DISSERTATION

Nutzung des Plasma-Proteoms zur Klassifizierung und Prognose von Patient:innen mit COVID-19

Leveraging the plasma proteome for classification and prognosis of patients with COVID-19

zur Erlangung des akademischen Grades

Doctor medicinae (Dr. med.)

vorgelegt der Medizinischen Fakultät
Charité – Universitätsmedizin Berlin

von

Pinkus Tober-Lau

Erstbetreuung: Prof. Dr. Florian Kurth

Datum der Promotion: 30.06.2024

Inhaltsverzeichnis

Tabellenverzeichnis

Abbildungsverzeichnis

Abkürzungsverzeichnis

95%-CI	95 %-Konfidenzintervall
ABCS	Age, Biomarkers, Clinical History, Sex Score
APACHE II	Acute Physiology And Chronic Health Evaluation II Score
APP	Akutphaseprotein
AUROC	Fläche unter der Receiver Operating Characteristic (ROC) -Kurve
BMI	Body-Mass-Index
CCI	Charlson-Komorbiditätsindex
COVID-19	Coronavirus Disease 2019
DIA	datenunabhängige Aufnahme (Data Independent Acquisition)
DNA	Desoxyribonukleinsäure
DNI/DNR	Nicht intubieren/nicht reanimieren (Do Not Intubate/Do Not Resuscitate)
DNN	tiefe neuronale Netzwerke (Deep Neural Networks)
ECMO	Extrakorporale Membranoxygenierung
EMA	Europäische Arzneimittel-Agentur (European Medicines Agency)
ESI	Elektrospray-Ionisation
FDA	US-Behörde für Lebens- und Arzneimittel (Food and Drug Administration)
HRAM	hochauflösender exakter Massenanalysatoren (High-Resolution Accurate Mass Analyzer)
IQR	Interquartilspanne
LASSO	Least Absolute Shrinkage and Selection Operator
LC	Flüssigkeitschromatographie bzw. -graph (Liquid Chromatography)
LLOQ	untere Bestimmungsgrenze (Lower Limit of Quantification)
ML	maschinelles Lernen (Machine Learning)
MRM	Multiple Reaction Monitoring; auch: SRM
MS	Massenspektrometrie
MS_1, MS_2	erster bzw. zweiter Massenfilter
MS/MS	Tandem-Massenspektrometrie
m/z	Masse-zu-Ladung-Verhältnis
OR	Chancenverhältnis (Odds Ratio)
PCR	Polymerase-Kettenreaktion

PHEIC	Gesundheitliche Notlage internationaler Tragweite (Public Health Emergency of International Concern)
QqQ	Triple-Quadrupol
PQP	Peptid-Suchparameter (Peptide Query Parameter)
RNA	Ribonukleinsäure
RPC	Umkehrphasen-Chromatographie (Reversed Phase Chromatography)
SARS-CoV-2	Schweres-akutes-Atemwegssyndrom-Coronavirus Typ 2 (Severe Acute Respiratory Syndrome Corona Virus 2)
SOFA	Sequential Organ Failure Assessment Score
SRM	Selective Reaction Monitoring; auch: MRM
SVM	Support Vector Machine
SWATH	Sequential Windowed Acquisition of All Theoretical Fragment Ion
TOF	Flugzeit (Time of Flight)
UHPLC	Ultra-Hochdruck-Flüssigkeitschromatographie (Ultra-High Pressure LC)
WHO	Weltgesundheitsorganisation (World Health Organization)

Zusammenfassung

Die Infektion mit SARS-CoV-2 weist ein hochvariables Krankheitsbild auf, von asymptomatischen Verläufen über Symptome eines grippalen Infekts bis hin zu respiratorischem Versagen und Tod. Höheres Alter und bestimmte Vorerkrankungen sind Risikofaktoren für schwere Krankheitsverläufe, dennoch stellte die (frühe) klinische Risikostratifizierung angesichts des variablen Krankheitsbildes eine besondere Herausforderung dar. Daher galt es, schnellstmöglich die Pathophysiologie von COVID-19 zu entschlüsseln und eine akkurate Klassifikation der Krankheitsschwere sowie Outcome-Prognose zu ermöglichen.

Die Tandem-Massenspektrometrie (MS/MS) -basierte Plasma-Proteomik ermöglicht die umfassende Charakterisierung der Wirtsantwort auf einen Erreger und erlaubt Rückschlüsse auf die zugrundeliegende Pathophysiologie. Im Rahmen der Pa-COVID-19-Studie analysierten wir insgesamt 881 Plasma-Proteome von 280 Patient:innen, ergänzt um umfangreiche Daten zu Demografie, Krankheitsschwere sowie 85 Routinelaborparameter.

Wir erfassten von 139 Patient:innen zu 687 longitudinalen Messzeitpunkten die ungerichteten Plasma-Proteome und quantifizierten 321 Proteingruppen, davon 189 in >99 % der Proben. Wir identifizierten vielfältige Korrelationen von proteomischen Akutphaseproteinen mit klinischen Laborparametern von Inflammation und Endorganschädigung (bspw. Kreatinin oder NT-proBNP). 113 Proteine und 55 Routinelaborparameter korrelierten mit der Krankheitsschwere. 20 Proteine, überwiegend Marker von Inflammation und Blutgerinnung, waren altersabhängig dysreguliert und bieten Einblick in die altersspezifischen Pathogenese von COVID-19.

Als nächstes untersuchten wir das prognostische Potenzial der Plasma-Proteomik mittels verschiedener Machine Learning-Modelle. Basierend auf proteomischen und klinischen Messungen, konnten wir akkurat die aktuelle Krankheitsschwere prädizieren. Bei kritisch kranken Patient:innen gelang es uns, anhand eines einzelnen frühen Messzeitpunktes deren Überleben bzw. Versterben zu prognostizieren, im Median 39 Tage vor dem Outcome (AUROC = 0,81); die Validierung des Modells an einer komplett unabhängigen Kohorte ergab eine fast perfekte Klassifikation (AUROC = 1,0).

Um einen klinischen Nutzen aus diesen Ergebnissen zu ziehen, identifizierten wir 50 relevante Peptide und entwickelten einen Multiple Reaction Monitoring (MRM) -basierten Peptid-Panelassay, den wir an zwei weiteren Kohorten, mit 30 (Citrat-Plasma) bzw. 164 (EDTA-Plasma) Patient:innen, validierten und der zur Anwendung in Routinelaboren geeignet ist.

Es gelang uns somit, mittels MS-basierter Proteomik relevante pathophysiologische Grundlagen von COVID-19 zu identifizieren und eine akkurate Klassifikation und Prognose von Patient:innen vorzunehmen. Die Ergebnisse der Studie sollten nun an größeren Kohorten untersucht werden, parallel erweitern wir das Peptid-Panelassay zur Anwendung auf andere Infektionserkrankungen wie Mpox, bakterielle Pneumonie und Malaria.

Abstract

Infection with SARS-CoV-2 presents with a highly variable clinical picture, ranging from asymptomatic cases over flu-like symptoms to severe organ damage, respiratory failure, and death. Higher age and different preexisting medical conditions are risk factors for severe disease, but early clinical risk stratification posed a special challenge due to the high variability in clinical presentation. Thus, it was crucial to quickly identify the pathophysiology of COVID-19 and enable accurate classification of disease severity and outcome prognosis.

Tandem mass spectrometry (MS/MS)-based plasma proteomics allows for a comprehensive characterization of the host response to a pathogen, enabling inference about the underlying pathophysiology. Within the Pa-COVID-19 study, we analyzed a total of 881 plasma proteomes from 280 patients, complemented by extensive data on demographics, disease severity, and 85 routine laboratory parameters.

We collected unbiased plasma proteomes from 139 patients at 687 longitudinal sampling timepoints and quantified a total of 321 protein groups, 189 of which were present in >99% of samples. We identified various correlations between acute-phase proteins and clinical laboratory markers of inflammation and organ damage (e.g., creatinine or NT-proBNP). 113 proteins and 55 routine laboratory markers correlated with disease severity. Several proteins, mainly indicating inflammation and coagulation, were differentially expressed depending on age, offering insight into the age-dependent pathogenesis of severe COVID-19.

Next, we examined the prognostic potential of plasma proteomics using different machine learning models. Based on proteomic and clinical measurements, we accurately predicted current disease severity. For critically ill patients, we were able to predict their survival or death based on a single measurement at an early timepoint, with a median of 39 days before the outcome (AUROC = 0.81). Validation of the model on a completely independent cohort yielded almost perfect classification (AUROC = 1.0).

To translate these results to the clinical routine, we identified 50 relevant peptides and developed a multiple reaction monitoring (MRM)-based peptide panel assay, which we

validated in two additional cohorts of 30 (citrate plasma) and 164 (EDTA plasma) patients, and which can be applied in routine laboratories.

In summary, applying MS-based proteomics we successfully identified relevant pathophysiological mechanisms of COVID-19, allowing for accurate disease severity classification and outcome prognosis. These results should now be validated in larger cohorts, simultaneously we are expanding the peptide panel assay for use in other infectious diseases including mpox, bacterial pneumonia, and malaria.

1 Einleitung

1.1 COVID-19

1.1.1 Hintergrund und Klinik

Im Dezember 2019 gelangten die ersten Berichte über Fälle einer neuartigen Pneumonie, verursacht durch einen unbekannten Erreger, aus Wuhan, Provinz Hubei, China, in die Welt [1]. Die folgende Pandemie führte global zu einer beispiellosen gesellschaftlichen, ökonomischen und politischen Herausforderung [2–4]. Nur wenige Wochen nach den initialen Berichten war der Auslöser der sich inzwischen rasch ausbreitenden Krankheit identifiziert, ein Betacoronavirus und naher Verwandter des Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV), der dementsprechend von der Weltgesundheitsorganisation (WHO) als SARS-CoV-2 bezeichnet wurde [5]; die Krankheit erhielt den Namen COVID-19 (**Co**rona**vi**rus **D**isease 20**19**)[6]. Innerhalb weniger Wochen verbreitete sich das neuartige Coronavirus auch außerhalb Chinas und infizierte bald Millionen von Menschen. Am 30.01.2020 erklärte die WHO den Ausbruch zur Gesundheitlichen Notlage Internationaler Tragweite (Public Health Emergency of International Concern, PHEIC) und sah sich nur wenige Wochen später gezwungen, am 11.03.2020 eine Pandemie auszurufen [7,8]. Schätzungen gehen davon aus, dass sich bis zum 5. Mai 2023, als die WHO die pandemische Notlage offiziell für beendet erklärte [9], über 675 Millionen Menschen infiziert hatten, von denen über 6,8 Millionen verstarben [10].

Während initial die Tröpfcheninfektion als Hauptübertragungsweg erachtet wurde, konnte bald bestätigt werden, dass eine Übertragung auch über Aerosole erfolgt und diese insbesondere in geschlossenen, schlecht belüfteten Räumen teils Stunden infektiös sind [11]. Die Infektion mit SARS-CoV-2 weist ein hochvariables Krankheitsspektrum auf, von asymptomatischen Überträgern über milde Verläufe mit Symptomen eines grippalen Infekts bis hin zu respiratorischem Versagen und Tod [12,13]. Frühe Schätzungen gingen von einem Manifestationsindex, also dem Anteil symptomatischer an allen Infizierten, von 55-85 % aus, wobei die mittlere Inkubationszeit auf ca. 6 Tage, die 95 %-Perzentile auf 12 Tage geschätzt wurde [13]. Mit dem Auftreten neuer SARS-CoV-2-Varianten ab Anfang 2021 wurden tendenziell etwas kürzere Inkubationszeiten beobachtet.

Zu Beginn der Erkrankung dominieren Symptome des oberen Respirationstrakts, u.a. Husten und Atemnot, Schnupfen sowie Geruchs- und Geschmacksverlust, ggf. in Kombination mit Allgemeinsymptomen wie Abgeschlagenheit, Kopf- und Gliederschmerzen und Fieber [13]. Bei den meisten Erkrankten kommt es innerhalb der ersten Woche zur Symptomresolution, wobei verschiedenartige Spätfolgen, zusammengefasst als Long-COVID oder Post-COVID-19-Syndrom bezeichnet, auftreten können [14].

Bei einem Teil der Erkrankten kann es im weiteren Krankheitsverlauf zu einer ausgeprägten Beteiligung der unteren Atemwege und weiterer Organsysteme, darunter Herz, Nieren und Gehirn, kommen, mit potenziell tödlichem Verlauf [13]. Als besondere Herausforderung für das klinische Management zeigte sich, dass Erkrankte in dieser Krankheitsphase bei ärztlicher Vorstellung oftmals trotz subjektiv nur leichter Symptomatik bereits eine schwere Hypoxämie aufwiesen und innerhalb weniger Stunden eine invasive Beatmung benötigten, ein Phänomen, das auch gelegentlich als „happy hypoxaemia“ (fröhliche Hypoxämie) bezeichnet wurde [15].

Bis Mai 2021, also vor einem maßgeblich durch Immunität und möglicherweise milderen SARS-CoV-2-Varianten veränderten Infektionsgeschehen, bedurften in Deutschland ca. 10 % der Fälle einer stationären Krankenhausbehandlung, 33 % davon auf einer Intensivstation. Insgesamt verstarben 3,1 % der gemeldeten Fälle, darunter 26 % der stationär behandelten Patient:innen [16]. Die große Anzahl von Erkrankten, die einer stationären Behandlung bedurften, führte weltweit wiederholt zu Überlastungen von Gesundheitssystemen, mit Aufnahmestopps in Krankenhäusern, überlasteten Rettungsdiensten und unzureichenden Kapazitäten zum Umgang mit Verstorbenen [17,18].

Während sich insbesondere höheres Alter, aber auch Vorerkrankungen wie Diabetes mellitus, Adipositas und Arteriosklerose als relevante Risikofaktoren für einen schweren Krankheitsverlauf abzeichneten [19–21], blieb die frühe klinische Risikostratifizierung nicht zuletzt aufgrund des so variablen Krankheitsbildes herausfordernd [22]. Altbewährte klinische Risiko-Scores boten bei dieser Erkrankung nur einen begrenzten prognostischen Nutzen [23], gleichzeitig bestand vor dem Hintergrund einer drohenden Überlastung des Gesundheitssystems, mit ausgeschöpften normal- und intensivstationären Behandlungskapazitäten, wie sie u.a. aus Bergamo (Italien) [17] und New York City (NY, USA) [18] berichtet wurden, ein dringender Bedarf, die klinische Prognose von

Patient:innen mit COVID-19 frühzeitig möglichst präzise im Sinne einer bedarfsorientierten, optimalen Ressourcenallokation zu bestimmen.

1.1.2 Bedeutung für die Wissenschaft

Wohl keine andere Krankheit ist jemals zuvor Gegenstand so rascher und umfangreicher Forschung geworden wie COVID-19 [24,25]. Mit den ersten Berichten aus Wuhan galt es, schnellstmöglich die Übertragungswege und Pathophysiologie der Erkrankung zu entschlüsseln, um zielgerichtete Infektionspräventionsmaßnahmen durchzuführen und wirksame Therapien sowie sichere und effektive Impfstoffe zu entwickeln. Die Welt sah eine außergewöhnliche nationale und internationale Kooperationsbereitschaft zwischen Forschenden, ermöglicht durch eine umfangreiche Bereitstellung von Forschungsmitteln durch Regierungen und Nichtregierungsorganisationen [25,26]. Binnen weniger Wochen wurde das gesamte SARS-CoV-2-Genom entschlüsselt und veröffentlicht [27], sodass bereits im ersten Quartal 2020 mit der Entwicklung von Impfstoffen gegen COVID-19 begonnen und nach wirksamen Virostatika und unterstützenden Therapien gesucht werden konnte [28].

Frühe Fallberichte und Beobachtungsstudien bestätigten in der klinischen Routine etablierten Biomarkern wie C-reaktives Protein (CRP), D-Dimeren und Leukozytenzahlen einen prognostischen Wert, und einzeln oder in Kombination u.a. mit Vitalparametern wurden erste Risikoscores postuliert [29–32]. Charakteristische Milchglasinfiltrate in der Computertomographie boten bei limitierten PCR-Testkapazitäten eine initiale diagnostische Einschätzungsmöglichkeit [33], Ausmaß und Kombination mit Konsolidierungen korrelierten mit Krankheitsschwere und Outcome [34]. Insgesamt hatten jedoch sowohl etablierte Risikoscores wie SOFA (Sequential Organ Failure Assessment [35]) oder APACHE II (Acute Physiology And Chronic Health Evaluation II [36]), als auch später spezifisch für COVID-19 entwickelte Scores wie ABCS (Age, Biomarkers, Clinical History, Sex [37]) oder 4C-Mortality Score [38], nur eine eingeschränkte Aussagekraft [23,32,37,38].

Um diesem dringenden klinischen Bedarf zu begegnen, galt es daher, schnellstmöglich die Grundlagen der Pathophysiologie von COVID-19, insbesondere der schweren Verlaufsformen, zu entschlüsseln, um einerseits die präzise Klassifikation und Prognose der

Erkrankung zu ermöglichen und andererseits neue Therapieansätze zu entwickeln und in die klinische Praxis zu translatieren [28,39]. Eine in den letzten Jahren zunehmend weiterentwickelte Technologie war und ist hier besonders geeignet, umfangreiche physiologische und pathophysiologische Prozesse abzubilden und eine umfassende Datengrundlage für systemische biologische Prozesse zu liefern: die Plasma-Proteomik.

1.2 Das Plasma-Proteom

1.2.1 Hintergrund

Das Proteom bezeichnet die Gesamtheit aller Proteine bzw. einer definierten Untereinheit eines Organismus, die zu einem definierten Zeitpunkt vorhanden sind. Das Ziel der Proteomik ist die „*umfassende, quantitative Beschreibung der Proteinexpression und ihrer Veränderungen unter dem Einfluss biologischer Störungen wie Krankheit oder medikamentöser Behandlung*“ (Übersetzung durch den Autor[1]) [40]. Während sich die Genomik und Transkriptomik mit der Gesamtheit der kodierten bzw. transkribierten Gene befassen, untersucht die Proteomik deren biologisch funktionelle Produkte, die Proteine, und erlaubt somit direkte Rückschlüsse auf die tatsächlich zu einem gegebenen Zeitpunkt ablaufenden biologischen Prozesse [41] (Abbildung 1).

Frühe Untersuchungen des Proteoms erfolgten mittels sogenannter 2-dimensionaler Gelelektrophorese (2-DE). Dabei werden Proteine zunächst anhand ihres spezifischen isoelektrischen Punktes sortiert und im Anschluss mittels „klassischer“ Elektrophorese entsprechend ihres Molekulargewichts aufgetrennt. Auf diese Weise ließen sich zwei Proben direkt miteinander vergleichen und einzelne hoch- oder herunterregulierte Proteine identifizieren. Dieses Verfahren ist jedoch aufwendig und ermöglicht nur eine geringe Tiefenauflösung, also dynamischen Konzentrationsbereich von Zielproteinen. Bei komplexen Proben wie bspw. humanem Plasma, in dem zwischen den am niedrigsten und den am höchsten abundanten Proteinen 7-12 Größenordnungen liegen [42], wird daher nur eine geringe Zahl verschiedener Proteine erfasst. Bereits um die Jahrtausendwende setzten sich daher Massenspektrometrie (MS) -basierte Verfahren durch und erfuhren eine rasche Weiterentwicklung [43].

[1]Original: “[…] comprehensive, quantitative description of protein expression and its changes under the influence of biological perturbations such as disease or drug treatment.” Anderson & Anderson 1998 [40]

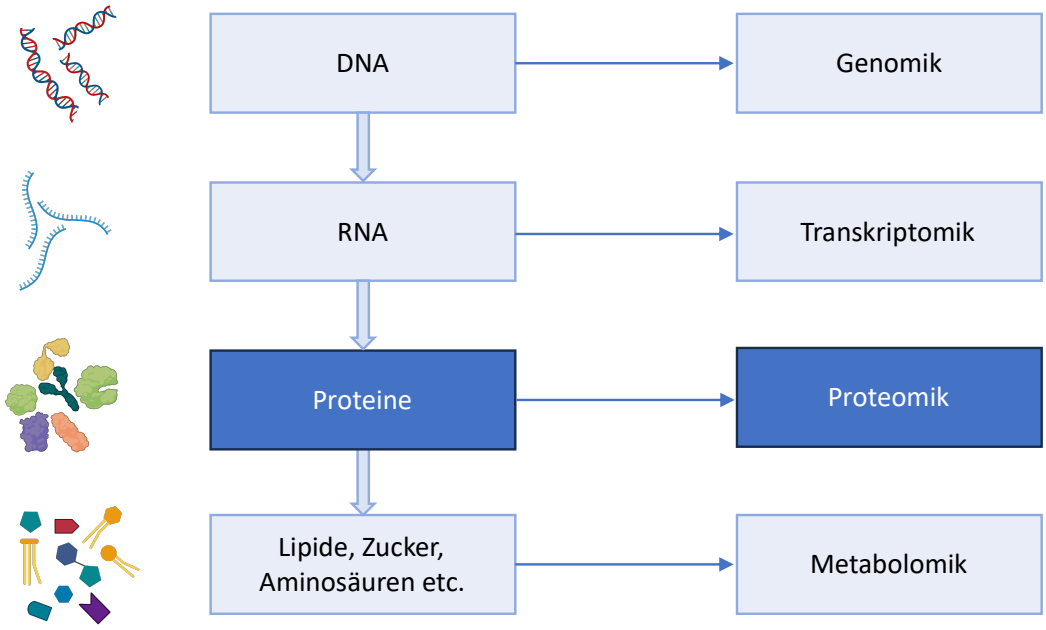


Abbildung 1: Einordnung der Proteomik in die -omics Wissenschaften. Die Genomik befasst sich mit der Gesamtheit der Gene (DNA-Ebene), die Transkriptomik mit der Gesamtheit der transkribierten Gene (RNA-Ebene) und die Proteomik mit den biologisch funktionellen Produkten, Proteinen, und erlaubt somit direkte Rückschlüsse auf die tatsächlich zu einem gegebenen Zeitpunkt ablaufenden biologischen Prozesse. Die Metabolomik schließlich befasst sich mit den Stoffwechselprodukten und -prozessen eines Organismus. Eigene Darstellung.

1.2.2 Flüssigkeitschromatographie – Massenspektrometrie (LC-MS) Proteomik

Zur Analyse komplexer Proben wie komplette Zellen oder Blutplasma, die eine breite dynamische Konzentrationsspanne von Proteinen sowie potenziell überlappende MS-Signale aufweisen, wird heutzutage in der Regel Flüssigkeitschromatographie (liquid chromatography, LC) mit Tandem-Massenspektrometrie (tandem mass spectrometry, MS/MS) kombiniert (Abbildung 2). Aufgrund der hohen Anzahl verschiedener Proteine erfolgt die Messung entsprechend dem sogenannten „bottom-up“ Ansatz [43]. Fortschritte in der Genomik haben ermöglicht, basierend auf genetischen Informationen Peptide und Proteine zu prädizieren; daher erfolgt zunächst die Messung einzelner Peptide, die nachfolgend mit theoretischen oder empirischen Proteindatenbanken abgeglichen werden. Theoretisch ist ein proteotypisches Peptid ausreichend, um das Vorhandensein eines bestimmten Proteins sicher nachzuweisen, wobei die Quantifizierung in der Praxis meist über mindestens zwei Peptide erfolgt. Die Messung mittels LC-MS/MS umfasst folgende Schritte:

1.2.2.1 Probenvorbereitung

Zunächst werden die Proteine einer Probe mittels spezifischer Proteasen verdaut, d.h. in Peptide aufgespalten. Meist wird dazu Trypsin genutzt: Aufgrund der spezifischen Spaltung an Carboxylenden von Lysin- und Argininresten (außer wenn Prolin folgt), können prädizierte Peptide definiert und entsprechende theoretische Datenbanken erstellt werden [43,44].

1.2.2.2 Flüssigkeitschromatographie

Die durch enzymatische Verdauung entstandenen Peptide werden nun mittels Flüssigkeitschromatographie entsprechend ihrer Moleküleigenschaften aufgetrennt (Abbildung 2, oberes Panel). In der Proteomik wird meist eine sog. „reverse phase“ Chromatographie (RPC) mit einer hydrophoben stationären Phase angewandt [43,44]. Die zu untersuchenden Peptide werden in einem Puffer gelöst und dieser in die Flüssigkeitssäule gegeben. Dabei binden hydrophobe Peptide stärker an die stationäre Phase und werden entsprechend stark retardiert, während hydrophile Peptide weniger Interaktionen mit dem Säulenmaterial haben. Im Anschluss wird die Säule mit einem Gradienten aus Lösemittel und Puffer eluiert und dadurch die Peptide allmählich von der stationären Phase gelöst. Somit verlassen verschiedene Peptide mit ähnlicher Molekularmasse, jedoch unterschiedlichen hydrophobischen und isoelektrischen Eigenschaften, die Säule des Chromatographen zu verschiedenen Zeitpunkten (Retentionszeit) [43,44]. Mittels Verbindungsstück (Interface) wird die nun zeitlich aufgetrennte flüssige Phase dem Massenspektrometer zugeführt.

1.2.2.3 Massenspektrometrie

Grundsätzlich umfasst ein Massenspektrometer eine Ionenquelle, einen Massenanalysator, der das Masse-zu-Ladung-Verhältnis (*m/z*) misst, und einen Detektor, der die Anzahl von Ionen bei jedem *m/z* Wert erfasst [44,45] (Abbildung 2, mittleres Panel). Die mittels LC aufgetrennte Probe wird im MS zunächst vernebelt und die Peptide mittels Elektrospray-Ionisation (ESI) ionisiert. Auf dem Weg durch den Massenspektrometer werden die nun geladenen Peptide, abhängig von dem verwendeten Gerät, entsprechend ihres *m/z*-Verhältnisses aufgetrennt und detektiert.

Bei der Tandem-Massenspektrometrie (MS/MS) werden in einem ersten Massenfilter (MS_1) spezifisch erfasste Vorläufer-Ionen in einer Kollisionszelle durch Zuführung von Kollisionsgas (z.B. Stickstoff) fragmentiert; die daraus entstehenden Produkt-Ionen passieren einen zweiten Massenfilter oder -analysator (MS_2) und werden von einem Detektor erfasst (Abbildung 2, unteres Panel). Häufig werden Quadrupole als Massenfilter und Fragmentierungszellen eingesetzt. Als MS_2 können, anstatt eines dritten Quadrupols (QqQ-MS), hochauflösende exakte Massenanalysatoren (HRAM, high-resolution accurate mass analyzers) wie Time-of-Flight (TOF-MS) oder Ionenfallen (z.B. Orbitrap) genutzt werden und so die datenunabhängige Detektion einer größeren Anzahl verschiedener Produkt-Ionen erfolgen. Durch die kombinierte Erfassung von Vorläufer- und Produkt-Ionen kann eine höhere Ionen-Spezifität erreicht und ein größerer dynamischer Konzentrationsbereich abgedeckt werden [44].

Im Anschluss an die LC-MS/MS erfolgt die primäre Analyse der Messungen, für die es grundsätzlich die folgenden zwei Ansätze gibt:

1.2.3 Ungerichtete (untargeted) Proteomik

Erfolgt die Messung einer Probe ohne die vorherige Festlegung zu untersuchender Ziel-peptide, spricht man von sogenannter Shotgun- (Schrotschuss-), ungerichteter (untargeted) oder auch Such- (Discovery-) Proteomik; so kann eine unvoreingenommene Analyse bspw. des Plasma-Proteoms erfolgen und neue potenzielle Proteinbiomarker identifiziert werden. Dies kann durch die kontinuierliche Aufzeichnung einer großen Anzahl von MS_1 und MS_2 Spektren mittels TOF- oder Orbitrap-MS erfolgen, bspw. im Rahmen von sog. DIA-SWATH-MS (Data Independent Acquisition - Sequential Windowed Acquisition of All Theoretical Fragment Ion - Mass Spectra) [46]. Die so entstandenen Massenspektren werden mit Datenbanken abgeglichen und entsprechenden Peptiden zugeordnet, aus denen wiederum spezifische Proteine abgeleitet werden. Mittels bioinformatischer Methoden können Rückschlüsse auf die Abundanz verschiedener Proteine geschlossen und diese so ins Verhältnis zueinander gesetzt werden.

1.2.4 Gerichtete (targeted) Proteomik

Im Gegensatz dazu erfolgt bei der gerichteten (targeted) Proteomik die Messung definierter Ziel-Peptide, die bspw. im Rahmen von Discovery-Proteomik identifiziert wurden.

Beim Selected-Reaction-Monitoring (SRM, auch: Multiple Reaction Monitoring, MRM) wird bei MS_1 lediglich ein vorher definiertes Vorläufer-Ion (i.d.R. zwischen 10-20 Aminosäuren lang) zur weiteren Fragmentierung zugelassen; die resultierenden Produkt-Ionen können im MS_2 sehr sensitiv und spezifisch erfasst und mittels Isotopen-markierter Peptidstandards exakt quantifiziert werden [47]. Im Vergleich zur ungerichteten Proteomik ist die MRM-Proteomik also für die definierten Proteine sensitiver und spezifischer, dafür in der Anzahl verschiedener Proteine limitiert.

1.2.5 Bioinformatische Analysetechniken

Um aus den umfangreichen massenspektrometrischen Rohdaten einen statistisch auswertbaren Datensatz zu generieren, sind verschiedene Schritte nötig (Abbildung 2, unteres Panel). Prinzipiell müssen die gemessenen Chromatogramme und Massenspektren spezifischen Peptiden zugeordnet und im nächsten Schritt die ihnen entsprechenden Proteine abgeleitet werden, um schlussendlich deren relative oder absolute Abundanz funktionell analysieren zu können [46,48,49].

Die Zuordnung von Massenspektren zu Peptiden kann grundsätzlich Spektren-zentriert oder Peptid-zentriert erfolgen. Bei Spektren-zentrierten Ansätzen werden zunächst ähnliche Spektren einer analysierten Probe gruppiert und dann mit einer entsprechenden Datenbank abgeglichen und Peptiden zugeordnet. Bei Peptid-zentrierten Ansätzen hingegen werden zunächst die in einer spezifischen Probe erwarteten Peptide und deren zugehörigen MS-Daten (Peptide Query Parameters (PQP): Retentionszeit, Vorläufer- und Produkt-Ion Massen und Signal-Intensitäten [Peaks]) definiert und dann mit den tatsächlich gemessenen Daten abgeglichen [46]. Neue Software-Tools kombinieren die Vorteile beider Ansätze – datenunabhängige Analyse bzw. größere Spezifität – bspw. unter Anwendung selbstlernender Algorithmen und tiefer neuronaler Netze (vgl. Demichev et al., 2020 [48]).

Aus den identifizierten Peptiden werden mittels Datenbankabgleichs die zugehörigen Proteine abgeleitet und, bspw. durch die teils gewichtete Aufsummierung der integrierten Peak-Flächen mehrerer Produkt- bzw. Vorläufer-Ionen, quantifiziert. Bevor die funktionelle Analyse erfolgen kann, müssen Normalisierungsschritte durchgeführt und ggf. feh-

lende Peptiddaten imputiert werden, ohne tatsächliche biologische Informationen zu verlieren. Sowohl bei der Präprozessierung als auch der funktionellen Auswertung des fertigen Datensatzes kommen zunehmend auch Machine Learning-Techniken zum Einsatz [48,50].

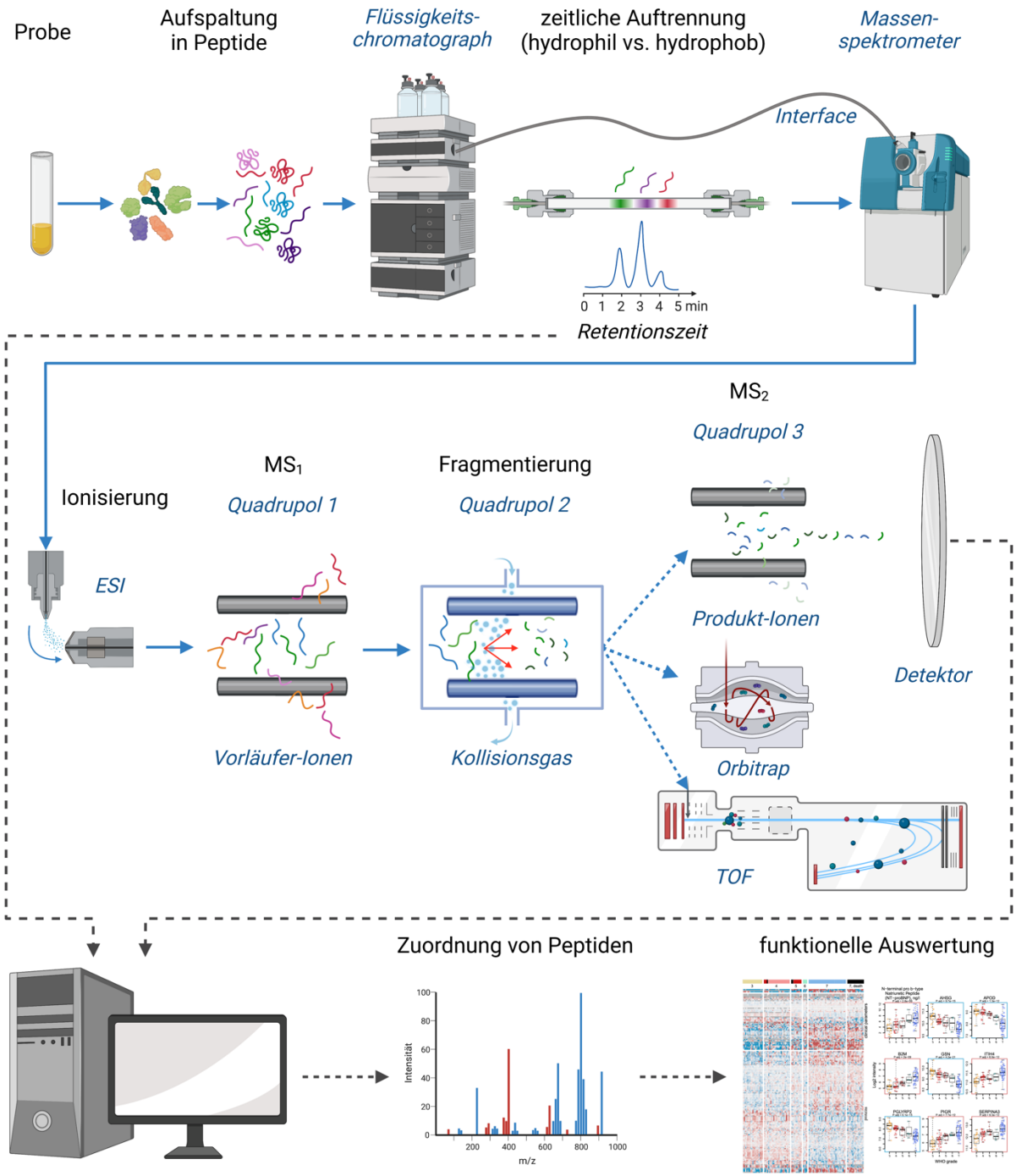


Abbildung 2: Schematischer Aufbau eines LC-MS/MS Systems. Die Proteine werden zunächst mittels Trypsin in Peptide aufgespalten. Im Anschluss erfolgt die Auftrennung der Peptide mittels (reverser) Flüssigkeitschromatographie (LC) entsprechend ihrer hydrophoben und isoelektrischen Eigenschaften. Die so zeitlich aufgetrennte Probe gelangt via Verbindungsstück (Interface)

in den Massenspektrometer, wird vernebelt und die Peptide mittels Elektrospray-Ionisation (ESI) ionisiert. Bei der Tandem-Massenspektrometrie werden im ersten Massenfilter (MS_1, Quadrupol 1) Vorläufer-Ionen selektiert, in einer Kollisionszelle durch Zuführung von Kollisionsgas spezifisch fragmentiert (Quadrupol 2) und die Produkt-Ionen im Anschluss detektiert, wobei als zweiter Massenfilter (MS_2) entweder ein dritter Quadrupol oder ein hochauflösender exakter Massenanalysator (Orbitrap oder Time-of-Flight [TOF]) eingesetzt werden kann. Es folgt die bioinformatische Aufbereitung, Zuordnung von Spektren zu Peptiden und Generierung eines Datensatzes zur statistischen Auswertung. Abbildung angelehnt an https ://healthcare-in-europe.com/de/news/massenspektrometrie-analytik-trifft-diagnostik.html.

1.3 Fragestellung

Die COVID-19-Pandemie hat Gesundheitssysteme weltweit vor besondere Herausforderungen gestellt. Insbesondere vor dem Hintergrund des hohen Patient:innenaufkommens während der verschiedenen Infektionswellen, mit der Gefahr einer Überlastung von Versorgungskapazitäten, war und ist eine rasche Identifikation von prognostischen Parametern von hoher Relevanz, um Ressourcen zielgerichtet einzusetzen und Patient:innen einer adäquate medizinische Versorgung zuzuführen. Mittels Plasma-Proteomik lässt sich innerhalb kurzer Zeit ein umfassendes Abbild des (patho-) physiologischen Ist-Zustandes von Patient:innen erfassen, auf dessen Basis neue Biomarker identifiziert und potenziell prognostische Modelle generiert werden können.

Im Rahmen der vorliegenden Arbeit soll untersucht werden, inwiefern das Plasma-Proteom zur Klassifikation von Krankheitsschwere und Prognose des Outcomes bei Patient:innen mit COVID-19 geeignet ist. Dazu werden folgende Teilaspekte analysiert:

1. Können mittels Plasma-Proteomik Rückschlüsse auf die pathophysiologischen Prozesse bei Patient:innen mit COVID-19 gezogen werden?
2. Ermöglicht das Plasma-Proteom eine Klassifikation der Krankheitsschwere bei COVID-19 und ist es hierbei etablierten klinischen und Routine-Laborparametern überlegen?
3. Kann das Plasma-Proteom zur Prognose individueller Krankheitsverläufe genutzt werden?

2 Methodik

2.1 Klinische Kohorten

2.1.1 Charité-Kohorte

Patient:innen wurden im Rahmen der Pa-COVID-19-Studie rekrutiert, einer prospektiven Beobachtungsstudie zur Pathophysiologie von Patient:innen mit COVID-19 an der Charité – Universitätsmedizin Berlin [51]. Die Studie wurde durch die Ethikkommission der Charité – Universitätsmedizin Berlin bewilligt (EA2/066/20) und entsprechend der Deklaration von Helsinki sowie den Richtlinien für Gute Klinische Praxis (ICH 1996) durchgeführt. Sie ist im Deutschen Register für Klinische Studien registriert (DRKS00021688). Grundsätzlich konnten alle Patient:innen mit PCR-bestätigter SARS-CoV-2-Infektion in die Studie eingeschlossen werden. Ausschlusskriterien waren a) die Nichteinwilligung durch den:die Patient:in oder dessen gesetzliche:n Vertreter:in oder b) das Vorliegen von Gründen, die gegen die Entnahme von zusätzlichem Studienblut sprachen. Im Rahmen der vorliegenden Arbeit wurden ausschließlich stationär aufgenommene Patient:innen untersucht.

Bei Einschluss wurden umfangreiche demografische, soziale und klinische Parameter erhoben. Während des stationären Aufenthalts erfolgten i.d.R. dreimal wöchentlich (Montag, Mittwoch, Freitag) die Erhebung klinischer Daten und Scores im Rahmen von Studienvisiten sowie eine Blutentnahme. Ethylendiamintetraessigsäure- (EDTA-) und Citrat-Blut wurde für 15 Minuten mit 2000 G bei 18°C zentrifugiert, das Plasma aliquotiert und bis zur weiteren Verwendung bei -80°C eingefroren. Die Erfassung der Krankheitsschwere erfolgte tagesgenau entsprechend der WHO Ordinal Scale for Clinical Improvement [52], von 3, stationäre Behandlung ohne zusätzlichen Sauerstoffbedarf, bis 8, Tod (Tabelle 1). Klinische Scores für Vorerkrankungen (CCI [Charlson Comorbidity Index] [53]), Mortalitätsrisiko (SOFA [Sequential Organ Failure Assessment] [54], APACHE II [Acute Physiology And Chronic Health Evaluation II] [36]) und COVID-19 (ABCS [Age, Biomarkers, Clinical history, Sex] [37]) wurden entweder aus dem klinischen Informationssystem (COPRA Version 6.0, COPRA System, Berlin, Deutschland) extrahiert oder manuell berechnet.

Die medizinische Behandlung erfolgte unabhängig von der Studienteilnahme entsprechend den lokalen und nationalen Leitlinien.

2.1.2 Innsbruck-Kohorte

Die Validierung der an der Pa-COVID-19-Kohorte trainierten prädiktiven Modelle erfolgte an einer zweiten, vollständig unabhängigen Kohorte. Patient:innen mit schwerer COVID-19 wurden im Rahmen einer Studie der Universitätsklinik für Innere Medizin II Innsbruck, Österreich, rekrutiert (EK-Nr. 1107/2020). Patient:innen mit PCR-bestätigter SARS-CoV-2-Infektion konnten eingeschlossen werden, wenn a) der erste positive PCR-Test maximal 7 Tage zurücklag, b) respiratorisches Versagen vorlag (SaO_2 <60mmHg oder SpO_2 <90 %) und c) typische Infiltrate in der Computertomographie nachgewiesen wurden. Eine schriftliche Einwilligung wurde bei Einschluss oder, bei intubierten Patient:innen, retrospektiv nach Extubation eingeholt. Die medizinische Behandlung erfolgte entsprechend den nationalen Leitlinien. Im Rahmen dieser Studie wurde ausschließlich die erste entnommene Probe bei maximaler Krankheitsschwere entsprechend WHO-Skala verwendet.

2.1.3 Generation Scotland-Kohorte

Als Vergleichskohorte der Normalbevölkerung wurden 199 zufällig ausgewählte Proben der prospektiven epidemiologischen Generation Scotland-Studie (GS) verwendet [55]. Die GS-Kohorte umfasst ca. 24.000 Menschen aus ca. 7000 Familien in Schottland.

