reactivity was evaluated for each antigen singularly and defined as the mean +6SD of the intensity signals of 12 negative controls. The 12 controls were carefully selected among pre-pandemic samples as representative of the background range for each single antigen included in the test. The performance of this serology assay was previously evaluated on 2090 negative samples (pre-pandemic samples collected in 2019 or earlier and including 26 samples from individuals infected by other Coronaviruses) and 331 samples from COVID-19 PCR-confirmed cases (sampled collected at least 17 days after disease onset), showing 99.7% sensitivity and 100% specificity.

Neutralizing antibodies

Spike pseudotype neutralization assay was conducted using vesicular stomatitis virus (VSV) where glycoprotein gene (G) has been deleted and substituted in trans with SARS-CoV-2 Spike protein (D614G variant) lacking terminal eighteen amino acids of the cytoplasmic domain. The pseudotype VSV-ΔG SARS-CoV-2 S D614G is cytopathic and expresses the firefly luciferase. For the neutralization assay, VSV-ΔG SARS-CoV-2 S D614G was incubated

with serial dilutions of patient sera starting at 1:40 dilution for 1 h at room temperature. A monolayer of Vero cells seeded at 60000 cells/well in black-well plates was infected with the pseudotype virus (virus only) or virus and serum mixture overnight at 37°C - 5% CO₂. The following day, cells were lysed using passive lysis buffer (Promega, Madison, WI, USA) and luciferase activity was measured upon addition of substrate (Promega® CellTiterGlo®, Madison, WI, USA) using Biotek plate reader and Gen5 software. Neutralization was calculated by comparing relative luciferase units of serum treated versus the virus only control. To determine neutralizing titers, the concentration of the antibody resulting in 50% neutralization (NT₅₀) was determined using XLFit and graphed using Prism (GraphPad, San Diego, CA, USA).

Autoantibodies against type I IFNs

The blocking activity of anti-IFN α and anti-IFN ω autoantibodies was determined by assessing a reporter luciferase activity as previously described (30). Briefly, HEK293T cells were transfected with the firefly luciferase plasmids under the control of human ISRE promoters in the pGL4.45 backbone, and a constitutively expressing Renilla luciferase plasmid for normalization (pRL-SV40). Cells were transfected in the presence of the X-tremeGene 9 transfection reagent (Sigma Aldrich, Saint-Louis, MI, USA) for 36 hours. Then, Dulbecco's modified Eagle medium (DMEM, Thermo Fisher Scientific, Waltham, MA, USA) supplemented with 10% healthy control or patient serum and were either left unstimulated or were stimulated with IFN α or IFN ω (10 ng/mL) for 16 hours at 37°C. Each sample was tested once. Finally, Luciferase concentrations were measured with the Dual-Glo reagent, according to the manufacturer's protocol (Promega, Madison, WI, USA). Firefly luciferase values were normalized against Renilla luciferase values, and fold induction is shown relative to controls transfected with empty plasmids.

Whole-genome sequencing (WGS)

WGS data was generated from DNA isolated from whole blood. Novaseq™ 6000 (Illumina Inc., San Diego, CA, USA) machines were used for DNA sequencing to a mean 30X coverage. The raw sequencing reads from FASTQ files were aligned using Burrows-Wheeler Aligner (BWA) (77), and GVCF 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 based on kinship (king cutoff of 0.3) and retained 24476739 SNPs that were given a 'PASS' filter status by VQSR. The analysis of the 72 samples from the critical and non-critical groups identified 15870076 variants with MAF < 5%. The first two principal components were generated using plink2 on LD-pruned variants with Hardy-Weinberg equilibrium in the controls with a P -value $\geq 1 \times 10^{-6}$ and MAF > 5% and were used as covariates to correct for population stratification.

Analysis of expression quantitative trait loci (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 MatrixEQTL R package (81) where we selected a linear model and a maximum distance for gene-SNP pairs of 1×10^6 . The top two principal components identified from the genotype principal component analysis were used as covariates to control for population stratification. We selected 304044 significant eQTLs with FDR ≤ 0.05 .

RNA sequencing (RNA-seq)

RNA extraction

Whole-blood RNA was extracted from PAXgene tubes with the PAXgene Blood RNA Kit following the manufacturer's instructions (Qiagen, Hilden, Germany). 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 RIN) (Agilent Technologies,

Santa Clara, CA, USA) and RiboGreen™ (for the concentration) (Thermo Fisher Scientific, Waltham, MA, USA). RNA sequencing libraries were generated using the TruSeq Stranded Total RNA with Ribo-Zero Globin kit (Illumina, San Diego, CA, USA) and sequenced on the Illumina NovaSeq 6000 instrument with S4 flow cells and 151-bp 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 (<http://www.gencodegenes.org>). Raw and processed datasets have been deposited in GEO with identifier GSE172114.

Differential gene expression (DGE) analysis

DGE analysis was performed for two different purposes: 1) for the combined omics analysis of differentially expressed genes and proteins, and 2) as step to determine feature selection for classification in the *in silico* computational intelligence approach. For the combined omics analysis, we first removed low 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, in order to avoid data leakage. Briefly, we removed low 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 (QLF)-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 structural causal modeling

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 structural causal modeling 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. In order 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 structural causal modeling (600 features is the maximum that the structural causal modeling can finish in a reasonable amount of time). More details about the classification algorithms used, the feature ranking, and the structural causal modeling are provided below.

Ensemble artificial intelligence

We used seven distinct ML approaches for our classification models. The relevant hyperparameters for each method are mentioned in their respective sections. Hyperparameters were selected by using 10-fold cross-validation of the training data, and the performance was evaluated using the held-out test data.

Least Absolute Shrinkage and Selection Operator (LASSO), and Ridge Regression

LASSO (86) is an L1-penalized linear regression model defined as:

$$\hat{\boldsymbol{\beta}}(\lambda) = \underset{\boldsymbol{\beta}, \beta_0}{\operatorname{argmin}} [-\log[L(\boldsymbol{y}; \boldsymbol{\beta}, \beta_0)]] + \lambda \|\boldsymbol{\beta}\|_1 \quad (1)$$

Ridge (87, 88) is an L2-penalized linear regression model defined as:

$$\hat{\boldsymbol{\beta}}(\lambda) = \underset{\boldsymbol{\beta}, \beta_0}{\operatorname{argmin}} [-\log[L(y; \boldsymbol{\beta}, \beta_0)]] + \lambda \|\boldsymbol{\beta}\|_2^2 \quad (2)$$

where the loss function is

$$L = \frac{1}{N} \sum_{i=1}^N (y_i(\beta_0 + \mathbf{x}_i \cdot \boldsymbol{\beta}) - \log(1 + e^{\beta_0 - \mathbf{x}_i \cdot \boldsymbol{\beta}}))$$