Tabelle 1: WHO Ordinal Scale for Clinical Improvement

WHO-Grad	Definition
WHO-1	keine Einschränkungen im Alltag
WHO-2	Einschränkungen im Alltag, ambulante Behandlung
WHO-3	Stationäre Behandlung ohne Sauerstofftherapie
WHO-4	… mit low-flow Sauerstofftherapie
WHO-5	… mit high-flow Sauerstofftherapie
WHO-6	… mit invasiver maschineller Beatmung (IMV)
WHO-7	… mit IMV und zusätzlich Katecholaminen, Hämodialyse oder ECMO
WHO-8	Tod

Die Klassifikation der Krankheitsschwere erfolgt entsprechend dem Bedarf von Sauerstofftherapie sowie ggf. zusätzlichen Organersatzverfahren. Modifiziert nach WHO [52].

2.2 Klinische Labormessungen

Klinische Routinelaborparameter der Charité-Kohorte wurden auf validierten Point-of-Care Blutgasanalyse-Geräten bzw. bei der Labor Berlin – Charité Vivantes GmbH (Sylter Straße 2, 13353 Berlin, Deutschland) durchgeführt. Messparameter umfassten u.a. Differentialblutbild, klinische Chemie, Enzymaktivitäten und Blutgasanalysen [56]. Lagen mehrere Messungen an einem Tag vor, so wurde diejenige ausgewählt, die zeitlich am nächsten an der Studienvisite lag (6:00 Uhr morgens auf Intensivstationen, 8:00 Uhr morgens auf Normalstationen; Abweichungen waren aufgrund von Stationsabläufen oder am Aufnahmetag möglich).

In der vorliegenden Arbeit werden die Namen klinischer Laborparameter ausgeschrieben und proteomische Parameter mit ihrem jeweiligen UniProtKB (www.uniprot.org) Gen-Namen (ausschließlich Majuskeln, ggf. in Kombination mit Ziffern) benannt. Eine Zuordnung von im Rahmen dieser Arbeit erwähnten Gen-Namen zu den Proteinen findet sich in Tabelle 2.

2.3 Discovery-Proteomik

Die Probenvorbereitung erfolgte semi-automatisiert. Nach dem Auftauen wurden 5µl EDTA- bzw. Citrat-Plasma (nur Kohorte 2, Abschnitt 3.3.1) auf vorbereitete 96-Well Platten mit 55µl Denaturierungs-/Reduktionsmedium (8M Urea, 100mM Ammonium-Bicarbonat (ABC), 50mM Dithiothretiol) gegeben und für 60 Minuten (min) bei 30°C inkubiert. Im Anschluss wurden 5µl Iodoacetamid als Alkylans hinzugegeben und die Mischung weitere 30 min bei 23°C inkubiert, bevor die Lösung mit 340µl Ammonium-Bicarbonat- (ABC) Puffer verdünnt wurde. Die enzymatische Verdauung erfolgte mittels Trypsin (220µl Proben-Gemisch auf 12,5µl [1µg/µl] Trypsin, anschließend Inkubation für 17 Std. bei 37°C) und wurde mit 25µl 10 %-iger Methansäure unterbrochen. Es folgte die Aufreinigung und Resuspension der Proben in 60µl 0,1 %-iger Methansäure; unlösliche Partikel wurden mittels Zentrifugation entfernt und die Probenüberstände auf eine neue 96-Well Plate übertragen. Jede 96-Well Plate umfasste u.a. 8 Plasma- und 4 Serumkontrollen zur Normalisierung zwischen Plates.

LC-MS/MS wurde auf dem Agilent 1290 Infinity II System (Agilent Technologies, Inc., Santa Clara, USA), gekoppelt an einen TripleTOF 6600 (SCIEX, Danaher Corporation,

Washington, D.C., USA) durchgeführt. MS_1 wurde auf *m/z* 100-1500 eingestellt, gefolgt von 25-mal MS_2 (mit variablen Vorläufer-Ionen von *m/z* 450-800).

2.4 Peptid-Panelassay

Für die selektierten Zielpeptide wurden native (Reinheit ≥95 %) und Isotopen-markierte (SIL, stable isotope labeled) Peptide synthetisiert und analog zu den klinischen Plasma-Proben weiterverarbeitet [57], wie unter Abschnitt 2.3 beschrieben.

LC-MRM wurde auf zwei verschiedenen Systemen durchgeführt: Agilent 1290 Infinity II System gekoppelt an einen Agilent 6495C Massenspektrometer (beide Agilent Technologies, Inc., Santa Clara, USA) sowie (für eine Auswahl von Proben) zusätzlich auf einem ExionLD AD UHPLC System gekoppelt an einen TripleQuad 7500 Massenspektrometer (beide SCIEX, Danaher Corporation, Washington, D.C., USA).

Tabelle 2: Zuordnung von Gen- zu Protein-Namen entsprechend UniProtKB

Genname	Proteinname
A2M	Alpha-2-macroglobulin
ACTA1	Actin, alpha skeletal muscle
ACTB	Actin Beta
ACTBL2	Beta-actin-like protein 2
ACTG1	Actin, cytoplasmic 1
AFM	Afamin
AGT	Angiotensinogen, Serpin A8
AHSG	Alpha-2-HS-glycoprotein, Alpha-2-Z-globulin, Fetuin-A
APOA1	Apolipoprotein A1
APOB	Apolipoprotein B
APOC3	Apolipoprotein C3
B2M	Beta-2-microglobulin
C1QA	Complement C1q subcomponent subunit A
C1QB	Complement C1q subcomponent subunit B
C1QC	Complement C1q subcomponent subunit C
C1R	Complement C1r subcomponent
C3	Complement component C3
C8A	Complement component C8 alpha chain
C9	Complement component C9
CD14	Cluster of Differentiation 14
CFD	Complement factor D
CRP	C reactive protein
CST3	Cystatin-C
ECM1	Extracellular matrix protein 1
EFEMP1	EGF-containing fibulin-like extracellular matrix protein 1
F12	Coagulation factor XII
F2	Prothrombin
FCGR3A	Low affinity immunoglobulin gamma Fc region receptor III-A, CD16a
FGA	Fibrinogen Alpha Chain
GPLD1	Phosphatidylinositol-glycan-specific phospholipase D
GSN	Gelsolin
HP	Haptoglobin
HPX	Hemopexin
HRG	Histidine-rich glycoprotein
IGFALS	Insulin-like growth factor binding protein, acid labile subunit
IGHV	Immunoglobulin heavy chain variable region

Genname	Proteinname
ITIH1	Inter-alpha-trypsin inhibitor heavy chain H1
ITIH2	Inter-alpha-trypsin inhibitor heavy chain H2
ITIH3	Inter-alpha-trypsin inhibitor heavy chain H3
KLKB1	Plasma kallikrein
LBP	Lipopolysaccharide Binding Protein
LRG1	Leucin-rich Alpha-2-Glykoprotein
LYZ	Lysozyme
ORM2	Alpha-1-acid glycoprotein 2
PGLYRP2	Peptidoglycan recognition protein 2
PIGR	Polymeric immunoglobulin receptor
PLG	Plasminogen
PRG4	Proteoglycan 4
SAA1	Serum amyloid A-1 protein
SAA2	Serum amyloid A-2 protein
SERPINA1	Alpha-1 antitrypsin
SERPINA3	Alpha-1-antichymotrypsin
SERPINA4	Kallistatin
SERPINC1	Antithrombin III
SERPIND1	Heparin cofactor 2
SERPING1	Plasma protease C1 inhibitor
TF	Transferrin
TFRC	Transferrin receptor protein 1
TTR	Transthyretin
VWF	Von-Willebrand-Factor

Zuordnung der im Rahmen dieser Arbeit benannten Gen- zu den jeweiligen Proteinnamen, entsprechend der englischsprachigen Originalbezeichnung. Eigene Darstellung. Quelle: www.uniprot.org.

2.5 Datenverarbeitung und statistische Analyse

2.5.1 Aufbereitung von MS-Daten

Massenspektrometrische Daten wurden mittels DIA-NN aufbereitet, einem Programmpaket das, mittels tiefen neuronalen Netzen (deep neural networks, DNN), automatisiert DIA Proteom-Datensätze prozessiert (Abgrenzung von Signal zu Hintergrundrauschen, Korrektur überlappender Signale sowie Quantifizierung) [48]. Eine Korrektur zwischen Platten erfolgte anhand der Kontrollproben (vgl. Abschnitt 2.3); die Korrektur zwischen verschiedenen Messchargen (Batches) erfolgte mittels linearer Regression. Korrekturen erfolgten auf Basis von Vorläufer-Ion-Daten. Es wurden keine Vorläufer-Ionen oder Proteine aufgrund fehlender Messdaten ausgeschlossen.

2.5.2 Statistische Analyse

Die statistische Analyse erfolgte mit JMP Pro Version 15 (SAS Institute Inc, Cary, NC, USA), R Version 3.6.0 (R Core Team, www.R-project.org) und Python Version 3.8.1 bzw. 3.8.5 (Python Software Foundation, Wilmington, DE, USA). Klinische Labor- und proteomische Daten wurden $\log_2$ transformiert (mit Ausnahme von Basenüberschuss, Oxyhämoglobin und sO_2). Es wurden keine Daten imputiert (außer für Machine Learning-Modelle, vgl. Abschnitte 2.5.2.8 und 2.5.2.9) oder aufgrund von Unvollständigkeit ausgeschlossen.

Die Signifikanztestung gegen einen Nullmedian (Verläufe) oder binäres Outcome erfolgte mittels Wilcoxon-W- bzw. Mann-Whitney-U-Test. Der Zusammenhang stetiger Variablen (bspw. Protein-Protein Korrelationen) wurde mittels Kendall-Tau-Test untersucht. Bei Einbeziehung von Co-Variablen wurde ein moderierter T-Test entsprechend dem „limma“ Paket für R [58] angewandt. Mehrfachtestung wurde mittels der Benjamini-Hochberg-Falscherkennungsrate (false discovery rate, FDR) -Methode korrigiert. P-Werte <0,05 bzw. FDR <5 % für Korrelations-Heatmaps wurden als signifikant gewertet.

2.5.2.1 Marker für Erkrankungsschwere

Es wurde der jeweils erste Messpunkt bei maximaler Erkrankungsschwere entsprechend WHO-Skala (außer Tod) ausgewählt und auf Korrelation mit WHO-Schweregrad untersucht (Wilcoxon-W-Test).

2.5.2.2 Altersabhängige Marker

Um altersabhängige Veränderungen im Proteom zu identifizieren und bei den weiteren Analysen entsprechend zu korrigieren, wurde der jeweils erste Messzeitpunkt eines bzw. einer Patient:in verwendet (Kendall-Tau-Test; Korrektur für WHO-Schweregrad mittels „limma“).

2.5.2.3 Marker von Hämodialyse und extrakorporaler Membranoxygenierung

Es erfolgte der Vergleich aller Messzeitpunkte bei WHO-Grad 7 von Patient:innen ohne Hämodialyse bzw. extrakorporaler Membranoxygenierung (ECMO) mit Messzeitpunkten nach Initiierung der jeweiligen Therapie mittels Mann-Whitney-U-Test.

2.5.2.4 Marker-Veränderungen im zeitlichen Verlauf

Es wurden Proben während des Zeitraums der stärksten Krankheitsschwere, d.h. WHO-Grad 6 oder 7 für invasiv beatmete Patient:innen, WHO-Grad 5, 4 oder 3 für alle weiteren Patient:innen, verwendet, um die Veränderungen des Plasma-Proteoms im zeitlichen Verlauf zu untersuchen. So konnten die Einflüsse möglicher „Schweregradwechsel“ sowie spezifischer Therapien minimiert werden. Es wurden ausschließlich Patient:innen mit ≥2 Messzeitpunkten und einem Abstand zwischen dem ersten und letzten Messzeitpunkt von ≥2 Tagen analysiert.

Mittels linearer Regression wurde für jeden Messwert versus [Tagzahl seit Krankenhausaufnahme] die Steigung (nicht-parametrische Theil-Sen-Methode) und die Mengensteigung ([Regressionssteigung] x [Anzahl der Tage zwischen erster und letzter Messung]) berechnet. Mittels Wilcoxon-W-Test wurde die Abweichung der medianen Mengensteigung vom Nullmedian getestet.

2.5.2.5 Korrelations-Heatmaps

Allgemeine Korrelationen wurden auf Basis des ersten Messzeitpunkts bei der jeweils maximalen Krankheitsschwere (WHO-Grad) eines:r individuellen Patient:in errechnet. Analog erfolgte die Berechnung von Korrelationen im zeitlichen Verlauf auf Basis der oben berechneten Mengensteigung. Als COVID-19-spezifische Proteinkorrelationen wurden all jene gewertet, die in der Generation Scotland Kohorte [55] (vgl. Abschnitt 2.1.3) nicht signifikant waren ($P \geq 0{,}05$) oder einen entgegengesetzten Trend aufwiesen.

2.5.2.6 Marker für Zunahme der Krankheitsschwere

Es wurde der jeweils erste Messzeitpunkt verwendet. Eine Zunahme der Krankheitsschwere wurde definiert als Zunahme um mindestens einen WHO-Grad bzw. Tod bei Patient:innen mit WHO-Grad 7. Proteome von Patient:innen mit und ohne Zunahme der Krankheitsschwere wurden verglichen, unter Berücksichtigung von Alter und aktuellem WHO-Schweregrad mittels linearem Modell.

2.5.2.7 Prädiktion von invasiver mechanischer Beatmung

Die Prädiktion, ob ein:e Patient:in zum ersten Messzeitpunkt bei maximaler Krankheitsschwere invasiv mechanisch beatmet wurde, erfolgte mittels Gradient Boosted Trees (in Python 3.8.1). Proteine mit ≥3 quantifizierten Peptiden wurden berücksichtigt, es wurden keine fehlenden Werte imputiert. Eine Kreuzvalidierung wurde durchgeführt, indem beim Trainieren jeweils eine Probe zurückgehalten, das Modell auf Basis der verbliebenen Proben trainiert und dieser Schritt für jede individuelle Probe wiederholt wurde (leave-1-out). Das so an der Charité-Kohorte trainierte Modell wurde direkt auf die Innsbruck-Kohorte angewendet.

2.5.2.8 Prädiktion von maximaler Krankheitsschwere

Die Prädiktion des WHO-Schweregrads zum ersten Messzeitpunkt bei maximaler Krankheitsschwere erfolgte mittels Elastic Nets in Python 3.8.1. Proteine mit ≥3 quantifizierten Peptiden wurden berücksichtigt. Fehlende Werte wurden mittels k-Nearest Neighbor (kNN) -Imputation ersetzt. Analog zur Prädiktion von invasiver mechanischer Beatmung erfolgte die Kreuzvalidierung mittels leave-1-out-Methode und das so trainierte Modell wurde direkt auf die Innsbruck-Kohorte angewendet.

2.5.2.9 Prädiktion von Überleben bei kritisch kranken Patient:innen

Es wurde der erste Messzeitpunkt bei maximaler Krankheitsschwere (WHO-Grad 7) ausgewählt, um das Outcome zu prädizieren, d.h. Überleben (Entlassung oder Verlegung an ein peripheres Krankenhaus zur Beatmungsentwöhnung) oder Versterben. Das Machine Learning (ML) -Modell wurde auf Basis von 57 Proteinen trainiert, für die es FDA-zugelassene MRM-Assays gibt und die mit mindestens 3 Peptiden quantifiziert wurden. Fehlende Werte wurden mittels Minimal Value Imputation imputiert.

Machine Learning wurde mittels Parenclitic Networks [59,60] durchgeführt. Parenclitic Networks sind ein Graphen-basierter Ansatz, bei dem Netzwerke, die die Abweichung eines Individuums von der Gesamtpopulation repräsentieren, abgeleitet werden. Dazu wird jedes Protein-Paar zueinander ins Verhältnis gesetzt (mittels radialem Support Vector Machine (SVM) Classifier) und ein Kantengewicht berechnet, das die Wahrscheinlichkeit für ein definiertes Outcome (hier: versterben) widerspiegelt. Es werden fünf Merkmale (features) generiert: Maximum, Mittelwert und Standardabweichung der Kantengewichte sowie Anzahl von Kantengewichten > 0,5 (d.h. prädiziertes Versterben) und Anzahl der Nodi (d.h. Proteine) mit mindestens einer Kante mit einem Gewicht > 0,5. Basierend auf topographischen Unterschieden individueller Netzwerke wird im Anschluss ein LASSO- (Least Absolute Shrinkage And Selection Operator) Klassifikationsmodell zur Prädiktion generiert.

Eine Kreuzvalidierung erfolgte, indem jeweils eine Probe (zusammen mit zwei weiteren, zufällig selektierten Proben, insgesamt 2 Überlebende und 1 Verstorbene:r) beim Trainieren des Modells zurückgehalten und das generierte Modell auf die eine zurückgehaltene Probe angewendet wurde (leave-3-out). Insgesamt wurde dieser Schritt 49-mal, d.h. einmal für jede Probe, wiederholt und am Ende ein Mittelwert der Prädiktion für jede Probe errechnet. Das so generierte Modell wurde direkt auf die „Innsbruck-Kohorte“ angewendet.

Um die Robustheit des Parenclitic Network-Modells zu verifizieren, wurde ein SVM-Modell in Python 3.8.5 auf Basis der gleichen Daten und selektierten Proteine wie für das Parenclitic Modell generiert.

2.5.3 Abbildungen

Abbildungen wurden in Microsoft PowerPoint Version 16.73 (Microsoft Corporation, Redmond, WA, USA), GraphPad Prism Version 9.4 (GraphPad Software, Inc, San Diego, CA, USA) und mit Hilfe von BioRender.com erstellt.

3 Ergebnisse

Im Zeitraum 1. März bis 19. November 2020 wurden 280 stationär behandelte Patient:innen mit PCR-bestätigter SARS-CoV-2-Infektion in die Pa-COVID-19 Studie eingeschlossen, von denen insgesamt 881 Plasma-Proteome vermessen und im Rahmen der folgenden drei Publikationen analysiert wurden: In Demichev, Tober-Lau et al. 2021 [56] wurden Korrelationen zwischen Proteom- und Routinelaborparametern untersucht sowie potenzielle neue Biomarker für Krankheitsschwere und Prognose bei COVID-19 beschrieben. Demichev, Tober-Lau et al. 2022 [61] befasst sich mit dem prognostischen Potenzial des Plasma-Proteoms bei kritisch kranken Patient:innen mit COVID-19 und untersucht, inwiefern sich das Überleben dieser speziellen Patient:innenpopulation vorhersagen lässt unter Zuhilfenahme von Machine Learning-Modellen. Wang, Cryar et al. 2022 [57] beschreibt die methodischen Anpassungen und erweiterten Messungen in zusätzlichen Patient:innengruppen als Basis für die Translation der Ergebnisse der beiden vorher beschriebenen Publikationen in die klinische Routineanwendung.

3.1 Discovery-Proteomik bei Patient:innen mit COVID-19

Der erste Teil der Ergebnisse der vorliegenden Arbeit geht auf die Identifikation potenzieller Biomarker von Erkrankungsschwere und Prognose bei COVID-19 ein. Die Ergebnisse wurden im peer-reviewed Fachjournal Cell Systems publiziert [56].

3.1.1 Studienpopulation

Vom 1. März bis 30. Juni 2020 wurden 139 stationär behandelte Patient:innen mit PCR-bestätigter SARS-CoV-2-Infektion an der Charité – Universitätsmedizin Berlin in die Studie eingeschlossen. Die Patient:innen wiesen variable Krankheitsverläufe mit Schweregraden von 3 bis 8 auf der WHO-Skala auf (vgl. Tabelle 1): 23 (16,5 %) blieben durchgehend ohne Sauerstoffbedarf (WHO-Grad 3); 47 (33,8 %) benötigten nicht-invasive Sauerstofftherapie (WHO-Grad 4 oder 5); 69 (49,6 %) benötigten im Verlauf ihrer Hospitalisierung eine invasive mechanische Beatmung (WHO-Grad 6 oder 7). Insgesamt erhielten 46 (33 %) Patient:innen Hämodialyse und 22 (16 %) benötigten ECMO. Zwanzig (14 %) Patient:innen verstarben; davon hatten drei Patient:innen eine Intubation bzw. Reanimation abgelehnt (DNI/DNR) und ein Patient verstarb an einem COVID-19-unabhängigen

Leiden (Ablehnung von Therapie bei Karzinom). In vier Fällen wurden im Verlauf der stationären Behandlung bei infauster Prognose entsprechend dem mutmaßlichen Patient:innenwillen die lebenserhaltenden Maßnahmen beendet.

Von den 139 Patient:innen waren 95 (68 %) männlich, der mediane Body-Mass-Index (BMI) betrug 27,8 kg/m^2 (IQR: 24,7 – 31,9 kg/m^2). Innerhalb der Kohorte gab es keine Korrelation zwischen Geschlecht bzw. Gewicht mit der Krankheitsschwere (P=0,06 bzw. P=0,11). Höheres Alter als Risikofaktor für schwere Krankheitsverläufe spiegelte sich wider: Patient:innen mit WHO-Grad 3 waren signifikant jünger als jene mit WHO-Grad 7 (49 [IQR: 35-70] vs. 62 [IQR: 53-72] Jahre, P=0,02); Patient:innen mit einem Alter ≥65 Jahre hatten ein signifikant erhöhtes Risiko im Krankenhaus zu versterben (OR 4,1 [95%-CI: 1,5-11,5]). Im Median dauerte die stationäre Behandlung 20 Tage (IQR: 9-48) und korrelierte mit der Krankheitsschwere (7 Tage bei WHO-Grad 3 vs. 46 Tage bei WHO-Grad 7). Die mediane Dauer zwischen Krankenhausaufnahme und Versterben trotz Maximaltherapie (d.h. ausgenommen Patient:innen mit DNI/DNR oder nicht-COVID-19-Todesursache) betrug 28 Tage (IQR: 16-46). Basischarakteristika der Kohorte sind in Tabelle 3 dargestellt.

Von den 139 Patienten wurden insgesamt 687 longitudinale Plasma-Proteome gemessen. Es konnten 321 verschiedene Proteingruppen in mindestens 75 % der Proben nachgewiesen werden, davon 200 Proteine in 98 % und 189 Proteine in 99 % der Proben. Zur besseren Veranschaulichung wurden verschiedene Immunglobulinketten zusammengruppiert, sodass insgesamt 177 Plasma-Proteingruppen für die Auswertung zur Verfügung standen.

Tabelle 3: Basischarakteristika der Discovery-Proteomik-Kohorte

			keine invasive Beatmung						invasive Beatmung					
	alle Patient:innen		max. WHO-Grad 3		max. WHO-Grad 4		max. WHO-Grad 5		max. WHO-Grad 6		max. WHO-Grad 7		verstorben**	
Anzahl	139	100 %	23	17 %	33	24 %	14	10 %	6	4 %	63	45 %	17	12 %
Geschlecht														
weiblich	44	32 %	13	57 %	8	24 %	3	21 %	1	17 %	19	30 %	3	18 %
männlich	95	68 %	10	43 %	25	76 %	11	79 %	5	83 %	44	70 %	14	82 %
Alter, Jahre (Median, IQR)	61	50 - 71	49	35 - 70	63	48 - 71	62	49 - 85	64	59 - 71	62	53 - 72	69	55 - 77
≥65	57	41 %	6	26 %	16	48 %	6	43 %	3	50 %	26	41 %	11	65 %
BMI, kg/m^2 (Median, IQR)	27,8	24,7 - 31,9	25,4	23,0 - 30,6	27,2	23,3 - 30,3	27,1	23,3 - 34,9	26,6	24,6 - 28,1	29,4	25,7 - 34,1	29,0	24,8 - 31,0
< 25 kg/m^2	40	29 %	10	43 %	13	39 %	3	21 %	2	33 %	12	20 %	4	24 %
≥ 25 kg/m^2	97	71 %	13	57 %	20	61 %	11	79 %	4	66 %	49	80 %	13	76 %
Vorerkrankungen														
CCI (Median, IQR)	3	1 - 4	1	0 - 3	4	1 - 5	3	1 - 6	2,5	2 - 5	3	1 - 4	3	3 - 6
<3	65	47 %	16	70 %	13	39 %	7	50 %	3	50 %	26	41 %	2	12 %
≥3	74	53 %	7	30 %	20	61 %	7	50 %	3	50 %	37	59 %	15	88 %
Dauer stationärer Aufenthalt*, Tage (Median, IQR)	20	9 - 48	7	4 - 10	14	9 - 17	20	13 - 28	38	24 - 71	46	32 - 75	28	16 - 46
Bauchlagerung**	51	38 %	-	-	-	-	-	-	2	33 %	49	78 %	13	76 %
invasive Beatmung**	69	50 %	-	-	-	-	-	-	6	100 %	63	100 %	17	100 %
Hämodialyse	46	33 %	0	0 %	1	3 %	0	0 %	1	17 %	44	70 %	15	88 %
ECMO**	22	16 %	-	-	-		-	-	-	-	22	35 %	8	47 %
Outcome														
verstorben (incl. DNI/DNR)	20***	14 %	0	0 %	1	3 %	2	14 %	0	0 %	17	27 %	-	-
DNI/DNR	3	2 %	0	0 %	1	3 %	2	14 %	0	0 %	0	0 %	-	-
sekundäre DNI/DNR	4	3 %	0	0 %	0	0 %	0	0 %	0	0 %	4	6 %	4	24 %

Die Daten stellen Anzahl (n) und Prozent (%) dar, sofern nicht anders angegeben. IQR: Interquartilspanne; CCI: Charlson's Comorbidity Index; ECMO: extrakorporale Membranoxygenierung; DNI/DNR: nicht intubieren/nicht reanimieren; sekundäre DNI/DNR: Therapielimitation bei infauster Prognose entsprechend dem mutmaßlichen Patient:innenwillen. Modifiziert nach Demichev, Tober-Lau, et al., 2021 [56].

* exklusive verstorbene Patient:innen

** exklusive Patient:innen mit DNI/DNR

*** inkl. einem Patienten mit COVID-19 unabhängiger Todesursache

3.1.2 Kovariation von Routine-Laborparametern und Proteommessungen

Im ersten Schritt untersuchten wir Korrelationen zwischen Routinelabor- und proteomischen Messungen (vgl. Tabelle 2 zur Zuordnung von UniProtKB Gen-Namen). Wir beobachteten robuste positive Korrelationen zwischen den etablierten klinischen Entzündungsmarkern Interleukin-6, C-reaktivem Protein und Procalcitonin mit proteomisch gemessenen Akutphaseproteinen (APPs) wie SAA1, SAA2, CRP, SERPINA1, SERPINA3, sowie (invers) GSN und AHSG. Auch fanden wir Korrelationen mit Proteinen der Blutgerinnung (u.a. FGA, F2, F12, KLKB1, PLG) und des Komplement-Systems (u.a. C1R, C8A, C9, CFD) sowie diverse Korrelationen zwischen proteomischen APPs und klinischen Laborparametern von Organschädigung (z.B. NT-proBNP, Kreatinin, Alanin-Aminotransferase) und Dyserythropoiese (u.a. Erythrozytenzahl und -verteilungsbreite).

Wir identifizierten darüber hinaus einige spezifische Korrelationen zwischen klinischen Laborwerten und proteomischen Markern. So korrelierten SERPINA1 (Alpha-1-Antitrypsin) und SERPINA3 (Alpha-1-Antichymotrypsin), zwei Neutrophilen-Serinproteaseninhibitoren, stark (Spearman R=0,72 bzw. R=0,79) mit der Neutrophilen-zu-Lymphozyten-Ratio, einem prognostischen Marker von COVID-19 [62]. Auch gab es eine starke Korrelation zwischen alkalischer Phosphatase bzw. Gamma-Glutamyltransferase, beides klinischen Markern u.a. für Erkrankungen der Gallenwege, und PIGR (polymerischer Immunglobulin-Rezeptor), der infolge viraler Infektion von Cholangiozyten vermehrt von diesen exprimiert wird [63].

Insgesamt analysierten wir die Zusammenhänge zwischen 85 klinischen Routineparametern und 177 proteomischen Plasma-Proteinen und stellten somit einen Gesamtdatensatz zur Verfügung, der zum Zeitpunkt seiner Veröffentlichung in Art und Umfang einzigartig war.

3.1.3 Korrelation mit Krankheitsschwere

Als nächstes untersuchten wir, welche Proteine und Laborparameter abhängig von der Krankheitsschwere (WHO-Grad) exprimiert waren. Wir identifizierten 113 Proteine, die entsprechend der Krankheitsschwere hoch- bzw. herunterreguliert waren, von denen unseres Wissens nach 30 im Rahmen dieser Arbeit erstmalig beschrieben wurden. Diese umfassten u.a. Mediatoren der Entzündungsreaktion und Immunantwort sowie Proteine

des Komplementsystems und Apolipoproteine. Ebenfalls identifizierten wir 55 klinische Routineparameter, die mit dem Schweregrad korrelierten, darunter diverse Marker für Organschädigungen (bspw. Herz: NT-proBNP und Troponin T; Niere: Kreatinin und Harnstoff; Leber: Aspartat- und Alanin-Aminotransferasen, Gamma-Glutamyltransferase und Bilirubin) sowie, invers korreliert, Hämoglobin, Erythrozytenzahl und Hämatokrit. Spezifische Veränderungen bei Patient:innen mit Organersatzverfahren waren u.a. erniedrigtes HP (Haptoglobin) und HPX (Hämopexin) unter Hämodialyse und ECMO, möglicherweise aufgrund von Hämolyse im extrakorporalen Kreislauf, sowie erhöhtes SERPINC1 (Antithrombin-III) als Zeichen der Substitution unter ECMO-Therapie (zusammengefasst in Tabelle 8).

Da sich höheres Alter schon früh als einer der wichtigsten Risikofaktoren für schwere Krankheitsverläufe bei COVID-19 darstellte [21], untersuchten wir den Datensatz mit Hinblick auf altersspezifische Veränderungen im Plasma. Tatsächlich beobachteten wir eine deutliche Überlappung zwischen Markern, die mit Krankheitsschwere, und solchen, die mit dem Alter korrelierten. Nach Korrektur für WHO-Schweregrad [56] verblieben 20 Proteine, die bei Patient:innen mit COVID-19, nicht jedoch in der Allgemeinbevölkerung (vgl. Abschnitt 2.1.3) mit dem Alter korrelierten. Diese ließen sich vornehmlich Inflammation, Lipidstoffwechsel und Blutgerinnung zuordnen und könnten auf die besondere Rolle dieser Kompartimente in der altersspezifischen Pathogenese von COVID-19 hinweisen.

3.1.4 Korrelation im zeitlichen Verlauf

Auch im longitudinalen Verlauf beobachteten wir Korrelationen zwischen Routinelabor- und proteomischen Markern von Entzündung, Akutphasereaktion sowie Blutgerinnung. Unabhängig vom maximalen WHO-Grad beobachteten wir bei fast allen Parametern eine allgemeine „Verbesserung“ des Proteoms während der maximalen Erkrankungsschwere (vgl. Abschnitt 2.5.2.4): initial mit zunehmender Krankheitsschwere hochregulierte Proteine nahmen mit der Zeit ab, initial herunterregulierte stiegen umgekehrt im zeitlichen Verlauf an. Dies betraf u.a. Proteine der Entzündungs- und Akutphasereaktion, Blutgerinnung, aber auch Immunregulation, des Lipidstoffwechsels und der extrazellulären Matrix.

Insgesamt stellten wir fest, dass sich der Phänotyp von einer initial ausgeprägten systemischen Entzündungsreaktion im zeitlichen Verlauf hin zu Immunomodulation, metabolischer Rekonstitution und Gewebereparatur wandelte – insbesondere die frühe Phase könnte daher entscheidend für den weiteren Krankheitsverlauf bei COVID-19 sein.

3.1.5 Machine Learning-Modelle zu Klassifikation von Krankheitsschwere

Der umfangreiche Datensatz ermöglichte es, verschiedene prädiktive Machine ML-Modelle zu generieren. Basierend auf Gradient Boosted Trees ließ sich mit hoher Genauigkeit ausschließlich auf Basis von klinischen Labor- bzw. proteomischen Messungen bestimmen, ob ein:e Patient:in zum Messzeitpunkt maschinell beatmet wurde (WHO-Grad 6 oder 7; nur klinisch: AUROC = 0,97, nur proteomisch: AUROC = 0,98; klinisch und proteomisch kombiniert: AUROC = 0,99). Dieses Modell ermöglichte eine deutlich höhere Präzision als etablierte Risikofaktoren bzw. -scores wie Alter, BMI, Vorerkrankungen (CCI) oder C-reaktives Protein. Die Anwendung des anhand der Charité-Kohorte trainierten Modells auf eine komplett unabhängige Kohorte von 99 Patient:innen (Innsbruck-Kohorte) lieferte vergleichbare Ergebnisse (AUROC = 0,97). Auch der WHO-Grad zum Messzeitpunkt ließ sich mittels ML-Modell zuverlässig bestimmen, wobei die Kombination aus klinischen Labor- und proteomischen Messungen die beste Korrelation aufwies, sowohl für die Charité- als auch die Innsbruck-Kohorte.

3.1.6 Prognose von Krankheitsverläufen

Aufgrund der Beobachtung von Veränderungen im zeitlichen Verlauf untersuchten wir, inwieweit das Proteom geeignet ist, um eine Verschlechterung der Krankheitsschwere (Zunahme des WHO-Grades) und, im Fall von kritisch kranken Patienten, Versterben vorherzusagen. Unter Berücksichtigung von aktuellem WHO-Grad und Alter identifizierten wir 11 Proteine und 9 klinische Laborparameter, die auf eine zukünftig klinische Verschlechterung hinwiesen. Diese umfassten u.a. Marker für Entzündung (CRP, ITIH2, SERPINA3, AHSG und B2M), Blutgerinnung (HRG und PLG) und Komplementaktivierung (C1R und CFD) sowie AGT und CST3 (Infokasten 1).

höhere Konzentration bei zukünftiger Zunahme der Krankheitsschwere

AGT: **Angiotensinogen** wird durch Renin sowie die Angiotensin-konvertierenden Enzyme ACE oder alternativ ACE2 aktiviert. ACE aktiviert Angiotensin I (Ang1) zu Angiotensin II (Ang2), mit pro-inflammatorischem, vasokonstriktivem und pro-fibrotischem Effekt [64]. ACE2 hingegen wandelt Ang1 und Ang2 zu Angiotensin 1-9 (Ang1-9) und 1-7 (Ang1-7) um (anti-inflammatorisch, vasodilatativ, anti-fibrotisch, anti-oxidativ) [65]. SARS-CoV-2 dringt über ACE2 in die Wirtszelle ein und führt zu dessen Internalisierung [64]. Folglich wird Ang1 überwiegend über ACE zu Ang2 aktiviert und der Abbau durch ACE2 verringert, sodass die pro-inflammatorischen Peptide akkumulieren [66,67]. Die Konzentration von Ang2 korreliert linear mit der Viruslast und Lungenschädigung bei COVID-19 [68].

B2M: **Beta-2-Mikroglobulin** bildet den extrazellulären Teil von MHC I (Haupthistokompatibilitätskomplex I) und ist somit auf allen kernhaltigen Zellen exprimiert. B2M wird auch von aktivierten Thrombozyten sezerniert und induziert in Makrophagen eine pro-inflammatorische M1-Polarisierung [69]. Bei Patient:innen mit chronischer Niereninsuffizienz ist erhöhtes B2M ein Marker für Versterben [70].

C1R: Die Serinprotease C1r bildet im Komplex mit C1q und C1s den Komplementfaktor C1, den Aktivator des klassischen Wegs der Komplementkaskade [71].

CFD: **Komplementfaktor D** aktiviert den alternativen Weg der Komplementkaskade durch Spaltung von Faktor B (CFB), der somit den Komplex C3bBb bilden kann („C3-Konvertase des alternativen Weges“) [71].

CRP: **C-reaktives Protein** ist ein Akutphaseprotein, das bei Entzündungsreaktionen und Infektionen stark hochreguliert ist [72].

CST3: **Cystatin C** ist ein Proteaseinhibitor und etablierter Nierenfunktionsparameter [73].

SERPINA3: **Alpha-1-Antichymotrypsin** ist ein Proteaseinhibitor, der vor Cathepsin G vermittelter Gewebeschädigung durch neutrophile Granulozyten schützt [74]. Die Spaltprodukte sind wesentlich stabiler gegenüber enzymatischem Abbau [75] und wirken stark chemotaktisch auf neutrophile Granulozyten [76]; sie könnten eine Rolle bei der dysregulierten Neurophilenaktivierung bei schwerer COVID-19 spielen [77].

niedrigere Konzentration bei zukünftiger Zunahme der Krankheitsschwere

AHSG: Alpha-2-HS-Glykoprotein, auch Fetuin-A, ist ein negatives Akutphaseprotein, das die Makrophagenaktivierung und Neutrophilendegranulation attenuiert [78].

HRG: **Histidin-reiches Glykoprotein** ist ein negatives Akutphaseprotein mit vielfältigen Funktionen, u.a. Regulation von Entzündungsreaktion, Immunantwort, Elimination von Erregern und Zelldetritus sowie Blutgerinnung und Fibrinolyse [79,80].

ITIH2: **Inter-Alpha-Trypsin-Inhibitor, schwere Kette 2**, liegt kovalent an Bikunin (AMBP) gebunden vor und bildet mit Hyaluronan das sogenannte „SHAP“ (**s**erum-derived **h**yaluronan-**a**ssociated **p**rotein) mit Matrix-stabilisierenden und immunmodulatorischen Effekten [81,82].

PLG: **Plasminogen bzw. Plasmin** ist das Schlüsselenzym der Fibrinolyse, besitzt jedoch auch vielfältige immunologische Funktionen, u.a. Dämpfung der Neutrophilenaktivität, Induktion von Efferozytose durch Makrophagen sowie deren Polarisierung von M1- (pro-inflammatorisch) hin zu M2-Phänotyp (Gewebe-reparierend) [83].

Infokasten 1: Prognostische Proteine für zukünftige Zunahme der Krankheitsschwere. Modifiziert nach Demichev, Tober-Lau, et al. 2021 [56].

Ebenfalls untersuchten wir, ob sich Korrelationen zwischen proteomischen bzw. klinischen Messungen und der verbleibenden Dauer der stationären Behandlung identifizieren ließen. In der Tat fanden wir 26 Proteine und 14 klinische Parameter, die bei Patient:innen mit WHO-Grad 3, also ohne Sauerstoffbedarf, mit der verbleibenden Zeit im Krankenhaus korrelierten. Da es sich bei diesen größtenteils auch um Schweregrad-Marker handelte, stellten wir die Hypothese auf, dass die Korrelation mit der Zeit bis zur Entlassung Unterschiede in der Krankheitsschwere innerhalb der Patient:innen mit WHO-Grad 3 widerspiegelt. Analog zur den Modellen zur Prädiktion des aktuellen WHO-Grads trainierten wir daher ein ML-Modell am ersten Messzeitpunkt; der prädizierte WHO-Grad korrelierte tatsächlich mit der prädizierten verbleibenden Zeit im Krankenhaus – ein möglicher Hinweis dafür, dass mittels proteomischer Messungen und ML eine graduiertere Einteilung der Krankheitsschwere vorgenommen werden kann als ausschließlich auf Basis klinischer Parameter.

3.2 Outcome-Prädiktion bei kritisch kranken Patient:innen

Der zweite Teil der vorliegenden Arbeit befasst sich mit der Prädiktion des Outcomes im Sinne von Überleben vs. Versterben von kritisch kranken Patient:innen mit COVID-19, d.h. WHO-Grad 7 (invasive Beatmung mit zusätzlichem Organersatzverfahren: Katecholamintherapie, Hämodialyse, ECMO). Die Ergebnisse dieser Arbeit wurden in der Erstausgabe des peer-reviewed Fachjournals „PLOS Digital Health“ publiziert [61].

3.2.1 Studienpopulation

Für die folgenden Analysen wurden die Subgruppe der 50 kritisch kranken Patient:innen der oben beschriebenen Kohorte ausgewählt. Von diesen Patient:innen lagen insgesamt 349 longitudinale Plasma-Proteome vor. Während des Beobachtungszeitraums gab es keine klinischen Versorgungsengpässe, sodass alle Patient:innen entsprechend medizinischer Indikation Maximaltherapie erhielten. Patient:innen mit DNI/DNR wurden von der Analyse ausgeschlossen.

Insgesamt erhielten 36 (72 %) Hämodialyse und 19 (38 %) ECMO-Therapie; 16 (32 %) erhielten sowohl Hämodialyse als auch ECMO. Fünfzehn (30 %) Patient:innen verstarben. Die mediane Zeitspanne zwischen erstem Messzeitpunkt und Outcome betrug 39 Tage (IQR: 16-64): für Überlebende 63 Tage (IQR: 44-89), für Verstorbene 28 Tage (IQR: 15-43). Basischarakteristika der Kohorte sind in Tabelle 4 aufgeführt.

3.2.2 Longitudinale Veränderungen des Plasma-Proteoms

Wir identifizierten 78 Proteine, deren Konzentrationen sich im zeitlichen Verlauf während der Hospitalisierung veränderten. Bei 14 dieser Proteine unterschied sich die longitudinale Veränderung zwischen Patient:innen, die überlebten, und solchen, die verstarben. Letztere wiesen eine Zunahme pro-inflammatorischer Proteine auf (u.a. SAA1, SAA2, CRP, SERPINA1, LBP), während diese bei den Überlebenden abnahmen. Umgekehrt nahm die Konzentration anti-inflammatorischer Proteine (SERPINA4 und A2M) bei den später verstorbenen Patient:innen im zeitlichen Verlauf ab, ein Hinweis auf andauernde Entzündungssignatur. Analog dazu nahm auch die Konzentration von zwei Schlüsselproteinen der Blutgerinnung, KLKB1 (Kallikrein) und F2 (Thrombin) bei verstorbenen Patient:innen weiter ab, bei überlebenden zu.