In both cases, $\lambda > 0$ is the regularization parameter that controls model complexity, $\boldsymbol{\beta}$ are the regression coefficients, β_0 is the intercept term, y represents an indicator function for the critical patients (i.e., $y_i = 1$ if the i -th training sample is a critical patient; otherwise $y_i = 0$), the vector \mathbf{x}_i is the i -th training sample, and the goal of the training procedure is to determine $\hat{\boldsymbol{\beta}}$ and β_0 , the optimal regression coefficients and the optimal intercept, that minimize the quantities defined in Eqs. (1) and (2). The predicted label is given by $y = \beta_0 + \mathbf{x} \cdot \boldsymbol{\beta}$, where a threshold of 0.5 is introduced to binarize the label for classification problems. In LASSO, the constraint placed on the norm of $\boldsymbol{\beta}$ (the strength of which is given by λ) causes coefficients of uninformative features to shrink to zero. This leads to a simpler model that contains only a few nonzero coefficients. We used the 'glmnet' function from the caret (89) R package to train all Lasso and Ridge models. For Ridge, the constraint placed on the norm of $\boldsymbol{\beta}$ plays a similar role in determining model complexity, except that coefficients for uninformative features do not necessarily shrink to zero.

For both Lasso and Ridge, we opted to implement the function over a custom tuning grid of λ from 2^{-8} to 2^2 . λ was selected via 10-fold cross-validation as the value that gave the minimum mean cross-validated error.

Support Vector Machines (SVM)

SVMs (90, 91) are a set of supervised learning models used for classification and regression analyses. The primal form of the optimization problem is:

$$\mathbf{w}_{opt} = \underset{\mathbf{w}, b, \mathbf{a}}{\operatorname{argmin}} \left[L_p = \frac{1}{2} \|\mathbf{w}\|_2^2 - \sum_{i=1}^N a_i y_i (\mathbf{x}_i \cdot \mathbf{w} + b) + \sum_{i=1}^N a_i \right] \quad (3)$$

where L_p is the loss function in its primal form (p for primal), \mathbf{w} represents the weights to be determined in the optimization, \mathbf{x}_i is the i -th training sample, y_i is the label of the i -th training sample, $a_i \geq 0$ are Lagrange multipliers, N is the number of training points, and b is the intercept term. Labels are predicted by thresholding $\mathbf{x}_i \cdot \mathbf{w} + b$.

The optimization problem in its dual form is defined as

$$\mathbf{a}_{opt} = \underset{\mathbf{a}}{\operatorname{argmax}} \left[L_D(\mathbf{a}) = \sum_{i=1}^N a_i - \frac{1}{2} \sum_{i,j=1}^N a_i a_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j) \right]$$

where L_D is the Lagrangian dual of the primal problem, a_i are the Lagrange multipliers, y_i and \mathbf{x}_i are the i -th label and training sample, respectively, and $K(\cdot, \cdot)$ is the kernel function. Maximization takes place subject to the constraints $\sum_i a_i y_i = 0$ and $a_i \geq C \geq 0, \forall i$. Here C is a hyperparameter that controls the degree of misclassification of the model for nonlinear classifiers. The optimal values of \mathbf{w} and b can be found in terms of the a_i 's, and the label of a new data point x can be found by thresholding the output $\sum_i a_i y_i K(\mathbf{x}_i, x) + b$.

In most cases, many of the a_i 's are zero, and evaluating predictions can be faster using the dual form. We used SVM with a linear kernel ('svmLinear2') (i.e., $K(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i \cdot \mathbf{x}_j$, the inner product of \mathbf{x}_i and \mathbf{x}_j) function from the caret (89) R package to train all SVM models. C ranged from 2^{-2} to 2^3 , and a 10-fold cross-validation was used to tune and select the hyperparameters with the best cross-validation accuracy for training the model.

Random Forest (RF)

RF (92, 93) is an ensemble learning method for classification and regression that builds a set (or forest) of decision trees. In RF, n samples are selected (typically two-thirds of all the training data) with replacement from the training data m times, which yields m different decision trees. Each tree is grown by considering 'mtry' of the total features, and the tree is split depending on which features yield the smallest Gini impurity. In the event of multiple training samples in

a terminal node of a particular tree, the predicted label is given by the mode of all the training samples in a terminal node. The final prediction for a new sample x is determined by taking the majority vote over all the trees in the forest. We used the 'rf' function from the caret (89) R package to train all RF models. Ten-fold cross-validation was used to tune the parameters for training the model. A tune grid with 44 values from 1 to 44 for 'mtry', the number of random variables considered for a split in each iteration during the construction of each tree, was used for the tuning model.

XGBoost (XGB)

XGB (94) is a distributed gradient boosting library for classification and regression by building an ensemble of decision trees. In contrast to RF, XGB uses an additive strategy to add new trees one at a time based on whether they optimize the objective function. The objective function for the t -th tree is

$$obj^{(t)} = \sum_{j=1}^T \left[G_j w_j + \frac{1}{2} (H_j + \lambda) w_j^2 \right] + \gamma T$$

where $G_j = 2 \sum_{i \in I_j} (\hat{y}_i^{(t-1)} - y_i)$, $H_j = 2 |I_j|$, λ and γ are hyperparameters controlling model complexity, T is the number of leaves in the trees, w_j is the combined score across all the data points for the j -th leaf. Here, I_j refers to the set of indices of data points assigned to the j -th leaf, $|I_j|$ is the size of the set I_j , $\hat{y}_i^{(t-1)}$ is the predicted score (without the t -th tree) of the i -th data point, and y_i is the actual label of the i -th data point. The default parameter tuning grid in R was used, and a 10-fold cross-validation was used to tune and select the hyperparameters with the best cross-validation accuracy for training the model.

Quantum Support Vector Machines (qSVM)

qSVM is a quantum adaptation of SVM that can be used for classification designed to be run with a quantum annealer (QA) (95). The advantage of running the optimization problem on a QA is that, since the QA samples from the quantum distribution, it retains both the lowest

energy solution and some of the next lowest-energy solutions (Albash and Lidar, 2018). Because of the suboptimal solutions, we expect qSVM to perform worse on the training data than classical SVM (which only includes the optimal solution). However, suboptimal solutions can capture different aspects of the training data and generate different decision boundaries. As such, a suitable combination of the suboptimal solutions in qSVM might outperform cSVM on the test data.