Tabelle 4: Basischarakteristika der Kohorte kritisch kranker Patient:innen

	alle Patient:innen		überlebt		verstorben	
Anzahl	50	100 %	35	70 %	15	30 %
Geschlecht						
weiblich	15	30 %	12	34 %	3	20 %
männlich	35	70 %	23	66 %	12	80 %
Alter, Jahre (Median, IQR)	62	54 - 73	61	54 - 69	69	56 - 76
≥65	22	44 %	12	34 %	10	67 %
BMI, kg/m^2 (Median, IQR)	29,4	27,4 - 34,7	31,0	27,4 – 35,1	29,0	25,7 - 31,2
< 25 kg/m^2	9	18 %	6	18 %	3	20 %
≥ 25 kg/m^2	40	82 %	28	82 %	12	80 %
Vorerkrankungen						
CCI (Median, IQR)	3	1 - 4	3	1 - 4	3	3 - 5
<3	18	36 %	16	46 %	2	13 %
≥3	32	64 %	19	54 %	13	87 %
Dauer stationärer Aufenthalt, Tage (Median, IQR)	50	33 - 79	63	44 - 89	28	16 - 43
Bauchlagerung	42	84 %	30	86 %	12	80 %
invasive Beatmung	50	100 %	35	100 %	15	100 %
Hämodialyse	36	72 %	23	66 %	13	87 %
ECMO	19	38 %	12	34 %	7	47 %
Outcome						
verstorben (incl. sekundäre DNI/DNR)	15	30 %	-	-	15	100 %
sekundäre DNI/DNR	3	6 %	-	-	3	6 %

Die Daten stellen Anzahl (n) und Prozent (%) dar, sofern nicht anders angegeben. IQR: Interquartilspanne; CCI: Charlson's Comorbidity Index; ECMO: extrakorporale Membranoxygenierung; DNI/DNR: nicht intubieren/nicht reanimieren; sekundäre DNI/DNR: Therapielimitation bei infauster Prognose entsprechend dem mutmaßlichen Patient:innenwillen. Modifiziert nach Demichev, Tober-Lau, et al., 2022 [61].

3.2.3 Prädiktion von Überleben vs. Versterben mittels Machine Learning

In der klinischen Praxis sind Einzelmessungen einfacher durchführbar als longitudinale Datenerhebungen und können frühzeitig, bspw. noch in der Rettungsstelle, bei der medizinischen Entscheidungsfindung unterstützen. Daher versuchten wir auf Basis eines einzelnen Messpunktes mittels Parenclitic Networks (vgl. Abschnitt 2.5.2.9) vorherzusagen, ob ein:e schwerstkranke:r Patient:in überleben oder versterben wird. Ein Patient mit prolongiertem Krankenhausverlauf und unklarem Outcome wurde für die Generierung

des Modells ausgeschlossen. Als mögliche Merkmale (features) für das Modell wählten wir aus unserem Datensatz die 57 Proteine aus, für die zu dem Zeitpunkt bereits FDA-zugelassene und validierte MRM-Assays öffentlich verfügbar waren (vgl. MRMAssayDB [84]). Unter Anwendung einer Leave-3-Out Kreuzvalidierung berechneten wir die Wahrscheinlichkeit für Überleben bzw. Versterben und klassifizierten 28 von 34 (82 %) Überlebende und 10 von 15 (67 %) Verstorbene richtig (AUROC = 0,81 [95%-CI: 0,68-0,94], P = 0,00038). Von den 25 wichtigsten Proteinen des Modells lassen sich 15 der Blutgerinnung zuordnen und acht gehören zum Komplementsystem. Um die Robustheit unserer Prädiktion zu überprüfen, wiederholten wir die Analyse mit einer SVM. Auch mittels SVM ließ sich das Outcome prädizieren, wenngleich etwas schlechter (AUROC = 0,66 [95%-CI: 0,49-0,84]).

Um das Modell zu validieren, wendeten wir das an der Charité-Kohorte trainierte Parenclitic Networks-Modell auf eine komplett unabhängige Kohorte aus Innsbruck, Österreich, an. Von 24 Patient:innen mit kritischer COVID-19 überlebten 19 (79,2 %) und 5 (20,8 %) verstarben. Die Zeit zwischen Messzeitpunkt und Outcome (Entlassung oder Versterben) betrug im Median 22 Tage (IQR: 15-42). Das an der Charité-Kohorte trainierte Modell konnte auch in dieser unabhängigen Kohorte das Outcome sehr präzise Prädizieren (AUROC = 1,0; P = 0,000047) und klassifizierte 18 der 19 überlebenden und alle 5 verstorbenen Patient:innen korrekt (Abbildung 3). Die Prädiktion mittels SVM gelang ebenfalls, wenngleich wieder etwas weniger genau (AUROC = 0,88 [95%-CI: 0,67-1,0]).

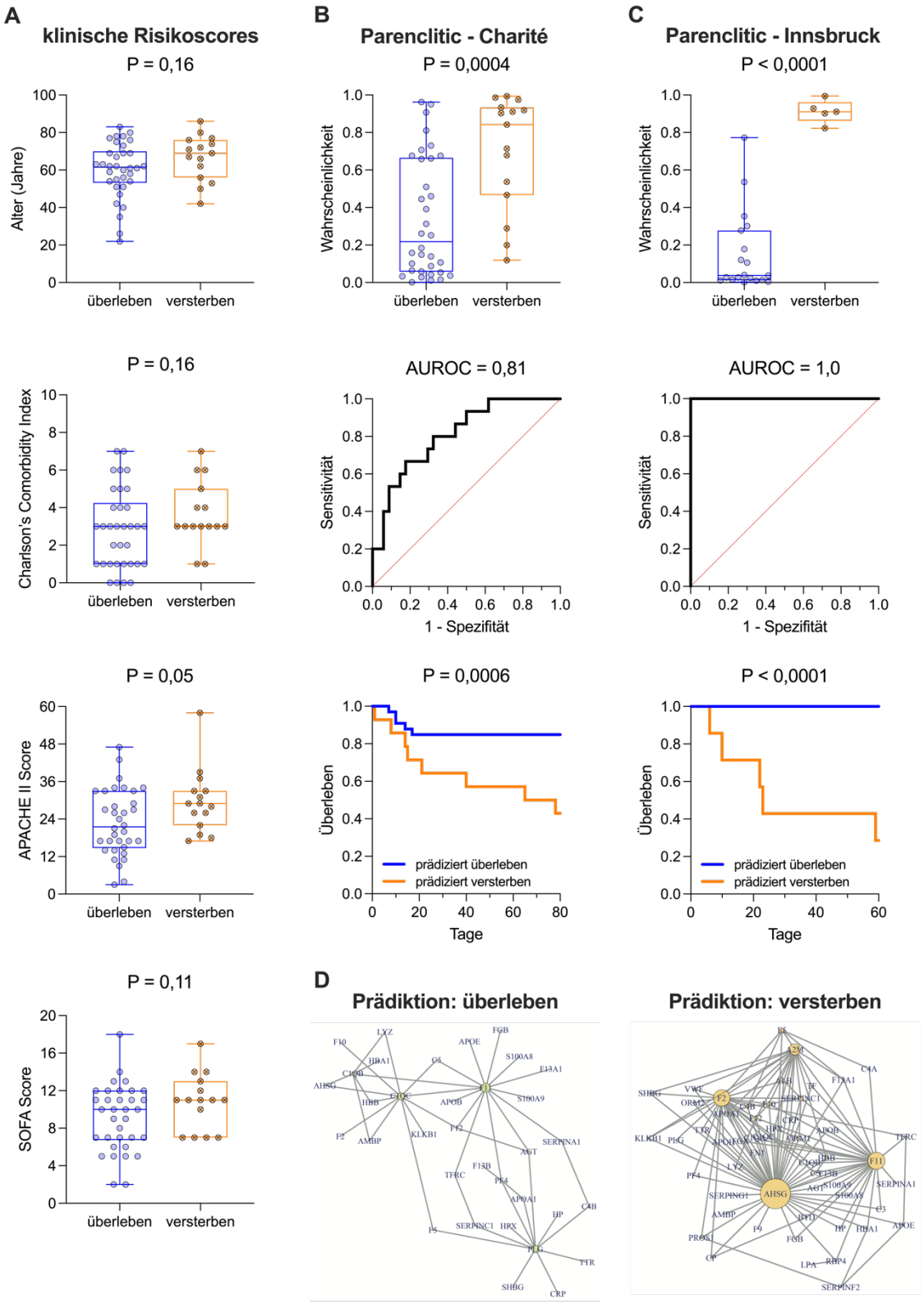


Abbildung 3: Prognose des Outcomes bei kritisch kranken Patient:innen. **A**: Prognostische Aussagekraft von Alter, Vorerkrankungen (CCI) und etablierten klinischen Risiko-Scores (SOFA und APACHE II). CCI und APACHE II wurden zum Zeitpunkt der Aufnahme, SOFA zum ersten Zeitpunkt bei Krankheitsschwere WHO-Grad 7 erhoben. **B**: Prognose von Überleben bzw. Versterben

mittels Parenclitic Networks-Modell, basierend auf dem Plasma-Proteom des ersten Messzeitpunktes bei WHO-Grad 7. Oben: Prognostizierte Wahrscheinlichkeit für Überleben bzw. Versterben. Ein Score von 0,0 entspricht sicherem Überleben, ein Score von 1,0 sicherem Versterben. Mitte: Receiver Operating Characteristic (ROC) -Kurve des Parenclitic Network-Modells. Unten: Kaplan-Meier-Überlebenskurven für Patient:innen mit prognostiziertem Überleben (blau) bzw. Versterben (orange); der Schwellenwert wurde entsprechend der Maximierung des Youden-Index auf 0,678 gesetzt. Die Gruppen wurden mittels Log-Rank-Test verglichen. **C**: Anwendung des an der Charité-Kohorte trainierte Modells auf eine vollständig unabhängige Kohorte kritisch kranker Patient:innen in einem anderen Land; Darstellung entsprechend B. **D**: Beispielhafte Abbildung zweier Parenclitic Networks, wobei ausschließlich Kanten mit einem Gewicht >0,5 abgebildet sind. Links dargestellt ist ein geringeres Sterberisiko bei einem überlebenden, rechts das hohe Sterberisiko bei einem verstorbenen Patienten. Abbildung modifiziert nach Demichev, Tober-Lau, et al. 2022 [61].

3.3 Translation in die klinische Routine: der Peptid-Panelassay

Die in den Abschnitten 3.1 und 3.2 gewonnenen Erkenntnisse zur Bedeutung des Plasma-Proteoms für die Klassifikation und Prognose von Patient:innen mit COVID-19 stellten die Grundlage für die Entwicklung eines Peptid-basierten Panelassays dar, der für die Translation in klinische Routinelabore geeignet ist und aktuell bei der Labor Berlin – Charité Vivantes GmbH validiert wird. Die Entwicklung dieses Assays wurde im peer-reviewed Fachjournal EClinicalMedicine publiziert [57].

3.3.1 Klinische Kohorten

Die Selektion der Peptide basierte auf der in Abschnitt 3.1 der Ergebnisse beschriebenen Kohorte von 139 Patient:innen der Pa-COVID-19 Studie (Tabelle 4). Die Validierung des Assays erfolgte an zwei weiteren im Rahmen der Pa-COVID-19 Studie rekrutierten Kohorten: MRM-Validierungskohorte 1 mit 15 gesunden Kontrollen und 30 sehr frühen COVID-19-Patient:innen (März 2020, Citrat-Plasma, Tabelle 5) sowie MRM-Validierungskohorte 2 mit 164 späteren Patient:innen (bis November 2020, EDTA-Plasma, Tabelle 6).

Tabelle 5: Basischarakteristika MRM-Validierungskohorte 1 (März 2020)

	Patient:innen		**gesunde Kontrollen**	
Anzahl	30	67 %	15	33 %
Geschlecht				
weiblich	10	33 %	11	73 %
männlich	30	67 %	4	27 %
Alter, Jahre (Median, IQR)	62	50-73	30	27-34
BMI, kg/m^2 (Median, IQR)	25,7	23,7-28,5	n/a	n/a
Erkrankungsschwere				
mild: WHO-Grad 3	10	33 %	-	-
moderat: WHO-Grad 4 oder 5	7	23 %	-	-
schwer: WHO-Grad 6 oder 7	13	43 %	-	-
Dauer stationärer Aufenthalt, Tage (Median, IQR)	17	9-38	-	-
Dexamethasontherapie	0	0 %	-	-
invasive Beatmung	13	43 %	-	-
Hämodialyse	8	73 %	-	-
ECMO	6	20 %	-	-

Wenn nicht anders angegeben, stellen die Daten als Anzahl (n) und Prozent (%) dar. IQR: Interquartilsspanne; BMI: Body Mass Index; ECMO: extrakorporale Membranoxygenierung. Eigene Darstellung.

Tabelle 6: Basischarakteristika MRM-Validierungskohorte 2 (bis November 2020)

			keine invasive Beatmung						invasive Beatmung			
	alle Patient:innen		max. WHO-Grad 3		max. WHO-Grad 4		max. WHO-Grad 5		max. WHO-Grad 6 oder 7		verstorben**	
Anzahl	164	100 %	23	14 %	42	26 %	38	23 %	61	37 %	29	18 %
Geschlecht												
weiblich	39	24 %	7	30 %	10	24 %	10	26 %	12	20 %	5	17 %
männlich	125	76 %	16	70 %	32	76 %	28	74 %	49	80 %	24	83 %
Alter, Jahre (Median, IQR)	60	51-69	52	41-61	57	51-66	63	57-76	62	54-70	64	56-71
≥65	57	35 %	3	13 %	12	29 %	18	47 %	24	39 %	13	45 %
BMI, kg/m^2 (Median, IQR), n=139	29,4	24,7-32,5	25,6	23,4-30,7	28,8	24,9-33,0	29,6	24,8-33,2	29,4	25,8-32,6	29,4	25,6-30,9
< 25 kg/m^2	38	27 %	9	45 %	9	25 %	8	31 %	12	21 %	6	22 %
≥ 25 kg/m^2	101	73 %	11	55 %	27	75 %	18	69 %	45	79 %	21	78 %
Vorerkrankungen												
CCI (Median, IQR)	2	1-4	2	0-4	2	1-3	3	2-4	3	1-3	3	2-5
<3	83	53 %	13	59 %	28	70 %	16	43 %	26	45 %	8	30 %
≥3	74	47 %	9	41 %	12	30 %	21	57 %	32	64 %	19	70 %
Dauer stationärer Aufenthalt*, Tage (Median, IQR)	12	8-29	7	5-10	9	6-12	15	11-20	48	30-79	34	23-53
Dexamethasontherapie	117	71 %	3	13 %	36	86 %	32	84 %	46	75 %	23	79 %
invasive Beatmung**	61	37 %	-	-	-	-	-	-	61	100 %	29	100 %
Hämodialyse	33	20 %	0	0 %	0	0 %	1	3 %	32	52 %	22	76 %
ECMO**	24	15 %	-	-	-	-	-	-	24	39 %	19	66 %
Outcome												
verstorben (incl. DNI/DNR)	34	21 %	0	0 %	0	0 %	4	11 %	30	49 %	29	100 %
DNI/DNR	5	3 %	0	0 %	0	0 %	4	11 %	1	2 %	-	-
sekundäre DNI/DNR	7	4 %	0	0 %	0	0 %	0	0 %	7	11 %	2	24 %
klinische Risikoscores												
APACHE II (Median, IQR)	21	14-28	-	-	-	-	12	10-15	26	18-31	28	21-33
SOFA (Median, IQR)	6	2-11	-	-	-	-	2	0-3	9	6-12	12	8-13
ABCS (Aufnahme; Median, IQR)	0,09	0,02-0,2	0,02	0-0,09	0,02	0,02-0,09	0,09	0,02-0,2	0,2	0,09-0,43	0,3	0,09-0,6
ABCS (Messzeitpunkt; Median, IQR)	0,09	0,02-0,25	0,02	0-0,02	0,02	0,02-0,09	0,09	0,02-0,14	0,2	0,09-0,58	0,3	0,2-0,6

Die Daten stellen Anzahl (n) und Prozent (%) dar, sofern nicht anders angegeben. APACHE II wurde bei Aufnahme erhoben, SOFA zum Messzeitpunkt , ABCS zu beiden Zeitpunkten. IQR: Interquartilspanne; CCI: Charlson's Comorbidity Index; ECMO: extrakorporale Membranoxygenierung; DNI/DNR: nicht intubieren/nicht reanimieren; sekundäre DNI/DNR: Therapielimitation bei infauster Prognose entsprechend dem mutmaßlichen Patient:innenwillen; SOFA: Sequential Organ Failure Assessment; APACHE II: Acute Physiology And Chronic Health Evaluation; ABCS: Age, Biomarkers, Clinical history, Sex score. Eigene Darstellung.

* exklusive verstorbene Patient:innen

** exklusive Patient:innen mit DNI/DNR

3.3.2 Peptidselektion und Panel-Aufbau

Insgesamt wurden 50 Peptide selektiert, die 30 Proteinen entsprechen. Die Selektion erfolgte aus dem Pool der mittels Discovery-Proteomik (vgl. Abschnitt 3.1) identifizierten Peptide bzw. Proteine, unter Berücksichtigung verschiedener Faktoren. Peptide galten als prinzipiell geeignet, wenn sie

1. prognostisch für die verbleibende Zeit im Krankenhaus,
2. signifikant unterschiedlich exprimiert zwischen verschiedenen WHO-Schwergraden oder
3. prognostisch für eine zukünftige Verschlechterung des WHO-Schweregrades

waren. Bei einem Teil der ausgewählten Peptide bzw. deren korrespondierenden Proteinen handelte es sich um bereits in der klinischen Routine etablierte Marker (z.B. CRP – C-reaktives Protein, SERPINC1 – Antithrombin-III) bzw. bei MRMAssayDB [84] hinterlegte Peptide. Zusätzlich zur medizinisch-statistischen Auswahl der Marker-Peptide wurden deren physikochemischen Eigenschaften berücksichtigt, um die Eignung zur chemischen Synthese (evaluiert mit dem Peptide Synthesis and Proteotypic Peptide Analyzing Tool, Thermo Fisher Scientific Inc., Waltham, MA, USA [85]) und eine breite chromatographische Verteilung der Peptide sicherzustellen. Von den 30 final selektierten Proteinen waren 18 (60 %) prognostisch für die verbleibende Zeit bis zur Entlassung aus dem Krankenhaus, 22 (73 %) korrelierten mit der Krankheitsschwere und 6 (20 %) waren prognostisch für eine zukünftige Zunahme der Krankheitsschwere, d.h. Anstieg des WHO-Grades (Tabelle 7). Von den 50 selektierten Peptiden sind 45 (90 %) proteotypisch für ein Protein; 5 Peptide können mehreren ähnlichen Protein-Isoformen zugeordnet werden.

Für jedes selektierte Peptid wurde zur Optimierung der Datenerhebung sowie der Erstellung einer Standardkurve ein nativer Peptidstandard (Reinheit ≥95 %) synthetisiert. Benchmarks ergaben eine geringe Varianz innerhalb und zwischen Batches (2,6 % innerhalb, 10,9 % zwischen). Des Weiteren wurden als interner Standard zur absoluten Quantifizierung Isotopen-markierte Peptide (SIL, stable isotope labeled) synthetisiert. Das untere Detektionslimit (LLOQ, lower limit of quantification) betrug im Median 143,6 ng/ml, eine lineare Quantifizierung gelang meist über 3-4 Größenordnungen.

Tabelle 7: Peptide im MRM-Panelassay

Peptid-Sequenz	Gen-Name	aktueller Schweregrad	Zunahme des Schweregrads	Zeit bis Entlassung
EITALAPSTMK	ACTA1			
HFQNLGK, TINPAVDHCCK	AFM	*		
TVVQPSVGAAAGPVVPPCPGR, CNLLAEK	AHSG			
AHVDALR, ATEHLSTLSEK	APOA1			
IAELSATAQEIIK, EQHLFLPFSYK	APOB			
FNAVLTNPQGDYDTSTGK, TNQVNSGGVLLR	C1QC			
VEGTAFVIFGIQDGEQR, VHQYFNVELIQPGAVK	C3	*		
VLDLSCNR	CD14			
ESDTSYVSLK, GYSIFSYATK	CRP			
ALDFAVGEYNK, LVGGPMDASVEEEGVRR	CST3			
ADQVCINLR	EFEMP1	*		
DSGSYFCR	FCGR3A	*		
GGEGTGYFVDFSVR, IADAHLDR	HRG			
DFALQNPSAVPR, LAELPADALGPLQR	IGFALS			
ASDTAMYYCAR	IGHV5-10-1; IGHV5-51	*		
GHMLENHVER	ITIH1			
DSVTGTLPK, IAYGTQGSSGYSLR	KLKB1			
STDYGIFQINSR, YWCNDGK	LYZ			
SDVMYTDWK	ORM2	*		
GCPDVQASLPDAK, TFTLLDPK	PGLYRP2			
EAQLPVIENK, CQSWSSMTPHR	PLG			
GLPNVVTSAISLPNIR	PRG4	*		
WEMPFDPQDTHQSR, EQLSLLDR	SERPINA3			
ANRPFLVFIR	SERPINC1			
TSCLLFMGR	SERPIND1			
LLDSLPSDTR, LVLLNAIYLSAK	SERPING1			
GDVAFVK, WCALSHHER	TF			
ILNIFGVIK, VSASPLLYTLIEK	TFRC			
GSPAINVAVHVFR, AADDTWEPFASGK	TTR			
ILTSDVFQDCNK, YAGSQVASTSEVLK	VWF			

Tabelle modifiziert nach Wang, Cryar, et al., 2022 [57]. Die Auswahl der Peptide erfolgte auf Grundlage ihres Potenzials zur Klassifikation der aktuellen oder Prognose zukünftiger Zunahme der Krankheitsschwere bzw. verbleibender Zeit im Krankenhaus bei Patient:innen mit WHO-Grad 3. Rot bzw. blau zeigt die Hoch- bzw. Herunterregulation eines Proteins bei COVID-19 an.

* signifikante Assoziation auf Peptidebene

Die absolute Quantifizierung von Peptiden in Plasma-Pools von COVID-19-Patient:innen mit verschiedener Krankheitsschwere zeigte, dass das gesamte Spektrum möglicher Peptidkonzentrationen im linearen Bereich der Standardkurve abgedeckt war, der Panelassay somit grundsätzlich zur absoluten Quantifizierung von Patient:innenproben geeignet ist.

3.3.3 Korrelation mit Krankheitsschwere

Zunächst analysierten wir MRM-Validierungskohorte 1 (30 Patient:innen mit variabler Krankheitsschwere, 15 gesunde Kontrollen, Tabelle 5) mit dem Panelassay. Wir quantifizierten 40 (80 %) der 50 Peptide zuverlässig in allen Proben; 32 Peptide wiesen in Abhängigkeit vom WHO-Grad verschiedene Konzentrationen auf. Der Großteil dieser Proteine ermöglichte eine zuverlässige Differenzierung zwischen gesunden Kontrollen und mild Erkrankten, wobei die Konzentration bei Zunahme der Krankheitsschwere den Trend fortsetzte und entsprechend weiter anstieg (z.B. CRP) bzw. abfiel (z.B. AHSG). Die Messung von MRM-Validierungskohorte 2 (164 Patient:innen, Tabelle 6) bestätigte diese Ergebnisse und erlaubte die Klassifikation von Patient:innen mit verschiedener Krankheitsschwere.

3.3.4 Machine Learning-Prädiktionsmodelle

Um dem Ziel eines prognostischen Panels näherzukommen, versuchten wir mittels SVM-Modell bei einem gegebenen Sample die aktuelle Krankheitsschwere (mild: WHO-Grad 3; moderat: WHO-Grad 4 und 5; schwer: WHO-Grad 6 und 7) zu prädizieren. Von den 164 Patient:innen der MRM-Validierungskohorte 2 wurden 109 (66,5 %) korrekt klassifiziert; 3 (1,8 %) Patient:innen mit milder Erkrankung wurden fälschlicherweise als kritisch erkrankt klassifiziert und 2 (1,2 %) mit kritischer Erkrankung als mild. Bei den verbleibenden 50 (30,5 %) Patient:innen wich die prädizierte um eine Stufe von der tatsächlichen Krankheitsschwere ab (Abbildung 4).

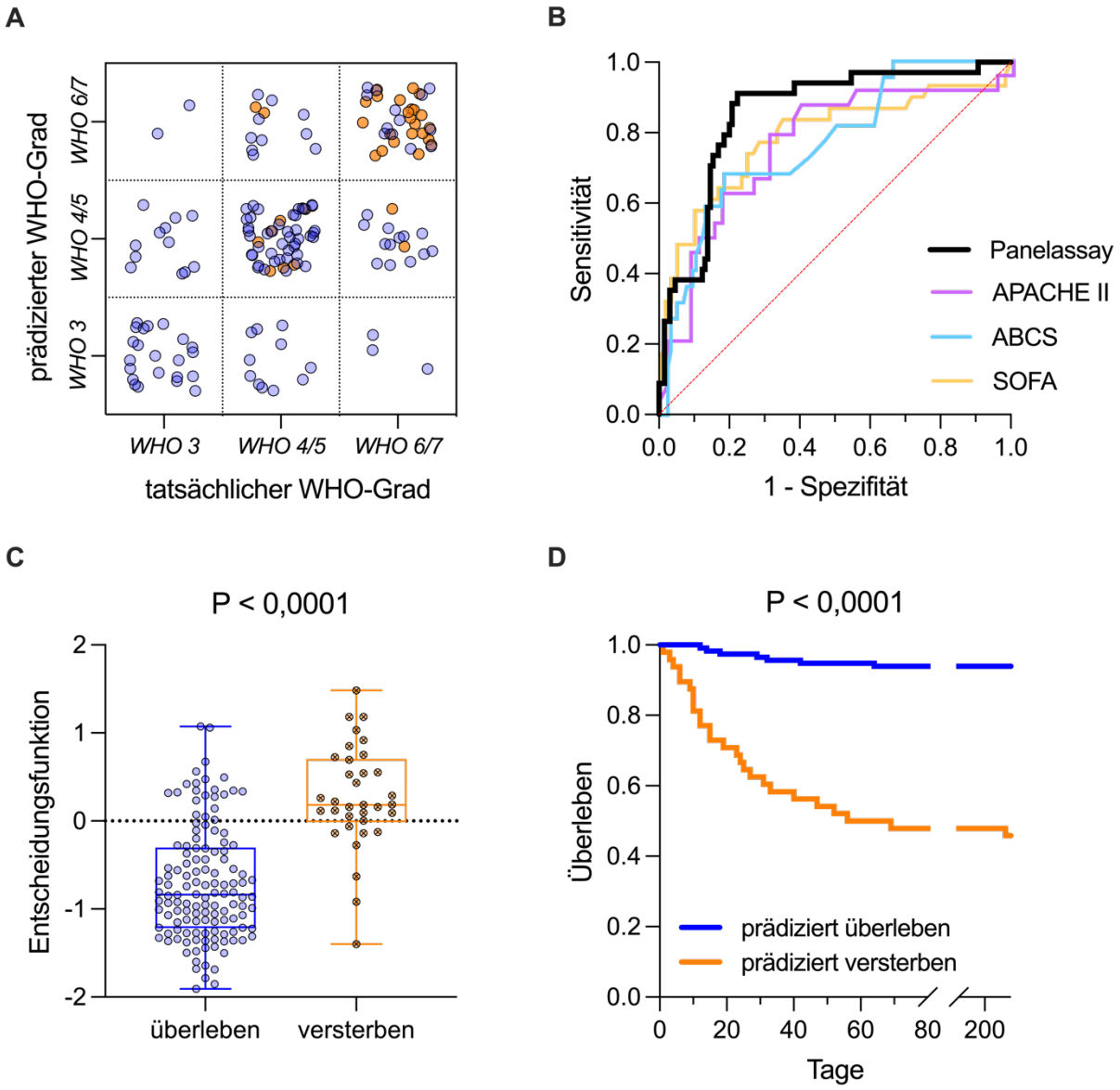
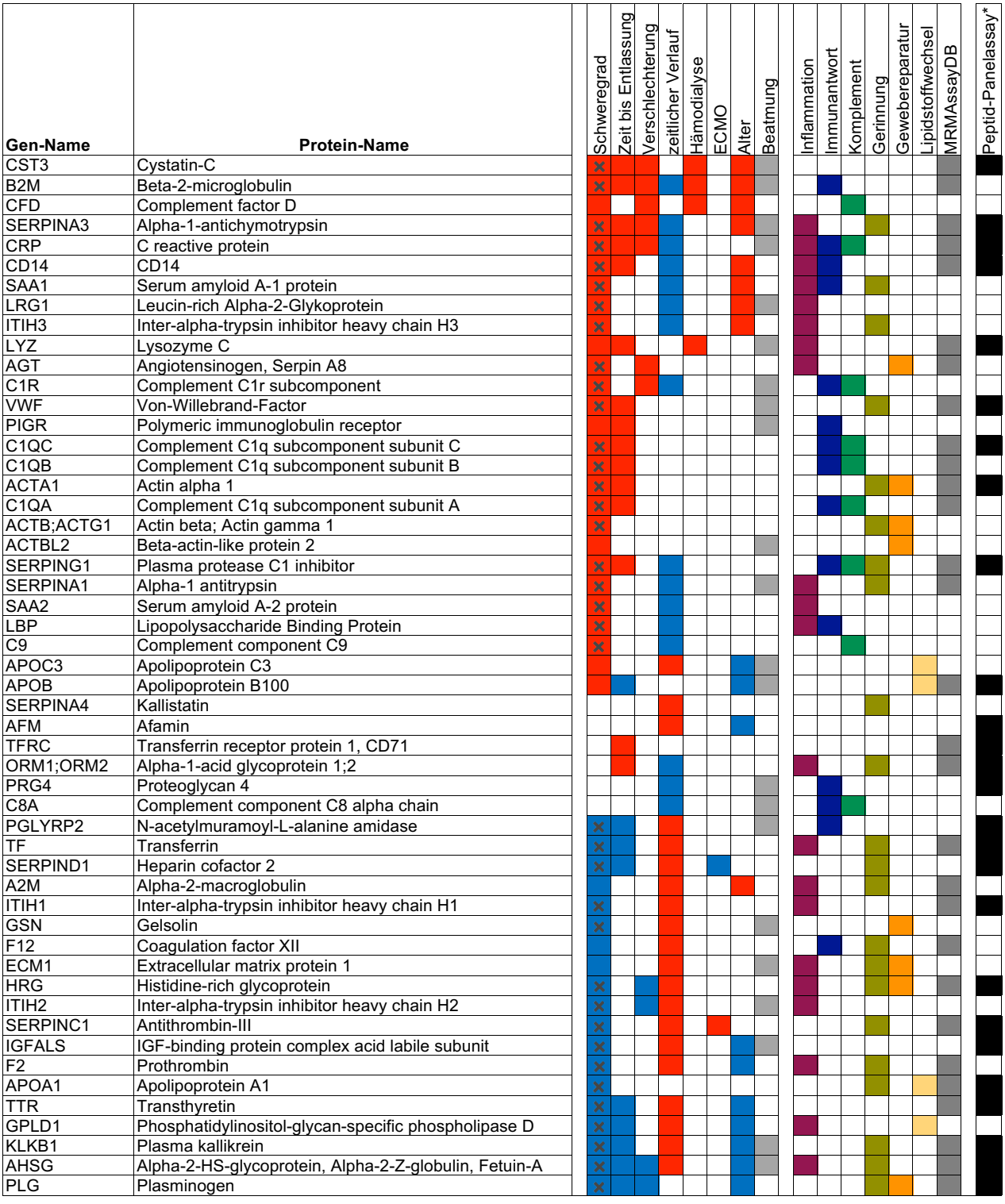


Abbildung 4: Prädiktion von Krankheitsschwere und Überleben bzw. Versterben mittels Peptid-Panelassay. **A**: Von den 164 Patient:innen der Kohorte 3 wurden 109 (66,5 %) korrekt klassifiziert; 3 (1,8 %) Patient:innen mit milder Erkrankung (WHO-Grad 3) wurden fälschlicherweise als kritisch erkrankt (WHO-Grad 6/7) klassifiziert und 2 (1,2 %) mit kritischer Erkrankung als mild. Bei den verbleibenden 50 (30,5 %) Patient:innen wich die prädizierte Krankheitsschwere um eine Stufe von der tatsächlichen ab. Blaue Punkte: überlebende Patient:innen; orange Punkte: verstorbene Patient:innen. **B**: ROC-Kurve für die Prädiktion von Überleben vs. Versterben mittels SVM-Classifier. Der Peptid-Panelassay bot eine bessere prognostische Aussagekraft als etablierte klinische Risiko-Scores. **C**: Boxplot der Entscheidungsfunktion des SVM-Modells für überlebende bzw. verstorbene Patient:innen. Werte <0 prognostizieren Überleben, Werte >0 prognostizieren Versterben. **D**: Kaplan-Meier-Überlebenskurve für prädiziertes Überleben (blau) bzw. Versterben (orange). Abbildung modifiziert nach Wang, Cryar, et al. 2022 [57].

Abschließend versuchten wir, wieder unter Anwendung einer SVM, ein Modell zur Prognose des Outcomes, d.h. Überleben oder Versterben, von der frühesten Probe nach Krankenhausaufnahme zu generieren. Dieser Zeitpunkt bietet in der klinischen Routine den potenziell größten Nutzen, bspw. bei Vorstellung in der Rettungsstelle. Unter Anwendung einer Leave-3-Out Kreuzvalidierung prädizierten wir 26 von 34 (76,5 %) verstorbene und 108 von 130 (83,1 %) überlebende Patient:innen korrekt (AUROC = 0,86 [95%-CI: 0,78 – 0,92]). Auch zwei weitere Methoden (logistische Regression, extra-trees) zeigten vergleichbare Ergebnisse (AUROC = 0,85 [95%-CI: 0,78-0,92] bzw. AUROC = 0,84 [95%-CI: 0,76 – 0,91]). Etablierte klinische Scores boten allesamt eine geringere diagnostische Aussagekraft als der Peptid-Panelassay (APACHE II: AUROC = 0,76 [95%-CI: 0,64 – 0,89]; ABCS: AUROC = 0,77 [95%-CI: 0,66 – 0,87]; SOFA: AUROC = 0,79 [95%-CI: 0,68 - 0,90]; Abbildung 4).

Zusammenfassend konnten wir zeigen, dass die Klassifikation der Krankheitsschwere von COVID-19 sowie die frühzeitige Prognose von Überleben bzw. Versterben keineswegs der Erfassung des gesamten Plasma-Proteoms bedarf, sondern auch mit einer Selektion zuverlässig quantifizierbarer Peptide auf bereits in klinischen Routinelaboren genutzten Geräten gelingen kann.

Tabelle 8: Zusammenfassung der Ergebnisse (Auswahl)



Darstellung modifiziert nach Demichev, Tober-Lau, et al., 2021 [56]. Zusammenfassung der Ergebnisse sowie funktionelle Zuordnung der im Rahmen dieser Arbeit beschriebenen Proteine. Rote Kästchen: positive Korrelation; blaue Kästchen: negative Korrelation mit dem jeweiligen Endpunkt. Assoziation mit Beatmung bezieht sich auf das in Abschnitt 3.1.5 beschriebene Machine Learning-Modell.

✖ bekannte Assoziationen mit Krankheitsschwere

* exklusive ausschließlich auf Peptidebene signifikanter Proteine

4 Diskussion

4.1 Zusammenfassung der Ergebnisse

Moderne LC-MS/MS-basierte Proteomik ermöglicht die präzise Quantifizierung von über 300 Plasma-Proteinen und somit die umfassende Charakterisierung der Wirtsantwort auf einen Erreger. Indem wir longitudinal 687 Plasma-Proteome von 139 COVID-19-Patient:innen mit verschiedenster Krankheitsschwere erfassten und mit umfangreichen klinischen und Routinelaborparametern ergänzten, identifizierten wir robuste Kovariationen von klinischen Entzündungsmarkern mit Plasma-Proteinen der Immunantwort, Entzündungsreaktion und Blutgerinnung. Auch konnten wir spezifische Signaturen von Therapien wie Hämodialyse und ECMO nachweisen. Insgesamt näherten sich bei den meisten Patient:innen die Plasma-Proteomprofile im zeitlichen Verlauf der Baseline an, d.h. initial erhöhte Proteinkonzentrationen nahmen ab, erniedrigte stiegen an, und zwar unabhängig von der Krankheitsschwere. Dies galt nicht für Patient:innen, die verstarben: hier konnten wir die charakteristische Linderung des Proteoms nicht beobachten, im Gegenteil, die pro-inflammatorische Signatur nahm im zeitlichen Verlauf sogar zu.

Diese umfangreichen Informationen zur Wirtsantwort auf SARS-CoV-2 veranlasste uns zu untersuchen, ob auch einzelne, frühe Messzeitpunkte prognostisch aussagekräftig sind. Tatsächlich ließ sich bei mild Erkrankten die Zeit bis zur Entlassung aus dem Krankenhaus vorhersagen; bei kritisch kranken Patient:innen war es möglich, teils Wochen im Voraus deren Überleben bzw. Versterben prädizieren.

Um die Ergebnisse der Discovery-Proteomik in klinische Routinelabore zu translatieren und somit einen direkten klinischen Nutzen zu ermöglichen, identifizierten wir 50 klassifizierende bzw. prognostische, zuverlässig quantifizierbaren Peptide, entsprechend 30 Proteinen, und entwickelten bzw. validierten an zwei weiteren klinischen Kohorten mit 30 bzw. 164 Patient:innen mit COVID-19 einen MRM-basierten Peptid-Panelassay. Der Panelassay ermöglichte bereits zum frühesten Messzeitpunkt nach Aufnahme die präzise Klassifikation der Krankheitsschwere und erlaubte die Prognose von Überleben bzw. Versterben, wobei er etablierten klinischen Mortalitäts-Risikoscores wie SOFA oder APACHE II überlegen war.

4.2 Interpretation und Einbettung in den bisherigen Forschungsstand

Das Auftreten und die rasche Ausbreitung der COVID-19-Pandemie hat uns ins Bewusstsein gerufen, dass selbst als robust geltende Gesundheitssysteme im Angesicht eines neuartigen Infektionserregers rasch an ihre Grenzen stoßen können. Insbesondere bei sehr variablen Krankheitsverläufen, wie denen von COVID-19, und in Situationen von Ressourcenknappheit ist daher eine fundierte und akkurate Einschätzung von Krankheitsschwere und -prognose entscheidend für die adäquate und effektive Versorgung von Patient:innen. So konnte bspw. gezeigt werden, dass eine frühzeitige Verlegung auf die Intensivstation das Outcome von Patient:innen mit schwerer COVID-19 verbessert [86]; jedoch führte deren Überlastung während der Pandemie zu einer erhöhten Mortalität [87].

Anders als bspw. bei schwerer bakterieller Pneumonie, spiegelt der klinische Phänotyp insbesondere in der frühen Phase von COVID-19 oftmals nicht das tatsächliche Ausmaß der Erkrankung wider. So können sich klinisch lediglich mit Abgeschlagenheit präsentierende Patient:innen zwar noch selbständig in ärztliche Versorgung begeben, eine Pulsoxymetrie offenbart jedoch eine deutlich reduzierte Sauerstoffsättigung als Zeichen einer fortgeschrittenen respiratorischen Insuffizienz im Sinne einer „happy hypoxaemia“ [15]. Schon früh wurden verschiedene klinische Charakteristika und Routine-Laborparameter für die Klassifikation von Krankheitsschwere sowie Prädiktion klinischer Outcomes wie Hospitalisierung, Intubationspflichtigkeit oder Versterben identifiziert. Diese basierten zumeist auf einzelnen oder Kombinationen weniger Parameter, bspw. Vitalzeichen, Ergebnisse bildgebender Verfahren oder Entzündungsmarker wie Blutzellzahlen, Interleukin-6 oder Ferritin [31,32,88]. Bereits etablierte Risikoscores wie APACHE II zur Prognose von Mortalität bei Intensivpatient:innen, aber auch neu entwickelte Modelle und Risikoscores wie der ABCS oder 4C-Mortality Score blieben bei dieser neuartigen Erkrankung in ihrer Aussagekraft eingeschränkt [23,32,37,38,89].

Im Gegensatz dazu ermöglicht das Plasma-Proteom mit mehreren hundert Parametern eine umfassende Charakterisierung des klinisch möglicherweise noch inapparenten molekularen Phänotyps und erlaubt somit direkte Rückschlüsse auf die zugrundeliegenden pathophysiologischen Prozesse bei Patient:innen mit COVID-19.

So beobachteten wir, dass sich sowohl bei den Korrelationen als auch den Machine Learning-Modellen viele der relevantesten Marker, die zur Klassifikation der Krankheitsschwere sowie zur Prognose des weiteren Krankheitsverlaufs geeignet waren, frühen Schritten der Entzündungsaktivierung zuordnen ließen. Diese umfassen u.a. die Komplementaktivatoren C1QA-C, C1R und CFD sowie Schlüsselproteine der Blutgerinnung wie PLG (Plasminogen), F2 (Prothrombin) und KLKB1 (Kallikrein), einem entscheidenden Bindeglied zwischen den Systemen [90]. Die Bedeutung der Dysregulation von Komplement- und Gerinnungssystem für die Pathophysiologie von COVID-19 wurde inzwischen vielfach bestätigt und nimmt mit Dexamethason und therapeutischer Antikoagulation eine zentrale Rolle bei der Therapie stationär behandelter Patient:innen ein [91–93].

Im Gegensatz dazu lassen sich viele Proteine, die ausschließlich zur Klassifikation der Krankheitsschwere, nicht jedoch zur Prädiktion des weiteren Krankheitsverlaufs geeignet sind, Downstream-Effektoren sowie Gewebeschädigung zuordnen, darunter GSN (Gelsolin), zirkulierende Actine (ACTB, ACTBL2, ACTG1) und ECM1. Sie scheinen also die pathophysiologischen Auswirkungen der dysregulierten Entzündungsreaktion widerzuspiegeln.

Bei der Prädiktion von Überleben bzw. Versterben kritisch kranker Patient:innen mittels Machine Learning-Modellen stellte sich AHSG (Fetuin-A, Alpha-Heremann-Schmidt-Glykoprotein) als Protein mit der höchsten Relevanz im Modell dar; es ist bei schwerer COVID-19 herunterreguliert [56,94]. Bei SARS-CoV-1 wurde beobachtet, dass Polymorphismen, die zu einer erhöhten Konzentration von AHSG führen, protektiv wirken [95]. Eine Schlüsselfunktion von AHSG ist die Dämpfung der Makrophagenaktivität [78]; deren Dysregulation spielt eine entscheidende Rolle bei der Lungendysfunktion von COVID-19 [77,96,97]. Niedrige AHSG-Spiegel könnten somit frühzeitig auf eine relevante Lungenbeteiligung hinweisen.