The objective function is the same as for classical SVM up to a change in sign, i.e.,

$$\alpha_{opt} = \underset{\alpha}{\operatorname{argmin}} \left[L_D(\alpha) = \frac{1}{2} \sum_{i,j=1}^N \alpha_i \alpha_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j) - \sum_{i=1}^N \alpha_i \right] []$$

subject to the constraints $\sum_i \alpha_i y_i = 0$ and $0 \leq \alpha_i \leq C, \forall i$.

qSVM was run on physical quantum annealers manufactured by D-Wave Inc. (96). The D-Wave ‘‘Advantage’’ device used in this work had 5436 qubits with 15 couplers per qubit, featuring the ‘‘Pegasus’’ connectivity graph between qubits. Since D-Wave can only produce binary solutions, we used the encoding as defined in (95) to convert the continuous variables α_i into K binary variables using base B :

$$\alpha_i = \sum_{k=0}^{K-1} B^k a_{Ki+k}, \quad a_{Ki+k} \in (0,1).$$

Using this encoding and also adding a penalty ξ to the loss function, the optimization problem becomes a quadratic unconstrained binary optimization (QUBO) problem, which can be run on a QA:

$$\begin{aligned} E &= \frac{1}{2} \sum_{i,j,k,l} a_{Ki+k} a_{Kj+l} B^{k+l} y_i y_j K(\mathbf{x}_i, \mathbf{x}_j) - \sum_{i,k} B^k a_{Ki+k} + \xi \left(\sum_{i,k} B^k a_{Ki+k} y_i \right)^2 \\ &= \sum_{i,j=0}^{N-1} \sum_{k,l=0}^{K-1} Q_{Ki+l,Kj+l} a_{Ki+k} a_{Kj+l}, \end{aligned}$$

where $Q_{Ki+k,Kj+l} = \frac{1}{2} B^{k+l} y_i y_j (K(\mathbf{x}_i, \mathbf{x}_j) + \xi) - \delta_{i,j} \delta_{k,l} B^k$. As the objective function above might necessitate connections between any pair of qubits, an embedding is necessary (97). Hyperparameters were selected using a custom 3-fold Monte-Carlo cross-validation on the

training data. Hyperparameters included the type of kernel (linear versus Gaussian), B (between 2 and 10), K (between 2 and 6), ξ (between 0 and 5), and γ (between 2^{-3} and 2^3). Closely related quantum ML approaches deploying QA have been used in classifying transcription factor binding to DNA (98) and in classification of multiomics human cancer data (39).

Deep artificial neural network (DANN)

We adapted common deep learning methodologies to analyze genomic datasets (99). Typical deep neural networks use a series of nonlinear transformations (termed layers), with the final output considered a prediction of class or regression variables. Each layer consists of a set of weights (W) and biases (b) that are tuned during the training phase to learn which nonlinear combinations of input features are most important for the prediction task. These types of models “automatically” learn patterns in the data and combine them in some abstract nonlinear fashion, to gain an ability to make predictions about the dataset.

The basic formulation of a fully connected DANN is given as

$$\text{For } m \text{ layers ... } \left\{ \begin{array}{l} f_1 = \rho_1 \left(\sum_{j=1}^{d_1} (W_{1,j} \times X_j) + b_{d_1+1} \right) \\ f_2 = \rho_2 \left(\sum_{j=1}^{d_2} (W_{2,j} \times f_1) + b_{d_2+1} \right) \\ f_m = \rho_m \left(\sum_{j=1}^{d_m} (W_{m,j} \times f_{m-1}) + b_{d_m+1} \right) \end{array} \right.$$

where the dimensions of W and b are determined by the number of neurons in each layer (d_1, d_2, \dots, d_m). Each layer used rectified linear units as activation functions:

$$\rho_l(z) = \max(0, z).$$

The final layer used a softmax function, with the number of neurons equal to the number of classes (K), to convert the logits to probabilities:

$$\Phi(f_m)_j = \frac{e^{f_{m,j}}}{\sum_{k=1}^K e^{f_{m,k}}} \text{ for } j = 1, \dots, K,$$

where $f_{m,j}$ is the output of the j -th neuron of the m -th layer. In addition, we used the concept of “dropout,” which randomly sets a portion of input values (η) to the layer to zero during the training phase (100). This has a strong regularization effect (essentially by injecting random noise) that helps prevent models from overfitting. Layers that included dropout were formulated as

$$f = \rho \left(\sum_{j=1}^d (W_j \times X_j) + b_{d+1} \right) \times m_l,$$

where $m_l \sim \text{Bernoulli}(\eta)$.

When evaluating models on test datasets, the dropout mask was not used. We used the categorical cross-entropy loss function to train DANNs, where (B_n) is the minibatch size, t_i is the correct class index, and p_i is the class probability from the softmax layer:

$$L_T = - \sum_{i=1}^{B_n} t_i \log(p_i).$$

We used minibatch stochastic gradient descent with Nesterov momentum to update the DANN parameters based on the above-described loss function (101). We used the TensorFlow (102) Python package to construct the DANNs.

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 based on 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, 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 600000 possible features were considered for each feature set (“up to” since they might not be unique, as the top 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; i.e., 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, 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). 600 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 directed acyclic graph (DAG), where the directionality of the arcs represents conditional dependencies between the nodes. BBNs were used to assess the conditional dependence and probabilistic relationships between the most informative genes and thereby identify putative driver genes. We relied on a set of common

assumptions to determine the causal structure: (1) causal sufficiency assumption, where there are no unobserved cofounders; (2) causal Markov assumption, where all d-separations in the graph (G) imply conditional independence in the observed probability distribution; and (3) causal faithfulness assumption, where all of the conditional independences in the observed probability distribution imply d-separations in the graph (G). We acknowledge that our data might not strictly meet all of these assumptions, however, the generated BBNs provide useful biological hypotheses that could be experimentally validated.

We determined BBNs using the *bnlearn* R package with the score-based hill-climbing algorithm that heuristically searched the space of all possible DAGs (Scutari, 2010). As the hill-climbing algorithm can get trapped in local optima and is quite dependent on the starting structure, we initialized 100 BBNs starting from different network seeds. During the hill-climbing process, each candidate BBN was assessed with the Bayesian information criterion (BIC) score (103, 104):

$$\text{BIC} = \log L(X_1, \dots, X_v) - \frac{d}{2} \log n,$$

where X_1, \dots, X_v is the node set, d is the number of free parameters, n is the sample size of the dataset, and L is the likelihood. Note that this definition of the BIC, which is the version implemented in the *bnlearn* package, rescales the classic definition by -2. The penalty term was used to prevent overly complicated structures and overfitting. Each run of the hill-climbing algorithm returns a structure that maximizes the BIC score (including evaluating the directions of edges). A caveat is that these structures might be partially oriented graphs (i.e., situations in which the directionality of some edges cannot be effectively determined). We use the *cextend* function from the *bnlearn* package to construct a DAG that is a consistent extension of X . We then generated a consensus network based on the 100 networks after hill-climbing and selected to keep the edges that were present in graphs at least 30% of the time. Any residual undirected edges contained in the consensus network were discarded. We assessed the statistical significance of the edges within the imposed consensus network by randomly permuting the dataset 10000 times and evaluating the consensus structure on these

scrambled datasets (thus providing an estimate of the null distribution). BBN edges with a FDR of at least 5% (i.e., the edge occurred in ≥ 500 of the random BBNs) were removed from the final network.