Insgesamt lassen sich viele der differentiell exprimierten Proteine der Akutphasereaktion zuordnen und sind auch bei anderen Entzündungszuständen wie bspw. Sepsis dysreguliert [98]. LRG1 und LBP sowie erniedrigtes A2M weisen auf eine andauernde Immunantwort hin und exazerbieren den pro-inflammatorischen Zustand [98,99]. Die Dysregulation verschiedener Proteine des Lipidmetabolismus, u.a. APOC3 und GPLD1, ist auch bei Patient:innen mit bakterieller Pneumonie bekannt und prognostisch ungünstig [100]. Es

scheint sich hier also nicht um eine COVID-19-spezifische, sondern eine allgemeine pro-inflammatorische Signatur zu handeln. Beim Vergleich mit einer gesunden Vergleichskohorte aus der prä-COVID-19-Zeit (Generation Scotland Kohorte [55]) beobachteten wir zudem eine altersspezifische Zunahme der basalen Inflammation, ein möglicher Faktor für das erhöhte Risiko von schwerer COVID-19 bei älteren Menschen [101].

Im Rahmen dieser Arbeit konnten wir demonstrieren, wie mittels Hochdurchsatz-Proteomik bei einer neuartigen Erkrankung rasch umfassende Rückschlüsse auf die zugrundeliegende Pathophysiologie gezogen werden. Besonders vielversprechend ist, dass eine Arbeitsgruppe in den Niederlanden ebenfalls einen Großteil der bei uns signifikanten Proteine als prognostisch relevant für das Outcome bei COVID-19 identifizierte [102].

Darüber hinaus erlauben longitudinale Messungen dieser klassifizierenden und prognostischen Marker ein effizientes und umfassendes Therapiemonitoring. Insbesondere bei der Überwachung experimenteller Interventionen im Rahmen klinischer Studien können so im Idealfall bereits frühzeitig Interventionen auf der molekularen Ebene als Abweichung vom prädizierten Verlauf erkennbar gemacht und so individuelle Rückschlüsse auf deren Effektivität gezogen werden.

Besonders wertvoll für die Anwendung in der klinischen Routine sind jedoch einzelne Messungen zur Unterstützung von Schlüsselentscheidungen im Krankheitsverlauf: in der Rettungsstelle um Patient:innen mit stationärem Behandlungsbedarf zu identifizieren; auf der Normalstation, um die Notwendigkeit der Therapieeskalation im Sinne einer invasiven Sauerstofftherapie und intensivmedizinischen Überwachung zu bestimmen; und auf der Intensivstation, um den Bedarf zusätzlicher Organersatzverfahren zu eruieren sowie, bei Ressourcenknappheit, deren effektivsten Einsatz zu prognostizieren.

Mittels Machine Learning-Modell ließ sich, basierend auf klinischen bzw. proteomischen Markern sowie einer Kombination aus beiden, eine graduierte Klassifikation der zum Messzeitpunkt vorliegenden Krankheitsschwere treffen. Die Prognostizierung des weiteren Krankheitsverlaufs stellt eine größere Herausforderung dar, insbesondere bei der relativ homogenen Gruppe der schwerstkranken Patient:innen mit COVID-19, die invasiv mechanisch beatmet werden und zusätzliche Organersatztherapien benötigen [19,103,104]. Diese Gruppe benötigt jedoch die meisten Ressourcen, sodass die

Entscheidung für oder gegen eine entsprechende Therapie auf Basis bestmöglicher Informationen erfolgen sollte und nicht ausschließlich auf Faktoren wie Alter, Vorerkrankungen oder Einzelmesswerten bzw. klinischen Scores wie SOFA mit eingeschränkter Übertragbarkeit auf COVID-19 [105,106]. Dies gilt umso mehr in Situationen, in denen Ressourcenknappheit droht oder bereits herrscht. Ebenfalls mittels verschiedener Machine Learning-Modelle konnten wir bei dieser Patient:innengruppe eine akkurate Prädiktion hinsichtlich Überleben treffen, im Median 39 Tage vor dem jeweiligen Outcome. Die gute Performance der Machine Learning-Modelle auch bei einer komplett unabhängigen Kohorte von einem anderen Krankenhaus in einem anderen Land weist auf das robuste translationale Potenzial dieser Technologie hin.

Die Translation der Discovery-Proteomik in die klinische Routine stellte daher den für uns nächsten logischen Schritt dar. Für die Etablierung in der Routinediagnostik bedarf es standardisierter Assays, die eine hohe Vergleichbarkeit und Reproduzierbarkeit aufweisen und idealerweise die absolute Quantifizierung von Proteinen erlauben. Methoden wie (Multiplex-) ELISA sind zeit- und materialaufwendig sowie, aufgrund hersteller- und chargenabhängiger Unterschiede der Affinitätsreagenzien, nur eingeschränkt vergleichbar. Auch deswegen sind sie in der Anzahl verschiedener Zielmoleküle eingeschränkt und daher wenig geeignet, die Breite der mittels Discovery-Proteomik identifizierten Marker nachzuweisen [107]. Moderne MS-basierte Assays hingegen bieten eine hohe Sensitivität und Spezifität, auch kann durch Hinzufügen interner Standards eine hohe Reproduzierbarkeit und absolute Quantifizierung über eine breite Konzentrationsspanne von 3-4 Größenordnungen [47,108,109] und aus verschiedenen Materialien (EDTA-, Heparin-, Citratplasma, Serum) [110,111] erreicht werden. Schon heute werden QqQ-Massenspektrometer in klinischen Routinelaboren, bspw. zum Nachweis kleiner Moleküle im Rahmen des Neugeborenenscreenings, angewendet. Auf diesen ließen sich Peptid-Panelassays verhältnismäßig einfach in bestehende Arbeitsabläufe integrieren [107,112,113] und Unterschiede bei Affinitätsreagenzien, wie sie bspw. bei der Bestimmung von Tumormarkern vorkommen, vermeiden.

Basierend auf einer Auswahl von 50 der bei der Discovery-Proteomik identifizierten Peptide ermöglichte unser Panelassay die zuverlässige Klassifikation und Prognose von Patient:innen mit COVID-19 verschiedener Krankheitsschwere und in verschiedenen Kohorten vor und nach Einführung von Dexamethason, aus verschiedenen Plasmen (EDTA,

Citrat) sowie auf zwei verschiedenen LC/MRM-MS-Systemen, und war dabei allen untersuchten klinischen Risikoscores überlegen.

4.3 Stärken und Schwächen der Studie

Mit insgesamt 280 Patient:innen und 881 Messzeitpunkten umfasst die vorliegende Arbeit eine der bis dahin größten proteomisch charakterisierten COVID-19-Kohorten. Entsprechend dem Protokoll der Pa-COVID-19 Studie [51], wurden umfangreiche klinische Daten zu Demographie, Risikofaktoren und medizinischer Vorgeschichte erhoben sowie tagesaktuell die Krankheitsschwere standardisiert erfasst und um Routinelaborparameter ergänzt. Jedoch gab es, trotz der umfangreichen Kohorte, verhältnismäßig wenige Patient:innen, die sich von einem niedrigen WHO-Grad 3 oder 4 bis hin zur beatmungspflichtigkeit verschlechterten. Ebenfalls wurden hier ausschließlich stationär behandelte Patient:innen untersucht; die Übertragbarkeit auf mild erkrankte, ambulant behandelte Patient:innen sollte in weiteren Studien untersucht werden.

Die Charité – Universitätsmedizin Berlin ist im Rahmen des „Save Berlin@COVID-19"-Konzepts der Senatsverwaltung für Gesundheit, Pflege und Gleichstellung Berlin als Berlins einziges Level-1 Zentrum, d.h. ARDS/ECMO-Zentrum, designiert worden [114]; folglich wurden viele Patient:innen bereits intubiert ans Zentrum verlegt, sodass frühe Proben, vor Erreichen der kritischen Erkrankung, häufig nicht verfügbar waren. Auch wurden einige Patient:innen zur Beatmungsentwöhnung an periphere Häuser und Rehakliniken verlegt, sodass ein nach Verlegung eintretender Tod theoretisch unbekannt bleiben konnte. Im Zuge dieser Arbeit wurde die Verlaufsdokumentation erneut auf Hinweise zum Outcome durchsucht (z.B. Notiz über Versterben des Patienten nach Verlegung). Da die Verlegung stets in einer späten Erkrankungsphase erfolgte und dezidiert zur Beatmungsentwöhnung, kann jedoch davon ausgegangen werden, dass diese Patient:innen überlebten.

Obwohl die Gruppe der kritisch kranken Patient:innen mit n=50 und auch die für die Entwicklung des Peptid-Panelassays verwendete Kohorte mit n=164 aus klinischer Perspektive verhältnismäßig groß sind, werden zur Validierung der Machine Learning-Modelle selbstverständlich deutlich größere Kohorten benötigt. Vielversprechend ist die Beobachtung, dass der an der Charité-Kohorte trainierte Prädiktor bei der komplett unabhängigen

Innsbruck-Kohorte ebenfalls akkurat zwischen Überleben und Versterben unterscheiden konnte. Des Weiteren sei hier hervorzuheben, dass eine andere Studie mit eigener Kohorte, eigenen Messverfahren und eigenen Algorithmen ähnliche Schlüsselproteine zur Prädiktion von Überleben identifizierte wie wir [102].

Die vorliegende Arbeit bezieht sich auf die frühe Phase der Pandemie, vor dem Auftreten von SARS-CoV-2-Varianten, der Anwendung COVID-19-spezifischer antiviraler Therapien oder Impfungen. Der Peptid-Panelassay wurde mit MRM-Validierungskohorte 2 bereits nach der Einführung von Dexamethason als Therapiestandard validiert, es bleibt jedoch abzuwarten, inwieweit sich die früheren Ergebnisse auch auf die heutige Situation übertragen lassen. Diese frühen Messungen erlauben jedoch einen „ungefilterten“ Einblick in das Plasma-Proteom bei einer neuartigen Infektionskrankheit und stehen als Vergleichskohorten für zukünftige Untersuchungen zur Verfügung: zum Einfluss neuer Varianten, Impfimmunität, und zu den Effekten von monoklonalen Antikörpern oder spezifischer antiviraler Therapien.

Die zugrundeliegende bottom-up Proteomik erlaubt nur eingeschränkte Rückschlüsse auf exakte Protein-Isoformen sowie -vorstufen bzw. aktive Produkte. So lässt sich auf Basis der gemessenen Peptide bspw. nicht sicher unterscheiden, ob es sich bei AGT um den Vorläufer Angiotensinogen handelt oder um die aktiven Formen Ang1, Ang2, Ang1-9, Ang1-7. Insofern kann die Proteomik zunächst als Screening-Instrument gesehen werden – für die klinische Applikation und für die sichere Unterscheidung verschiedener Isoformen und Produkte bedarf es gerichteter Verfahren wie MRM. Um die klinische Translation zu ermöglichen, entwickelten wir einen Peptid-Panelassay mit zunächst 50 Peptiden. Durch eine Ausweitung auf 100 oder 200 Peptide ließe sich die Präzision des Panelassays noch erhöhen; dies ist Gegenstand unserer aktuellen und zukünftigen Forschung.

Letztlich bleibt festzuhalten, dass Machine Learning-Modelle noch keine Zulassung zur medizinischen Entscheidungsfindung haben. Jedoch mehren sich mögliche Anwendungsfelder von Machine Learning und wir konnten mit der vorliegenden Arbeit eine weitere klinische Applikation aufweisen.

5 Fazit und Ausblick

Die Identifikation neuer Biomarker und deren Translation in die klinische Routine dauert häufig viele Jahre oder gar Jahrzehnte. Die COVID-19-Pandemie hat gezeigt, dass diese Zeit durch enge Kooperationen zwischen Grundlagenwissenschaftler:innen und Kliniker:innen sowie der adäquaten Finanzierung deutlich verkürzt werden kann. Mittels Hochdurchsatz-Proteomik gelang es uns, innerhalb von weniger als zwei Jahren nach Auftreten einer neuen Erkrankung eine Reihe relevanter Marker der Wirtsantwort auf einen neuartigen Erreger zu identifizieren, prognostische Modelle zu entwickeln und daraus einen in klinischen Routinelaboren anwendbaren Peptid-Panelassay zu etablieren. Die vorliegende Arbeit kann als Proof-of-Concept betrachtet werden, dass eine präzise Krankheitsprognose auf Basis einzelner Messpunkte möglich ist. Nachdem der Peptid-Panelassay auch nach der Einführung von Dexamethason in die Standardtherapie von Patient:innen mit respiratorischer Insuffizienz durch COVID-19 noch robuste Ergebnisse produzierte, erfolgt aktuell die Validierung mit Proben der klinischen Routine bei der Labor Berlin – Charité Vivantes GmbH.

Da viele der identifizierten Marker nicht COVID-19 spezifisch sind, sondern vielmehr die allgemeine Immunantwort widerspiegeln, untersuchen wir darüber hinaus, inwieweit unsere Technik auch zur Klassifikation und Prognose anderer Infektionskrankheiten geeignet ist. Als im Mai 2022 vermehrt Fälle des Mpox-Virus außerhalb der Endemiegebiete in Afrika auftraten, konnten wir rasch den von uns entwickelten Peptid-Panelassay auf Mpox-Fälle, die stationär an der Charité – Universitätsmedizin Berlin behandelt wurden, anwenden und wichtige Gemeinsamkeiten, aber auch Unterschiede im Plasma-Proteom bei dieser grundverschiedenen Infektionskrankheit identifizieren [115].

Derzeit führen wir Discovery-Proteomik-Analysen mit Plasma-Proben von weiteren Infektionskrankheiten, u.a. bakterieller Pneumonie, aber auch Tropenkrankheiten wie Malaria und Loiasis durch und planen den Peptid-Panelassay auf bis zu 200 Peptide zu erweitern. So nähern wir uns Schritt für Schritt dem übergeordneten Ziel: Der Entwicklung eines Pan-Infektion Peptid-Panelassays zur Klassifikation und frühzeitigen Prognose verschiedenster Infektionskrankheiten.

6 Literaturverzeichnis

1. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, Zhao X, Huang B, Shi W, Lu R, Niu P, Zhan F, Ma X, Wang D, Xu W, Wu G, Gao GF, Tan W, China Novel Coronavirus Investigating and Research Team. A novel Coronavirus from patients with pneumonia in China, 2019. N Engl J Med. 2020 Feb 20;382(8):727–33.

2. World Bank Group. Chapter 1. The economic impacts of the COVID-19 crisis [Internet]. World Bank Group; 2022 Feb [cited 2023 Apr 15]. Available from: https://www.worldbank.org/en/publication/wdr2022/brief/chapter-1-introduction-the-economic-impacts-of-the-covid-19-crisis

3. Impact of the Covid-19 pandemic on trade and development: Lessons learned [Internet]. UNCTAD. 2022 [cited 2023 Apr 15]. Available from: https://unctad.org/publication/impact-covid-19-pandemic-trade-and-development-lessons-learned

4. ILO, FAO, IFAD, and WHO. Impact of COVID-19 on people's livelihoods, their health and our food systems [Internet]. www.who.int. 2020 [cited 2023 Apr 15]. Available from: https://www.who.int/news/item/13-10-2020-impact-of-covid-19-on-people's-livelihoods-their-health-and-our-food-systems

5. Coronaviridae Study Group of the International Committee on Taxonomy of Viruses. The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. Nat Microbiol. 2020 Apr;5(4):536–44.

6. World Health Organization. Novel Coronavirus(2019-nCoV) - Situation Report - 22 [Internet]. www.who.int. 2020 [cited 2023 Apr 30]. Available from: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200211-sitrep-22-ncov.pdf

7. World Health Organization. Statement on the second meeting of the International Health Regulations (2005) Emergency Committee regarding the outbreak of novel coronavirus (2019-nCoV) [Internet]. www.who.int. 2020 [cited 2023 Apr 15]. Available from: https://www.who.int/news-room/detail/30-01-2020-statement-on-the-second-meeting-of-the-international-health-regulations-(2005)-emergency-committee-regarding-the-outbreak-of-novel-coronavirus-(2019-ncov)

8. World Health Organisation. WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020 [Internet]. www.who.int. 2020 [cited 2023 Apr 15]. Available from: https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020

9. World Health Organization. Statement on the fifteenth meeting of the IHR (2005) Emergency Committee on the COVID-19 pandemic [Internet]. www.who.int. 2023 [cited 2023 May 10]. Available from: https://www.who.int/news/item/05-05-2023-statement-on-the-fifteenth-meeting-of-the-international-health-regulations-(2005)-emergency-committee-regarding-the-coronavirus-disease-(covid-19)-pandemic

10. ArcGIS Dashboards [Internet]. [cited 2023 May 10]. Available from: https://gisanddata.maps.arcgis.com/apps/dashboards/bda7594740fd40299423467b48e9ecf6

11. Birgand G, Peiffer-Smadja N, Fournier S, Kerneis S, Lescure FX, Lucet JC. Assessment of air contamination by SARS-CoV-2 in hospital settings. JAMA Netw Open. 2020 Dec 1;3(12):e2033232.

12. Rothe C, Schunk M, Sothmann P, Bretzel G, Froeschl G, Wallrauch C, Zimmer T, Thiel V, Janke C, Guggemos W, Seilmaier M, Drosten C, Vollmar P, Zwirglmaier K, Zange S, Wölfel R, Hoelscher M. Transmission of 2019-nCoV Infection from an Asymptomatic Contact in Germany. N Engl J Med. 2020 Mar 5;382(10):970–1.

13. Robert Koch Institut. RKI - Coronavirus SARS-CoV-2 - Epidemiologischer Steckbrief zu SARS-CoV-2 und COVID-19 [Internet]. www.rki.de. 2021 [cited 2023 May 11]. Available from: https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Steckbrief.html?nn=13490888

14. Soriano JB, Murthy S, Marshall JC, Relan P, Diaz JV, WHO Clinical Case Definition Working Group on Post-COVID-19 Condition. A clinical case definition of post-COVID-19 condition by a Delphi consensus. Lancet Infect Dis. 2022 Apr;22(4):e102–7.

15. Stawicki SP, Jeanmonod R, Miller AC, Paladino L, Gaieski DF, Yaffee AQ, De Wulf A, Grover J, Papadimos TJ, Bloem C, Galwankar SC, Chauhan V, Firstenberg MS, Di Somma S, Jeanmonod D, Garg SM, Tucci V, Anderson HL, Fatimah L, Worlton TJ, Dubhashi SP, Glaze KS, Sinha S, Opara IN, Yellapu V, Kelkar D, El-Menyar A, Krishnan V, Venkataramanaiah S, Leyfman Y, Saoud Al Thani HA, Wb Nanayakkara P, Nanda S, Cioè-Peña E, Sardesai I, Chandra S, Munasinghe A, Dutta V, Dal Ponte ST, Izurieta R, Asensio JA, Garg M. The 2019-2020 novel Coronavirus (severe acute respiratory syndrome Coronavirus 2) pandemic: A joint American college of academic international medicine-world academic council of emergency medicine multidisciplinary COVID-19 working group consensus paper. J Glob Infect Dis. 2020 Apr;12(2):47–93.

16. Schilling J, Tolksdorf K, Marquis A, Faber M, Pfoch T, Buda S, Haas W, Schuler E, Altmann D, Grote U, Diercke M, RKI COVID-19 Study Group. The different periods of COVID-19 in Germany: a descriptive analysis from January 2020 to February 2021. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz. 2021 Sep;64(9):1093–106.

17. Nacoti M, Ciocca A, Giupponi A, Brambillasca P, Lussana F, Pisano M, Goisis G, Bonacina D, Fazzi F, Naspro R, Longhi L, Cereda M, Montaguti C. At the Epicenter of the Covid-19 Pandemic and Humanitarian Crises in Italy: Changing Perspectives on Preparation and Mitigation. NEJM Cataly [Internet]. 2020 Mar 21 [cited 2023 May 11];1(2). Available from: https://catalyst.nejm.org/doi/full/10.1056/CAT.20.0080

18. Thompson CN, Baumgartner J, Pichardo C, Toro B, Li L, Arciuolo R, Chan PY, Chen J, Culp G, Davidson A, Devinney K, Dorsinville A, Eddy M, English M, Fireteanu AM, Graf L, Geevarughese A, Greene SK, Guerra K, Huynh M, Hwang C, Iqbal M, Jessup J, Knorr J, Lall R, Latash J, Lee E, Lee K, Li W, Mathes R, McGibbon E, McIntosh N, Montesano M, Moore MS, Murray K, Ngai S, Paladini M, Paneth-Pollak R, Parton H, Peterson E, Pouchet R, Ramachandran J, Reilly K, Sanderson Slutsker J, Van Wye G, Wahnich A, Winters A, Layton M, Jones L, Reddy V, Fine A. COVID-19 outbreak - New York City, February 29-June 1, 2020. MMWR Morb Mortal Wkly Rep. 2020 Nov 20;69(46):1725–9.

19. Huang C, Soleimani J, Herasevich S, Pinevich Y, Pennington KM, Dong Y, Pickering BW, Barwise AK. Clinical Characteristics, Treatment, and Outcomes of Critically Ill Patients With COVID-19: A Scoping Review. Mayo Clin Proc. 2021 Jan;96(1):183–202.

20. Rajpal A, Rahimi L, Ismail-Beigi F. Factors leading to high morbidity and mortality of COVID-19 in patients with type 2 diabetes. J Diabetes. 2020 Dec;12(12):895–908.

21. CDC. Risk for COVID-19 infection, hospitalization, and death by age group [Internet]. Centers for Disease Control and Prevention. 2023 [cited 2023 Apr 15]. Available from: https://www.cdc.gov/coronavirus/2019-ncov/covid-data/investigations-discovery/hospitalization-death-by-age.html

22. Cevik M, Bamford CGG, Ho A. COVID-19 pandemic-a focused review for clinicians. Clin Microbiol Infect. 2020 Jul 1;26(7):842–7.

23. Beigmohammadi MT, Amoozadeh L, Rezaei Motlagh F, Rahimi M, Maghsoudloo M, Jafarnejad B, Eslami B, Salehi MR, Zendehdel K. Mortality predictive value of APACHE II and SOFA scores in COVID-19 patients in the intensive care unit. Can Respir J. 2022 Mar 28;2022:5129314.

24. Ioannidis JPA, Bendavid E, Salholz-Hillel M, Boyack KW, Baas J. Massive covidization of research citations and the citation elite. Proc Natl Acad Sci U S A. 2022 Jul 12;119(28):e2204074119.

25. Park JJH, Mogg R, Smith GE, Nakimuli-Mpungu E, Jehan F, Rayner CR, Condo J, Decloedt EH, Nachega JB, Reis G, Mills EJ. How COVID-19 has fundamentally changed clinical research in global health. Lancet Glob Health. 2021 May 1;9(5):e711–20.

26. Chinnery PF, Pearce JJ, Kinsey AM, Jenkinson JM, Wells G, Watt FM. How COVID-19 has changed medical research funding. Interface Focus. 2021 Dec 6;11(6):20210025.

27. Wu F, Zhao S, Yu B, Chen YM, Wang W, Song ZG, Hu Y, Tao ZW, Tian JH, Pei YY, Yuan ML, Zhang YL, Dai FH, Liu Y, Wang QM, Zheng JJ, Xu L, Holmes EC, Zhang YZ. Author Correction: A new coronavirus associated with human respiratory disease in China. Nature. 2020 Apr;580(7803):E7.

28. Dhama K, Sharun K, Tiwari R, Dadar M, Malik YS, Singh KP, Chaicumpa W. COVID-19, an emerging coronavirus infection: advances and prospects in designing and developing vaccines, immunotherapeutics, and therapeutics. Hum Vaccin Immunother. 2020 Jun 2;16(6):1232–8.

29. Danwang C, Endomba FT, Nkeck JR, Wouna DLA, Robert A, Noubiap JJ. A meta-analysis of potential biomarkers associated with severity of coronavirus disease 2019 (COVID-19). Biomark Res. 2020 Aug 31;8(1):37.

30. Henry BM, de Oliveira MHS, Benoit S, Plebani M, Lippi G. Hematologic, biochemical and immune biomarker abnormalities associated with severe illness and mortality in coronavirus disease 2019 (COVID-19): a meta-analysis. Clin Chem Lab Med. 2020 Jun 25;58(7):1021–8.

31. Gallo Marin B, Aghagoli G, Lavine K, Yang L, Siff EJ, Chiang SS, Salazar-Mather TP, Dumenco L, Savaria MC, Aung SN, Flanigan T, Michelow IC. Predictors of COVID-19 severity: A literature review. Rev Med Virol. 2021 Jan;31(1):1–10.

32. Wynants L, Van Calster B, Collins GS, Riley RD, Heinze G, Schuit E, Bonten MMJ, Damen JAA, Debray TPA, De Vos M, Dhiman P, Haller MC, Harhay MO, Henckaerts L, Kreuzberger N, Lohman A, Luijken K, Ma J, Andaur CL, Reitsma JB, Sergeant JC, Shi C, Skoetz N, Smits LJM, Snell KIE, Sperrin M, Spijker R, Steyerberg EW, Takada T, van Kuijk SMJ, van Royen FS, Wallisch C, Hooft L, Moons KGM, van Smeden M. Prediction models for diagnosis and prognosis of covid-19 infection: systematic review and critical appraisal. BMJ. 2020 Apr 7;369:m1328.

33. Ai T, Yang Z, Hou H, Zhan C, Chen C, Lv W, Tao Q, Sun Z, Xia L. Correlation of chest CT and RT-PCR testing for Coronavirus disease 2019 (COVID-19) in China: A report of 1014 cases. Radiology. 2020 Aug;296(2):E32–40.

34. Wasilewski PG, Mruk B, Mazur S, Półtorak-Szymczak G, Sklinda K, Walecki J. COVID-19 severity scoring systems in radiological imaging - a review. Pol J Radiol. 2020 Jul 17;85(1):e361–8.

35. Ferreira FL, Bota DP, Bross A, Mélot C, Vincent JL. Serial evaluation of the SOFA score to predict outcome in critically ill patients. JAMA. 2001 Oct 10;286(14):1754–8.

36. Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: a severity of disease classification system. Crit Care Med. 1985 Oct;13(10):818–29.

37. Jiang M, Li C, Zheng L, Lv W, He Z, Cui X, Dietrich CF. A biomarker-based age, biomarkers, clinical history, sex (ABCS)-mortality risk score for patients with coronavirus disease 2019. Ann Transl Med. 2021 Feb;9(3):230.

38. Knight SR, Ho A, Pius R, Buchan I, Carson G, Drake TM, Dunning J, Fairfield CJ, Gamble C, Green CA, Gupta R, Halpin S, Hardwick HE, Holden KA, Horby PW, Jackson C, Mclean KA, Merson L, Nguyen-Van-Tam JS, Norman L, Noursadeghi M, Olliaro PL, Pritchard MG, Russell CD, Shaw CA, Sheikh A, Solomon T, Sudlow C, Swann OV, Turtle LC, Openshaw PJ, Baillie JK, Semple MG, Docherty AB, Harrison EM, ISARIC4C investigators. Risk stratification of patients admitted to hospital with covid-19 using the ISARIC WHO Clinical Characterisation Protocol: development and validation of the 4C Mortality Score. BMJ. 2020 Sep 9;370:m3339.

39. Phua J, Weng L, Ling L, Egi M, Lim CM, Divatia JV, Shrestha BR, Arabi YM, Ng J, Gomersall CD, Nishimura M, Koh Y, Du B, Asian Critical Care Clinical Trials Group. Intensive care management of coronavirus disease 2019 (COVID-19): challenges and recommendations. Lancet Respir Med. 2020 May;8(5):506–17.

40. Anderson NL, Anderson NG. Proteome and proteomics: new technologies, new concepts, and new words. Electrophoresis. 1998 Aug;19(11):1853–61.

41. Ong SE, Mann M. Mass spectrometry-based proteomics turns quantitative. Nat Chem Biol. 2005 Oct;1(5):252–62.

42. Corthals GL, Wasinger VC, Hochstrasser DF, Sanchez JC. The dynamic range of protein expression: a challenge for proteomic research. Electrophoresis. 2000 Apr;21(6):1104–15.

43. Aebersold R, Mann M. Mass-spectrometric exploration of proteome structure and function. Nature. 2016 Sep 15;537(7620):347–55.

44. Karpievitch YV, Polpitiya AD, Anderson GA, Smith RD, Dabney AR. Liquid chromatography mass spectrometry-based proteomics: Biological and technological aspects. Ann Appl Stat. 2010;4(4):1797–823.

45. Aebersold R, Mann M. Mass spectrometry-based proteomics. Nature. 2003 Mar 13;422(6928):198–207.

46. Ludwig C, Gillet L, Rosenberger G, Amon S, Collins BC, Aebersold R. Data-independent acquisition-based SWATH-MS for quantitative proteomics: a tutorial. Mol Syst Biol. 2018 Aug 13;14(8):e8126.

47. Anderson L, Hunter CL. Quantitative mass spectrometric multiple reaction monitoring assays for major plasma proteins. Mol Cell Proteomics. 2006 Apr;5(4):573–88.

48. Demichev V, Messner CB, Vernardis SI, Lilley KS, Ralser M. DIA-NN: neural networks and interference correction enable deep proteome coverage in high throughput. Nat Methods. 2020 Jan;17(1):41–4.

49. Chen C, Hou J, Tanner JJ, Cheng J. Bioinformatics methods for mass spectrometry-based proteomics data analysis. Int J Mol Sci [Internet]. 2020 Apr 20;21(8). Available from: http://dx.doi.org/10.3390/ijms21082873

50. Neely BA, Palmblad M. Machine learning in proteomics and metabolomics. J Proteome Res. 2022 Nov 4;21(11):2553–4.

51. Kurth F, Roennefarth M, Thibeault C, Corman VM, Müller-Redetzky H, Mittermaier M, Ruwwe-Glösenkamp C, Heim KM, Krannich A, Zvorc S, Schmidt S, Kretzler L, Dang-Heine C, Rose M, Hummel M, Hocke A, Hübner RH, Opitz B, Mall MA, Röhmel J, Landmesser U, Pieske B, Knauss S, Endres M, Spranger J, Mockenhaupt FP, Tacke F, Treskatsch S, Angermair S, Siegmund B, Spies C, Weber-Carstens S, Eckardt KU, Schürmann D, Uhrig A, Stegemann MS, Zoller T, Drosten C, Suttorp N, Witzenrath M, Hippenstiel S, von Kalle C, Sander LE. Studying the pathophysiology of coronavirus disease 2019: a protocol for the Berlin prospective COVID-19 patient cohort (Pa-COVID-19). Infection. 2020 Aug;48(4):619–26.

52. World Health Organization. WHO R&D Blueprint-Novel Coronavirus COVID-19 Therapeutic Trial Synopsis. Geneva: World Health Organization. 2020;

53. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chronic Dis. 1987;40(5):373–83.

54. Vincent JL, Moreno R, Takala J, Willatts S, De Mendonça A, Bruining H, Reinhart CK, Suter PM, Thijs LG. The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. On behalf of the Working Group on Sepsis-Related Problems of the European Society of Intensive Care Medicine. Intensive Care Med. 1996 Jul;22(7):707–10.

55. Smith BH, Campbell A, Linksted P, Fitzpatrick B, Jackson C, Kerr SM, Deary IJ, Macintyre DJ, Campbell H, McGilchrist M, Hocking LJ, Wisely L, Ford I, Lindsay RS, Morton R, Palmer CNA, Dominiczak AF, Porteous DJ, Morris AD. Cohort Profile: Generation Scotland: Scottish Family Health Study (GS:SFHS). The study, its participants and their potential for genetic research on health and illness. Int J Epidemiol. 2013 Jun;42(3):689–700.

56. Demichev V, Tober-Lau P, Lemke O, Nazarenko T, Thibeault C, Whitwell H, Röhl A, Freiwald A, Szyrwiel L, Ludwig D, Correia-Melo C, Aulakh SK, Helbig ET, Stubbemann P, Lippert LJ, Grüning NM, Blyuss O, Vernardis S, White M, Messner CB, Joannidis M, Sonnweber T, Klein SJ, Pizzini A, Wohlfarter Y, Sahanic S, Hilbe R, Schaefer B, Wagner S, Mittermaier M, Machleidt F, Garcia C, Ruwwe-Glösenkamp C, Lingscheid T, Bosquillon de Jarcy L, Stegemann MS, Pfeiffer M, Jürgens L, Denker S, Zickler D, Enghard P, Zelezniak A, Campbell A, Hayward C, Porteous DJ, Marioni RE, Uhrig A, Müller-Redetzky H, Zoller H, Löffler-Ragg J, Keller MA, Tancevski I, Timms JF, Zaikin A, Hippenstiel S, Ramharter M, Witzenrath M, Suttorp N, Lilley K, Mülleder M, Sander LE, PA-COVID-19 Study group, Ralser M, Kurth F. A time-resolved proteomic and prognostic map of COVID-19. Cell Syst. 2021 Aug 18;12(8):780-794.e7.

57. Wang Z, Cryar A, Lemke O, Tober-Lau P, Ludwig D, Helbig ET, Hippenstiel S, Sander LE, Blake D, Lane CS, Sayers RL, Mueller C, Zeiser J, Townsend S, Demichev V, Mülleder M, Kurth F, Sirka E, Hartl J, Ralser M. A multiplex protein panel assay for severity prediction and outcome prognosis in patients with COVID-19: An observational multi-cohort study. EClinicalMedicine. 2022 Jul;49(101495):101495.

58. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015 Apr 20;43(7):e47.

59. Zanin M, Alcazar JM, Carbajosa JV, Paez MG, Papo D, Sousa P, Menasalvas E, Boccaletti S. Parenclitic networks: uncovering new functions in biological data. Sci Rep. 2014 May 29;4(1):5112.

60. Whitwell HJ, Blyuss O, Menon U, Timms JF, Zaikin A. Parenclitic networks for predicting ovarian cancer. Oncotarget. 2018 Apr 27;9(32):22717–26.

61. Demichev V, Tober-Lau P, Nazarenko T, Lemke O, Kaur Aulakh S, Whitwell HJ, Röhl A, Freiwald A, Mittermaier M, Szyrwiel L, Ludwig D, Correia-Melo C, Lippert LJ, Helbig ET, Stubbemann P, Olk N, Thibeault C, Grüning NM, Blyuss O, Vernardis S, White M, Messner CB, Joannidis M, Sonnweber T, Klein SJ, Pizzini A, Wohlfarter Y, Sahanic S, Hilbe R, Schaefer B, Wagner S, Machleidt F, Garcia C, Ruwwe-Glösenkamp C, Lingscheid T, Bosquillon de Jarcy L, Stegemann MS, Pfeiffer M, Jürgens L, Denker S, Zickler D, Spies C, Edel A, Müller NB, Enghard P, Zelezniak A, Bellmann-Weiler R, Weiss G, Campbell A, Hayward C, Porteous DJ, Marioni RE, Uhrig A, Zoller H, Löffler-Ragg J, Keller MA, Tancevski I, Timms JF, Zaikin A, Hippenstiel S, Ramharter M, Müller-Redetzky H, Witzenrath M, Suttorp N, Lilley K, Mülleder M, Sander LE, Kurth F, Ralser M, PA-COVID-19 Study group. A proteomic survival predictor for COVID-19 patients in intensive care. PLOS Digit Health. 2022 Jan 18;1(1):e0000007.

62. Lian J, Jin C, Hao S, Zhang X, Yang M, Jin X, Lu Y, Hu J, Zhang S, Zheng L, Jia H, Cai H, Zhang Y, Yu G, Wang X, Gu J, Ye C, Yu X, Gao J, Yang Y, Sheng J. High neutrophil-to-lymphocyte ratio associated with progression to critical illness in older patients with COVID-19: a multicenter retrospective study. Aging (Albany NY). 2020 Jul 30;12(14):13849–59.

63. Turula H, Wobus CE. The role of the polymeric immunoglobulin receptor and secretory immunoglobulins during mucosal infection and immu- nity. Viruses. 2018;10.

64. Zhang H, Penninger JM, Li Y, Zhong N, Slutsky AS. Angiotensin-converting enzyme 2 (ACE2) as a SARS-CoV-2 receptor: molecular mechanisms and potential therapeutic target. Intensive Care Med. 2020 Apr;46(4):586–90.

65. Turner AJ. Chapter 25 - ACE2 Cell Biology, Regulation, and Physiological Functions. In: Unger T, Steckelings UM, dos Santos RAS, editors. The Protective Arm of the Renin Angiotensin System (RAS). Boston: Academic Press; 2015. p. 185–9.

66. Silhol F, Sarlon G, Deharo JC, Vaïsse B. Downregulation of ACE2 induces overstimulation of the renin-angiotensin system in COVID-19: should we block the renin-angiotensin system? Hypertens Res. 2020 Aug;43(8):854–6.

67. Batlle D, Wysocki J, Soler MJ, Ranganath K. Angiotensin-converting enzyme 2: enhancing the degradation of angiotensin II as a potential therapy for diabetic nephropathy. Kidney Int. 2012 Mar;81(6):520–8.

68. Liu Y, Yang Y, Zhang C, Huang F, Wang F, Yuan J, Wang Z, Li J, Li J, Feng C, Zhang Z, Wang L, Peng L, Chen L, Qin Y, Zhao D, Tan S, Yin L, Xu J, Zhou C, Jiang C, Liu L. Clinical and biochemical indexes from 2019-nCoV infected patients linked to viral loads and lung injury. Sci China Life Sci. 2020 Mar;63(3):364–74.

69. Hilt ZT, Pariser DN, Ture SK, Mohan A, Quijada P, Asante AA, Cameron SJ, Sterling JA, Merkel AR, Johanson AL, Jenkins JL, Small EM, McGrath KE, Palis J, Elliott MR, Morrell CN. Platelet-derived β2M regulates monocyte inflammatory responses. JCI Insight [Internet]. 2019 Mar 7;4(5). Available from: http://dx.doi.org/10.1172/jci.insight.122943

70. Makridakis M, Kontostathi G, Petra E, Stroggilos R, Lygirou V, Filip S, Duranton F, Mischak H, Argiles A, Zoidakis J, Vlahou A. Multiplexed MRM-based protein quantification of putative prognostic biomarkers for chronic kidney disease progression in plasma. Sci Rep. 2020 Mar 16;10(1):4815.

71. Hajishengallis G, Reis ES, Mastellos DC, Ricklin D, Lambris JD. Novel mechanisms and functions of complement. Nat Immunol. 2017 Nov 16;18(12):1288–98.

72. Pepys MB, Hirschfield GM. C-reactive protein: a critical update. J Clin Invest. 2003 Jun;111(12):1805–12.

73. Peralta CA, Shlipak MG, Judd S, Cushman M, McClellan W, Zakai NA, Safford MM, Zhang X, Muntner P, Warnock D. Detection of chronic kidney disease with creatinine, cystatin C, and urine albumin-to-creatinine ratio and association with progression to end-stage renal disease and mortality. JAMA. 2011 Apr 20;305(15):1545–52.

74. Benarafa C. Regulation of Neutrophil Serine Proteases by Intracellular Serpins. In: Geiger M, Wahlmüller F, Furtmüller M, editors. The Serpin Family: Proteins with Multiple Functions in Health and Disease. Cham: Springer International Publishing; 2015. p. 59–76.

75. Gettins PGW, Olson ST. Inhibitory serpins. New insights into their folding, polymerization, regulation and clearance. Biochem J. 2016 Aug 1;473(15):2273–93.

76. Potempa J, Fedak D, Dubin A, Mast A, Travis J. Proteolytic inactivation of alpha-1-anti-chymotrypsin. Sites of cleavage and generation of chemotactic activity. J Biol Chem. 1991 Nov 15;266(32):21482–7.

77. Schulte-Schrepping J, Reusch N, Paclik D, Baßler K, Schlickeiser S, Zhang B, Krämer B, Krammer T, Brumhard S, Bonaguro L, De Domenico E, Wendisch D, Grasshoff M, Kapellos TS, Beckstette M, Pecht T, Saglam A, Dietrich O, Mei HE, Schulz AR, Conrad C, Kunkel D, Vafadarnejad E, Xu CJ, Horne A, Herbert M, Drews A, Thibeault C, Pfeiffer M, Hippenstiel S, Hocke A, Müller-Redetzky H, Heim KM, Machleidt F, Uhrig A, Bosquillon de Jarcy L, Jürgens L, Stegemann M, Glösenkamp CR, Volk HD, Goffinet C, Landthaler M, Wyler E, Georg P, Schneider M, Dang-Heine C, Neuwinger N, Kappert K, Tauber R, Corman V, Raabe J, Kaiser KM, Vinh MT, Rieke G, Meisel C, Ulas T, Becker M, Geffers R, Witzenrath M, Drosten C, Suttorp N, von Kalle C, Kurth F, Händler K, Schultze JL, Aschenbrenner AC, Li Y, Nattermann J, Sawitzki B, Saliba AE, Sander LE, Deutsche COVID-19 OMICS Initiative (DeCOI). Severe COVID-19 Is Marked by a Dysregulated Myeloid Cell Compartment. Cell. 2020 Sep 17;182(6):1419-1440.e23.

78. Ombrellino M, Wang H, Yang H, Zhang M, Vishnubhakat J, Frazier A, Scher LA, Friedman SG, Tracey KJ. Fetuin, a negative acute phase protein, attenuates TNF synthesis and the innate inflammatory response to carrageenan. Shock. 2001 Mar;15(3):181–5.

79. Poon IKH, Patel KK, Davis DS, Parish CR, Hulett MD. Histidine-rich glycoprotein: the Swiss Army knife of mammalian plasma. Blood. 2011 Feb 17;117(7):2093–101.

80. Wakabayashi S. New insights into the functions of histidine-rich glycoprotein. Int Rev Cell Mol Biol. 2013;304:467–93.