After deriving a final consensus network structure, we performed a series of *in silico* tests to determine the importance of each gene to the network. For each of the 600 genes, we removed all incident edges (both incoming and outgoing) and recalculated the BIC of the entire network. Doing so resulted in a lower BIC, and the magnitude of the change in BIC is a measure of how important a gene is to the network. We also experimented with permuting the data corresponding to a single gene; the results for the mean change in BIC using the permutation test and removing all the incident edges did not significantly differ (Pearson's correlation > 0.999). Having derived a measure for the importance of each gene to the network, we can compare the mean change in BIC of the top five driver genes to 1000 random sets of five genes from the network.

Real-time reverse transcription quantitative PCR (RT-qPCR)

Total RNA was extracted from cells using the RNeasy Mini Kit (Qiagen, Hilden, Germany), and the RNA quality was assessed using an Agilent 2100 BioAnalyzer before reverse transcription into cDNA with Maxima™ H Minus Mastermix and following the manufacturer's instructions (Thermo Fisher Scientific, Waltham, MA, USA). RT-qPCR was performed using QuantStudio3 (Thermo Fisher Scientific, Waltham, MA, USA) 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'-GGACTCAGAGGATTGCTGCATTTAG-3', reverse 5'-CTTCGAAGTAGCTGAGTCATGCTGG-3'; and GAPDH (housekeeping gene), forward 5'-GGTGAAGGTCGGAGTCAACGGA-3' and 5'-GAGGGATCTCGCTCCTGGAAGA-3' (Integrated DNA Technologies, Coralville, IO, USA). The RT-qPCR 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 $2^{-\Delta\Delta C_t}$ 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 (ELISA)

The concentrations of soluble ADAM9 (sADAM9) and soluble MICA (sMICA) in the serum of critical and non-critical patients and healthy controls were quantified by ELISA. For soluble ADAM9, we used the Human sADAM9 DuoSet ELISA kit (R&D Systems, Minneapolis, MN, USA) 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 mg/ml and 0.2 mg/ml, respectively) and one biotinylated monoclonal mouse antibody for detection (A15-C199B9 at 60 pg/ml). Coating of MaxiSorp ELISA plates (Thermo Fisher Scientific, Waltham, MA, USA) was performed in PBS at 4°C overnight. After three washing steps with PBS, the wells were blocked with 200 µl of 10% BSA in PBS for 1 h 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 sera. 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 h, followed by incubation with 100 µl of biotinylated detection antibody for 1 h. The plates were subsequently incubated for 1 h with 100 µl of a 5000-fold dilution of streptavidin poly-HRP (Thermo Fisher Scientific, Waltham, MA, USA) per well. The reactions were finally revealed using TMB Ultra (Thermo Fisher Scientific, Waltham, MA, USA) at 100 µl/well for 15 min and stopped with 100 µl of 1 M HCl. The absorbance was measured at 450 nm.

Cell culture

Vero 76 cell lines were grown at 37 °C under 5% CO₂ and maintained in DMEM (Thermo Fisher Scientific, Waltham, MA, USA) containing 100 units/ml penicillin and supplemented with 10%

fetal bovine serum (Pan Biotech, Aidenbach, Germany). ACE2-expressing A549 cells (A549-ACE2) were grown at 37 °C under 5% CO₂ and maintained in DMEM (Thermo Fisher Scientific, Waltham, MA, USA) containing 10 µg/ml of blasticidine S (Invitrogen, Carlsbad, CA, USA).

Silencing and cell transfection

The cells were transfected with predesigned Stealth siRNA directed against ADAM9 (HSS112867) or the control Stealth RNAi Negative Control Duplex medium GC (45-55%) (Thermo Fisher Scientific, Waltham, MA, USA) using Lipofectamine™ RNAiMAX Transfection Reagent (Thermo Fisher Scientific, Waltham, MA, USA). One day prior to transfection, the cells were seeded in a 24-well plate at 0.05×10^6 cells per well. First, 1.5 µl of Lipofectamine™ RNAiMAX Transfection Reagent was added to 25 µl of Opti-MEM™ medium, followed by addition of the mix containing 5 pmoles of siRNA in 25 µl of Opti-MEM™ medium (Thermo Fisher Scientific, Waltham, MA, USA). The mixture was incubated at room temperature for 5 min and then added to the cells. The cells were collected or infected after 48 h.

Western blot

After collection and centrifugation, the cells were washed once in Dulbecco's phosphate buffered saline (D-PBS, Sigma Aldrich, Saint-Louis, MI, USA). The pellet was resuspended in 60 µl of RIPA lysis buffer (150 mM NaCl, 5 mM EDTA, 1% NP40, 50 mM Tris pH 8, 0.5% sodium deoxycholate, and 0.1% SDS) including protease inhibitors (cOmplete, Roche Diagnostics, Rotkreuz, Switzerland) and maintained on ice for 20 min. The total cellular extract was then centrifuged for 30 min at 13000 g to remove all cell debris. A Bradford assay was used for protein quantification (Bio-Rad protein Assay, Bio-Rad Laboratories, Hercules, CA, USA). For Western blotting analysis, 20 µg of total cell extract was loaded on an 8% SDS-polyacrylamide gel. After migration, the proteins were transferred onto a PVDF membrane with a semidry transfer system (Trans-Blot, Bio-Rad Laboratories, Hercules, CA, USA). The membranes were blocked for 1 h in 5% skimmed milk/TBS 0.05%/Tween 20 and then incubated with the anti-ADAM9 antibody (ab218242; Abcam, Cambridge, UK) for 2 h at 4°C in

5% BSA/TBS 0.1% Tween at 1/1000 dilution. The membrane was then incubated with the secondary antibody coupled to HRP (Bio-Rad Laboratories, Hercules, CA, USA). Bound antibodies were revealed with an enhanced chemiluminescence detection system using ChemiDoc XRS (Bio-Rad Laboratories, Hercules, CA, USA). An anti-GAPDH antibody (MAB374, Merck Millipore, Burlington, MA, USA) was used for loading control.

***In vitro* viral infections**

Vero 76 and A549-ACE2 cell lines were infected with wild-type SARS-CoV-2 virus at Multiplicities Of Infections (MOIs) of 10 and 400, respectively. The percentage of infected cells was determined by staining with SARS-CoV-2 nucleocapsid (% 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.

Flow cytometry staining

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, Irvine, CA, USA) at a 1/200 dilution in Perm/Wash™ (Becton, Dickinson and Company, Franklin Lakes, NJ, USA) for 45 min at room temperature. The antibody was then revealed by incubation with an Alexa 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.

Viral RT-qPCR

RNA was extracted from the supernatant of infected cells using the NucleoSpin Dx Virus Kit (Macherey-Nagel GmbH & Co.KG, Düren, Germany). RT-qPCR was performed using TaqPath™ 1-Step RT-qPCR Master Mix (CG) on the Quanstudio3 instrument (Thermo Fisher Scientific, Waltham, MA, USA). The primer/probe mix used for absolute quantification of the virus was N1 and N2 from the 2019-nCoV RUO Kit (Integrated DNA Technologies, Coralville, IO, USA), and the positive control for the standard curve was 2019-nCoV N Positive Control

(Integrated DNA Technologies, Coralville, IO, USA). The reaction was performed in 20 μ l, which included 5 μ l of eluted RNA, 5 μ l of TaqPath Master Mix and 1.5 μ l of the primer/probe. The RT-qPCR 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 groups comparisons, data were analyzed by unpaired, two-sided Mann-Whitney or student's *t*-test. For three or more groups comparisons, data were analyzed by unpaired, two-sided Kruskal-Wallis test, followed by Dunn's post-test; * *P* < 0.05, ** *P* < 0.01, *** *P* < 0.001, **** *P* < 0.0001. In figures and tables, "n" represents the number of biological replicates and "N" the number of times an experiment was independently performed.

Acknowledgments

We thank Drs. Benjamin Terrier, and Philippe Georgel for critical reading of this manuscript, Olivier Schwartz and his team for sharing the A549-ACE2 cell line and Christel Thauvin-Robinet and Franck Lethimonnier (AVIESAN) for providing access to the AURAGEN PFMG2025 platform. We would like to acknowledge the Cytometry Core Facility Hyperion (Brest, France) for their technical assistance as well as 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 Sweta R. Bajaj, Chen-Hao Chen, and Allison R. McLean for the bioinformatics processing of patient data, Emmanuelle Jouanguy for help with anti-IFN autoantibodies assessments, Cecilia Hellström for expert technical and bioinformatics assistance and Léa 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. Bahram) and SFRI-STRAT'US (ANR-20-SFRI-0012 to S. Bahram); 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. Bahram); the University of Strasbourg (including Initiative d'Excellence IDEX UNISTRA; to S. Bahram and R. Carapito); the European Regional Development Fund (European Union) INTERREG V program PERSONALIS (to R. Carapito and S. Bahram); the French Proteomic Infrastructure (ProFI; ANR-10-INBS-08-03; to C. Carapito); France Médecine Génomique 2025 (to D. Sanlaville); and Genuity Science. This work was partially supported by a DOE/HEP QuantISED program grant, QCCFP/Quantum Machine Learning and Quantum Computation Frameworks (QCCFP-QMLQCF) for HEP (Grant No. DE-SC0019219 to DAL).

Author Contributions

Conceptualization: RC, TWC, SB, ; Methodology: RC, RL, CC, SG, RG, RM, DL, CMo, TWC, SB, ; Software: RL, SG, RG, JML, RM, HW, NP, SM, TS, AV, DL, TWC, ; Validation: RC, RL, CC, SG, RG, RM, HW, TWC, SB, ; Formal analysis: RC, RL, CC, SG, RG, JML, RM, PB, HW, NP, SM, TS, SK, MJA, EP, AV, PLT, DL, JRG, CMo, TWC, SB, ; Investigation: RC, CC, VR, PS, Ale, AH, LM, PB, AG, QZ, AM, AP, AB, OT, SBT, AL, GL, JL, SS, Aha, CM, SK, MJA, EP, PN, AT, PLT, MR, CMo, SB, ; Resources: RC, JH, CC, RM, FD, YR, MSD, OC, TNCT, XM, OH, SFK, BG, BD, KK, JP, PMM, SK, MJA, EP, PN, AT, DS, FS, PLT, JLC, YH, DL, JRG, FM, CMo, TWC, SB, ; Data Curation: RC, RL, CC, SG, RG, JML, RM, TWC, SB, ; Writing - Original Draft: RC, RL, TWC, SB, ; Writing - Review & Editing: RC, RL, JH, CC, VR, RG, PS, RM, AH, LM, FD, YR, TNCT, AP, MS, JP, SK, MJA, EP, JS, JLC, DL, MR, JRG, CMo, TWC, SB, ; Visualization: RC, RL, RG, TWC, ; Supervision: RC, JRG, CMo, TWC, SB, ; Project

administration: RC, AM, MR, CMO, TWC, SB, ; Funding acquisition: RC, JRG, TWC, SB. All authors have read and agreed to the content of the manuscript.

Declaration of Interests

RL, RG, SG, JL, HW, JRG and TWC are employees of Genuity Science. TWC, RC and SB, are, thru their employers, named as inventors on two patent applications covering findings reported in this work. The remaining authors have no conflicts of interest to declare.

Data and materials availability

RNAseq datasets have been deposited on Gene Expression Omnibus (GEO) under the GEO with identifier GSE172114. Proteomics data have been deposited in the ProteomeXchange Consortium database with the identifier PXD025265.

TABLES

Table 1. Patients description

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
Age – median, IQR	40 [33; 46]	38 [31; 45]	41 [34; 46]	0.24
Male - n (%)	53 (73.6)	17 (68.0)	36 (76.6)	0.61
BMI (kg/m²) – median, IQR	30.0 [26.8; 35.0]	29.7 [23.8; 33.0]	30.2 [27.1; 35.6]	0.54
Time since first symptoms (days) – median, IQR	8.0 [6.0; 11.0]	9.5 [7.2; 13.5]	7.0 [6.0; 10.0]	0.08
COVID-19 treatments (during hospital stay) - n (%)				
Lopinavir/Ritonavir	21 (29.1)	3 (12.0)	18 (38.3)	0.02
Remdesivir	3 (4.1)	1 (4.0)	2 (4.2)	1.00
Hydroxychloroquine	19 (26.4)	2 (8.0)	17 (36.2)	0.01
Corticosteroids	6 (8.3)	1 (4.0)	6 (12.8)	0.25
Neurological symptoms - n (%)	26 (50.0)	10/25 (40.0)	16/27 (59.2)	0.27
Outcome - n (%)				
In-hospital and day-28-mortality	6 (8.3)	0	6 (12.8)	0.09
Characteristics of ICU patients				
		Critical Group (n=47)		
Baseline severity scores				
	SAPS II – median, IQR	38 [33; 47]		
	SOFA – median, IQR	6 [4; 9]		
ARDS - n (%)				
		45 (95.7)		
	Moderate	21 (46.7)		
	Severe	24 (53.3)		
Supportive treatments				
	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)		

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.