81. Fries E, Blom AM. Bikunin--not just a plasma proteinase inhibitor. Int J Biochem Cell Biol. 2000 Feb 1;32(2):125–37.

82. Zhuo L, Itano N, Nonogaki T, Shen LI, Wu J, Watanabe H, Kimata K. Biological Function of SHAP–Hyaluronan Covalent Complex [Internet]. Chemistry and Biology of Hyaluronan. 2004. p. 205–22. Available from: http://dx.doi.org/10.1016/b978-008044382-9/50040-6

83. Heissig B, Salama Y, Takahashi S, Osada T, Hattori K. The multifaceted role of plasminogen in inflammation. Cell Signal. 2020 Nov;75:109761.

84. Bhowmick P, Roome S, Borchers CH, Goodlett DR, Mohammed Y. An update on MRMAssayDB: A comprehensive resource for targeted proteomics assays in the community. J Proteome Res. 2021 Apr 2;20(4):2105–15.

85. Thermo Fisher Scientific. Peptide Synthesis and Proteotypic Peptide Analyzing Tool [Internet]. [cited 2023 May 8]. Available from: https://www.thermofisher.com/de/de/home/life-science/protein-biology/peptides-proteins/custom-peptide-synthesis-services/peptide-analyzing-tool.html

86. Sun Q, Qiu H, Huang M, Yang Y. Lower mortality of COVID-19 by early recognition and intervention: experience from Jiangsu Province. Ann Intensive Care. 2020 Mar 18;10(1):33.

87. Wilcox ME, Rowan KM, Harrison DA, Doidge JC. Does unprecedented ICU capacity strain, as experienced during the COVID-19 pandemic, impact patient outcome? Crit Care Med. 2022 Jun 1;50(6):e548–56.

88. Cavallazzi R, Bradley J, Chandler T, Furmanek S, Ramirez JA. Severity of illness scores and biomarkers for prognosis of patients with Coronavirus disease 2019. Semin Respir Crit Care Med. 2023 Feb;44(1):75–90.

89. Fernandes S, Sérvio R, Patrício P, Pereira C. Validation of the Acute Physiology and Chronic Health Evaluation (APACHE) II score in COVID-19 patients admitted to the intensive care unit in times of resource scarcity. Cureus. 2023 Feb;15(2):e34721.

90. Kashuba E, Bailey J, Allsup D, Cawkwell L. The kinin–kallikrein system: physiological roles, pathophysiology and its relationship to cancer biomarkers. Biomarkers. 2013 Jun 1;18(4):279–96.

91. REMAP-CAP Investigators, ACTIV-4a Investigators, ATTACC Investigators, Goligher EC, Bradbury CA, McVerry BJ, Lawler PR, Berger JS, Gong MN, Carrier M, Reynolds HR, Kumar A, Turgeon AF, Kornblith LZ, Kahn SR, Marshall JC, Kim KS, Houston BL, Derde LPG, Cushman M, Tritschler T, Angus DC, Godoy LC, McQuilten Z, Kirwan BA, Farkouh ME, Brooks MM, Lewis RJ, Berry LR, Lorenzi E, Gordon AC, Ahuja T, Al-Beidh F, Annane D, Arabi YM, Aryal D, Baumann Kreuziger L, Beane A, Bhimani Z, Bihari S, Billett HH, Bond L, Bonten M, Brunkhorst F, Buxton M, Buzgau A, Castellucci LA, Chekuri S, Chen JT, Cheng AC, Chkhikvadze T, Coiffard B, Contreras A, Costantini TW, de Brouwer S, Detry MA, Duggal A, Džavík V, Effron MB, Eng HF, Escobedo J, Estcourt LJ, Everett BM, Fergusson DA, Fitzgerald M, Fowler RA, Froess JD, Fu Z, Galanaud JP, Galen BT, Gandotra S, Girard TD, Goodman AL, Goossens H, Green C, Greenstein YY, Gross PL, Haniffa R, Hegde SM, Hendrickson CM, Higgins AM, Hindenburg AA, Hope AA, Horowitz JM, Horvat CM, Huang DT, Hudock K, Hunt BJ, Husain M, Hyzy RC, Jacobson JR, Jayakumar D, Keller NM, Khan A, Kim Y, Kindzelski A, King AJ, Knudson MM, Kornblith AE, Kutcher ME, Laffan MA, Lamontagne F, Le Gal G, Leeper CM, Leifer ES, Lim G, Gallego Lima F, Linstrum K, Litton E, Lopez-Sendon J, Lother SA, Marten N, Saud Marinez A, Martinez M, Mateos Garcia E, Mavromichalis S, McAuley DF, McDonald EG, McGlothlin A, McGuinness SP, Middeldorp S, Montgomery SK, Mouncey PR, Murthy S, Nair GB, Nair R, Nichol AD, Nicolau JC, Nunez-Garcia B, Park JJ, Park PK, Parke RL, Parker JC, Parnia S, Paul JD, Pompilio M, Quigley JG, Rosenson RS, Rost NS, Rowan K, Santos FO, Santos M, Santos MO, Satterwhite L, Saunders CT, Schreiber J, Schutgens REG, Seymour CW, Siegal DM, Silva DG Jr, Singhal AB, Slutsky AS, Solvason D, Stanworth SJ, Turner AM, van Bentum-Puijk W, van de Veerdonk FL, van Diepen S, Vazquez-Grande G, Wahid L, Wareham V, Widmer RJ, Wilson JG, Yuriditsky E, Zhong Y, Berry SM, McArthur CJ, Neal MD, Hochman JS, Webb SA, Zarychanski R. Therapeutic anticoagulation with heparin in critically ill patients with Covid-19. N Engl J Med. 2021 Aug 26;385(9):777–89.

92. Afzali B, Noris M, Lambrecht BN, Kemper C. The state of complement in COVID-19. Nat Rev Immunol. 2022 Feb 15;22(2):77–84.

93. Kluge S, Janssens U, Welte T, Weber-Carstens S, Schälte G, Spinner C, Malin JJ, Gastmeier P, Langer F, Bracht H, Westhoff M, Pfeifer M, Rabe KF, Hoffmann F, Böttiger BW, Weinmann-Menke J, Kersten A, Berlit P, Krawczyk M, Nehls W, Haase R, Müller OJ,

Specker C, Nothacker M, Skoetz N, Marx G, Karagiannidis C. S3-Leitlinie - Empfehlungen zur stationären Therapie von Patienten mit COVID-19 [Internet]. 2022 [cited 2023 May 7]. Available from: https://register.awmf.org/assets/guidelines/113-001LGl_S3_Empfehlungen-zur-stationaeren-Therapie-von-Patienten-mit-COVID-19_2022-09_1.pdf

94. D'Alessandro A, Thomas T, Dzieciatkowska M, Hill RC, Francis RO, Hudson KE, Zimring JC, Hod EA, Spitalnik SL, Hansen KC. Serum Proteomics in COVID-19 Patients: Altered Coagulation and Complement Status as a Function of IL-6 Level. J Proteome Res [Internet]. 2020 Aug 14; Available from: http://dx.doi.org/10.1021/acs.jproteome.0c00365

95. Zhu X, Wang Y, Zhang H, Liu X, Chen T, Yang R, Shi Y, Cao W, Li P, Ma Q, Zhai Y, He F, Zhou G, Cao C. Genetic variation of the human α-2-Heremans-Schmid glycoprotein (AHSG) gene associated with the risk of SARS-CoV infection. PLoS One. 2011 Aug 17;6(8):e23730.

96. Chua RL, Lukassen S, Trump S, Hennig BP, Wendisch D, Pott F, Debnath O, Thürmann L, Kurth F, Völker MT, Kazmierski J, Timmermann B, Twardziok S, Schneider S, Machleidt F, Müller-Redetzky H, Maier M, Krannich A, Schmidt S, Balzer F, Liebig J, Loske J, Suttorp N, Eils J, Ishaque N, Liebert UG, von Kalle C, Hocke A, Witzenrath M, Goffinet C, Drosten C, Laudi S, Lehmann I, Conrad C, Sander LE, Eils R. COVID-19 severity correlates with airway epithelium-immune cell interactions identified by single-cell analysis. Nat Biotechnol. 2020 Aug;38(8):970–9.

97. Merad M, Martin JC. Pathological inflammation in patients with COVID-19: a key role for monocytes and macrophages. Nat Rev Immunol. 2020 Jun;20(6):355–62.

98. Thavarajah T, Dos Santos CC, Slutsky AS, Marshall JC, Bowden P, Romaschin A, Marshall JG. The plasma peptides of sepsis. Clin Proteomics. 2020 Jul 2;17:26.

99. Rehman AA, Ahsan H, Khan FH. α-2-Macroglobulin: a physiological guardian. J Cell Physiol. 2013 Aug;228(8):1665–75.

100. Sharma NK, Tashima AK, Brunialti MKC, Ferreira ER, Torquato RJS, Mortara RA, Machado FR, Assuncao M, Rigato O, Salomao R. Proteomic study revealed cellular assembly and lipid metabolism dysregulation in sepsis secondary to community-acquired pneumonia. Sci Rep. 2017 Nov 15;7(1):15606.

101. Bartleson JM, Radenkovic D, Covarrubias AJ, Furman D, Winer DA, Verdin E. SARS-CoV-2, COVID-19 and the ageing immune system. Nat Aging. 2021 Sep 14;1(9):769–82.

102. Völlmy F, van den Toorn H, Zenezini Chiozzi R, Zucchetti O, Papi A, Volta CA, Marracino L, Vieceli Dalla Sega F, Fortini F, Demichev V, Tober-Lau P, Campo G, Contoli M, Ralser M, Kurth F, Spadaro S, Rizzo P, Heck A Jr. A serum proteome signature to predict mortality in severe COVID-19 patients. Life Sci Alliance [Internet]. 2021 Sep;4(9). Available from: http://dx.doi.org/10.26508/lsa.202101099

103. Wang ZH, Shu C, Ran X, Xie CH, Zhang L. Critically ill patients with Coronavirus disease 2019 in a designated ICU: Clinical features and predictors for mortality. Risk Manag Healthc Policy. 2020 Jul 20;13:833–45.

104. Gupta S, Hayek SS, Wang W, Chan L, Mathews KS, Melamed ML, Brenner SK, Leonberg-Yoo A, Schenck EJ, Radbel J, Reiser J, Bansal A, Srivastava A, Zhou Y, Sutherland A, Green A, Shehata AM, Goyal N, Vijayan A, Velez JCQ, Shaefi S, Parikh CR, Arunthamakun J, Athavale AM, Friedman AN, Short SAP, Kibbelaar ZA, Abu Omar S, Admon AJ, Donnelly JP, Gershengorn HB, Hernán MA, Semler MW, Leaf DE, STOP-COVID Investigators. Factors associated with death in critically ill patients with Coronavirus disease 2019 in the US. JAMA Intern Med. 2020 Nov 1;180(11):1436–47.

105. Hirner S, Pigoga JL, Naidoo AV, Calvello Hynes EJ, Omer YO, Wallis LA, Bills CB. Potential solutions for screening, triage, and severity scoring of suspected COVID-19 positive patients in low-resource settings: a scoping review. BMJ Open. 2021 Sep 15;11(9):e046130.

106. Ehni HJ, Wiesing U, Ranisch R. Saving the most lives-A comparison of European triage guidelines in the context of the COVID-19 pandemic. Bioethics. 2021 Feb;35(2):125–34.

107. Hartl J, Kurth F, Kappert K, Horst D, Mülleder M, Hartmann G, Ralser M. Quantitative protein biomarker panels: a path to improved clinical practice through proteomics. EMBO Mol Med, in press [Internet]. 2023; Available from: http://dx.doi.org/10.15252/emmm.202216061

108. Gerber SA, Rush J, Stemman O, Kirschner MW, Gygi SP. Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. Proc Natl Acad Sci U S A. 2003 Jun 10;100(12):6940–5.

109. Yang JJ, Han Y, Mah CH. Streamlined MRM method transfer between instruments assisted with HRMS matching and retention- time prediction. Anal Chim Acta. 2020;1100:88–96.

110. Xu RN, Fan L, Rieser MJ. Recent advances in high-through- put quantitative bioanalysis by LC-MS/MS. J Pharm Biomed Anal. 2007;44:342–55.

111. Hall TG, Smukste I, Bresciano KR. Identifying and overcom- ing matrix effects in drug discovery and development. In: Tandem Mass Spectrometry−Applications and Principles Intech Open. London, UK.; 2012. p. 390–419.

112. Jans J, Broeks MH, Verhoeven-Duif NM. Metabolomics in diag- nostics of inborn metabolic disorders. Current Opinion in Systems Biology. 2021.

113. Van Der Gugten JG. Tandem mass spectrometry in the clinical labora- tory: A tutorial overview. Clinical Mass Spectrometry. 2020;15:36–43.

114. Senatsverwaltung für Gesundheit, Pflege und Gleichstellung. SenGPG Pandemieplan - COVID 19 [Internet]. Jun, 2020. Available from: https://www.berlin.de/sen/gesundheit/_assets/themen/gesundheitsschutz-und-umwelt/infektionsschutz/sengpg-pandemieplan-covid-19.pdf

115. Wang Z, Tober-Lau P, Farztdinov V, Lemke O, Schwecke T, Steinbrecher S, Muenzner J, Kriedemann H, Sander LE, Hartl J, Mülleder M, Ralser M, Kurth F. The human host response to monkeypox infection: a proteomic case series study. EMBO Mol Med. 2022 Nov 8;14(11):e16643.

7 Eidesstattliche Versicherung

Ich, Pinkus Tober-Lau, versichere an Eides statt durch meine eigenhändige Unterschrift, dass ich die vorgelegte Dissertation mit dem Thema: Nutzung des Plasma-Proteoms zur Klassifizierung und Prognose von Patient:innen mit COVID-19 (engl.: Leveraging the plasma proteome for classification and prognosis of patients with COVID-19) selbstständig und ohne nicht offengelegte Hilfe Dritter verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel genutzt habe.

Alle Stellen, die wörtlich oder dem Sinne nach auf Publikationen oder Vorträgen anderer Autor:innen beruhen, sind als solche in korrekter Zitierung kenntlich gemacht. Die Abschnitte zu Methodik (insbesondere praktische Arbeiten, Laborbestimmungen, statistische Aufarbeitung) und Resultaten (insbesondere Abbildungen, Graphiken und Tabellen) werden von mir verantwortet.

Ich versichere ferner, dass ich die in Zusammenarbeit mit anderen Personen generierten Daten, Datenauswertungen und Schlussfolgerungen korrekt gekennzeichnet und meinen eigenen Beitrag sowie die Beiträge anderer Personen korrekt kenntlich gemacht habe (siehe Anteilserklärung). Texte oder Textteile, die gemeinsam mit anderen erstellt oder verwendet wurden, habe ich korrekt kenntlich gemacht.

Meine Anteile an etwaigen Publikationen zu dieser Dissertation entsprechen denen, die in der untenstehenden gemeinsamen Erklärung mit dem Erstbetreuer, angegeben sind. Für sämtliche im Rahmen der Dissertation entstandenen Publikationen wurden die Richtlinien des ICMJE (International Committee of Medical Journal Editors; www.icmje.org) zur Autorenschaft eingehalten. Ich erkläre ferner, dass ich mich zur Einhaltung der Satzung der Charité – Universitätsmedizin Berlin zur Sicherung Guter Wissenschaftlicher Praxis verpflichte.

Weiterhin versichere ich, dass ich diese Dissertation weder in gleicher noch in ähnlicher Form bereits an einer anderen Fakultät eingereicht habe.

Die Bedeutung dieser eidesstattlichen Versicherung und die strafrechtlichen Folgen einer unwahren eidesstattlichen Versicherung (§§156, 161 des Strafgesetzbuches) sind mir bekannt und bewusst.

_____ _____

Datum Unterschrift

8 Anteilserklärung an den erfolgten Publikationen

Die dieser Arbeit zugrundeliegenden Publikationen mit geteilter Erstautorenschaft entstanden im Rahmen eines Kooperationsprojektes zwischen der Medizinischen Klinik m.S. Infektiologie und Pneumologie und dem Institut für Biochemie, beide Charité – Universitätsmedizin Berlin. Herr Pinkus Tober-Lau war federführend seitens der Infektiologie für das Projekt verantwortlich, d.h. die Selektion von Patient:innen für die Analysen, Erhebung und Bereitstellung von klinischen Proben und Daten sowie die Interpretation der Analysen aus biomedizinischer und klinischer Sicht, Herr Dr. Vadim Demichev übernahm als Bioinformatiker die Aufbereitung der Proteomdaten sowie die induktive Statistik und Modellierung. Die Studienleitung lag bei Herrn PD Dr. med. Florian Kurth, Herrn Prof. Dr. Markus Ralser, und Herrn Prof. Dr. med. Leif Erik Sander. Es bestand stets die Möglichkeit zur Rücksprache mit der Studienleitung.

Die Reihenfolge der Erstautoren wurde entsprechend der Seniorität festgelegt.

Pinkus Tober-Lau hatte folgenden Anteil an den folgenden Publikationen:

Publikation 1: Demichev V*, **Tober-Lau P***, Lemke O, Nazarenko T, Thibeault C, Whitwell H, Röhl A, Freiwald A, Szyrwiel L, Ludwig D, Correia-Melo C, Aulakh SK, Helbig ET, Stubbemann P, Lippert LJ, Grüning NM, Blyuss O, Vernardis S, White M, Messner CB, Joannidis M, Sonnweber T, Klein SJ, Pizzini A, Wohlfarter Y, Sahanic S, Hilbe R, Schaefer B, Wagner S, Mittermaier M, Machleidt F, Garcia C, Ruwwe-Glösenkamp C, Lingscheid T, Bosquillon de Jarcy L, Stegemann MS, Pfeiffer M, Jürgens L, Denker S, Zickler D, Enghard P, Zelezniak A, Campbell A, Hayward C, Porteous DJ, Marioni RE, Uhrig A, Müller-Redetzky H, Zoller H, Löffler-Ragg J, Keller MA, Tancevski I, Timms JF, Zaikin A, Hippenstiel S, Ramharter M, Witzenrath M, Suttorp N, Lilley K, Mülleder M, Sander LE, PA-COVID-19 Study group, Ralser M, Kurth F. A time-resolved proteomic and prognostic map of COVID-19. *Cell Syst*. 2021;12(8):780-794.e7.

Beitrag im Einzelnen:
Im Detail war Herr Tober-Lau von Beginn an maßgeblich am konzeptionellen und strukturellen Aufbau der Pa-COVID-19 Studie beteiligt und erstellte und pflegte u.a. auch die bis heute genutzte Datenbank (MS Access, JMP). Er selektierte geeignete Proben und

unterstützte bei deren Verarbeitung. Er erhob die klinischen Daten aus SecuTrial® bzw. den Klinikinformationssystemen SAP und COPRA 6, führte eigenständig Plausibilitätskontrollen und Datenbereinigung durch, errechnete klinische Scores und bereitete die Daten zur statistischen Analyse durch den Bioinformatiker, Herrn Demichev, auf. Ebenfalls führte er die deskriptive Datenanalyse durch und erstellte Supplementary Tables 2 und 9 (Patientencharakteristika und klinische Metadaten).

Herr Demichev übernahm die bioinformatische Aufarbeitung der Proteomdaten sowie die Durchführung statistischen Analysen der von Herrn Tober-Lau entwickelten Fragestellungen. Ebenfalls entwickelte er die eingesetzten Machine Learning-Modelle. Herr Demichev erstellte die Figure 5 und setzte die von Herrn Tober-Lau inhaltlich und konzeptionell erarbeiteten Figures 1-4 graphisch um. Herr Demichev erstellte die Supplementary Tables 3-8 und 10.

Herr Tober-Lau analysierte die durch Herrn Demichev berechneten Korrelationen, Verlaufsdaten und Modelle aus klinischer und biologischer Sicht, und untersuchte diese selbstständig mit Hinblick auf relevante Zusammenhänge. Er führte die umfassende Literaturrecherche durch, interpretierte eigenständig funktionelle Zusammenhänge und entwickelte weiterführende Fragestellungen, bspw. zum Einfluss von Alter, ECMO, Dialyse auf das Proteom (Supplementary Notes 1-3). Auch führte Herr Tober-Lau die funktionelle Annotation der Proteine durch (Figure 5, Box 1, Supplementary Table 3). Herr Tober-Lau erstellte federführend Einleitung, klinischen Methodenteil, Ergebnisse und Diskussion des Manuskripts, wobei er stets Rücksprache Herrn Demichev sowie den Studienleitern hielt. Er führte die Ersteinreichung des Manuskripts durch und arbeitete gemeinsam mit Herrn Demichev die Anmerkungen der Reviewer ein.

Publikation 2: Demichev V*, **Tober-Lau P***, Nazarenko T, Lemke O, Kaur Aulakh S, Whitwell HJ, Röhl A, Freiwald A, Mittermaier M, Szyrwiel L, Ludwig D, Correia-Melo C, Lippert LJ, Helbig ET, Stubbemann P, Olk N, Thibeault C, Grüning NM, Blyuss O, Vernardis S, White M, Messner CB, Joannidis M, Sonnweber T, Klein SJ, Pizzini A, Wohlfarter Y, Sahanic S, Hilbe R, Schaefer B, Wagner S, Machleidt F, Garcia C, Ruwwe-Glösenkamp C, Lingscheid T, Bosquillon de Jarcy L, Stegemann MS, Pfeiffer M, Jürgens L, Denker S, Zickler D, Spies C, Edel A, Müller NB, Enghard P, Zelezniak A, Bellmann-Weiler R, Weiss G, Campbell A, Hayward C, Porteous DJ, Marioni RE, Uhrig A, Zoller H,

Löffler-Ragg J, Keller MA, Tancevski I, Timms JF, Zaikin A, Hippenstiel S, Ramharter M, Müller-Redetzky H, Witzenrath M, Suttorp N, Lilley K, Mülleder M, Sander LE, Kurth F, Ralser M, PA-COVID-19 Study group. A proteomic survival predictor for COVID-19 patients in intensive care. *PLOS Digit Health*. 2022;1(1):e0000007.

Beitrag im Einzelnen:
Diese Publikation ging aus der oben genannten hervor und erfolgte wieder in enger Zusammenarbeit zwischen der Infektiologie und Pneumologie – Pinkus Tober-Lau, Prof. Dr. med. Leif Erik Sander, PD Dr. med. Florian Kurth – und dem Institut für Biochemie – Dr. Vadim Demichev und Prof. Dr. Markus Ralser.

Auch bei dieser Publikation war Herr Tober-Lau verantwortlich für die Erhebung und Bereitstellung der klinischen Daten sowie deren klinische und biologische Auswertung. Er selektierte geeignete Proben und unterstützte bei deren Verarbeitung. Er erhob die klinischen Daten aus SecuTrial® bzw. den Klinikinformationssystemen SAP und COPRA 6, führte eigenständig Plausibilitätskontrollen und Datenbereinigung durch, errechnete klinische Scores und bereitete die Daten zur statistischen Analyse durch den Bioinformatiker, Herrn Demichev, auf. Ebenfalls führte er die deskriptive Datenanalyse durch und erstellte Supplementary Table 1 (Patientencharakteristika), die funktionelle Annotation von Supplementary Table 2 sowie die Supplementary Tables „Charite – metadata.tsv“ und „Charlson_APACHE_SOFA.txt“.

Herr Demichev übernahm wieder die bioinformatische Aufbereitung der Proteomdaten und führte die statistische Analyse der von Herrn Tober-Lau entwickelten Fragestellungen durch. Er erstellte die Machine Learning-Modelle. Herr Tober-Lau erstellte den Graphical Abstract und erarbeitete auch hier Inhalt und Layout der Figures, die graphische Umsetzung erfolgte durch Herrn Demichev.

Herr Tober-Lau analysierte die von Herrn Demichev generierten Korrelationen und Modelle aus biomedizinischer Sicht, und führte eine umfassende Literaturrecherche sowie die klinische und biologische Interpretation der Daten und Ergebnisse durch, in Rücksprache mit der Studienleitung. Er verfasste im Austausch mit Herrn Demichev sowie der Studienleitung Einleitung, klinische Methoden, Ergebnisse und Diskussion des Manu-

skripts und führte selbstständig die Einreichung durch. Gemeinsam mit den Kollegen arbeitete er die Anmerkungen der Reviewer ein.

Publikation 3: Wang Z*, Cryar A*, Lemke O, **Tober-Lau P**, Ludwig D, Helbig ET, Hippenstiel S, Sander LE, Blake D, Lane CS, Sayers RL, Mueller C, Zeiser J, Townsend S, Demichev V, Mülleder M, Kurth F, Sirka E, Hartl J, Ralser M. A multiplex protein panel assay for severity prediction and outcome prognosis in patients with COVID-19: An observational multi-cohort study. *EClinicalMedicine*. 2022;49(101495):101495.

Beitrag im Einzelnen:
Die Publikation erfolgte unter Federführung des Instituts für Biochemie. Pinkus Tober-Lau war primärer Ansprechpartner für klinische Datenerhebung und Fragestellungen. Konkret führte er die Erhebung auf Aufbereitung der klinischen Daten aus SecuTrial bzw. SAP und COPRA durch und erstellte die deskriptiven Kohortenbeschreibungen (Tabelle 1, Supp. Tabelle 1). Er identifizierte geeignete Proben zur Vermessung und unterstütze bei der Probenvorbereitung. Die klinische Literaturrecherche inkl. funktioneller Annotation der Proteine sowie die Interpretation der Daten erfolgte durch Pinkus Tober-Lau. Er begleitete die Erstellung des Manuskripts federführend von klinischer Seite und verfasste u.a. die Abschnitte zu den klinische Methoden und klinischer Dateninterpretation.

_____

Unterschrift, Datum und Stempel des erstbetreuenden Hochschullehrers

_____

Unterschrift des Doktoranden/der Doktorandin

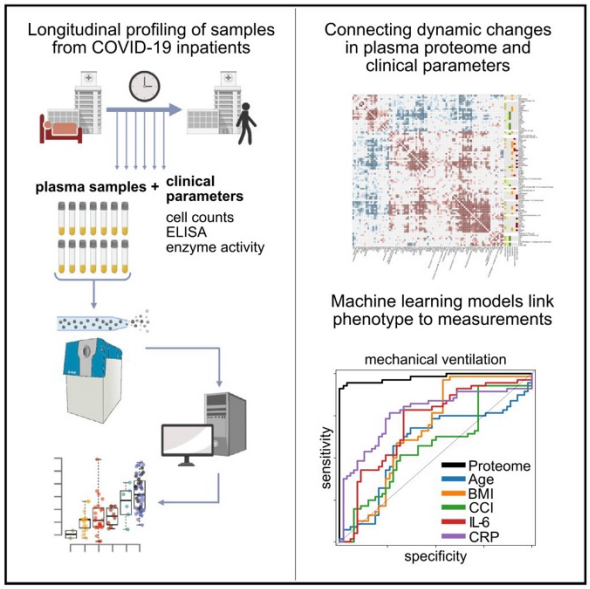
9 Druckexemplare der Publikationen

Article

Cell Systems

A time-resolved proteomic and prognostic map of COVID-19

Graphical abstract



Authors

Vadim Demichev, Pinkus Tober-Lau, Oliver Lemke, ..., PA-COVID-19 Study group, Markus Ralser, Florian Kurth

Correspondence

markus.ralser@charite.de

In brief

Demichev, Tober-Lau et al., present a time-resolved molecular map of the COVID-19, measuring plasma proteomes of patients with COVID-19 along with an extensive panel of clinical diagnostic parameters at 687-time points. They describe the specificity and dynamics, as well as the predictive and prognostic power of the molecular signatures in COVID-19.

Highlights

- Plasma proteomes combined with clinical parameters characterize COVID-19 progression
- Machine learning models allow highly precise prediction of the disease phenotype
- The early molecular host response is predictive of COVID-19 progression
- The molecular response to COVID-19 is age specific

Demichev et al., 2021, Cell Systems *12*, 780–794
August 18, 2021
https://doi.org/10.1016/j.cels.2021.05.005

CellPress

CellPress
OPEN ACCESS

Cell Systems

Article

A time-resolved proteomic and prognostic map of COVID-19

Vadim Demichev,[1,2,3,27] Pinkus Tober-Lau,[4,27] Oliver Lemke,[1] Tatiana Nazarenko,[8,11] Charlotte Thibeault,[4] Harry Whitwell,[9,10,26] Annika Röhl,[1] Anja Freiwald,[1] Lukasz Szyrwiel,[2] Daniela Ludwig,[1] Clara Correia-Melo,[2] Simran Kaur Aulakh,[2] Elisa T. Helbig,[4] Paula Stubbemann,[4] Lena J. Lippert,[4] Nana-Maria Grüning,[1] Oleg Blyuss,[10,12,13] Spyros Vernardis,[2] Matthew White,[2] Christoph B. Messner,[1,2] Michael Joannidis,[25] Thomas Sonnweber,[19] Sebastian J. Klein,[25] Alex Pizzini,[19] Yvonne Wohlfarter,[21] Sabina Sahanic,[19] Richard Hilbe,[19] Benedikt Schaefer,[20] Sonja Wagner,[20] Mirja Mittermaier,[4,22] Felix Machleidt,[4] Carmen Garcia,[4] Christoph Ruwwe-Glösenkamp,[4]

(Author list continued on next page)

[1]Charité Universitätsmedizin Berlin, Department of Biochemistry, 10117 Berlin, Germany
[2]The Francis Crick Institute, Molecular Biology of Metabolism Laboratory, London NW11AT, UK
[3]The University of Cambridge, Department of Biochemistry and Cambridge Centre for Proteomics, Cambridge CB21GA, UK
[4]Charité Universitätsmedizin Berlin, Department of Infectious Diseases and Respiratory Medicine, 10117 Berlin, Germany
[5]Charité Universitätsmedizin Berlin, Medical Department of Hematology, Oncology & Tumor Immunology, Virchow Campus & Molekulares Krebsforschungszentrum, 13353 Berlin, Germany
[6]Charité Universitätsmedizin Berlin, Department of Nephrology and Internal Intensive Care Medicine, 10117 Berlin, Germany
[7]Bernhard Nocht Institute for Tropical Medicine, Department of Tropical Medicine, and University Medical Center Hamburg-Eppendorf, Department of Medicine, 20359 Hamburg, Germany
[8]University College London, Department of Mathematics, London WC1E 6BT, UK
[9]National Phenome Centre and Imperial Clinical Phenotyping Centre, Department of Metabolism, Digestion and Reproduction, Imperial College London, London SW72AZ, UK
[10]Lobachevsky University, Department of Applied Mathematics, Nizhny Novgorod 603105, Russia
[11]University College London, Department of Women's Cancer, EGA Institute for Women'S Health, London WC1E 6BT, UK
[12]University of Hertfordshire, School of Physics, Astronomy and Mathematics, Hatfield AL10 9AB, UK
[13]Sechenov First Moscow State Medical University, Department of Paediatrics and Paediatric Infectious Diseases, Moscow 119435, Russia
[14]Lobachevsky University, Laboratory of Systems Medicine of Healthy Ageing, Nizhny Novgorod 603105, Russia
[15]Chalmers Tekniska Högskola, Department of Biology and Biological Engineering, SE-412 96 Gothenburg, Sweden
[16]University of Edinburgh, Centre for Genomic and Experimental Medicine, Institute of Genetics and Cancer, Edinburgh EH4 2XU, UK
[17]University of Edinburgh, Usher Institute, Edinburgh EH16 4UX, UK
[18]University of Edinburgh, MRC Human Genetics Unit, Institute of Genetics and Cancer, Edinburgh EH4 2XU, UK

(Affiliations continued on next page)

SUMMARY

COVID-19 is highly variable in its clinical presentation, ranging from asymptomatic infection to severe organ damage and death. We characterized the time-dependent progression of the disease in 139 COVID-19 inpatients by measuring 86 accredited diagnostic parameters, such as blood cell counts and enzyme activities, as well as untargeted plasma proteomes at 687 sampling points. We report an initial spike in a systemic inflammatory response, which is gradually alleviated and followed by a protein signature indicative of tissue repair, metabolic reconstitution, and immunomodulation. We identify prognostic marker signatures for devising risk-adapted treatment strategies and use machine learning to classify therapeutic needs. We show that the machine learning models based on the proteome are transferable to an independent cohort. Our study presents a map linking routinely used clinical diagnostic parameters to plasma proteomes and their dynamics in an infectious disease.

INTRODUCTION

The coronavirus disease 2019 (COVID-19) has created unprecedented societal challenges, particularly for public health and the global economy (Alwan et al., 2020; Blumenthal et al., 2020; Rosenbaum, 2020). Efficient management of these challenges is hampered by the variability of clinical manifestations, ranging from asymptomatic infection with severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) to death, despite maximum intensive care. Biomarkers and molecular signatures enabling accurate prognosis of future disease courses are needed to optimize resource allocation and personalize treatment strategies. Patients likely to progress to severe disease and organ failure and those likely to remain stable could be identified early, which

Cell Systems
Article

CellPress
OPEN ACCESS

Tilman Lingscheid,[4] Laure Bosquillon de Jarcy,[4] Miriam S. Stegemann,[4] Moritz Pfeiffer,[4] Linda Jürgens,[4] Sophy Denker,[5,22] Daniel Zickler,[6] Philipp Enghard,[6] Aleksej Zelezniak,[2,15] Archie Campbell,[16,17] Caroline Hayward,[18] David J. Porteous,[16,17] Riccardo E. Marioni,[16] Alexander Uhrig,[4] Holger Müller-Redetzky,[4] Heinz Zoller,[20] Judith Löffler-Ragg,[19] Markus A. Keller,[21] Ivan Tancevski,[19] John F. Timms,[11] Alexey Zaikin,[8,11,14] Stefan Hippenstiel,[4,24] Michael Ramharter,[7] Martin Witzenrath,[4,24] Norbert Suttorp,[4,24] Kathryn Lilley,[3] Michael Mülleder,[23] Leif Erik Sander,[4,24] PA-COVID-19 Study group, Markus Ralser,[1,2,28,*] and Florian Kurth[4,7]

[19]Medical University of Innsbruck, Department of Internal Medicine II, 6020 Innsbruck, Austria
[20]Medical University of Innsbruck, Christian Doppler Laboratory for Iron and Phosphate Biology, Department of Internal Medicine I, 6020 Innsbruck, Austria
[21]Medical University of Innsbruck, Institute of Human Genetics, 6020 Innsbruck, Austria
[22]Berlin Institute of Health, 10178 Berlin, Germany
[23]Charité – Universitätsmedizin Berlin, Core Facility - High-Throughput Mass Spectrometry, 10117 Berlin, Germany
[24]German Centre for Lung Research, 35392 Gießen, Germany
[25]Medical University Innsbruck, Division of Intensive Care and Emergency Medicine, Department of Internal Medicine, 6020 Innsbruck, Austria
[26]Imperial College London, Section of Bioanalytical Chemistry, Division of Systems Medicine, Department of Metabolism, Digestion and Reproduction, London SW7 2AZ, UK
[27]These authors contributed equally
[28]Lead contact
*Correspondence: markus.ralser@charite.de
https://doi.org/10.1016/j.cels.2021.05.005

is particularly valuable in scenarios where health care systems reach capacity limits. Prognostic panels would also optimize the monitoring of novel treatments, thereby accelerating clinical trials (Phua et al., 2020; Saxena, 2020; Wu et al., 2020). Knowledge of factors that differentiate recovery from deterioration throughout the disease will further enhance our understanding of the inflammatory host response as well as the underlying pathophysiology and provide new therapeutic targets.

A number of biomarkers that classify COVID-19 severity have recently been described. These are based on clinical chemistry, enzyme activities, immune profiling, single-cell sequencing, proteomics, and metabolomics (D'Alessandro et al., 2020; Laing et al., 2020; Liu et al., 2020b; Messner et al., 2020; Overmyer et al., 2020; Schulte-Schrepping et al., 2020; Shen et al., 2020; Shu et al., 2020; Wynants et al., 2020). As severity classifiers, the molecular signatures recorded in blood, serum, plasma, or immune cells characterize the COVID-19 pathology and host responses. Furthermore, markers of dysregulated coagulation, inflammation, and other organ dysfunction have been established as risk factors for severe illness, including low platelet count, elevated levels of D-dimer, C-reactive protein (CRP), interleukin 6 (IL-6), ferritin, troponin, and markers of kidney injury (Danwang et al., 2020; Henry et al., 2020). Proteomic investigations that characterize the comprehensive host response have revealed the activation of the complement cascade and acute phase response, both of which center around IL-6-driven pathways. In turn, these systematic studies have revealed that other common antiviral pathways, such as type I interferons (IFN), do not dominate the early response to COVID-19, probably reflecting evasion of the IFN system by SARS-CoV-2 and the subsequent activation of inflammatory cascades (Hadjadj et al., 2020; Yang et al., 2020). Furthermore, proteomic data and diagnostic parameters have pointed to underlying pathological mechanisms and possible therapeutic targets. For instance, using high-throughput proteomics, we reported a decline in plasma levels of gelsolin (GSN) in patients with severe COVID-19 in a previous study (Messner et al., 2020), and recombinant human GSN is currently undergoing clinical testing for COVID-19 pneumonia in a phase II trial (ClinicalTrials.gov identifier: NCT04358406).

The severity of the disease, and the biomarker signatures that indicate severity, correlate with the outcome, but the highest diagnostic need is to stratify within therapeutically homogeneous patients. For instance, to identify those among the mildly affected individuals with the highest risk for deterioration, or among the most severely affected, those with the highest chance to respond positively to an augmentation of therapy. Predicting future trajectories on an individualized basis would also help accelerate therapeutic developments to judge the impact of the treatment on an individual disease course. To obtain a comprehensive picture of how the molecular COVID-19 phenotype develops over time, we deeply phenotyped a group of 139 COVID-19 inpatients at 687 sampling points. On the one hand, we measured a compendium of 86 clinical parameters, routine diagnostic markers, and clinically established risk scores using gold standard accredited clinical tests. On the other hand, we captured the patient's molecular phenotype by measuring plasma proteomes in an untargeted fashion. For this, we made use of liquid chromatography coupled with tandem mass spectrometry, using a recently developed platform technology that utilizes analytical flow rate chromatography, data-independent acquisition mass spectrometry (SWATH-MS), and deep-neural network-based data processing (Demichev et al., 2020; Messner et al., 2020) (Figure S1). By combining the compendium of diagnostic parameters with the proteomes in a time- and patient-resolved fashion, we obtained a comprehensive molecular picture that captures changes in the patient's molecular phenotype as they depend on the severity, age, and disease progression. We identify prognostic biomarkers and depict their distinct trajectories. We exemplify the power of our resource by showing that the biomarker profiles and diagnostic parameters classify treatment requirements, in particular, the need for mechanical ventilation. Furthermore, we report the future prediction of recovery time in mildly ill patients as well as the individual risk of

CellPress
OPEN ACCESS

Cell Systems
Article

clinical deterioration. Our study demonstrates the predictability of COVID-19 disease trajectories based on the molecular phenotype of the early disease stage.

RESULTS

Covariation of clinical diagnostic parameters and the plasma proteome characterizes the host response to COVID-19

We longitudinally phenotyped 139 patients admitted to Charité University Hospital, Berlin, Germany, between March 01, 2020, and June 30, 2020, due to PCR-confirmed SARS-CoV-2 infection (Figure S2). The patients exhibited highly variable disease courses, graded according to the World Health Organization (WHO) ordinal scale for clinical improvement (Table S1), which reflects the treatment that the patient is receiving as a measure of disease severity. The patients included in our study range from WHO grade 3, which includes patients who require inpatient care without supplemental oxygen therapy, to WHO grade 7, which includes patients with severe COVID-19 who require invasive mechanical ventilation and additional organ support therapies such as renal replacement therapy (RRT) and extracorporeal membrane oxygenation (ECMO) (WHO, 2020). In total, 23 out of 139 (17%) patients in the WHO grade 3 category were stable, without requiring supplemental oxygen therapy and could be discharged after a median of 7 days of inpatient care (Table S2 and Figure S2); 47 (34%) patients required either low-flow or high-flow supplemental oxygen therapy; 69 (50%) patients either presented with severe COVID-19 (WHO grade 6 or 7, i.e., requiring invasive mechanical ventilation) or deteriorated and required invasive mechanical ventilation during their hospitalization; 46 patients (33%) required RRT; and 22 (16%) were treated with ECMO. A total of 20 (13%) patients died, including three patients with *do not intubate/do not resuscitate (DNI/DNR)* orders in place and one patient who died due to a non-COVID-19-related cause. Common risk factors for severe COVID-19 were reflected in the outcomes: patients with a severe course of disease were older than those with mild disease (49 years [IQR 35–70] for WHO grade 3 versus 62 years [IQR 53–72] for WHO grade 7, $p = 0.02$), and an age of 65 years or older was associated with a higher risk of death -(OR 4.1 [95% CI 1.5–11.5]). Our cohort further reflected that men and individuals with a high BMI have an increased likelihood to be hospitalized upon a COVID-19 infection; 68% of the patients were men, and the median BMI was 27.8 (IQR 24.7–31.9). However, we noted that within the group of patients hospitalized with COVID-19, sex and BMI were not further associated with disease severity or an increased risk of death. The median duration of hospitalization was 20 days (IQR 9–48) and correlated with severity (7 days for WHO grade 3 versus 46 days for WHO grade 7). The median time from admission to death despite receiving maximum treatment was 28 days (IQR 16–46).

To capture the diverse disease trajectories on a molecular and biochemical level, we systematically collected 86 clinical and accredited diagnostic parameters as measured with certified tests. Moreover, we monitored the development of risk scores such as the "sequential organ failure assessment" (SOFA) score, blood gas analyses, blood cell counts, enzyme activities, and inflammation biomarkers (Table S3). To complement these parameters with an untargeted analysis, we employed a recently developed high-throughput proteomics platform (Messner et al., 2020). This platform makes use of the data-independent acquisition technique SWATH-MS (Gillet et al., 2012), a sample preparation pipeline designed to ISO13485 reporting standards, which is optimized for reducing batch effects, high-flow rate chromatography to provide highly consistent peptide separation in large sample series, and uses DIA-NN (Data-Independent Acquisition by Neural Networks) to analyze proteomics data recorded with 5-min chromatography (Messner et al., 2020; Demichev et al., 2020) (Figure S1 for a detailed overview of the proteomic workflow). In total, we measured 1,169 plasma proteome samples to determine 687 human proteomes, in which we quantified 321 plasma protein groups. Owing to the nature of the high-flow proteomics platform, data completeness was high; thus, we decided against the use of imputation strategies in the analysis of differential protein abundance. Total data completeness was 75%, with 200 proteins consistently quantified with 98% completeness, and 189 proteins with 99% completeness (Figure S1).

To identify interdependencies of the diagnostic parameters that are routinely used in clinical decision making and the plasma proteomes, we characterized their covariation and present a direct correlation map (Figures 1B and S3–S4; Tables S4, S5, and S6). We report a robust positive or negative correlation of IL-6 levels and other inflammatory markers (CRP, procalcitonin) with acute phase proteins (APPs) (APOA2, APOE, CD14, CRP, GSN, ITIH3, ITIH4, LYZ, SAA1, SAA2, SERPINA1, SERPINA3, and AHSG; the protein names corresponding to the gene identifiers are provided in Table S3), coagulation factors and related proteins (FGA, FGB, FGG, F2, F12, KLKB1, PLG, and SERPINC1), and the complement system (C1R, C1S, C8A, C9, CFB, CFD, and CFHR5). Our data, therefore, link the prominent role of the IL-6 response in COVID-19 (D'Alessandro et al., 2020) to coagulation and the complement cascade. Consistently, in our data, markers of cardiac (troponin T, NT-proBNP) and renal (creatinine, urea) function, as well as anemia and dyserythropoiesis (hemoglobin, hematocrit, erythrocytes, and red blood cell distribution width) correlate with various APPs (APOA2, APOE, CD14, GSN, LYZ, SAA1, SAA2, and SERPINA3; Figure 1B and Table S4) supporting the role of inflammation in COVID-19-related organ damage and its impact on erythropoiesis.

Increased levels of neutrophils and the occurrence of immature granulocyte precursors as markers of emergency myelopoiesis have been linked to severe COVID-19 (Schulte-Schrepping et al., 2020). Our data reveal covariation between neutrophil counts and the levels of two inhibitors of neutrophil serine proteases, SERPINA1 and SERPINA3 (Figure 1C). These two proteins show the highest correlation (0.72 and 0.79 Spearman R, respectively) with the neutrophil-to-lymphocyte ratio (NLR), a prognostic marker for COVID-19 (Lian et al., 2020; Liu et al., 2020a). We further report a strong correlation (Figure 1C) of alkaline phosphatase and gamma-glutamyl transferase activities, both characteristic of biliary disorders (Poynard and Imbert-Bismut, 2012), with plasma levels of the polymeric immunoglobulin receptor (PIGR). We notice that cholangiocytes (bile duct epithelium cells) express ACE-2 and can be directly infected with SARS-CoV-2 (Zhao et al., 2020), potentially leading to host viral

Cell Systems
Article

CellPress
OPEN ACCESS

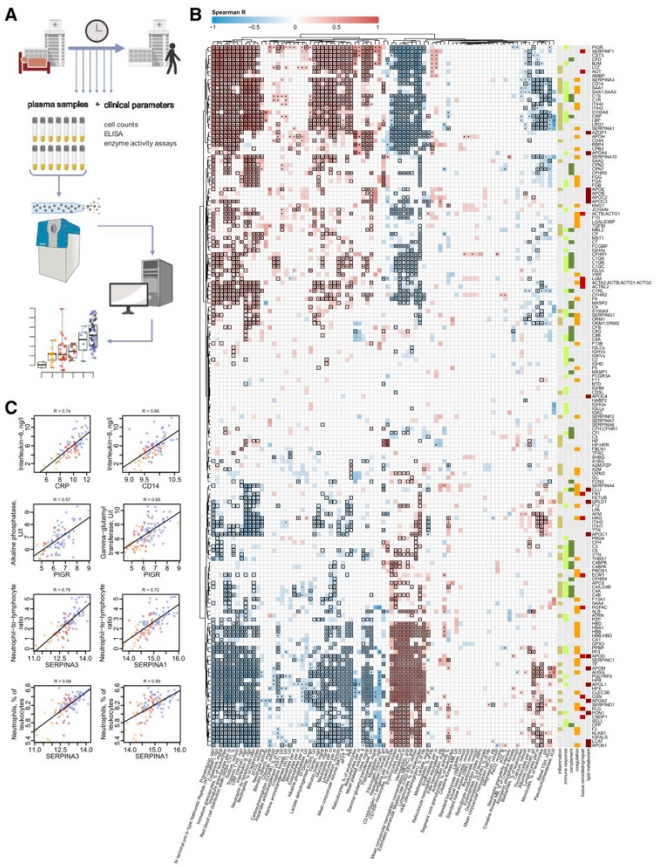


Figure 1. Interdependence of clinical, diagnostic, physiological and proteomic parameters during the clinical progression of COVID-19

(A) Study design. Schematic of the cohort of 139 patients with PCR-confirmed SARS-CoV-2 infection treated at Charité University Hospital Berlin. Plasma proteomics and accredited diagnostic tests were applied at 687 sampling points to generate high-resolution time series data for 86 routine diagnostic parameters and 321 protein quantities (see also Figures S1 and S2).

(B) Covariation map for plasma proteins and routine diagnostic and physiological parameters. Statistically significant correlations (Spearman; $p < 0.05$) are colored. Dots indicate statistical significance after row-wise multiple-testing correction (false discovery rate [FDR] < 0.05), black rectangles—column-wise. The panel on the right of the heatmap provides manual functional annotation for the proteins (see also Figures S3 and S4, and Tables S4, S5, and S6).

(C) Covariation of key diagnostic parameters and plasma protein markers (log_2-transformed) in COVID-19 (see also Tables S4, S5, and S6). Dots colors correspond to the WHO grade of the patient, see Figure 2B.

response-induced expression of PIGR and cell destruction (Schneeman et al., 2005; Turula and Wobus, 2018).

A map of plasma proteins and diagnostic parameters that depend on age and disease severity

113 proteins and 55 accredited diagnostic parameters responded in accordance to an increase in the WHO score as a measure of progressing COVID-19 severity (Figures 2, S5, and S6; STAR methods). To the best of our knowledge, more than 30 of these proteins have not been associated with COVID-19 severity previously (Table S3). The proteins that change dependent on disease severity include mediators of inflammation and immune response (CD44, B2M, PIGR, and A2M), components of the complement cascade (CFD, and CFHRs), and apolipoproteins (APOA2, APOC3, APOD, APOE, and APOL1). Furthermore, numerous markers of organ dysfunction (cardiac: NT-proBNP, troponin T; renal: creatinine, urea; liver: aspartate aminotransferase, alanine aminotransferase, gamma-glutamyl transferase, and total bilirubin) and, inversely, markers of anemia (hemoglobin, erythrocytes, and hematocrit) were correlated with the WHO grade of the patient. In order to further dissect the proteomic signatures of the most severely ill patients requiring maximum treatment (WHO grade 7), we specifically characterized the impact of organ support treatments (RRT and ECMO) on the patients' molecular phenotype (Figures S7 and S8). We showed, for instance, that HP and HPX are reduced in patients on RRT and ECMO as a sign of hemolysis in the extracorporeal circuit, whereas elevated SERPINC1 levels mirror substitution of antithrombin during ECMO. We discuss these findings in Note S1.

A total of 61 proteins and 18 diagnostic parameters varied with patients' age (Figure S9). Out of these, 37 proteins do not change with age in a pre-COVID-19 general population baseline (Generation Scotland cohort [Smith et al., 2006]), for which proteomes have been measured with the same proteomic technology (Messner et al., 2020) (Figure S10). We observed that a number of markers that increase with age in COVID-19 patients also correlated with a high WHO grade (Figures S6 and S9). To identify markers that are upregulated or downregulated in older patients in comparison with younger patients with a comparable therapy need, i.e. WHO grade, we tested the relationship between omics feature levels and age by accounting for WHO severity grade as a covariate using linear modeling

CellPress
OPEN ACCESS

Cell Systems
Article

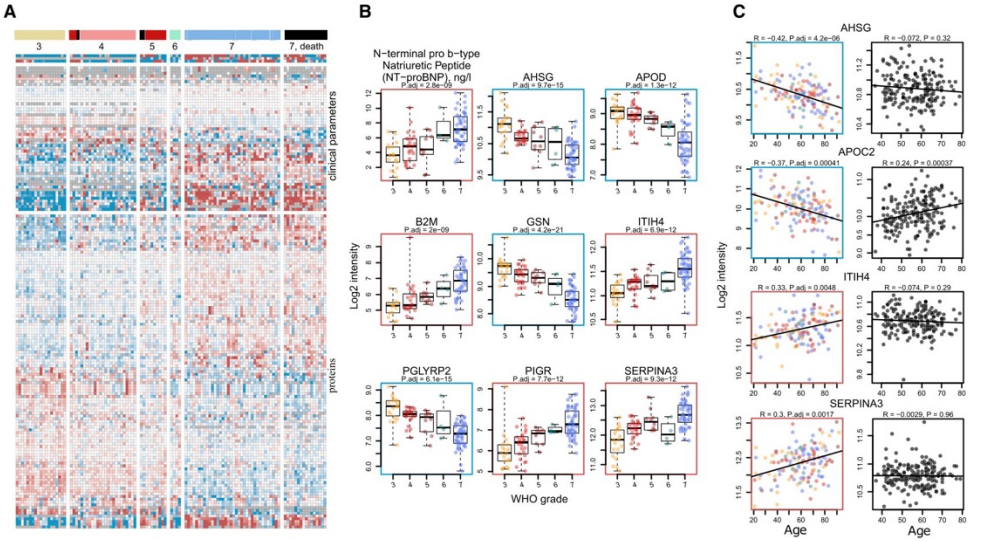


Figure 2. The molecular phenotype of patients with COVID-19 and its dependency on severity and age

(A) Plasma proteome and clinical diagnostic parameters in dependency of COVID-19 severity irrespective of age. The patients are grouped according to the maximum clinical treatment received (WHO ordinal scale), used as an indicator of disease severity (Table S1). 113 proteins and 55 routine diagnostic parameters vary significantly (FDR < 0.05) between patients of the different WHO groups upon accounting for age as a covariate using linear modeling (Ritchie et al., 2015). A fully annotated heatmap is provided in Figure S5 (see also Figure S6 and Table S3).

(B) Selected protein markers and routine diagnostic parameters (log_2-transformed) plotted against the WHO ordinal scale.

(C) Selected proteins differentially abundant depending on age (FDR < 0.05). Left, colored: this data set (log_2-transformed levels; statistical testing was performed by accounting for the WHO grade as a covariate Ritchie et al., 2015 and STAR methods; for visualization only, the data were corrected for the WHO grade); right, black: general population (log_2-transformed levels; Generation Scotland cohort).

(Ritchie et al., 2015) (STAR methods). This analysis identified 36 proteins and 12 clinical laboratory markers that are up- or are downregulated with age in COVID-19 patients within the same level of care, i.e., one WHO grade (Figures 2C and S11, summarized in Figure 5). Out of these, 20 proteins do not change with age in the pre-COVID-19 population baseline (Generation Scotland cohort proteome data, Messner et al., 2020; Figure S10), or show the opposite correlation with age in the general population (e.g., APOC2, Figure 2C). These proteins that only show an age-dependency in COVID-19 patients but not in the general population point toward age-dependent differences in host response patterns to SARS-CoV-2, and include markers involved in inflammation (SERPINA3, ITIH4, SAA1, SAA1, SAA2, ITIH3, CFB, C7, and AHSG), lipid metabolism (APOC1, APOC2, APOC3, APOB, and APOD), and coagulation (KLKB1, and FBLN1). We consider the implications of these findings in Note S2.

Time-dependent alleviation of severity indicators highlights the role of the early host response in COVID-19 progression

The time-resolved nature of our study facilitated a covariation analysis of protein levels and accredited diagnostic parameters along the patient trajectory over time (Figure S12; Table S7). Correlating the dynamics of omics features during the peak period of the disease (STAR methods), we noted covariation of inflammatory markers, APPs, fibrinogen precursor proteins, and the NLR. The correlation between APPs and the markers of cardiac and renal impairment observed across different patients at the earliest time points (Figure 1B; Table S4) was not reflected as a trend over time (Figure S13).

To further dissect the dynamics of the patients' molecular phenotype during the course of COVID-19, we determined the longitudinal trend for all protein and diagnostic parameters during the peak period of the disease (i.e., while receiving maximum treatment; STAR methods). In total, 89 proteins and 37 clinical parameters significantly changed over time (Figure 3B, trends across all time points at the maximum WHO grade are provided in Figure S14; STAR methods). In general, we found that most proteins and diagnostic parameters that correlate with disease severity return toward baseline during the peak period of the disease. Many of these were most prominently changed in the early samples (Figure S6) but alleviated with time, irrespective of the outcome (Figure S14; summarized in Figure 5). For example, components of the coagulation cascade with known acute phase activity, such as fibrinogen, and many complement factors, significantly decreased over time. Proteins indicative of inflammatory response (e.g., ORM1, SERPINA1 and SERPINA3, SAA1, SAA2 [Luo et al., 2015; Sack, 2018; Wu et al., 2015]) and markers of inflammation,

Cell Systems
Article

CellPress
OPEN ACCESS

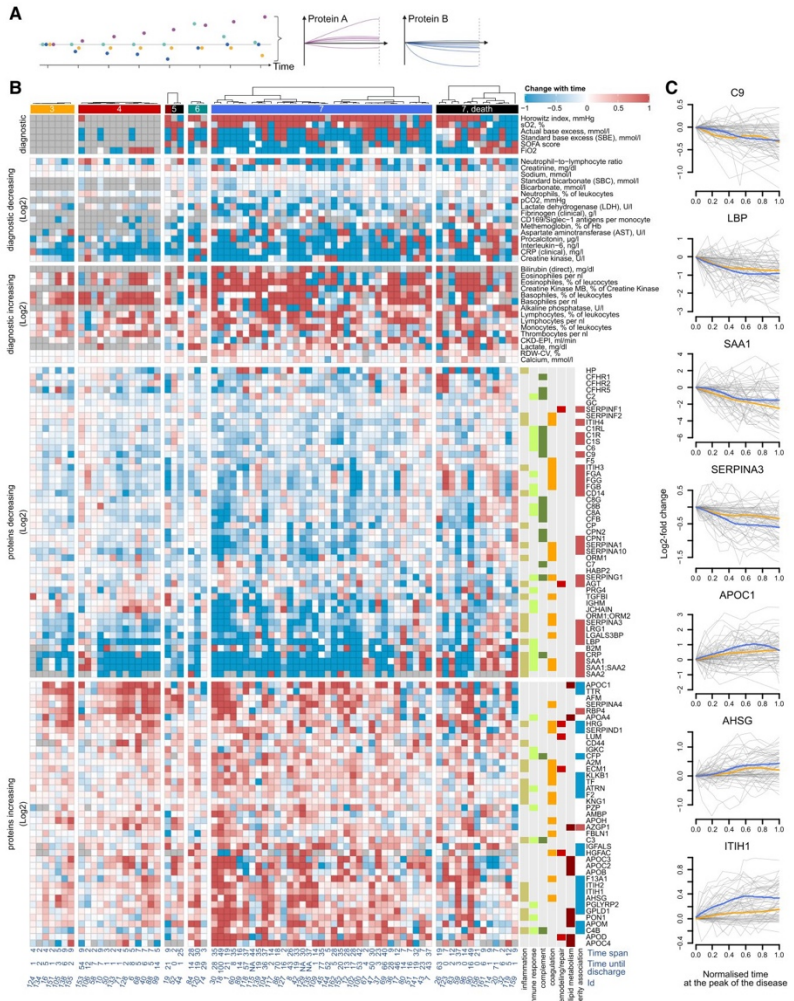


Figure 3. The progression of the COVID-19 molecular patient phenotype over time

(A) Schematic: each patient is followed during inpatient care by repetitive sampling, and the "trajectory" of each of the proteins and the routine diagnostic features is analyzed (points of different colors at each time point) (see also Figure S2).

(B) Protein levels and routine diagnostic parameters that change significantly (FDR < 0.05) over time during the peak of the disease, shown for individual patients stratified by their maximum treatment received (WHO grade): 89 proteins, 37 clinical diagnostic markers show time dependency during the disease course (illustrated as $\log_2$-fold changes or absolute value changes, as indicated). The panel to the right of the heatmap provides manual functional annotation for the proteins. Known associations with COVID-19 severity are indicated (blue - downregulated in severe COVID-19, and red - upregulated) (D'Alessandro et al., 2020; Laing et al., 2020; Messner et al., 2020; Shen et al., 2020). Below the heatmap, the time span between the first and the last sampling time point at the peak of the disease is indicated as well as the remaining time until the discharge (see also Figures S14 and S15, and Table S3).

(C) Trajectories (change of $\log_2$-transformed levels with time) for selected proteins. Sampling points during the peak period of the disease (STAR methods) are considered. x axis: 0 – first time point measured at the peak of the disease, 1 – last. The y axis reflects the change relative to the first valid measurement during the peak of the disease. Loess approximations are shown for patients, which did (blue), and did not (orange), require invasive mechanical ventilation. See also Figure S16.

CellPress
OPEN ACCESS

Cell Systems
Article

such as CRP or IL-6, also declined over time. Conversely, extracellular matrix (ECM) proteins, such as ECM1, LUM, and immunoregulatory factors (e.g. AHSG, A2M Rehman et al., 2013, and HRG Wakabayashi, 2013) and proteins involved in lipid metabolism (e.g., APOC1, APOD, APOM, GPLD1, and PON1), and negative APPs (e.g., ITIH1, Figure 3C), which are downregulated in severe COVID-19 (Figure S6, summarized in Figure 5), increased over time, approaching the baseline. This general alleviation of the initial molecular phenotype of COVID-19 was consistently detected in both mildly and severely ill patients (outlier trajectories discussed in Note S3). Indeed, only 13 proteins showed differences in trend depending on the WHO score (Figure S15). We provide visualization of individual trajectories for all omics features measured between the first and the last time points sampled at the peak of the disease (Figures 3C and S16).

Overall, the molecular patient phenotype reflected an initial spike in the systemic inflammatory response, which alleviated gradually, followed by a protein signature indicative of tissue repair, metabolic reconstitution, and immunomodulation. This was observed in both mildly and severely ill patients, highlighting the early disease phase as a major molecular determinant of the COVID-19 phenotype.

Proteomes and diagnostic clinical markers allow for prediction of disease severity by machine learning

Using a machine learning algorithm based on gradient boosted trees (STAR methods), we first evaluated the extent to which diagnostic parameters and proteomes characterize treatment requirements, as reflected by the WHO grade. Both proteomes and clinical diagnostic parameters were highly discriminative of the patient receiving invasive mechanical ventilation (WHO grade 6 or 7, clinical laboratory values AUROC = 0.97, proteomic data AUROC = 0.98, combined data AUROC = 0.99; Figure 4C). The machine learning models significantly outperformed the predictive scores derived from established COVID-19 risk factors such as age, BMI, Charlson comorbidity index (CCI), or molecular predictors such as CRP or IL-6 levels (Figure 4C). In order to assess the transferability of the proteomic predictors, we tested our model in an independent cohort of 99 hospitalized patients with COVID-19 from another hospital in a different healthcare system (Innsbruck cohort, STAR methods). The proteomic model trained on the main Charité cohort demonstrated a comparably high patient stratification performance when applied to this validation cohort (Figure 4D; AUROC = 0.97). Scores reflecting the contribution of individual proteins and clinical parameters to the machine learning model are provided in Table S3. Of note, we were able to establish machine learning models that not merely classified patients based on severity but were able to predict the current WHO severity grade from the proteome, from clinical measurements, and both (Figure 4E). Again, combined proteomic and clinical laboratory data performed best.

Having observed clear time trajectories for many proteins and diagnostic parameters, we hypothesized that the molecular signature of the initial host response can be exploited for the prediction of the future disease course. We started by investigating the potential of using the levels of proteins and diagnostic parameters for prediction of future clinical worsening, defined as progression to a higher severity grade on the WHO scale, i.e., a requirement for supplemental low-flow oxygen therapy, high-flow oxygen therapy, or invasive mechanical ventilation. Upon using a linear model to account for current therapy (WHO grade) and age as covariates, 11 proteins and 9 clinical laboratory markers were identified as predictors of future worsening of the clinical condition, across all treatment groups (STAR methods) (Figures 4A and S17; Box 1). Increased or decreased plasma levels of these proteins functioning in inflammation (CRP, ITIH2, SERPINA3, AHSG, and B2M), coagulation (HRG, and PLG), and complement activation (C1R, and CFD), as well as levels of AGT and CST3, were predictive of future clinical deterioration.

Next, we investigated the predictability of the remaining time needed in the hospital for mildly ill patients with maximum WHO grade 3. We identified 26 protein biomarkers and 14 routine diagnostic markers (Figures 4B and S18) that correlate with the time between the first sampling point and discharge from inpatient care. The proteomic signature associated with a longer need for inpatient treatment is characterized by proteins of the complement system (C1QA, C1QB, and C1QC) and reflects altered coagulation (KLKB1, PLG, and SERPIND1) and inflammation (CD14, B2M, SERPINA3, CRP, GPLD1, PGLYRP2, and AHSG). As most of these proteins are also predictors of the required treatment (Figure S6; Table S3), we hypothesized that the time of inpatient care for mild (WHO grade 3) cases correlates with the severity of the disease in these patients. To test this hypothesis, we generated machine learning models for WHO grade prediction, similar to those shown in Figure 4E, but trained the model only on the first time point data measured for each patient (to avoid using any future information with respect to that time point). We observed that the predictions derived from the first time point data correlated with the remaining time in the hospital (Figure 4F). We conclude that machine learning allows us to finely distinguish between more and less severe patients within a single treatment group, i.e. WHO grade.

DISCUSSION

Upfront clinical decision making is essential for optimum treatment allocation to patients as well as for efficient resource management within the hospital. For instance, early referral to intensive care treatment units has been shown to improve prognosis and outcome for patients with severe COVID-19 (Sun et al., 2020). One of the peculiarities of COVID-19 is that the examinable clinical conditions of patients often do not reflect the true severity of the disease, e.g., with respect to respiratory insufficiency. In contrast to patients with severe bacterial pneumonia, patients with COVID-19 often clinically appear to be only slightly affected, despite being in severe respiratory failure, a phenomenon termed "happy hypoxemia" (Stawicki et al., 2020). Clinical decisions therefore need to be supported by objective, molecular diagnostics. These diagnostic analyses help further in the monitoring of therapies and clinical trials as they allow for determining the extent to which a given patient has deviated from the disease trajectory that would be achieved without therapy.

Several recent investigations have identified protein biomarkers and clinical parameters that classify patients with COVID-19 according to disease severity and/or received treatment (D'Alessandro et al., 2020; Laing et al., 2020; Liu et al., 2020b; Messner et al., 2020; Overmyer et al., 2020; Schulte-Schrepping et al., 2020; Shen et al., 2020; Shu et al., 2020;

Cell Systems
Article

CellPress
OPEN ACCESS

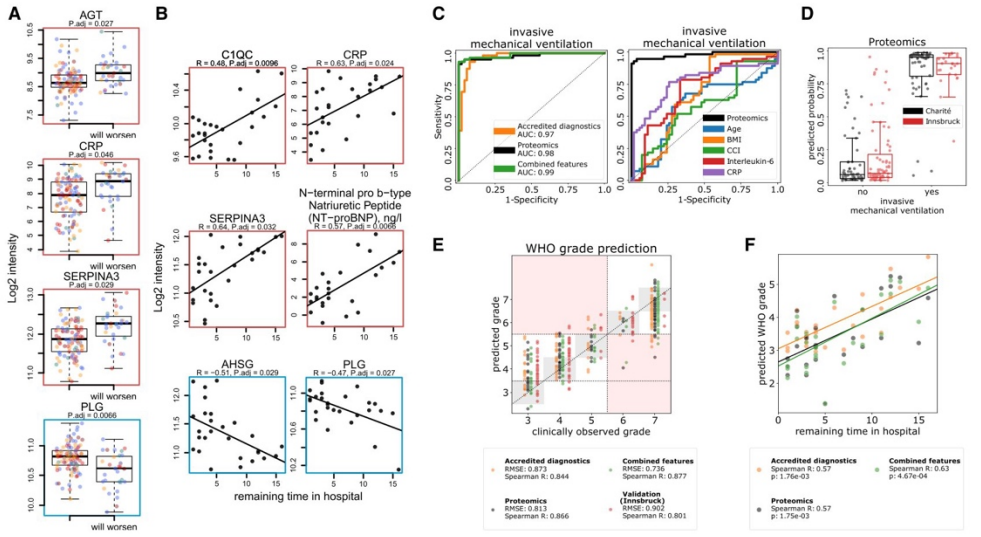


Figure 4. Predicting COVID-19 treatment requirement and future disease progression from the early molecular phenotype by using machine learning.

(A) Selected proteins that are predictive (FDR < 0.05) of the future clinical deterioration of the disease (that is progression to a higher WHO grade in the future; STAR methods). Illustrated are the $\log_2$-transformed levels of the proteins at the first sampling point upon correction (for visualization only) for the impact of the WHO grade and age as covariates (Ritchie et al., 2015) (see also Figure S17).

(B) Selected proteins and routine diagnostic parameters predictive (FDR < 0.05) of the remaining time in hospital for patients receiving mild treatment (WHO grade 3). Statistical testing was performed by including patient's age as a covariate (STAR methods). Illustrated are the $\log_2$-transformed levels of the proteins (upon correction for age as a covariate, for visualization only) at the first sampling point, plotted against the remaining time in hospital (days) (see also Figure S18).

(C) Left: performance of a machine learning model characterizing the need for invasive mechanical ventilation, based on either the proteomic data, accredited diagnostic parameters, or both. Right: comparison of the performance of a machine learning model characterising the need for invasive mechanical ventilation based on proteomic data to established prognostic parameters.

(D) Prediction performance, based on the proteome, visualized as boxplots. Cross-validation predictions on the Charité cohort are shown in black, predictions of a model trained on the Charité data and then applied to an independent cohort from another hospital (Innsbruck cohort) are shown in red.

(E) Prediction of the WHO grade itself using machine learning (cross-validated, first time point at the maximum treatment level for each patient is used, STAR methods), based on either the proteome, clinical diagnostic parameters, or both. The performance of the proteomic model trained on the Charité cohort and applied to the Innsbruck cohort is also shown.

(F) A machine learning model was trained to predict the level of necessary treatment (WHO grade) using the data (proteomics, clinical, or both) from the first time point measured for each patient. Derived predictions for patients who did not receive supplemental oxygen at this time point (WHO = 3) were plotted against the remaining time (days) in hospital for these patients.

Wynants et al., 2020). In other studies, the potential prognostic value of several established and newly discovered markers for predicting the future course of the disease was investigated, e.g., for IL-6, ferritin, or resistin (Grifoni et al., 2020; Meizlish et al., 2020). Yet, it remained challenging so far, to put their prognostic value in relation to patient age and current level of care, the two most important apparent characteristics for prognosis in COVID-19. For instance, a patient at WHO grade 5 who requires high-flow oxygen therapy is significantly more likely to progress to mechanical ventilation and subsequently die than an inpatient at WHO grade 3 that does not require oxygen support. Likewise, a 90-year-old patient at WHO grade 3 is significantly more likely to progress to more severe disease and to stay in the hospital for a longer period of time than a 20-year-old patient at the same WHO grade.

To identify (1) which proteomic markers and laboratory parameters correlate with each other by being attributed to a common biological or physiological response, and (2) which markers reflect disease trajectories, we longitudinally phenotyped 139 individuals admitted at Charité University Hospital, Berlin, Germany, between March 01, 2020, and June 30, 2020, due to PCR-confirmed SARS-CoV-2 infection (Figure S2). We recorded a large panel of 86 parameters with accredited diagnostic tests. These tests capture the compendium of analytical parameters that are available for the clinical routine. In parallel, we record plasma proteomes using a recently developed mass spectrometry platform. This platform introduced the use of analytical (high-flow rate) chromatography to routine proteomics in order to increase throughput and measurement precision to the scale of clinical trials (Messner et al., 2020). The platform reaches a

CellPress
OPEN ACCESS

Cell Systems
Article

Box 1. Proteins predictive of future worsening, i.e., disease progression to higher WHO grade

HIGH LEVELS INDICATIVE OF POOR PROGNOSIS

AGT: Angiotensinogen: Conversion via angiotensin-converting enzymes ACE and ACE2 produces AngI/AngII (pro-inflammatory, vasoconstrictive, pro-fibrotic) and Ang1-7/Ang1-9 (anti-inflammatory, vasodilative, anti-fibrotic), respectively (Turner, 2015; Zhang et al., 2020). Increased AGT likely reflects increased AngI/AngII due to SARS-CoV-2 mediated inactivation of ACE2 (Tay et al., 2020) and subsequently predominant conversion of AGT to AngII. AngII correlates with viral load (Liu et al., 2020c) and has tissue damaging effects (Zhang et al., 2020).
B2M: Beta-2-microglobulin: Component of major histocompatibility complex (MHC I) on all nucleated cells and platelets. Released abundantly by activated platelets leading to pro-inflammatory M1-like macrophage polarization (Hilt et al., 2019). Increase of B2M has been associated with death in patients with chronic kidney disease (Makridakis et al., 2020).
C1R: Complement C1r: Initiator of the classical complement pathway (Hajishengallis et al., 2017).
CFD: Complement Factor D: Initiator of the alternative complement pathway by cleaving Factor B (CFB) to form the C3bBb alternative pathway convertase (Volanakis and Narayana, 1996).
CRP: C-reactive protein: Acute phase protein, strongly upregulated in inflammation and infection, including COVID-19.
CST3: Cystatin C: Biomarker of kidney function (Peralta et al., 2011).
SERPINA3: Alpha-1-antichymotrypsin: Protease inhibitor of neutrophil cathepsin G (Benarafa, 2015). When cleaved at reactive site loop, it becomes stable to degradation and becomes a strong neutrophil chemoattractant (Banda et al., 1988; Potempa et al., 1991)

LOW LEVELS INDICATIVE OF POOR PROGNOSIS

AHSG: Alpha-2-HS glycoprotein (Fetuin-A): Negative acute phase protein attenuating macrophage activation and neutrophil degranulation (Ombrellino et al., 2001).
HRG: Histidine-rich glycoprotein: Negative acute phase protein, regulator of inflammation and immune response, clearance of pathogens and cell detritus as well as coagulation and fibrinolysis through a variety of interactions (Poon et al., 2011; Wakabayashi, 2013).
ITIH2: Inter-alpha-trypsin inhibitor heavy chain H2: Covalently linked to bikunin (AMBP), the complex binds to hyalarunan (HA) to form serum-derived hyaluronan-associated protein (SHAP) which has matrix-stabilizing and immunomodulatory effects (Fries and Blom, 2000; Zhuo et al., 2004).
PLG: Plasminogen, Plasmin: Mediator of fibrinolysis (Chapin and Hajjar, 2015). More recently, immunological functions including neutrophil attenuation as well as macrophage efferocytosis and polarization from pro-inflammatory M1 to tissue-repairing M2 phenotype have been identified (Heissig et al., 2020).

similar proteomic depth as other contemporary mass spectrometry technologies that address undepleted human plasma that is constrained by its huge dynamic range (Anderson and Anderson, 2002) (Box 2 for the resources generated).

The comprehensive and time-resolved molecular phenotyping of this patient cohort is complemented by a comparison with a healthy population baseline (Generation Scotland [Smith et al., 2006]) measured with the same proteomic platform (Messner et al., 2020), and the characterization of an independent cohort from an unrelated health care system (Innsbruck cohort, Austria) for validating the created predictors. The measurements were performed on samples collected during the early period of COVID-19, i.e., before immunomodulatory treatments such as dexamethasone became standard of care for severe COVID-19 (RECOVERY Collaborative Group, 2020). Our data thus reflect treatment-naive trajectories, which are of major value as baseline data for future studies.

We report an initial spike in the early inflammatory host response as a determinant for the future course of the disease. As our results indicate, the patients in our cohort showed molecular marker signatures of higher basal inflammation with increasing age, which might be partially responsible for the higher risk of severe COVID-19 in older individuals. While several approaches of targeted anti-inflammatory treatment have not been successful in preventing clinical deterioration in COVID-19 so far (Stone et al., 2020), our study indicates that this special population of older patients might benefit particularly from treatments that mitigate the inflammatory host response. We report numerous interdependencies between clinical laboratory markers and alterations in proteomes, linking, for example, clinical inflammatory markers to components of the complement cascade and the coagulation system. Using machine learning, we show that both plasma proteomes and the compendium of established diagnostic parameters can be used for accurate characterization of disease severity, significantly outperforming established individual risk markers, such as CRP or IL-6 levels. Of note, the combination of proteomic features and clinical laboratory markers repeatedly showed the best performance in the machine learning models. Furthermore, the models generated could be transferred for prediction in an independent cohort from another hospital and healthcare system, highlighting the robustness of this approach and its translational potential.

We observed a considerable overlap between prognostic markers and those that classify treatment according to COVID-19 severity (Figure 5). Out of 49 prognostic markers, 41 correlated with the WHO severity score. As an example, SERPINA3

Cell Systems
Article

CellPress
OPEN ACCESS

Box 2. Overview of resources generated

We provide deep and time-resolved resources that characterize COVID-19 at the level of plasma proteomes and established diagnostic parameters. We demonstrate the extent to which proteomes and diagnostic parameters interdepend, in initial response to the disease and in dynamics during the disease course. We show how they change with age, differ depending on the disease severity, reflect the therapy received and evolve over time. Our data have been acquired for COVID-19 patients' samples and analyzed in the context of general population proteomics (Generation Scotland) for which samples have been measured with the same proteomic technology (Messner et al., 2020), but we also expect it to be of high value as a reference for studies of other types of viral pneumonia as well as any investigations involving both routine clinical phenotyping and plasma proteomics.

Summary of the resource data generated in the study.

1. Covariation maps. We provide a covariation map between plasma proteins measured with at least 3 peptides and clinical laboratory measurements (Figure 1B; Table S4). In addition, we provide a full covariation map between all features measured in the study (Figure S3; Table S5) as well as a COVID-19 specific protein-protein covariation map (Figure S4; Table S6). Finally, we also provide a correlation map for the changes of different omics features with time (Figure S12; Table S7).
2. A map of plasma protein levels and clinical laboratory measurements depending on disease severity (Figures S5 and S6; Table S3).
3. Characterization of age-dependency of plasma protein levels and clinical laboratory measurements in COVID-19, and in comparison with the general population (Figures 2C and S9–S11; Table S3).
4. Characterization of the dynamics of plasma protein levels and clinical laboratory measurements during the course of COVID-19 (Figures 3B, S14, and S15; Table S3).
5. Characterization of the predictive power of plasma protein levels and clinical laboratory measurements in COVID-19 (Figures 4, S17, and S18; Table S3).
6. Proteomic and clinical signatures observed in severe COVID-19 patients undergoing RRT and ECMO (Figures S7 and S8; Table S3).

(Alpha-1 antichymotrypsin) can be used for both the classification of severity and prediction of future disease course. Both SERPINA3 and SERPINA1, another classifier of severity, possess anti-inflammatory properties and are involved in the protection of tissues from neutrophil elastase- and cathepsin G-mediated tissue damage (Benarafa, 2015). Our data show a strong correlation of both serpins with levels of neutrophils and NLR in peripheral blood. SERPINA1 is mainly produced by the liver but it is also produced in epithelial cells, pulmonary alveolar cells, tissue macrophages, blood monocytes, and granulocytes. Hence, this finding presumably reflects a systemic response to the increased NLR. After binding to effector enzymes, SERPIN-proteinase complexes are normally rapidly cleared from the blood but become resistant to degradation when cleaved at the reactive site loop (Gettins and Olson, 2016). Cleaved SERPINA1 and SERPINA3 have been shown to act as strong neutrophil chemoattractants (Banda et al., 1988; Potempa et al., 1991). The observed increase in levels of SERPINA1 and SERPINA3 might therefore partly reflect the more stable, chemoattractant, pro-inflammatory cleaved forms, rather than the short-lived tissue-protective proteins in severe COVID-19. Given the prominent role of neutrophil activation in severe COVID-19 (Schulte-Schrepping et al., 2020), this finding merits further investigation.

Our data also highlight angiotensinogen (AGT) as a marker for future worsening. Activation of angiotensinogen occurs via the protease renin and the endogenous angiotensin-converting enzymes ACE or ACE2. ACE converts angiotensin I (AngI) to pro-inflammatory, vasoconstrictive, and pro-fibrotic angiotensin II (AngII) (Zhang et al., 2020). ACE2, in contrast, mediates conversion of angiotensins I and II to anti-inflammatory, vasodilative, anti-fibrotic, and anti-oxidant angiotensins 1–9 (Ang1-9) and 1–7 (Ang1-7) (Turner, 2015). SARS-CoV-2 invades host cells of the lung, heart, kidneys, and other organs via ACE2, resulting in the internalization and downregulation of ACE2 (Hoffmann et al., 2020; Tay et al., 2020; Zhang et al., 2020). Subsequently, angiotensinogen is converted predominantly via ACE to AngII and is less degraded by ACE2, resulting in AngII accumulation (Batlle et al., 2012; Silhol et al., 2020). We can thus assume that the higher plasma levels of AGT gene products in severely ill patients, as measured in our study, mainly reflect the higher levels of AngII. Importantly, we observed a strong correlation of AGT with markers of acute kidney injury (AKI; creatinine, urea; Figures S3 and S13; Table S5), a frequent complication of COVID-19 and a risk factor for poor prognosis and fatal outcome (Fu et al., 2020). Aggravated by the absence of tissue-protective Ang1-7, elevated levels of AngII lead to activation of the renin-angiotensin-system (RAS) and contribute to hypoxic kidney injury (Kasal et al., 2020). Of note, apart from tissue damaging effects, AngII has been shown to linearly correlate with viral load and lung injury in SARS-CoV-2 infection (Liu et al., 2020c).

Overall, many of the markers that are both classifiers and predictors of the future disease course are initiators of the inflammatory response. This group includes some of the key initiators of the complement cascade: C1QA, C1QB, C1QC, C1R, and CFD. In contrast, severity markers without prognostic value largely include downstream effectors of inflammation-associated damage, such as GSN and the circulating actins ACTBL2 and ACTB, ACTG1, and ECM1. Thus, this high-precision, high-throughput approach can help us understand mechanisms of immune-mediated organ damage on a molecular basis.

Despite the high resolution and high throughput of the mass spectrometry platform deployed in our study, the direct translation of our results into clinical practice will require the development of a clinical assay according to FDA or EMA standards.

CellPress
OPEN ACCESS

Cell Systems
Article

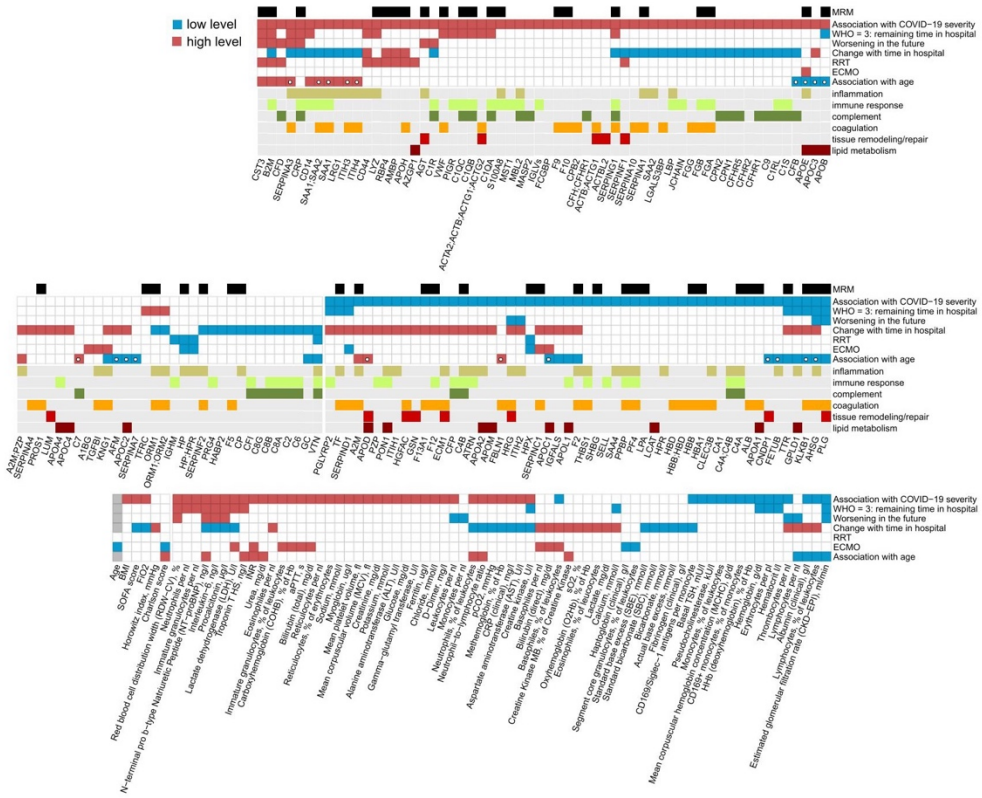


Figure 5. Summary: association of individual plasma proteins, routine diagnostic and physiological parameters with severity, necessary therapy, and progression of COVID-19.

For each statistical test considered (association with WHO grade, prediction of the remaining time in hospital for patients at WHO grade 3, prediction of worsening, i.e., progression to a higher WHO grade in the future, the trend during the peak period of the disease, association with RRT, association with ECMO and association with higher patient age), measurements, which show significant differences are highlighted, with the color indicating the trend, e.g., red for CST3 in the "Association with COVID-19 severity" test indicates higher levels of CST3 in severely ill patients. Proteins for which MRMAssayDB (Bhowmick et al., 2018) lists that a targeted proteomic assay has been developed are indicated with a black bar at the top. Proteins which change significantly with age in the Charité COVID-19 cohort (FDR < 0.05) but do not change significantly ($p < 0.05$) with age in the general population (Generation Scotland cohort), are highlighted with a white circle in the 7th row ("Association with age"). See also Figures S6–S8, S10, S11, S14, S17, and S18, and Table S3.