FIGURE LEGENDS

Figure 1. Multiomics analytical strategy

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 in DMSO/FCS until utilization for Helios mass cytometry (Maxpar Direct Immune Profiling System, Fluidigm) and whole proteomics. Plasma was used for cytokine profiling (IL-17 ELISA, V-PLEX Proinflammatory Panel and S-PLEX Human IFN- α 2a Kit, Mesoscale Discovery) 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 (PAXgene tubes, PreAnalytiX) and whole-genome sequencing (WGS). The number of treated samples per group and per omics is indicated below each omics designation. **B.** RNA-seq pipeline based on the NC vs. C comparison. The RNA-seq data were partitioned 100 times with 80% for training and the rest for testing. For each partition of the data, feature selection was performed based on differential expression and the genes that were significantly differentially expressed ($FDR \leq 0.05$) in each partition of the training data were selected for both the training and corresponding test data. Classification was performed using an ensemble computational approach with seven different algorithms. After classification and verification of the quality of the results on the test dataset, an ensemble feature ranking score across six of the seven algorithms and all 100 partitions of the data was determined. The top 600 of those features were used as the input for structural causal modeling to derive a putative causal network. **C.** Cytokines and immune cells were quantified following the manufacturer's instructions. 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 (B and C) were confirmed in a second cohort of 81 critical and 73 recovered critical patients. The differential expression of *ADAM9*, the main driver gene, was compared to publicly available bulk RNA-seq data. Finally, *ex vivo* infection experiments with SARS-CoV-2 were conducted to validate a driver gene candidate.

Figure 2. Immune profiling of healthy individuals, non-critical and critical COVID-19 patients

A. The concentrations of proinflammatory cytokines in plasma were quantified by cytokine profiling assays (V-PLEX Proinflammatory Panel and S-PLEX Human IFN- α 2a Kit, Mesoscale Discovery) or ELISA (IL-17, R&D Systems). **B.** Absolute lymphocyte counts. Each dot represents a single patient. **C.** viSNE map 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 (see Fig 1A, right panel). **D-G.** The proportions of modified lymphocyte subsets from COVID-19 patients and healthy controls as determined by mass cytometry. Proportions of T-cell subsets (**D**), B-cell subsets (**E**), dendritic cells (**F**) and nonclassical monocytes (**G**) are shown. The other cell subsets are presented in Figure S2. Each dot represents a single patient. In (A) and (D-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, *****P* < 0.0001. In (B), the *P*-value was determined by a two-tailed unpaired t-test; **P* < 0.05, ***P* < 0.01, ****P* < 0.001, *****P* < 0.0001. In (A), data are shown as box-and-whiskers plots with medians, 25th to 75th percentiles, maximal and minimal values, and include n=41 critical patients, n=24 non-critical patients and n=21 healthy controls. In (B), (D), (E-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.

Figure 3. Plasma and PBMCs proteomics of healthy individuals, non-critical and critical COVID-19 patients

A. Total number of proteins identified and used for quantification and differential analysis in the plasma of patients and healthy controls. Each dot represents a patient. Bars represent means \pm standard deviations. **B.** Multidimensional scaling plot of the normalized intensities of all patients/individuals in the three groups. **C.** Volcano plot representing the differentially

expressed proteins (DEPs) in critical versus non-critical patients. The orange dots represent the proteins that are differentially expressed with a corrected P -value < 0.05 . Proteins labeled in green and purple represent downregulated apolipoproteins and upregulated acute phase proteins, respectively. **D.** Normalized intensities of the proteins S100A8 and S100A9 in the three groups. Data are shown as box-and-whiskers plots with medians, 25th to 75th percentiles, 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; $*P < 0.05$, $** P < 0.01$, $*** P < 0.001$, $**** P < 0.0001$. **E.** Heatmap showing the expression of apolipoproteins involved in macrophage functions and acute phase proteins in the three groups. Upregulated proteins are shown in red and downregulated proteins are shown in light blue. **F.** Total number of proteins identified and used for quantification and differential analysis in PBMCs of patients and healthy controls. Each dot represents a patient. Bars represent means \pm standard deviations. **G.** Multidimensional scaling plot of the normalized intensities of all patients/individuals in the three groups. **H.** Volcano plot representing the DEPs in critical versus non-critical patients. The orange dots represent the proteins that are differentially expressed with a corrected P -value < 0.05 . Proteins labeled in green and purple are upregulated proteins involved in the regulation of blood coagulation and myeloid cell differentiation, respectively. **I.** Heatmap showing the expression of proteins involved in the regulation of blood coagulation and myeloid cell differentiation in the three groups. Upregulated proteins are shown in red and downregulated proteins are shown in light blue.

Figure 4. RNA-seq and combined omics analysis of critical patient-specific pathways

A. Volcano plot representing the differentially expressed genes in critical versus non-critical patients. The orange dots represent the genes that are differentially expressed with a corrected P -value < 0.05 . Proteins labeled in green and purple represent upregulated genes involved in blood pressure regulation and viral entry, respectively. **B.** Gene set enrichment analysis plots showing positive enrichment of inflammatory response, myeloid leukocyte activation and

neutrophil degranulation pathways. NES, normalized enrichment score. **C.** Enriched nested gene ontology (nGO) categories in critical vs. non-critical patients in RNA-seq, plasma proteomics and PBMC proteomics.

Figure 5. Integrated AI/ML and probabilistic programming of non-critical and critical COVID-19 patients

A. ROCs of the train and test sets for critical vs non-critical comparisons. All methods performed similarly. Other classification metrics are provided in Table S3. **B.** Putative network showing 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 showing the normalized gene counts of the five driver genes in critical and non-critical patients. The indicated values correspond to the FDR. Data are shown as box-and-whiskers plots with medians, 25th to 75th percentiles, maximal and minimal values, and include n=46 critical and n=23 non-critical patients.

Figure 6. Validation of ADAM9 as a key driver of viral infection and replication

A. Quantitative RT-PCR confirmation of the differential expression of ADAM9 in non-critical (n=19) vs. critical patients (n=38) and in healthy controls (n=20). **B.** Soluble ADAM9 (sADAM9) concentration in serum of healthy controls (n=15), non-critical (n=22) and critical patients (n=43) determined by ELISA. **C.** Soluble MICA concentration (sMICA) in serum of healthy controls (n=11), non-critical (n=22) and critical patients (n=43) determined by ELISA. **D.** Expression of ADAM9 according to the genotype of the eQTL rs7840270 in (n are indicated below the genotypes). **E.** Experimental approach for assessing viral uptake and viral replication in silenced Vero 76 or A549-ACE2 cells. **F.** Flow cytometry-based intracellular nucleocapsid staining in control and ADAM9-silenced Vero 76 and A549-ACE2 cells. One representative experiment of N=3 independent experiments with n=3 in each group is shown. **G.** Quantitative RT-PCR of SARS-CoV-2 in culture supernatant after the silencing of *ADAM9* in Vero 76 or A549-ACE2 cells. 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-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, **** *P* < 0.0001. In (F-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-C) and (F-G) bars represent means ± standard deviations.