We further note that the use of machine learning is currently not a certified method to inform clinical decisions. However, in addition to multiple works that have successfully used machine learning for clinical prognosis previously (see recent reviews [Kelly et al., 2019; Lee and Lee, 2020; Nagendran et al., 2020; Shah et al., 2019; Vollmer et al., 2020]), our results bear a strong implication of the future potential of machine learning for clinical applications, including personalized medicine. This calls for a worldwide effort aimed at developing procedures, which would allow reliable clinical validation of machine learning predictors, their approval, and their routine deployment in the clinic.

In summary, by following a deeply phenotyped COVID-19 patient cohort over time at the level of the proteome and established diagnostic biomarkers and physiological parameters, we have created a rich data resource for understanding the extent and progression of COVID-19. We have shown that an early spike in the inflammatory response is a key determinant of COVID-19, and that future disease progression is predictable by using panels of accredited diagnostic parameters as well as proteomic measurements from early time point samples. By using machine learning, we demonstrated that the proteome allows to precisely characterize the patients' phenotype and that the resulting machine learning models are robust and perform accurately when applied to samples from a different hospital and healthcare system. Our study provides comprehensive information about the key determinants of the varying COVID-19

Cell Systems
Article

trajectories as well as marker panels for early prognosis that can be exploited for clinical decision making, to devise personalized therapies, as well as for monitoring the development of much needed COVID-19 treatments.

CONSORTIA

Malte Kleinschmidt, Katrin M. Heim, Belén Millet, Lil Meyer-Arndt, Ralf H. Hübner, Tim Andermann, Jan M. Doehn, Bastian Opitz, Birgit Sawitzki, Daniel Grund, Peter Radünzel, Mariana Schürmann, Thomas Zoller, Florian Alius, Philipp Knape, Astrid Breitbart, Yaosi Li, Felix Bremer, Panagiotis Pergantis, Dirk Schürmann, Bettina Temmesfeld-Wollbrück, Daniel Wendisch, Sophia Brumhard, Sascha S. Haenel, Claudia Conrad, Philipp Georg, Kai-Uwe Eckardt, Lukas Lehner, Jan M. Kruse, Carolin Ferse, Roland Körner, Claudia Spies, Andreas Edel, Steffen Weber-Carstens, Alexander Krannich, Saskia Zvorc, Linna Li, Uwe Behrens, Sein Schmidt, Maria Rönnefarth, Chantip Dang-Heine, Robert Röhle, Emma Lieker, Lucie Kretzler, Isabelle Wirsching, Christian Wollboldt, Yinan Wu, Georg Schwanitz, David Hillus, Stefanie Kasper, Nadine Olk, Alexandra Horn, Dana Briesemeister, Denise Treue, Michael Hummel, Victor M. Corman, Christian Drosten, and Christof von Kalle

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
  - Experimental model and subject details
  - Innsbruck Patient cohort and clinical data
- METHOD DETAILS
  - Materials
  - Mass spectrometry
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Data analysis
  - Markers of the disease severity
  - Markers varying with age in COVID-19
  - Markers of RRT and ECMO
  - Markers predictive of time in hospital
  - Markers predictive of disease worsening
  - Peak period of the disease definition
  - Markers changing during the peak of disease
  - Correlation maps
  - Prediction of current mechanical ventilation
  - WHO grade prediction
  - Prediction of the remaining time in hospital
  - Supplementary Note 1. Diagnostic parameters and Proteome signatures that indicate therapeutic interventions
  - Supplementary Note 2. Age-specific response to COVID-19 in the context of severity markers
  - Supplementary Note 3. Diverging trends at the proteome level during the disease peak in individual patients

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cels.2021.05.005.

ACKNOWLEDGMENTS

We thank Robert Lane, Jean-Baptiste Vincedent and Nick Morrice (Sciex) for help with the TripleTOF 6600. This work was supported by the Berlin University Alliance (501_Massenspektrometrie, 501_Linklab, 112_PreEP_Corona_Ralser), by UKRI/NIHR through the UK Coronavirus Immunology Consortium (UK-CIC), the BMBF/DLR Projektträger (01KI20160A, 01ZX1604B, 01KI20337, 01KX2021), Charité-BIH Centrum für Therapieforschung (BIH_PA_covid-19_Ralser), the BBSRC (BB/N015215/1, BB/N015282/1), the Francis Crick Institute, which receives its core funding from Cancer Research UK (FC001134), the UK Medical Research Council (FC001134), and the Wellcome Trust (FC001134 and IA 200829/Z/16/Z), as well as the European Research Council (SyG 951475 to M.R.). This work was further supported by the Ministry of Education and Research (BMBF), as a part of the National Research Node 'mass spectrometry in Systems Medicine (MSCoresys), under grant agreement 031L0220A. This study was further supported by the German Federal Ministry of Education and Research (NaFoUniMedCovid19 – NUM-NAPKON, NUM-COVIM, FKZ: 01KX2021 and PROVID—FKZ: 01KI20160A) to F.K., L.E.S., M.W., N.S., and S.H.; L.E.S. is supported by the German Research Foundation (DFG, SFB-TR84 114933180) and by the Berlin Institute of Health (BIH), which receives funding from the Ministry of Education and Research (BMBF). M.W. is supported by grants from the German Research Foundation, SFB-TR84 C06 and C09, by the German Ministry of Education and Research (BMBF) in the framework of the CAPSyS (01ZX1304B), CAPSyS-COVID (01ZX1604B), SYMPATH (01ZX1906A) and PROVID project (01KI20160A) and by the Berlin Institute of Health (CM-COVID). S.H. is supported by the German Research Foundation (DFG, SFB-TR84 A04 and B06), and the BMBF (PROVID, and project 01KI2082). N.S. is supported by grants from the German Research Foundation, SFB-TR84 C09 und Z02, by the German Ministry of Education and Research (BMBF) in the framework of the PROGRESS 01KI07114. This study was further supported by Wellcome Trust (200829/Z/16/Z). The Generation Scotland study received core support from the Chief Scientist Office of the Scottish Government Health Directorates (CZD/16/6) and the Scottish Funding Council (HR03006), and is now supported by the Wellcome Trust (216767/Z/19/Z). A.C. is funded by HDR UK and the Wellcome Trust (216767/Z/19/Z). C.H. is supported by an MRC University Unit Programme grant (MC_UU_00007/10) (QTL in Health and Disease). R.M. is supported by an Alzheimer's Research UK project grant (ARUK-PG2017B-10). H.W., J.F.T., A.Z., and T.N. are supported by a Medical Research Council grant (MR/R02524X/1) and H.W., A.Z., and O.B. by the Ministry of Science and Higher Education agreement no. 075-15-2020-808. H.W. is supported by the National Institute for Health Research (NIHR) Imperial Biomedical Research Centre (BRC). J.F.T is supported by the National Institute for Health Research (NIHR) UCLH/UCL Biomedical Research Centre. M.M. is a participant in the Bih-Charité Digital Clinician Scientist Program funded by the Charité – Universitätsmedizin Berlin, the Berlin Institute of Health, and the German Research Foundation (DFG). M.A.K. is supported by the Austrian Science Funds (FWF; P33333) and the Austrian Research Promotion Agency (FFG, #878654). Figures were created with biorender.com

AUTHOR CONTRIBUTIONS

L.E.S., M.R., and F.K. designed the study. A.F., D.L., and M.M. conducted experiments. P.T.-L., C.T., A.F., D.L., E.T.H., P.S., C.B.M., M.J., T.S., S.J.K., A.P., Y.W., S.S., R.H., B.S., S.W., M.M., F.M., C.G., C.R.-G., T.L., L.B.J., M.S.S., M.P., L.J., S.D., L.J.L., D.Z., P.E., A.U., H.Z., J.L.-R., M.A.K., I.T., H.M.-R., M.W., N.S., L.E.S., and F.K. acquired clinical samples and data. V.D., P.T.-L., O.L., T.N., C.T., H.W., S.K.A., A.R., N.-M.G., L.S., S.V., M.W., C.B.M., O.B., A.Z., A.C., C.H., D.J.P., R.E.M., J.F.T., A.Z., K.L., S.H., M.M., M.R., and F.K. analyzed the data. V.D., P.T.-L., C.T., C.C.-M., E.T.H., P.S., S.H., M.R., M.M., T.L., L.J.L., K.L., L.E.S., M.R., and F.K. interpreted the data. V.D., P.T.-L., M.R., F.K., and L.E.S. wrote the first draft of the manuscript. All authors contributed to finalizing the manuscript.

CellPress
OPEN ACCESS

Cell Systems
Article

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: November 11, 2020
Revised: March 24, 2021
Accepted: May 7, 2021
Published: June 14, 2021

REFERENCES

Ağirbaşli, M., Song, J., Lei, F., Wang, S., Kunselman, A.R., Clark, J.B., Myers, J.L., and Ündar, A. (2015). Apolipoprotein E levels in pediatric patients undergoing cardiopulmonary bypass. Artif. Organs *39*, 28–33.

Alwan, N.A., Burgess, R.A., Ashworth, S., Beale, R., Bhadelia, N., Bogaert, D., Dowd, J., Eckerle, I., Goldman, L.R., Greenhalgh, T., et al. (2020). Scientific consensus on the COVID-19 pandemic: we need to act now. Lancet *396*, e71–e72.

Anderson, N.L., and Anderson, N.G. (2002). The human plasma proteome: history, character, and diagnostic prospects. Mol. Cell. Proteomics *1*, 845–867.

ARDS Definition Task Force, Ranieri, V.M., Rubenfeld, G.D., Thompson, B.T., Ferguson, N.D., Caldwell, E., Fan, E., Camporota, L., and Slutsky, A.S. (2012). Acute respiratory distress syndrome: the Berlin definition. JAMA *307*, 2526–2533.

Banda, M.J., Rice, A.G., Griffin, G.L., and Senior, R.M. (1988). Alpha 1-proteinase inhibitor is a neutrophil chemoattractant after proteolytic inactivation by macrophage elastase. J. Biol. Chem. *263*, 4481–4484.

Batlle, D., Wysocki, J., Soler, M.J., and Ranganath, K. (2012). Angiotensin-converting enzyme 2: enhancing the degradation of angiotensin II as a potential therapy for diabetic nephropathy. Kidney Int *81*, 520–528.

Benarafa, C. (2015). Regulation of neutrophil serine proteases by intracellular serpins. In The Serpin Family: Proteins with Multiple Functions in Health and Disease, M. Geiger, F. Wahlmüller, and M. Furtmüller, eds. (Springer International Publishing)), pp. 59–76.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society: Series B (Methodological) *57*, 289–300.

Bhowmick, P., Mohammed, Y., and Borchers, C.H. (2018). MRMAssayDB: an integrated resource for validated targeted proteomics assays. Bioinformatics *34*, 3566–3571.

Blumenthal, D., Fowler, E.J., Abrams, M., and Collins, S.R. (2020). Covid-19 - implications for the health care system. N. Engl. J. Med. *383*, 1483–1488.

Chapin, J.C., and Hajjar, K.A. (2015). Fibrinolysis and the control of blood coagulation. Blood Rev *29*, 17–24.

Chen, T., and Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD international conference on Knowledge Discovery and Data Mining, pp. 785–794.

Chung, H.Y., Kim, D.H., Lee, E.K., Chung, K.W., Chung, S., Lee, B., Seo, A.Y., Chung, J.H., Jung, Y.S., Im, E., et al. (2019). Redefining chronic inflammation in aging and age-related diseases: proposal of the Senoinflammation concept. Aging Dis *10*, 367–382.

Cox, J., Hein, M.Y., Luber, C.A., Paron, I., Nagaraj, N., and Mann, M. (2014). Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. Mol. Cell. Proteomics *13*, 2513–2526.

D'Alessandro, A., Thomas, T., Dzieciatkowska, M., Hill, R.C., Francis, R.O., Hudson, K.E., Zimring, J.C., Hod, E.A., Spitalnik, S.L., and Hansen, K.C. (2020). Serum proteomics in COVID-19 patients: altered coagulation and complement status as a function of IL-6 level. J. Proteome Res. *19*, 4417–4427.

Danwang, C., Endomba, F.T., Nkeck, J.R., Wouna, D.L.A., Robert, A., and Noubiap, J.J. (2020). A meta-analysis of potential biomarkers associated with severity of coronavirus disease 2019 (COVID-19). Biomark. Res. *8*, 37.

Dassati, S., Waldner, A., and Schweigreiter, R. (2014). Apolipoprotein D takes center stage in the stress response of the aging and degenerative brain. Neurobiol. Aging *35*, 1632–1642.

Demichev, V., Messner, C.B., Vernardis, S.I., Lilley, K.S., and Ralser, M. (2020). DIA-NN: neural networks and interference correction enable deep proteome coverage in high throughput. Nat. Methods *17*, 41–44.

Do Carmo, S., Jacomy, H., Talbot, P.J., and Rassart, E. (2008). Neuroprotective effect of apolipoprotein D against human coronavirus OC43-induced encephalitis in mice. J. Neurosci. *28*, 10330–10338.

Ferrucci, L., Corsi, A., Lauretani, F., Bandinelli, S., Bartali, B., Taub, D.D., Guralnik, J.M., and Longo, D.L. (2005). The origins of age-related proinflammatory state. Blood *105*, 2294–2299.

Figueroa, D.M., Gordon, E.M., Yao, X., and Levine, S.J. (2019). Apolipoproteins as context-dependent regulators of lung inflammation. In Mechanisms and Manifestations of Obesity in Lung Disease, R.A. Johnston and B.T. Suratt, eds. (Academic Press), pp. 301–326.

Franceschi, C., Garagnani, P., Parini, P., Giuliani, C., and Santoro, A. (2018). Inflammaging: a new immune-metabolic viewpoint for age-related diseases. Nat. Rev. Endocrinol. *14*, 576–590.

Fries, E., and Blom, A.M. (2000). Bikunin–not just a plasma proteinase inhibitor. Int. J. Biochem. Cell Biol. *32*, 125–137.

Fu, E.L., Janse, R.J., de Jong, Y., van der Endt, V.H.W., Milders, J., van der Willik, E.M., de Rooij, E.N.M., Dekkers, O.M., Rotmans, J.I., and van Diepen, M. (2020). Acute kidney injury and kidney replacement therapy in COVID-19: a systematic review and meta-analysis. Clin. Kidney J. *13*, 550–563.

Gettins, P.G.W., and Olson, S.T. (2016). Inhibitory serpins. New insights into their folding, polymerization, regulation and clearance. Biochem. J. *473*, 2273–2293.

Gillet, L.C., Navarro, P., Tate, S., Röst, H., Selevsek, N., Reiter, L., Bonner, R., and Aebersold, R. (2012). Targeted data extraction of the MS/MS spectra generated by data- independent acquisition: a new concept for consistent and accurate proteome analysis. Mol. Cell. Proteomics *11*, O111.016717.

Gordon, S.M. (2014). Proteomic diversity in HDL: a driving force for particle function and target for therapeutic intervention. In The HDL Handbook, Second Edition, T. Komoda, ed. (Academic Press), pp. 293–322.

Goronzy, J.J., and Weyand, C.M. (2013). Understanding immunosenescence to improve responses to vaccines. Nat. Immunol. *14*, 428–436.

Grifoni, E., Valoriani, A., Cei, F., Lamanna, R., Gelli, A.M.G., Ciambotti, B., Vannucchi, V., Moroni, F., Pelagatti, L., Tarquini, R., et al. (2020). Interleukin-6 as prognosticator in patients with COVID-19. J. Infect. *81*, 452–482.

Gu, Z., Eils, R., and Schlesner, M. (2016). Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics *32*, 2847–2849.

Hadjadj, J., Yatim, N., Barnabei, L., Corneau, A., Boussier, J., Smith, N., Péré, H., Charbit, B., Bondet, V., Chenevier-Gobeaux, C., et al. (2020). Impaired type I interferon activity and inflammatory responses in severe COVID-19 patients. Science *369*, 718–724.

Hajishengallis, George, Reis, E.S., Mastellos, D.C., Ricklin, D., and Lambris, J.D. (2017). Novel mechanisms and functions of complement. Nature Immunology *18*, 1288–1298. https://doi.org/10.1038/ni.3858.

Han, S., Yang, K., Zhu, H., Liu, J., Zhang, L., and Zhao, J. (2018). Proteomics investigation of the changes in serum proteins after high- and low-flux hemodialysis. Ren. Fail. *40*, 506–513.

Harris, S.E., Riggio, V., Evenden, L., Gilchrist, T., McCafferty, S., Murphy, L., Wrobel, N., Taylor, A.M., Corley, J., Pattie, A., et al. (2017). Age-related gene expression changes, and transcriptome wide association study of physical and cognitive aging traits, in the Lothian Birth Cohort 1936. Aging *9*, 2489–2503.

Heissig, B., Salama, Y., Takahashi, S., Osada, T., and Hattori, K. (2020). The multifaceted role of plasminogen in inflammation. Cell Signal *75*. https://doi.org/10.1016/j.cellsig.2020.109761.

Henry, B.M., de Oliveira, M.H.S., Benoit, S., Plebani, M., and Lippi, G. (2020). Hematologic, biochemical and immune biomarker abnormalities associated with severe illness and mortality in coronavirus disease 2019 (COVID-19): a meta-analysis. Clin. Chem. Lab. Med. *58*, 1021–1028.

Hilt, Z.T., Pariser, D.N., Ture, S.K., Mohan, A., Quijada, P., Asante, A.A., Cameron, S.J., Sterling, J.A., Merkel, A.R., Johanson, A.L., et al. (2019).

Platelet-derived β2M regulates monocyte inflammatory responses. JCI Insight *4*. https://doi.org/10.1172/jci.insight.122943.

Hoffmann, M., Kleine-Weber, H., Schroeder, S., Krüger, N., Herrler, T., Erichsen, S., Schiergens, T.S., Herrler, G., Wu, N.-H., Nitsche, A., et al. (2020). SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. Cell *181*, 271–280.e8.

Kasal, D.A., De Lorenzo, A., and Tibiriçá, E. (2020). COVID-19 and microvascular disease: pathophysiology of SARS-CoV-2 infection with focus on the renin-angiotensin system. Heart Lung Circ *29*, 1596–1602.

Kelly, C.J., Karthikesalingam, A., Suleyman, M., Corrado, G., and King, D. (2019). Key challenges for delivering clinical impact with artificial intelligence. BMC Med *17*, 195.

Kurth, F., Roennefarth, M., Thibeault, C., Corman, V.M., Müller-Redetzky, H., Mittermaier, M., Ruwwe-Glösenkamp, C., Heim, K.M., Krannich, A., Zvorc, S., et al. (2020). Studying the pathophysiology of coronavirus disease 2019: a protocol for the Berlin prospective COVID-19 patient cohort (Pa-COVID-19). Infection *48*, 619–626.

Laing, A.G., Lorenc, A., Del Molino Del Barrio, I., Das, A., Fish, M., Monin, L., Muñoz-Ruiz, M., McKenzie, D.R., Hayday, T.S., Francos-Quijorna, I., et al. (2020). A dynamic COVID-19 immune signature includes associations with poor prognosis. Nat. Med. *26*, 1623–1635.

Lee, C.S., and Lee, A.Y. (2020). Clinical applications of continual learning machine learning. Lancet Digit. Health *2*, e279–e281.

Lian, J., Jin, C., Hao, S., Zhang, X., Yang, M., Jin, X., Lu, Y., Hu, J., Zhang, S., Zheng, L., et al. (2020). High neutrophil-to-lymphocyte ratio associated with progression to critical illness in older patients with COVID-19: a multicenter retrospective study. Aging *12*, 13849–13859.

Liu, J., Liu, Y., Xiang, P., Pu, L., Xiong, H., Li, C., Zhang, M., Tan, J., Xu, Y., Song, R., et al. (2020a). Neutrophil-to-lymphocyte ratio predicts critical illness patients with 2019 coronavirus disease in the early stage. J. Transl. Med. *18*, 206.

Liu, Y., Gao, W., Guo, W., Guo, Y., Shi, M., Dong, G., Ge, Q., Zhu, J., and Lu, J. (2020b). Prominent coagulation disorder is closely related to inflammatory response and could be as a prognostic indicator for ICU patients with COVID-19. J. Thromb. Thrombolysis *50*, 825–832.

Liu, Y., Yang, Y., Zhang, C., Huang, F., Wang, F., Yuan, J., Wang, Z., Li, J., Li, J., Feng, C., et al. (2020c). Clinical and biochemical indexes from 2019-nCoV infected patients linked to viral loads and lung injury. Sci. China Life Sci. *63*, 364–374.

Liu, Y.C., Zou, X.B., Chai, Y.F., and Yao, Y.M. (2014). Macrophage polarization in inflammatory diseases. Int. J. Biol. Sci. *10*, 520–529.

Luo, Z., Lei, H., Sun, Y., Liu, X., and Su, D.F. (2015). Orosomucoid, an acute response protein with multiple modulating activities. J. Physiol. Biochem. *71*, 329–340.

Makridakis, M., Kontostathi, G., Petra, E., Stroggilos, R., Lygirou, V., Filip, S., Duranton, F., Mischak, H., Argiles, A., Zoidakis, J., and Vlahou, A. (2020). Multiplexed MRM-based protein quantification of putative prognostic biomarkers for chronic kidney disease progression in plasma. Sci. Rep. *10*, 4815.

McDonnell, T., Wincup, C., Buchholz, I., Pericleous, C., Giles, I., Ripoll, V., Cohen, H., Delcea, M., and Rahman, A. (2020). The role of beta-2-glycoprotein I in health and disease associating structure with function: more than just APS. Blood Rev *39*, 100610.

Meizlish, M.L., Pine, A.B., Bishai, J.D., Goshua, G., Nadelmann, E.R., Simonov, M., Chang, C.-H., Zhang, H., Shallow, M., Bahel, P., et al. (2020). A neutrophil activation signature predicts critical illness and mortality in COVID-19. medRxiv. https://doi.org/10.1182/bloodadvances.2020003568.

Messner, C.B., Demichev, V., Wendisch, D., Michalick, L., White, M., Freiwald, A., Textoris-Taube, K., Vernardis, S.I., Egger, A.-S., Kreidl, M., et al. (2020). Ultra-high-throughput clinical proteomics reveals classifiers of COVID-19 infection. Cell Syst *11*, 11–24.e4.

Millard, S.P. (2014). EnvStats, an RPackage for environmental statistics. In Wiley StatsRef: Statistics Reference Online, N. Balakrishnan, T. Colton, B. Everitt, W. Piegorsch, F. Ruggeri, and J.L. Teugels, eds. (Wiley).

Muffat, J., and Walker, D.W. (2010). Apolipoprotein D: an overview of its role in aging and age-related diseases. Cell Cycle *9*, 269–273.

Murphy, A.J., Akhtari, M., Tolani, S., Pagler, T., Bijl, N., Kuo, C.L., Wang, M., Sanson, M., Abramowicz, S., Welch, C., et al. (2011). ApoE regulates hematopoietic stem cell proliferation, monocytosis, and monocyte accumulation in atherosclerotic lesions in mice. J. Clin. Invest. *121*, 4138–4149.

Nagendran, M., Chen, Y., Lovejoy, C.A., Gordon, A.C., Komorowski, M., Harvey, H., Topol, E.J., Ioannidis, J.P.A., Collins, G.S., and Maruthappu, M. (2020). Artificial intelligence versus clinicians: systematic review of design, reporting standards, and claims of deep learning studies. BMJ *368*, m689.

WHO. (2020). R&D Blueprint-Novel Coronavirus COVID-19 Therapeutic Trial Synopsis (World Health Organization).

Ombrellino, M., Wang, H., Yang, H., Zhang, M., Vishnubhakat, J., Frazier, A., Scher, L.A., Friedman, S.G., and Tracey, K.J. (2001). Fetuin, a negative acute phase protein, attenuates TNF synthesis and the innate inflammatory response to carrageenan. Shock *15*, 181–185.

Overmyer, K.A., Shishkova, E., Miller, I.J., Balnis, J., Bernstein, M.N., Peters-Clarke, T.M., Meyer, J.G., Quan, Q., Muehlbauer, L.K., Trujillo, E.A., et al. (2020). Large-scale multi-omic analysis of COVID-19 severity. medRxiv. https://doi.org/10.1101/2020.07.17.20156513.

Page, C., Goicochea, L., Matthews, K., Zhang, Y., Klover, P., Holtzman, M.J., Hennighausen, L., and Frieman, M. (2012). Induction of alternatively activated macrophages enhances pathogenesis during severe acute respiratory syndrome coronavirus infection. J. Virol. *86*, 13334–13349.

Pascual, M., Steiger, G., Estreicher, J., Macon, K., Volanakis, J.E., and Schifferli, J.A. (1988). Metabolism of complement factor D in renal failure. Kidney Int *34*, 529–536.

Patricio, P., Paiva, J.A., and Borrego, L.M. (2019). Immune response in bacterial and candida sepsis. Eur. J. Microbiol. Immunol. (Bp) *9*, 105–113.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. (2011). Scikit-learn: machine learning in Python. J. Mach. Learn. Res. *12*, 2825–2830.

Peralta, C.A., Shlipak, M., Judd, S., Cushman, M., McClellan, W., Zakai, N.A., Safford, M.M., Zhang, X., Muntner, P., and Warnock, D. (2011). Detection of chronic kidney disease with creatinine, cystatin C, and urine albumin-to-creatinine ratio and association with progression to end-stage renal disease and mortality. JAMA *305*, 1545–1552.

Perez-Riverol, Y., Csordas, A., Bai, J., Bernal-Llinares, M., Hewapathirana, S., Kundu, D.J., Inuganti, A., Griss, J., Mayer, G., Eisenacher, M., et al. (2019). The PRIDE database and related tools and resources in 2019: improving support for quantification data. Nucleic Acids Res *47*, D442–D450.

Peters, M.J., Joehanes, R., Pilling, L.C., Schurmann, C., Conneely, K.N., Powell, J., Reinmaa, E., Sutphin, G.L., Zhernakova, A., Schramm, K., et al. (2015). The transcriptional landscape of age in human peripheral blood. Nat. Commun. *6*, 8570.

Pham, T.V., Henneman, A.A., and Jimenez, C.R. (2020). iq: an R package to estimate relative protein abundances from ion quantification in DIA-MS-based proteomics. Bioinformatics *36*, 2611–2613.

Phua, J., Weng, L., Ling, L., Egi, M., Lim, C.M., Divatia, J.V., Shrestha, B.R., Arabi, Y.M., Ng, J., Gomersall, C.D., et al. (2020). Intensive care management of coronavirus disease 2019 (COVID-19): challenges and recommendations. Lancet Respir. Med. *8*, 506–517.

Poon, I.K.H., Patel, K.K., Davis, D.S., Parish, C.R., and Hulett, M.D. (2011). Histidine-rich glycoprotein: the Swiss Army knife of mammalian plasma. Blood *117*, 2093–2101.

Potempa, J., Fedak, D., Dubin, A., Mast, A., and Travis, J. (1991). Proteolytic inactivation of alpha-1-anti-chymotrypsin. Sites of cleavage and generation of chemotactic activity. J. Biol. Chem. *266*, 21482–21487.

Poynard, T., and Imbert-Bismut, F. (2012). Laboratory testing for liver disease. In Zakim and Boyer's Hepatology, Sixth Edition, T.D. Boyer, M.P. Manns, and A.J. Sanyal, eds. (W.B. Saunders), pp. 201–215.

Ramasamy, I. (2014). Recent advances in physiological lipoprotein metabolism. Clin. Chem. Lab. Med. *52*, 1695–1727.

Rea, I.M., Gibson, D.S., McGilligan, V., McNerlan, S.E., Alexander, H.D., and Ross, O.A. (2018). Age and age-related diseases: role of inflammation triggers and cytokines. Front. Immunol. *9*, 586.

RECOVERY Collaborative Group (2020). Dexamethasone in hospitalized patients with Covid-19 - preliminary report. N. Engl. J. Med. *384*, 693–704.

Rehman, A.A., Ahsan, H., and Khan, F.H. (2013). α-2-macroglobulin: a physiological guardian. J. Cell. Physiol. *228*, 1665–1675.

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res *43*, e47.

Rosenbaum, L. (2020). The untold toll - the pandemic's effects on patients without Covid-19. N. Engl. J. Med. *382*, 2368–2371.

Sack, G.H., Jr. (2018). Serum amyloid A - a review. Mol. Med. *24*, 46.

Saxena, S.K. (2020). Coronavirus Disease 2019 (COVID-19): Epidemiology, Pathogenesis, Diagnosis, And Therapeutics (Springer nature).

Schneeman, T.A., Bruno, M.E.C., Schjerven, H., Johansen, F.E., Chady, L., and Kaetzel, C.S. (2005). Regulation of the polymeric Ig receptor by signaling through TLRs 3 and 4: linking innate and adaptive immune responses. J. Immunol *175*, 376–384.

Schulte-Schrepping, J., Reusch, N., Paclik, D., Baßler, K., Schlickeiser, S., Zhang, B., Krämer, B., Krammer, T., Brumhard, S., Bonaguro, L., et al. (2020). Severe COVID-19 is marked by a dysregulated myeloid cell compartment. Cell *182*, 1419–1440.e23.

Shah, P., Kendall, F., Khozin, S., Goosen, R., Hu, J., Laramie, J., Ringel, M., and Schork, N. (2019). Artificial intelligence and machine learning in clinical development: a translational perspective. NPJ Digit. Med. *2*, 69.

Shao, B., de Boer, I., Tang, C., Mayer, P.S., Zelnick, L., Afkarian, M., Heinecke, J.W., and Himmelfarb, J. (2015). A cluster of proteins implicated in kidney disease is increased in high-density lipoprotein isolated from hemodialysis subjects. J. Proteome Res. *14*, 2792–2806.

Sharma, N.K., Tashima, A.K., Brunialti, M.K.C., Ferreira, E.R., Torquato, R.J.S., Mortara, R.A., Machado, F.R., Assuncao, M., Rigato, O., and Salomao, R. (2017). Proteomic study revealed cellular assembly and lipid metabolism dysregulation in sepsis secondary to community-acquired pneumonia. Sci. Rep. *7*, 15606.

Shen, B., Yi, X., Sun, Y., Bi, X., Du, J., Zhang, C., Quan, S., Zhang, F., Sun, R., Qian, L., et al. (2020). Proteomic and metabolomic characterization of COVID-19 patient sera. Cell *182*, 59–72.e15.

Shu, T., Ning, W., Wu, D., Xu, J., Han, Q., Huang, M., Zou, X., Yang, Q., Yuan, Y., Bie, Y., et al. (2020). Plasma proteomics identify biomarkers and pathogenesis of COVID-19. Immunity *53*, 1108–1122.e5.

Silhol, F., Sarlon, G., Deharo, J.C., and Vaïsse, B. (2020). Downregulation of ACE2 induces overstimulation of the renin-angiotensin system in COVID-19: should we block the renin-angiotensin system? Hypertens. Res. *43*, 854–856.

Singer, M., Deutschman, C.S., Seymour, C.W., Shankar-Hari, M., Annane, D., Bauer, M., Bellomo, R., Bernard, G.R., Chiche, J.D., Coopersmith, C.M., et al. (2016). The third international consensus definitions for sepsis and septic shock (Sepsis-3). JAMA *315*, 801–810.

Smith, B.H., Campbell, H., Blackwood, D., Connell, J., Connor, M., Deary, I.J., Dominiczak, A.F., Fitzpatrick, B., Ford, I., Jackson, C., et al. (2006). Generation Scotland: the Scottish Family Health Study; a new resource for researching genes and heritability. BMC Med. Genet. *7*, 74.

Smyth, G.K. (2004). Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Stat. Appl. Genet. Mol. Biol. *3*, article3.

Soysal, P., Stubbs, B., Lucato, P., Luchini, C., Solmi, M., Peluso, R., Sergi, G., Isik, A.T., Manzato, E., Maggi, S., et al. (2016). Inflammation and frailty in the elderly: a systematic review and meta-analysis. Ageing Res. Rev. *31*, 1–8.

Stawicki, S.P., Jeanmonod, R., Miller, A.C., Paladino, L., Gaieski, D.F., Yaffee, A.Q., De Wulf, A., Grover, J., Papadimos, T.J., Bloem, C., et al. (2020). The 2019–2020 novel coronavirus (severe acute respiratory syndrome coronavirus 2) pandemic: a joint American College of Academic International Medicine-world academic council of emergency medicine multidisciplinary COVID-19 working group consensus paper. J. Glob. Infect. Dis. *12*, 47–93.

Stone, J.H., Frigault, M.J., Serling-Boyd, N.J., Fernandes, A.D., Harvey, L., Foulkes, A.S., Horick, N.K., Healy, B.C., Shah, R., Bensaci, A.M., et al. (2020). Efficacy of tocilizumab in patients hospitalized with Covid-19. N. Engl. J. Med. *383*, 2333–2344.

Sun, Q., Qiu, H., Huang, M., and Yang, Y. (2020). Lower mortality of COVID-19 by early recognition and intervention: experience from Jiangsu Province. Ann. Intensive Care *10*, 33.

Tavazoie, M.F., Pollack, I., Tanqueco, R., Ostendorf, B.N., Reis, B.S., Gonsalves, F.C., Kurth, I., Andreu-Agullo, C., Derbyshire, M.L., Posada, J., et al. (2018). LXR/ApoE activation restricts innate immune suppression in cancer. Cell *172*, 825–840.e18.

Tay, M.Z., Poh, C.M., Rénia, L., MacAry, P.A., and Ng, L.F.P. (2020). The trinity of COVID-19: immunity, inflammation and intervention. Nat. Rev. Immunol. *20*, 363–374.

Turner, A.J. (2015). ACE2 cell biology, regulation, and physiological functions. In The Protective Arm of the Renin Angiotensin System (RAS), T. Unger, U.M. Steckelings, and R.A.S. dos Santos, eds. (Academic Press), pp. 185–189.

Turula, H., and Wobus, C.E. (2018). The role of the polymeric immunoglobulin receptor and secretory immunoglobulins during mucosal infection and immunity. Viruses *10*, 237.

Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., et al. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat. Methods *17*, 261–272.

Volanakis, J.E., and Narayana, S.V. (1996). Complement factor D, a novel serine protease. Protein Sci *5*, 553–564.

Vollmer, S., Mateen, B.A., Bohner, G., Király, F.J., Ghani, R., Jonsson, P., Cumbers, S., Jonas, A., McAllister, K.S.L., Myles, P., et al. (2020). Machine learning and artificial intelligence research for patient benefit: 20 critical questions on transparency, replicability, ethics, and effectiveness. BMJ *368*, l6927.

Wakabayashi, S. (2013). New insights into the functions of histidine-rich glycoprotein. In International Review of Cell and Molecular Biology, K.W. Jeon, ed. (Academic Press), pp. 467–493.

Wermuth, P.J., and Jimenez, S.A. (2015). The significance of macrophage polarization subtypes for animal models of tissue fibrosis and human fibrotic diseases. Clin. Transl. Med. *4*, 2.

Wu, D., Koganti, R., Lambe, U.P., Yadavalli, T., Nandi, S.S., and Shukla, D. (2020). Vaccines and therapies in development for SARS-CoV-2 infections. J. Clin. Med. *9*, 1885.

Wu, Y., Potempa, L.A., El Kebir, D., and Filep, J.G. (2015). C-reactive protein and inflammation: conformational changes affect function. Biol. Chem. *396*, 1181–1197.

Wynants, L., Van Calster, B., Collins, G.S., Riley, R.D., Heinze, G., Schuit, E., Bonten, M.M.J., Dahly, D.L., Damen, J.A.A., Debray, T.P.A., et al. (2020). Prediction models for diagnosis and prognosis of covid-19 infection: systematic review and critical appraisal. BMJ *369*, m1328.

Yang, L., Liu, S., Liu, J., Zhang, Z., Wan, X., Huang, B., Chen, Y., and Zhang, Y. (2020). COVID-19: immunopathogenesis and immunotherapeutics. Signal Transduct. Target. Ther. *5*, 128.

Zhang, H., Penninger, J.M., Li, Y., Zhong, N., and Slutsky, A.S. (2020). Angiotensin-converting enzyme 2 (ACE2) as a SARS-CoV-2 receptor: molecular mechanisms and potential therapeutic target. Intensive Care Med *46*, 586–590.

Zhao, B., Ni, C., Gao, R., Wang, Y., Yang, L., Wei, J., Lv, T., Liang, J., Zhang, Q., Xu, W., et al. (2020). Recapitulation of SARS-CoV-2 infection and cholangiocyte damage with human liver ductal organoids. Protein Cell *11*, 771–775.

Zhuo, L., Itano, N., Nonogaki, T., Shen, L., Wu, J., Watanabe, H., and Kimata, K. (2004). Chapter 9 - Biological Function of SHAP–Hyaluronan Covalent Complex. Chemistry and Biology of Hyaluronan, 205–222. https://doi.org/10.1016/B978-008044382-9/50040-6.

Cell Systems
Article

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
Human Serum	Sigma-Aldrich	Cat# S7023-50MB
Human Plasma (EDTA, Pooled Donor)	Genetex	Cat# GTX73265
Chemicals, peptides, and recombinant proteins		
Water for chromatography (LC-MS Grade) LiChrosolv®	Merck	Cat# 115333
Acetonitrile (Acetonitrile, Optima™ LC/MS Grade, Fisher Chemical™)	Fisher Scientific	Cat# A955-212
Methanol (Optima LC-MS Grade, Fisher Chemical)	Fisher Scientific	Cat# A456-212
DL-Dithiothreitol (BioUltra)	Sigma-Aldrich	Cat# 43815
Iodoacetamide (BioUltra)	Sigma-Aldrich	Cat# I1149
Ammonium Bicarbonate (Eluent additive for LC-MS)	Sigma-Aldrich	Cat# 40867
Urea (puriss. P.a., reag. Ph. Eur.)	Honeywell Research Chemicals	Cat# 33247H
Formic Acid, LC-MS Grade (Eluent additive for LC-MS)	Thermo Scientific™ Pierce™	Cat# 85178
Trypsin (Sequence grade)	Promega	Cat# V511X
Mass Spec-Compatible Human Extract	Promega	Cat# V6951
Retention time peptides Biognosys iRT kit	Biognosys	Cat# Ki-30002-b
MS synthetic peptide calibration kit	SCIEX	Cat# 5045759
Deposited Data		
Raw mass spectrometry proteomics data (commercial plasma and serum control samples)	This study	PXD025752
Software and algorithms		
Proteomics data analysis via Deep Neural Networks, DIA-NN	Demichev et al., 2020	https://github.com/vdemichev/DiaNN
DIA-NN R package	Demichev et al., 2020	https://github.com/vdemichev/diann-rpackage
ComplexHeatmap R package	(Gu et al., 2016)	https://github.com/jokergoo/ComplexHeatmap
EnvStats R package	(Millard, 2014)	https://CRAN.R-project.org/package=EnvStats
Limma R package	(Ritchie et al., 2015)	https://bioconductor.org/packages/limma/
eBayes R package	(Smyth, 2004)	https://github.com/cran/limma/blob/master/R/ebayes.R
XGBoost 1.2.0 Python package	(Chen and Guestrin, 2016)	https://pypi.org/project/xgboost/1.2.0/
scikit-learn 0.23.2 Python package	(Pedregosa et al., 2011)	https://scikit-learn.org/0.23/
scipy 1.5.2 Python package	(Virtanen et al., 2020)	https://pypi.org/project/scipy/1.5.2/
Other		
Zorbax RRHD Eclipse Plus 95A C18, 2.1 x 50mm, 1.8 um, 1200 bar	Agilent	Cat# 959757-902
Infinitylab Poroshell 120 EC-C18, 2.1x50mm 1.9um	Agilent	Cat# 699675-902
BioPureSPE Macro 96-Well, 100mg PROTO 300 C18	The Nest Group, Inc.	HNS S18V-L

CellPress
OPEN ACCESS

Cell Systems
Article

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Markus Ralser (markus.ralser@charite.de).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- The processed proteomic and clinical source data is available in this paper's supplemental information.
- The raw mass spectrometry proteomics source data for the quality control plasma and serum acquisitions has been deposited to the ProteomeXchange Consortium via the PRIDE partner repository (Perez-Riverol et al., 2019) with the dataset identifier PXD025752.
- This paper does not report original code.
- The machine learning scripts used to generate the figures reported in this paper are available in this paper's supplemental information.
- Any additional information required to reproduce this work is available from the Lead Contact.

Experimental model and subject details

Charité patient cohort and clinical data

Patients were recruited within the Pa-COVID-19 study conducted at Charité - Universitätsmedizin Berlin, a prospective observational cohort study on the pathophysiology of COVID-19. The study protocol has been described in detail before (Kurth et al., 2020). All patients with PCR-confirmed SARS-CoV-2 infection were eligible for inclusion. Refusal to provide informed consent by the patient or a legal representative and any condition prohibiting supplemental blood collection for serial biosampling were exclusion criteria. Patients were treated according to current national and international guidelines. Three patients had *Do Not Intubate and Do Not Resuscitate* (DNI/DNR) orders in place, declining mechanical ventilation and other organ support or cardiopulmonary resuscitation. In 4 further cases, limitation of therapy was decided at a later time point according to the patient's presumed wish ("secondary DNR") and predictably unfavorable outcome. All other patients received maximum intensive care treatment including organ replacement therapies at the discretion of the responsible physicians.

Biosampling for proteome measurement was performed 3 times per week after inclusion. The WHO ordinal scale for clinical improvement (Table S1) was used to assess disease severity. ARDS was defined according to the Berlin ARDS criteria (ARDS Definition Task Force et al., 2012). Sepsis was defined according to sepsis-3 criteria (Singer et al., 2016). The study was approved by the ethics committee of Charité - Universitätsmedizin Berlin (EA2/066/20) and conducted in accordance with the Declaration of Helsinki and guidelines of Good Clinical Practice (ICH 1996). The study is registered in the German and the WHO international registry for clinical studies (DRKS00021688). Clinical data was captured in a purpose built electronic case report form data using the capture system SecuTrial®. All routine laboratory parameters were analyzed in accredited laboratories at Charité - Universitätsmedizin Berlin. Pseudonymized data exported from SecuTrial® were processed using JMP Pro 14 (SAS Institute Inc., Cary, NC, USA). If a laboratory value was missing for a given day, values from up to two preceding days were used for the analysis.