Supplementary Materials and Methods

Fig. S1 to S8

Table S1 to S4

Reference (48)

Data File S1

REFERENCES

1. R. Horton, Offline: After COVID-19-is not a pandemic. *Lancet* **396**, 874 (2020).
2. 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).
3. G. G. Severe Covid, 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. Wendorff, 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. Protti, A. Aghemo, A. Lleo, A. Biondi, A. Caballero-Garralda, A. Gori, A. Tanck, A. Carreras Nolla, A. Latiano, A. L. Fracanzani, A. Peschuck, A. Julia, A. Pesenti, A. Voza, D. Jimenez, B. Mateos, B. Nafria 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. Garcia 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. Aneli, 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. Ruhlemann, M. Schaefer, M. Carrabba, M. Riveiro-Barciela, M. E. Figuera Basso, M. G. Valsecchi, M. Hernandez-Tejero, M. Acosta-Herrera, M. D'Angio, M. Baldini, M. Cazzaniga, M. Schulzky, M. Cecconi, M. Wittig, M. Ciccarelli, M. Rodriguez-Gandia, M. Bocciolone, M. Miozzo, N. Montano, N. Braun, N. Sacchi, N. Martinez, O. Ozer, O. Palmieri, P. Faverio, P. Preatoni, P. Bonfanti, P. Omodei, P. Tentorio, P. Castro, P. M. Rodrigues, A. Blandino 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. Folseraas, L. Valenti, A. Franke, T. H. Karlsen, Genomewide Association Study of Severe Covid-19 with Respiratory Failure. *N Engl J Med* **383**, 1522-1534 (2020).
4. J. F. Shelton, A. J. Shastri, C. Ye, C. H. Weldon, T. Filshtein-Sonmez, D. Coker, A. Symons, J. Esparza-Gordillo, C.-T. andMe, S. Aslibekyan, A. Auton, Trans-ancestry analysis reveals genetic and nongenetic associations with COVID-19 susceptibility and severity. *Nat Genet* **53**, 801-808 (2021).

5. 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. *Jama-J Am Med Assoc* **324**, 663-673 (2020).
6. 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. Chbihi, L. Bonnet-Madin, K. Dorgham, N. Smith, W. M. Schneider, B. S. Razooky, 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. Seeleuthner, 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. Quiros-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. Karabela, C. Rodriguez-Gallego, G. Novelli, S. Hraiech, Y. Tandjaoui-Lambiotte, X. Duval, C. Laouenan, C.-S. Clinicians, C. Clinicians, C. G. Imagine, C. C. S. G. French, V. C. C. Co, U. M. C. C.-B. Amsterdam, C. H. G. Effort, N.-U. T. C. I. 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. Maniatis, V. Soumelis, A. Amara, M. Nussenzweig, A. Garcia-Sastre, F. Krammer, A. Pujol, D. Duffy, R. P. Lifton, S. Y. Zhang, G. Gorochoy, 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**, (2020).
7. CDC, Scientific Evidence for Conditions that Increase Risk of Severe Illness | COVID-19. *Cent. Dis. Control Prev.*, (2021).
8. A. Silvin, N. Chapuis, G. Dunsmore, A. G. Goubet, A. Dubuisson, L. Derosa, C. Almiré, C. Henon, O. Kosmider, N. Droin, P. Rameau, C. Catelain, A. Alfaro, C. Dussiau, 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. Willekens, 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. Solary, Elevated Calprotectin and Abnormal Myeloid Cell Subsets Discriminate Severe from Mild COVID-19. *Cell* **182**, 1401-1418 e1418 (2020).
9. 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).
10. E. J. Giamarellos-Bourboulis, M. G. Netea, N. Rovina, K. Akinosoglou, A. Antoniadou, N. Antonakos, G. Damoraki, T. Gkavogianni, M. E. Adami, P. Katsaounou, M. Ntaganou, M. Kyriakopoulou, G. Dimopoulos, I. Koutsodimitropoulos, D. Velissaris, P. Koufargyris, A. Karageorgos, K. Katrini, V. Lekakis, M. Lupse, 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 e1003 (2020).
11. 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. Szwebel, S. H. Merklings, J. M. Treluyer, D. Veyer, L. Mouthon, C. Blanc, P. L. Tharaux, F. Rozenberg, A. Fischer, D. Duffy, F. Rieux-Laucat, 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, C. H. S. group, Type I IFN immunoprofiling in COVID-19 patients. *J Allergy Clin Immunol* **146**, 206-208 e202 (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. Hippenstiel, A. Hocke, H. Muller-Redetzky, K. M. Heim, F. Machleidt, A. Uhrig, L. Bosquillon de Jarcy, L. Jurgens, M. Stegemann, C. R. Glosenkamp, H. D. Volk, C. Goffinet, 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. Witzgenrath, C. Drosten, N. Suttrop, C. von Kalle, F. Kurth, K. Handler, J. L. Schultze, A. C. Aschenbrenner, Y. Li, J. Nattermann, B. Sawitzki, A. E. Saliba, L. E. Sander, C.-O. I. Deutsche, Severe COVID-19 Is Marked by a Dysregulated Myeloid Cell Compartment. *Cell* **182**, 1419-1440 e1423 (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. Hoepfer, I. Pink, T. Welte, M. Cornberg, M. Stoll, L. Goudeva, R. Blasczyk, A. Ganser, I. Prinz, R. Fo, C. Koenecke, C. R. Schultze-Florey, Reappearance of effector T cells is associated with recovery from COVID-19. *Ebiomedicine* **57**, (2020).
 20. J. Helms, C. Tacquard, F. Severac, I. Leonard-Lorant, M. Ohana, X. Delabranche, H. Merdji, R. Clere-Jehl, M. Schenck, F. Fagot Gandet, S. Fafi-Kremer, V. Castelain, F. Schneider, L. Grunebaum, E. Angles-Cano, L. Sattler, P. M. Mertes, F. Meziani, C. T. Group, 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, U. K. Hlh Across Speciality Collaboration, 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. Yuen, 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, C. China Medical Treatment Expert Group for, 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. Mastroiani, 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. Moldwawer, R. S. Hotchkiss, Severe immunosuppression and not a cytokine storm characterizes COVID-19 infections. *JCI Insight* **5**, (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. Bizien, 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. Rohrllich, A. Coaquette, F. Vuoto, K. Faure, N. Mahlaoui, P. Kotnik, T. Battelino, K. Trebusak Podkrajsek, K. Kisand, E. M. N. Ferre, T. DiMaggio, L. B. Rosen, P. D. Burbelo, M. McIntyre, 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**, (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.

- Bizien, S. Trouillet-Assant, K. Dobbs, A. A. de Jesus, A. Belot, A. Kallaste, E. Catherinot, Y. Tandjaoui-Lambiotte, J. Le Pen, G. Kerner, B. Bigio, Y. Seeleuthner, R. Yang, A. Bolze, A. N. Spaan, O. M. Delmonte, M. S. Abers, A. Aiuti, 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. Haerynck, 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-Laucat, 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, H. Lab, N.-U. I. R. t. C. Group, C. Clinicians, C.-S. Clinicians, C. G. Imagine, C. C. S. G. French, C. Milieu Interieur, V. C. C. Co, U. M. C. C.-B. Amsterdam, C. H. G. 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**, (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 e1105 (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 (N Y)* **2**, 99-112 e117 (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. Gabriele, L. Del Campo 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, C. Reinmun, E. 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 (N Y)* **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. Kratochvil, 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, 3rd, 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, Q. S. B. 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**, (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. Mac 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. Mathy, T. Perica, K. B. Pilla, S. J. Ganesan, D. J. Saltzberg, R. Rakesh, X. Liu, S. B. Rosenthal, L. Calviello, S. Venkataramanan, 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. Kyawe, 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**, (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**, (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. Coller, K. L. Pellegrini, D. Kazmin, G. Alaaeddine, 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. Roupael, M. Peiris, B. Pulendran, Systems biological assessment of immunity to mild versus severe COVID-19 infection in humans. *Science* **369**, 1210-+ (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. Witzernath, C. Goffinet, C. 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. Freiwald, K. Textoris-Taube, S. I. Vernardis, A. S. Egger, M. Kreidl, D. Ludwig, C. Kilian, F. Agostini, A. Zelezniak, C. Thibeault, M. Pfeiffer, S. Hippenstiel, 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. Witzernath, F. Kurth, L. E. Sander, M. Ralser, Ultra-High-Throughput Clinical Proteomics Reveals Classifiers of COVID-19 Infection. *Cell Syst* **11**, 11-24 e14 (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. Aderem, J. A. Bluestone, L. L. Lanier, P. D. Greenberg, R. Gottardo, M. M. Davis, J. D. Goldman, J. R. Heath, I.-S. C. B. Uni, Multi-Omics Resolves a Sharp Disease-State Shift between Mild and Moderate COVID-19. *Cell* **183**, 1479-+ (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**, (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**, (2020).
58. B.-C.-.-U. Brasileiras. (2021), vol. 2021.
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. C.-C. 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, C.-C. 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. Lamhamedi-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. Zuccolo, 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. Hergrueter, F. Polverino, M. E. Laucho-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. Costa Clemens, F. G. Naveca, V. Nascimento, F. Nascimento, C. Fernandes da Costa, P. C. Resende, A. Pauvolid-Correa, 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. Screaton, Antibody evasion by the P.1 strain of SARS-CoV-2. *Cell* **184**, 2939-2954 e2939 (2021).
71. S. Hicks, D. Loo, K. Sinkevicius, J. Scribner, B. Barat, N. Yoder, C. Espelin, M. Themeles, F. Chen, J. Lucas, J. Brown, B. Matin, M. Fuller, J. Lee, P. Salomon, J. Costoplus, S. Yancey, G. Diedrich, S. Gorlatov, T. Son, M. Chiechi, P. Li, M. Spliedt, V. Ciccarone, J. Hooley, N. Gantt, J. Tamura, K. Donahue, P. Moore, S. Johnson, T. Chittenden, R. Gregory, E. Bonvini, in *AACR Annual Meeting 2019*. (American Association for Cancer Research, Atlanta, GA, 2019), vol. 79, pp. 1533.
72. I. Pasteur, "Protocol: Real-time RT-PCR assays for the detection of SARS-CoV-2," (2021).

73. E. W. Deutsch, A. Csordas, Z. Sun, A. Jarnuczak, 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**, S96-104 (2002).
75. M. Kokla, J. Virtanen, M. Kolehmainen, 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 Research* **43**, e47-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, Sentieon DNaseq variant calling workflow demonstrates strong computational performance and accuracy. *Frontiers in genetics* **10**, (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 11-11 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 Biology* **11**, R25 (2010).
86. R. Tibshirani, Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)* **58**, 267-288 (1996).
87. A. E. Hoerl, R. W. Kennard, K. F. Baldwin, Ridge regression: some simulations. *Communications in Statistics-Theory and Methods* **4**, 105-123 (1975).
88. A. E. Hoerl, R. W. Kennard, Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics* **12**, 55-67 (1970).
89. M. Kuhn, Building predictive models in R using the caret package. *Journal of Statistical Software* **28**, 1-26 (2008).
90. 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.
91. C. Cortes, V. Vapnik, Support-vector networks. *Mach Learn* **20**, 273-297 (1995).
92. L. Breiman, Random forests. *Mach Learn* **45**, 5-32 (2001).
93. L. Breiman, J. Friedman, C. J. Stone, R. A. Olshen, *Classification and regression trees*. (Chapman & Hall, 1993), pp. 358-358.

94. T. Q. Chen, C. Guestrin, XGBoost: A Scalable Tree Boosting System. *Kdd'16: Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining*, 785-794 (2016).
95. D. Willsch, M. Willsch, H. De Raedt, K. Michielsen, Support vector machines on the D-Wave quantum annealer. *Computer Physics Communications* **248**, 107006 (2020).
96. M. W. Johnson, M. H. Amin, S. Gildert, T. Lanting, F. Hamze, N. Dickson, R. Harris, A. J. Berkley, J. Johansson, P. Bunyk, Quantum annealing with manufactured spins. *Nature* **473**, 194-198 (2011).
97. V. Choi, Minor-embedding in adiabatic quantum computation: I. The parameter setting problem. *Quantum Inf Process* **7**, 193-209 (2008).
98. 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**, (2018).
99. 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).
100. N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov, Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research* **15**, 1929-1958 (2014).
101. I. Sutskever, J. Martens, G. Dahl, G. Hinton, in *International conference on machine learning*. (2013), pp. 1139-1147.
102. 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)*. (2016), pp. 265-283.
103. W. Lam, F. Bacchus, Learning Bayesian belief networks: An approach based on the MDL principle. *Computational intelligence* **10**, 269-293 (1994).
104. M. Scutari, Learning Bayesian Networks with the bnlearn R Package. *2010* **35**, 22 (2010).