Innsbruck Patient cohort and clinical data

Serum samples from 99 patients admitted to the intensive care unit at the Department of Medicine, University Hospital of Innsbruck for the treatment of respiratory failure due to severe COVID-19 were collected within the first days (median 7.5, IQR 5-12) after admission. Written informed consent was either obtained before sampling or retrospectively after recovery, if patients were mechanically ventilated at the time of sampling. COVID-19 was diagnosed on the basis of a (i) positive SARS-CoV2 PCR within the last 7 days prior to study inclusion, (ii) respiratory failure defined as a partial pressure of oxygen < 60 mmHg on arterial blood gas analysis or a peripheral oxygen saturation of < 90% and (iii) typical infiltrates on computed tomography scanning of the chest. Patients were treated according to national guidelines. The study was approved by the local ethics research committee EK-Nr. 1107/2020, and EK-Nr. 1103/2020 for follow-up.

METHOD DETAILS

Materials

Water for chromatography (LC-MS Grade, LiChrosolv®, Merck; 115333), Acetonitrile (LC-MS Grade Optima; A955-212) and Methanol (LC-MS Grade, Optima; A456-212) were purchased from Fisher Chemicals. DL-Dithiothreitol (BioUltra, 43815), Iodoacetamide (BioUltra, I1149) and Ammonium Bicarbonate (Eluent additive for LC-MS, 40867) were purchased from Sigma Aldrich. Urea (puriss. P.a., reag. Ph. Eur., 33247H) and Formic Acid (Eluent additive for LC-MS, 85178) were purchased from Thermo Scientific. Trypsin

Cell Systems
Article

CellPress
OPEN ACCESS

(Sequence grade, V511X) was purchased from Promega. Control samples were prepared from Human Serum (Sigma Aldrich, S7023-50MB) and Human Plasma (EDTA, Pooled Donor, Genetex GTX73265).

Mass spectrometry

Mass spectrometry-based proteomics analysis was performed as described previously (Messner et al., 2020) with minor adjustments to the workflow (Figure S1). Semi-automated sample preparation was performed in 96-well format, using in advance prepared stock solution plates stored at -80°C. Briefly, 5µl of thawed plasma samples were transferred to the pre-made denaturation/reduction stock solution plates (55µl 8M Urea, 100mM ammonium bicarbonate (ABC), 50mM dithiothreitol) resuspended and incubated at 30°C for 60 minutes. 5µl was then transferred from the iodoacetamide stock solution plate (100mM) to the sample plate and incubated in the dark at 23°C for 30 minutes before dilution with 100mM ABC buffer (340µl). 220µl of this solution was transferred to the pre-made trypsin stock solution plate (12.5µl, 0.1µg/µl) and incubated at 37°C for 17 h (Benchmark Scientific Incu-Mixer MP4). The digestion was quenched by addition of formic acid (10% v/v, 25µl). The digestion mixture was cleaned-up using C18 96-well plates (BioPureSPE Macro 96-Well, 100mg PROTO C18, The Nest Group) and redissolved in 60µl 0.1% formic acid with shaking. Insoluble particles were removed by centrifugation and the samples transferred to a new plate.

Each 96-well plate contained 8 plasma and 4 serum sample preparation controls, and the acquisition workflow included a pooled quality control sample every ~10 injections. Liquid chromatography was performed using the Agilent 1290 Infinity II system coupled to a TripleTOF 6600 mass spectrometer (SCIEX) equipped with IonDrive Turbo V Source (Sciex). A total of 5µl was injected, and the peptides were separated in reversed phase mode using a C18 ZORBAX Rapid Resolution High Definition (RRHD) column 2.1mm x 50mm, 1.8µm particles or Infinitylab Poroshell 120 EC-C18, 2.1 x 50mm 1.9 µm particles. A linear gradient was applied which ramps from 1% B to 40% B in 5 minutes (Buffer A: 0.1% FA; Buffer B: ACN/0.1% FA) with a flow rate of 800µl/min. For washing the column, the organic solvent was increased to 80% B in 0.5 minutes and was kept for 0.2 minutes at this composition before going back to 1% B in 0.1 min. The mass spectrometer was operated in the high sensitivity mode. The DIA/SWATH method consisted of an MS1 scan from m/z 100 to m/z 1500 (20 ms accumulation time) and 25 MS2 scans (25ms accumulation time) with variable precursor isolation width covering the mass range from m/z 450 to m/z 850 (Messner et al., 2020). An IonDrive Turbo V Source (Sciex) was used with ion source gas 1 (nebulizer gas), ion source gas 2 (heater gas) and curtain gas set to 50, 40 and 25, respectively. The source temperature was set to 450 and the ion spray voltage to 5500V. System suitability was evaluated using synthetic peptides (Sciex 5045759, Biognosys Ki-30002-b) and human protein extracts (Promega V6951).

QUANTIFICATION AND STATISTICAL ANALYSIS

Data analysis

The data were processed with DIA-NN (Demichev et al., 2020), an open-source software suite for DIA / SWATH data processing (https://github.com/vdemichev/DiaNN, commit 4498bd7) using a two-step spectral library refinement procedure as described previously (Messner et al., 2020), with filtering at precursor level q-value (1%), library q-value (0.5%) and gene group q-value (1%). Highly hydrophobic peptides (reference retention time > 110 on the iRT scale) were discarded. Batch correction was performed at the precursor level as described previously (Messner et al., 2020), using linear regression for intra-batch correction (for each MS batch) and control samples for inter-plate correction. Protein quantification was subsequently carried out using the MaxLFQ algorithm (Cox et al., 2014; Pham et al., 2020) as implemented in the DIA-NN R package (https://github.com/vdemichev/diann-rpackage). One of the 96-well plates (#12) featured technical replicates of a number of samples that were also analysed on other plates: in an extra batch correction step, the median $\log_2$-protein levels across these replicates on plate 12 were matched to the respective median $\log_2$-levels (across the same biological samples) throughout other plates, to correct protein levels on plate 12. Further batch correction was performed for Innsbruck data, to match the mean $\log_2$-transformed protein levels in the respective control samples to $\log_2$-transformed protein levels in control samples acquired for the Charité cohort. The Generation Scotland cohort proteomics raw data, which we described previously (Messner et al., 2020), have been reanalyzed using the updated software pipeline, to ensure comparability. Exclusion of precursors or proteins based on the data completeness was not performed.

Statistical testing was performed in the R environment for statistical computing, version 3.6.0 (R core team, www.R-project.org). All protein and clinical laboratory measurements (except for standard and actual base excess, oxyhemoglobin and sO2) were first $\log_2$-transformed, to ensure optimal performance of linear models assuming Gaussian errors, as well as to reduce the impact of outliers. Imputation of the data was not performed, as all the statistical tests applied can accommodate missing values. Likewise, no data filtering based on missing value rates was applied. For differential abundance testing, only protein groups matched to at least three different unmodified peptide sequences were considered. Significance testing for a zero median (for analysing trajectories) or against binary variables (worsening, death) was performed using the Wilcoxon W test or Mann-Whitney U test, respectively, as implemented in the "wilcox.test" function of the "stats" R package. Testing against a continuous variable (e.g. when determining significance of pairwise correlations) was performed using the Kendall Tau test, with the slope estimated using the Theil-Sen method, as implemented in the "kendallTrendTest" function of the "EnvStats" (Millard, 2014) package. When covariates had to be taken into account, we used linear modelling with the "limma" (Ritchie et al., 2015) R package, with P-values obtained using "eBayes" (Smyth, 2004). Modelling with "limma" was likewise used to correct for these covariates for visualisation purposes. WHO grade was considered as a "factor-type" covariate (resulting in a "limma" design matrix with one-hot encoding for different WHO grades). Multiple-testing correction was performed using the Benjamini-Hochberg false discovery rate controlling procedure (Benjamini and Hochberg, 1995)

CellPress OPEN ACCESS

Cell Systems
Article

as implemented in the "p.adjust" function of the "stats" R package. The adjusted p-values below 0.05 were considered significant. Multiple-testing correction for differential abundance analysis was performed separately for proteins, for which MRMAssayDB lists a targeted assay (Bhowmick et al., 2018), the rest of proteins measured, the clinical laboratory measurements and the clinical factors (age, Charlson score, BMI, Horowitz index and FiO_2, SOFA score), to ensure that the false discovery rate stayed below 0.05 for each of these categories of features. Likewise, when determining the significance of correlations in correlation matrices, correction was performed for each row or each column separately, to ensure less than 5% false discoveries in each row or column, respectively. For correlation map visualisations, black points were used to indicate row-wise significant correlations, and black rectangles at the border of the respective cell - column-wise significant correlations.

Quantities of gene products corresponding to open reading frames named IGxx (i.e. different types of immunoglobulin chains) were summed together to generate quantities representative of the overall levels of immunoglobulin classes (IGHVs, IGLVs, etc). This does not affect any conclusions of this work and was done purely to improve visualization and simplify the interpretation of the heatmaps and correlation maps. Full protein level tables, including levels of individual immunoglobulin gene products, are provided in supplementary materials. For visualisation, different WHO grades were color-coded throughout the manuscript (see Figure S2).

Markers of the disease severity

The first time point measured at the maximum WHO grade was chosen for each patient. For each omics feature, its values (log_2-transformed when necessary, as described above) were tested for a trend depending on the WHO grade. Age was included as a covariate in the linear model as described above.

Markers varying with age in COVID-19

The first sampling time point measured was chosen per patient. For each omics feature, its values (log_2-transformed when necessary, as described above) were tested for a trend depending on age. The test was performed either using the Kendall Tau test (as described above; Figures S9 and S10), or by accounting for WHO grade as a covariate in the linear model (as described above; Figures S11 and 5).

Markers of RRT and ECMO

For each omics feature, the P-value was calculated using the Mann-Whitney test, comparing between the median levels (log_2-transformed when necessary, as described above) across all sampling time points at WHO grade 7 in patients who did not receive the therapy and the median levels (log_2-transformed when necessary, as described above) across all sampling time points at WHO grade 7 after initiation of the respective therapy in patients who did.

Markers predictive of time in hospital

Patients, for which the first sampling time point before the outcome corresponded to the WHO = 3 severity grade (that is the patient did not require supplemental oxygen on that day), were considered. Thus, no correction for disease severity was necessary. Testing of levels (log_2-transformed when necessary, as described above) of each omics feature (measured for the first sampling time point) vs the remaining time in hospital (days) was performed by including age as a covariate in the linear model as described above.

Markers predictive of disease worsening

The first sampling time point measured was chosen per patient. Future disease worsening was defined as a future increase in the WHO grade (for patients at WHO grade < 7) or death (for patients at WHO grade 7). For each omics feature, its levels (log_2-transformed when necessary, as described above) were compared between patients who did not worsen and patients who did, with age and current WHO grade (as factor) included as covariates in the linear model as described above.

Peak period of the disease definition

When studying the dynamic changes in omics values during the disease course, we focused on the time points sampled when the disease was the most severe for a particular patient. This allowed us to look at molecular changes over time without the need to take into account the potential impact of changes in disease severity and the level of treatment. For each patient, we thus defined the "peak period of the disease" as the time when the patient was receiving the most intensive treatment during their stay in hospital, that is the time when the patient was at WHO grade 6 or 7, for patients who received invasive mechanical ventilation at some point, or otherwise at their maximum WHO grade (3, 4 or 5).

Markers changing during the peak of disease

Only patients with at least two days between the first and last sampling time points at the peak of the disease (as defined above) were considered. For each omics feature, a linear regression model was fitted for its levels (log_2-transformed when necessary, as described above) vs the day number (with the slope estimated using the nonparametric Theil-Sen method, as implemented in the "kendallTrendTest" function of the "EnvStats" (Millard, 2014) R package), and the quantity $slope_{adj}$ = (regression slope) * (number of days between first and last time points) was calculated. A non-parametric approach was chosen because of its superior robustness to outliers. A Wilcoxon W test was then applied to compare the median of $slope_{adj}$ to zero. The values of $slope_{adj}$ for each feature are visualised in Figure S14. The non-parametric approach was chosen here due to its robustness with respect to outliers.

Correlation maps

General correlation maps were generated using the values (log_2-transformed when necessary, as described above) of features at the first time point measured at the maximum WHO grade for each patient. The correlation map between feature changes during the peak of the disease (as defined above) was generated by correlating the $slope_{adj}$ values (as defined above). The map of significant protein correlations not detected in the general population was generated by excluding all correlations which were either significant (P <= 0.05, without multiple-testing correction) with the same trend in the Generation Scotland cohort, or could not be calculated reliably therein (less than 20 valid points).

Prediction of current mechanical ventilation

To reflect the power of omics measurements in characterising the phenotype, a classifier was constructed to predict mechanical ventilation (WHO grade > 5) at the present time point using the proteomic and/or accredited diagnostic data. For the proteomic data only proteins characterized by at least 3 peptides were taken into account. The first time point measured at the maximum WHO grade was selected per patient. We used a gradient boosted tree algorithm implemented in the XGBoost 1.2.0 (Chen and Guestrin, 2016) under Python 3.8.1. The classifier was constructed using leave-one-out cross-validation. To circumvent overfitting a subsampling of 0.5 of the training data per boosting step and an L2 regularization term "lambda" of 20 were applied.

For the assessment of classifier performance, the leave-one-out method was applied in the following way: the prediction was made for each sample separately, by excluding (withholding) this sample from the dataset, training the classifier on the remaining (independent) samples and then predicting the withheld sample using the trained model. The source code is provided in supplementary materials. For the determination of the feature importances, one classifier was trained on all data points using the same setup as described above. The feature importances were then extracted directly from the trained classifier.

For the validation of the trained models, samples from an independent cohort (Innsbruck) were used. A model was trained on the data collected at the Charité using the same setup and parameters as described above and the proteins that were characterized in both cohorts. The evaluation was performed on the Innsbruck cohort that was not used for training. ROC-curves and AUC were calculated using scikit-learn 0.23.2 (Pedregosa et al., 2011). The machine learning scripts are provided in Data S1.

WHO grade prediction

For the prediction of the WHO grade an elastic net was applied as implemented in scikit-learn 0.23.2. The WHO grade was predicted for the first time point at maximum WHO grade per patient using a leave-one-out cross-validation procedure. A training/prediction based on proteomic (proteins with at least 3 peptides) and/or accredited diagnostic data from the Charité cohort was performed. Additionally, the proteomic model was validated using proteomic data set from the Innsbruck cohort that was not included in the training. Features with more than 10% missing values were removed. All data were log_2-transformed when necessary (as described above), standardized and kNN-imputed (5 neighbors). The latter two steps were fitted on the training data only. For the elastic net an "l1_ratio" of 0.05 was used coupled to a 5-fold cross-validated recursive feature-elimination algorithm ("step" = 10, "min_features" = 20). Calculations of metrics were performed using scikit-learn 0.23.2 and scipy 1.5.2 (Virtanen et al., 2020). The machine learning scripts are provided in Data S1.

Prediction of the remaining time in hospital

For the prediction of the remaining time in hospital a WHO grade predictor as described above was trained on the first data points for every patient. The predicted WHO grades for every patient at WHO grade 3 who stayed in hospital for at least 1 day after sample time were correlated to the remaining time in hospital. The Spearman correlation was calculated using scipy 1.5.2. The machine learning scripts are provided in Data S1.

Supplementary Note 1. Diagnostic parameters and Proteome signatures that indicate therapeutic interventions

We investigated to what extent specific organ replacement therapies in severely ill patients, (renal replacement therapy (RRT) and extracorporeal membrane oxygenation (ECMO)) were reflected in the proteome and at the level of accredited diagnostic parameters. HP and HPX were reduced in patients on RRT and ECMO, reflecting hemolysis in the extracorporeal circuits (Figures S7 and S8). Elevated SERPINC1 (Antithrombin III) levels mirror substitution of antithrombin during ECMO. The reason for elevated levels of APOE in patients with ECMO is unclear, but is in line with reports on increased levels of APOE in pediatric patients after cardiopulmonary bypass (Ağirbaşli et al., 2015). The proteins increased in patients receiving RRT mainly reflect impaired kidney function and have been associated with RRT before (AMBP, B2M, CST3, LYZ, RBP4, Figure S7) (Shao et al., 2015). Of note, increased levels of AMBP, B2M and LYZ have been associated with death in chronic kidney disease (Makridakis et al., 2020). Levels of CFD and APOH, both involved in the complement system, were also increased (McDonnell et al., 2020; Volanakis and Narayana, 1996). CFD is eliminated renally and accumulates in end stage renal disease, possibly leading to enhanced complement activation via the alternative pathway (Pascual et al., 1988). In contrast, levels of APOH have even been reported to be slightly lower following high-flux hemodialysis (Han et al., 2018).

We note that the analysis of the effect of treatments on the proteome has two limitations. First, some of the markers identified might be prognostic for the treatment rather than reflect its effect. Age and the Charlson comorbidity index belong to this category: patients receiving ECMO were significantly younger and had a lower number of pre-existing chronic conditions than those who did not.

CellPress
OPEN ACCESS

Cell Systems
Article

Second, the results might be partially confounded by the time elapsed from the onset of the disease, as we have shown (Figure 3) that omics signature changes with time in COVID-19 patients while on invasive mechanical ventilation.

Supplementary Note 2. Age-specific response to COVID-19 in the context of severity markers

Older age is one of the most significant risk factors for severe disease and adverse outcome in COVID-19. Enhanced understanding of underlying mechanisms for the age-specific response to SARS-CoV-2 infection is therefore important and needed for the development of effective age-specific strategies for prevention and treatment. Furthermore, dissecting the age-specific components of the host response will improve our knowledge of the pathogenicity of similar viruses, making the world better prepared for future pandemics. Current theories characterizing the link between the higher age and risk for severe disease include immunosenescence, elevated baseline inflammation, or altered protein glycosylation landscape leading to impaired antiviral response or reduced immune tolerance (Franceschi et al., 2018; Goronzy and Weyand, 2013; Rea et al., 2018; Tay et al., 2020). However, a detailed and mechanistic understanding of the relation between COVID-19 and aging is lacking. In this work, we leverage the large size and high precision of the proteomic data acquired to map the age-related response to COVID-19, to provide a reference dataset (Figures 2C, 5, and S11) for future studies addressing this problem.

We report elevation of several inflammatory and acute phase proteins such as SERPINA3, ITIH4, SAA1, and ITIH3 in older patients with COVID-19. SAA1 has been shown to induce macrophage polarization to the M2-type which promotes tissue repair but also possesses pro-fibrotic properties involved in the pathogenesis of pulmonary fibrosis (Liu et al., 2014; Page et al., 2012; Wermuth and Jimenez, 2015). Moreover, SAA1 mediates displacement of APOA1 from HDL leading to loss of the cardio- and vasoprotective properties of high density lipoprotein (HDL) (Gordon, 2014). SERPINA3, as discussed above, has an ambivalent role as a neutrophil proteinase inhibitor but also a powerful neutrophil chemoattractant. Upregulation of SERPINA3 with age in COVID-19, along with the higher neutrophil-to-lymphocyte ratio, suggests that excessive neutrophil response is one of the aggravating factors in older COVID-19 patients. Taken together, our findings point toward a disproportionately dysregulated inflammatory response to SARS-CoV-2 with age, which may be explained by an increased baseline inflammation and immunosenescence in older patients (Chung et al., 2019; Ferrucci et al., 2005; Soysal et al., 2016). Age-dependent increase of FBLN1 and decrease of KLKB1 reflect alterations in blood coagulation which may aggravate this effect by predisposing older patients to thromboembolic events, one of the key clinical characteristics of severe COVID-19.

Interestingly, a number of apolipoproteins displayed a strong age-specific signature in COVID-19. For instance, APOC2, a component of chylomicrons, very low density lipoprotein (VLDL) and high density lipoprotein (HDL), and activator of lipoprotein lipase involved in triglyceride metabolism (Ramasamy, 2014), was downregulated with age in COVID-19, but upregulated with age in the general population (Harris et al., 2017; Peters et al., 2015) (Figure 2C). Dysregulation of apolipoproteins has been observed in community acquired pneumonia and associated with unfavourable outcome (Sharma et al., 2017). Remarkably, contrary to the general trend, APOD, APOC3 and APOE show opposite trends in older COVID-19 patients and in severe disease (Figure 5). APOD is expressed by many tissues, including the brain (Dassati et al., 2014). An increase in APOD has been previously observed in ischemic stroke and CNS inflammation and may reflect (subclinical) involvement of the central nervous system especially in older patients with more severe inflammation and more comorbidities (Muffat and Walker, 2010). Conversely, high levels of APOD have been shown to temper coronavirus-mediated encephalitis in mice, indicating its role as a marker of CNS damage as well as tissue protection and repair (Carmo et al., 2008). APOE, involved in inflammation, immune response and lipid metabolism, is upregulated in severe COVID-19 but downregulated with age in this cohort. APOE typically mediates anti-inflammatory effects by downregulation of NFκB and inhibition of macrophage response to IFNγ and TLR3, both mediators of viral immune response. Moreover, it neutralizes bacterial LPS and enhances the adaptive immune response by facilitating antigen presentation (Figueroa et al., 2019). Downregulation with age may reflect a compromised immune response leading to over-activation of NFκB and insufficient pathogen clearance in older patients. Finally, APOE has been described to reduce proliferation of myeloid progenitor cells (Murphy et al., 2011) and to reduce myeloid derived suppressor cell (MDSC) survival in mice (Tavazoie et al., 2018). Thus, lower levels of APOE in the elderly may favor expansion of immature and dysfunctional neutrophils that have been described as a hallmark in severe COVID-19 (Schulte-Schrepping et al., 2020). This broad involvement of APOE merits further investigation in future studies.

Supplementary Note 3. Diverging trends at the proteome level during the disease peak in individual patients

Some patients (59, 90, 96, 123) who died exhibited protein concentration trajectories distinctly similar to "typical" survivors (Figure 3B). Two of them (59, 90) had a prolonged ICU stay with repeated septic episodes and finally defined limitations of therapy according to presumed patients' wishes ("secondary DNR"). Their protein signatures probably reflect the phenomenon of immune paralysis that can follow bacterial sepsis associated with a prolonged ICU treatment (Patricio et al., 2019). One patient (96) was receiving ongoing immunosuppressive therapy for an autoimmune disorder, and a fourth patient (123) had a history of kidney transplantation, both died of septic shock. Whether the particular group of solid organ recipients shows a distinct protein signature associated with the outcome requires further investigation.

Cell Systems
Article

We also note that some surviving patients do not show a trajectory characteristic of the typical 'alleviation' of the proteomic phenotype (WHO = 4: 58, 106, 153; WHO = 6 or 7: 43, 80). Specifically, the proteomic response in patients 106, 153 and 141 was indicative of overall 'worsening' of the proteome (Figure 3B). In contrast, patients 43 and 80 exhibited the overall 'alleviation' of the proteome, except for the spike in the levels of CRP and serum amyloid (Figure 3B). Shorter time spans between sampling days may explain these observations in four of these patients (43, 58, 80, 106), indicating that the host inflammatory response requires a certain time to resolve, especially in more severely ill patients, and some of the markers of systemic inflammation might linger, whereas a typical alleviation of the proteomic signature can be observed even within a few days in moderate disease courses. The unusual pattern of patient 153 was likely confounded by a skin infection that subsequently required antibiotic treatment.

PLOS DIGITAL HEALTH

RESEARCH ARTICLE

A proteomic survival predictor for COVID-19 patients in intensive care

Vadim Demichev[1,2,3☯], Pinkus Tober-Lau[4☯], Tatiana Nazarenko[5,6], Oliver Lemke[1], Simran Kaur Aulakh[2], Harry J. Whitwell[7,8,9], Annika Röhl[1], Anja Freiwald[1], Mirja Mittermaier[4,10], Lukasz Szyrwiel[2], Daniela Ludwig[1], Clara Correia-Melo[2], Lena J. Lippert[4], Elisa T. Helbig[4], Paula Stubbemann[4], Nadine Olk[4], Charlotte Thibeault[4], Nana-Maria Grüning[1], Oleg Blyuss[11,12,13], Spyros Vernardis[2], Matthew White[2], Christoph B. Messner[1,2], Michael Joannidis[14], Thomas Sonnweber[15], Sebastian J. Klein[14], Alex Pizzini[15], Yvonne Wohlfarter[16], Sabina Sahanic[15], Richard Hilbe[15], Benedikt Schaefer[17], Sonja Wagner[17], Felix Machleidt[4], Carmen Garcia[4], Christoph Ruwwe-Glösenkamp[4], Tilman Lingscheid[4], Laure Bosquillon de Jarcy[4], Miriam S. Stegemann[4], Moritz Pfeiffer[4], Linda Jürgens[4], Sophy Denker[18], Daniel Zickler[19], Claudia Spies[20], Andreas Edel[20], Nils B. Müller[19], Philipp Enghard[19], Aleksej Zelezniak[2,21], Rosa Bellmann-Weiler[15], Günter Weiss[15], Archie Campbell[22,23], Caroline Hayward[24], David J. Porteous[22,23], Riccardo E. Marioni[22], Alexander Uhrig[4], Heinz Zoller[17], Judith Löffler-Ragg[15], Markus A. Keller[16], Ivan Tancevski[15], John F. Timms[6†], Alexey Zaikin[5,6,8,25], Stefan Hippenstiel[4,26], Michael Ramharter[27], Holger Müller-Redetzky[4], Martin Witzenrath[4,26], Norbert Suttorp[4,26], Kathryn Lilley[3], Michael Mülleder[28], Leif Erik Sander[4,26], PA-COVID-19 Study group, Florian Kurth[4,27‡*], Markus Ralser[1,2‡]

1 Charité–Universitätsmedizin Berlin, Department of Biochemistry, Berlin, Germany, **2** The Francis Crick Institute, Molecular Biology of Metabolism Laboratory, London, United Kingdom, **3** The University of Cambridge, Department of Biochemistry and Cambridge Centre for Proteomics, Cambridge, United Kingdom, **4** Charité–Universitätsmedizin Berlin, Department of Infectious Diseases and Respiratory Medicine, Berlin, Germany, **5** University College London, Department of Mathematics, London, United Kingdom, **6** University College London, Department of Women's Cancer, EGA Institute for Women's Health, London, United Kingdom, **7** National Phenome Centre and Imperial Clinical Phenotyping Centre, Department of Metabolism, Digestion and Reproduction, Imperial College London, London, United Kingdom, **8** Lobachevsky University, Laboratory of Systems Medicine of Healthy Ageing, Nizhny Novgorod, Russia, **9** Imperial College London, Section of Bioanalytical Chemistry, Division of Systems Medicine, Department of Metabolism, Digestion and Reproduction, London, United Kingdom, **10** Berlin Institute of Health, Berlin, Germany, **11** Lobachevsky University, Department of Applied Mathematics, Nizhny Novgorod, Russia, **12** University of Hertfordshire, School of Physics, Astronomy and Mathematics, Hatfield, United Kingdom, **13** Sechenov First Moscow State Medical University, Department of Paediatrics and Paediatric Infectious Diseases, Moscow, Russia, **14** Medical University Innsbruck, Division of Intensive Care and Emergency Medicine, Department of Internal Medicine, Innsbruck, Austria, **15** Medical University of Innsbruck, Department of Internal Medicine II, Innsbruck, Austria, **16** Medical University of Innsbruck, Institute of Human Genetics, Innsbruck, Austria, **17** Medical University of Innsbruck, Christian Doppler Laboratory for Iron and Phosphate Biology, Department of Internal Medicine I, Innsbruck, Austria, **18** Charité–Universitätsmedizin Berlin, Medical Department of Hematology, Oncology & Tumor Immunology, Virchow Campus & Molekulares Krebsforschungszentrum, Berlin, Germany, **19** Charité–Universitätsmedizin Berlin, Department of Nephrology and Internal Intensive Care Medicine, Berlin, Germany, **20** Charité–Universitätsmedizin Berlin, Department of Anesthesiology and Intensive Care, Berlin, Germany, **21** Chalmers University of Technology, Department of Biology and Biological Engineering, Gothenburg, Sweden, **22** University of Edinburgh, Centre for Genomic and Experimental Medicine, Institute of Genetics and Cancer, United Kingdom, **23** University of Edinburgh, Usher Institute, Edinburgh, United Kingdom, **24** University of Edinburgh, MRC Human Genetics Unit, Institute of Genetics and Cancer, Edinburgh, United Kingdom, **25** Centre for Analysis of Complex Systems, Sechenov First Moscow State Medical University, Moscow, Russia, **26** German Centre for Lung Research, Germany, **27** Bernhard Nocht Institute for Tropical Medicine, Department of Tropical Medicine, and University Medical Center Hamburg-Eppendorf, Department of Medicine, Hamburg, Germany, **28** Charité–Universitätsmedizin Berlin, Core Facility—High-Throughput Mass Spectrometry, Berlin, Germany

OPEN ACCESS

Citation: Demichev V, Tober-Lau P, Nazarenko T, Lemke O, Kaur Aulakh S, Whitwell HJ, et al. (2022) A proteomic survival predictor for COVID-19 patients in intensive care. PLOS Digit Health 1(1): e0000007. https://doi.org/10.1371/journal.pdig.0000007

Editor: Reinhard Bauer, Universitatsklinikum Jena, GERMANY

Received: June 11, 2021

Accepted: November 18, 2021

Published: January 18, 2022

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: https://doi.org/10.1371/journal.pdig.0000007

Data Availability Statement: The protein quantities table along with the associated metadata are provided in supplementary materials. All scripts used to train and assess the machine learning

models are likewise provided (S1 Data.zip). The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) via the PRIDE partner repository with the dataset identifier PXD029009. All scripts used to train and assess the machine learning models as well as the respective input proteomics data are provided in supplementary materials (S1 Data.zip).

Funding: This work was supported by the Berlin University Alliance (501_Massenspektrometrie, 501_Linklab, 112_PreEP_Corona_Ralser), by UKRI/NIHR through the UK Coronavirus Immunology Consortium (UK-CIC), the BMBF/DLR Projektträger (01KI20160A, 01ZX1604B, 01KI20337, 01KX2021), Charité-BIH Centrum für Therapieforschung (BIH_PA_covid-19_Ralser), the BBSRC (BB/N015215/1, BB/N015282/1), the Francis Crick Institute, which receives its core funding from Cancer Research UK (FC001134), the UK Medical Research Council (FC001134), and the Wellcome Trust (FC001134 and IA 200829/Z/16/Z), as well as the European Research Council (SyG 951475 to M.R.). The work was further supported by the Ministry of Education and Research (BMBF), as part of the National Research Node 'Mass spectrometry in Systems Medicine (MSCoresys), under grant agreements 031L0220A and 161L0221. The study was further supported by the German Federal Ministry of Education and Research (NaFoUniMedCovid19 – NUM -NAPKON, NUM-COVIM, FKZ: 01KX2021 and PROVID—FKZ: 01KI20160A) to Florian Kurth, Leif E. Sander, Martin Witzenrath, Norbert Suttorp and Stefan Hippenstiel. Leif Erik Sander is supported by the German Research Foundation (DFG, SFB-TR84 114933180) and by the Berlin Institute of Health (BIH), which receives funding from the Ministry of Education and Research (BMBF). Martin Witzenrath is supported by grants from the German Research Foundation, SFB-TR84 C06 and C09, by the German Ministry of Education and Research (BMBF) in the framework of the CAPSyS (01ZX1304B), CAPSyS-COVID (01ZX1604B), SYMPATH (01ZX1906A) and PROVID project (01KI20160A) and by the Berlin Institute of Health (CM-COVID). Stefan Hippenstiel is supported by the German Research Foundation (DFG, SFB-TR84 A04 and B06), and the BMBF (PROVID, and project 01KI2082). Norbert Suttorp is supported by grants from the German Research Foundation, SFB-TR84 C09 und Z02, by the German Ministry of Education and Research (BMBF) in the framework of the PROGRESS 01KI07114. The study was further supported by Wellcome Trust (200829/Z/16/Z). The Generation Scotland study received core support from the Chief Scientist Office of the

☯ These authors contributed equally to this work.
† Deceased.
‡ FK and MR also contributed equally to this work.
* florian.kurth@charite.de

Abstract

Global healthcare systems are challenged by the COVID-19 pandemic. There is a need to optimize allocation of treatment and resources in intensive care, as clinically established risk assessments such as SOFA and APACHE II scores show only limited performance for predicting the survival of severely ill COVID-19 patients. Additional tools are also needed to monitor treatment, including experimental therapies in clinical trials. Comprehensively capturing human physiology, we speculated that proteomics in combination with new data-driven analysis strategies could produce a new generation of prognostic discriminators. We studied two independent cohorts of patients with severe COVID-19 who required intensive care and invasive mechanical ventilation. SOFA score, Charlson comorbidity index, and APACHE II score showed limited performance in predicting the COVID-19 outcome. Instead, the quantification of 321 plasma protein groups at 349 timepoints in 50 critically ill patients receiving invasive mechanical ventilation revealed 14 proteins that showed trajectories different between survivors and non-survivors. A predictor trained on proteomic measurements obtained at the first time point at maximum treatment level (i.e. WHO grade 7), which was weeks before the outcome, achieved accurate classification of survivors (AUROC 0.81). We tested the established predictor on an independent validation cohort (AUROC 1.0). The majority of proteins with high relevance in the prediction model belong to the coagulation system and complement cascade. Our study demonstrates that plasma proteomics can give rise to prognostic predictors substantially outperforming current prognostic markers in intensive care.

Author summary

Healthcare systems around the world are struggling to accommodate high numbers of the most severely ill patients with COVID-19. Moreover, the pandemic creates a pressing need to accelerate clinical trials investigating potential new therapeutics. While various biomarkers can discriminate and predict the future course of disease for patients of different disease severity, prognosis remains difficult for patient groups with similar disease severity, e.g. patients requiring intensive care. Established risk assessments in intensive care medicine such as the SOFA or APACHE II show only limited reliability in predicting future disease outcomes for COVID-19. In this study we hypothesized that the plasma proteome, which reflects the complete set of proteins that are expressed by an organism and are present in the blood, and which is known to comprehensively capture the host response to COVID-19, can be leveraged to allow for prediction of survival in the most critically ill patients with COVID-19. Here, we found 14 proteins, which over time changed in opposite directions for patients who survive compared to patients who do not survive on intensive care. Using a machine learning model which combines the measurements of multiple proteins, we were able to accurately predict survival in critically ill patients with COVID-19 from single blood samples, weeks before the outcome, substantially outperforming established risk predictors.

Scottish Government Health Directorates (CZD/16/6) and the Scottish Funding Council (HR03006), and is now supported by the Wellcome Trust (216767/Z/19/Z). Archie Campbell is funded by HDR UK and the Wellcome Trust (216767/Z/19/Z). Caroline Hayward is supported by an MRC University Unit Programme Grant (MC_UU_00007/10) (QTL in Health and Disease). Riccardo Marioni is supported by an Alzheimer's Research UK project grant (ARUK-PG2017B-10). HJ. Whitwell, JF. Timms, A. Zaikin and T. Nazarenko are supported by a Medical Research Council grant (MR/R02524X/1) and HJ. Whitwell, A. Zaikin and O. Blyuss by the Ministry of Science and Higher Education agreement No. 075-15-2020-808. HJ. Whitwell is supported by the National Institute for Health Research (NIHR) Imperial Biomedical Research Centre (BRC). J. Timms is supported by the National Institute for Health Research (NIHR) UCLH/UCL Biomedical Research Centre. Mirja Mittermaier is a participant in the BIH-Charité Digital Clinician Scientist Program funded by the Charité – Universitätsmedizin Berlin, the Berlin Institute of Health, and the German Research Foundation (DFG). Markus A. Keller is supported by the Austrian Science Funds (FWF; P33333) and the Austrian Research Promotion Agency (FFG, #878654). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors declare no competing interests. Author John F. Timms was unable to confirm their authorship contributions. On their behalf, the corresponding author has reported their contributions to the best of their knowledge.

Introduction

The COVID-19 pandemic has brought health systems around the globe to the brink of collapse. Capacities for intensive care treatment of patients with organ failure have reached their limits in many regions with intense SARS-CoV-2 transmission and were often central to political decisions regarding restrictions on public life, e.g. through contact restrictions or lockdowns. The global impact of the pandemic increases the pressures to devise new clinical approval strategies so that potential therapeutics can be identified and tested faster, at higher accuracy, and in clinical trials with smaller sample sizes [1]. Various models for classification of disease severity and for prediction of clinical trajectories and outcome have been developed for COVID-19, based on laboratory measurements, clinical scores, imaging, and omics technologies [2–5]. These pointed to the importance of specific immune cells, inflammatory and antiviral cytokines and chemokines, as well as the coagulation cascade in COVID-19 disease progression [5–13]. They predict the risk of the future need for mechanical ventilation in the heterogeneous group of patients at early time points, e.g. at admission to the hospital, when clinical parameters and biomarkers differ substantially between mildly affected and severely ill patients [2–5,14].

Treatment decisions within the most severely ill patients, for instance whether a patient should be treated with extracorporeal membrane oxygenation (ECMO), have a major impact on resources. Currently, such decisions are often based primarily on the patient's age, comorbidities, and established intensive care prognosis models, such as the Sequential Organ Failure Assessment (SOFA) or Acute Physiology and Chronic Health Evaluation (APACHE II), which assess the patient on the basis of a combination of established clinical and laboratory risk parameters [15,16]. However, the predictive values of both SOFA and APACHE II for the most critical forms of COVID-19 are limited [17–19], creating a diagnostic gap and imminent need for reliable predictors, specifically validated in severely ill COVID-19 patients, to guide and tailor efforts in treating these critically ill patients. Moreover, the lack of reliable predictors increases the challenge of interpreting the results of early phase clinical trials, which typically enroll low numbers of patients. Indeed, testing for the success of a clinical intervention requires classification of divergent clinical trajectories within more homogeneous groups, such as WHO grade 7 patients. At least in COVID-19, this is hampered by the fact that molecular signatures within a group of patients with comparable disease severity are considerably more similar when compared to the differences between mild and severe patients [6,7,14].

Plasma proteomics holds the promise of integrating the genetic background of an individual with their life history, physiological, nutritional, and demographic parameters, and hence, have the potential to form the foundation of a new generation of predictors [20–24]. Among the spectrum of proteomic technologies available, mass spectrometry has the appeal that once markers are identified, they allow for the direct generation of targeted panel assays measurable by selective reaction monitoring (SRM), simplifying their implementation into clinical routine. Recently, new mass spectrometry based proteomic technologies have been developed to increase throughput and measurement precision, so that the path from discovery to application is simplified [6,25–28].

We studied proteomes of two well characterized cohorts of the most severely ill patients with COVID-19 in two independent health care centers (Charité–Universitätsmedizin Berlin, Germany, and Medical University of Innsbruck, Austria) who gave informed consent to deep clinical and molecular phenotyping [14,19,29]. Using a recently published dataset from our group [14], we specifically assessed whether proteomic measurements can be used to predict the outcome (death vs. survival) of severe COVID-19 from time series data,

as well as from samples taken at key clinical decision points. We found 14 protein concentration trajectories that, over the timeline of disease progression, distinguish survivors from non-survivors. Moreover, a machine learning (ML) model, based on parenclitic networks, generated accurate prognosis on single time point samples that were collected once the patient reached the maximum treatment level. Emphasizing the prognostic potential of the proteome, this sample was, in median, taken 39 days before outcome. The ML predictor trained on these samples substantially outperformed established clinical risk scores and predicted the outcome among a group of severely ill patients with similar clinical presentation with high accuracy.

Results

The exploratory cohort used for marker identification and model generation consisted of the 50 most severely ill COVID-19 patients out of a cohort of 168 patients with varying disease severity, treated between 15 March and 16 September 2020 at Charité University Hospital, Berlin, Germany, a tertiary care referral centre for the treatment of ARDS with associated weaning centre (**Fig 1A**) [14,19,29]. There were no treatment restrictions due to shortages of intensive care capacity at the time of this patient cohort. The 50 patients selected for the study were treated in intensive care with invasive mechanical ventilation plus additional organ support such as renal replacement therapy (RRT), ECMO, or vasopressors, corresponding to grade 7 on the WHO Ordinal Scale for Clinical Improvement. Patients with limitations of therapy according to their wish were excluded. Thirty-six (72%) patients required RRT, 19 (38%) patients were treated with ECMO, and 16 (32%) patients were treated with both RRT and ECMO. Fifteen (30%) patients died. Median time of hospitalization in survivors was 63 days (n = 35, IQR 44–89). Median time from admission to death was 28 days (n = 15, IQR 16–43). Patient characteristics are shown in S1 Table. The details on the proteomic workflow, protein detection rates, as well as patient trajectories are provided in S1 and S2 Figs of our previous work [14].

Within this treatment group of critically ill COVID19 patients, the Charlson Comorbidity Index [30,31] performed poorly in classifying survivors from non-survivors by AUROC values of 0.63 (P = 0.16, **Fig 2A**). From a time-resolved data resource for the PA-COVID-19 study, spanning over a compendium of clinical parameters, plasma proteomes, cell counts, enzyme activities, and outcomes [14], we further determined the SOFA and APACHE II scores. These scores, too, could not confidently distinguish survivors from non-survivors (**Fig 2A**, AUROC = 0.68, P = 0.05 for APACHE II score at ICU admission, and AUROC = 0.65, P = 0.11 for SOFA score at the time of first sampling at WHO grade 7).

Studying the plasma proteomes [14] we found 78 proteins for which the concentration changed significantly during the patients' disease course. Out of these proteins, 14 were found to change differently over time for survivors and non-survivors (**Fig 1B**, **Fig 1C**). Patients with fatal outcomes were characterized by a significant increase in inflammatory proteins over time (SAA1, SAA2, CRP, ITIH3, LRG1, SERPINA1, SERPINA10 and LBP). Conversely, the levels of these proteins in plasma decreased over time in survivors. Moreover, anti-inflammatory proteins (SERPINA4, A2M) decreased over time in non-survivors, indicating a persistent pro-inflammatory signature. Similarly, two key proteins of the coagulation system, thrombin (F2) and plasma kallikrein (KLKB1), known to be decreased in severe COVID-19 [12,14], further decreased over time in non-survivors, while increasing in survivors.

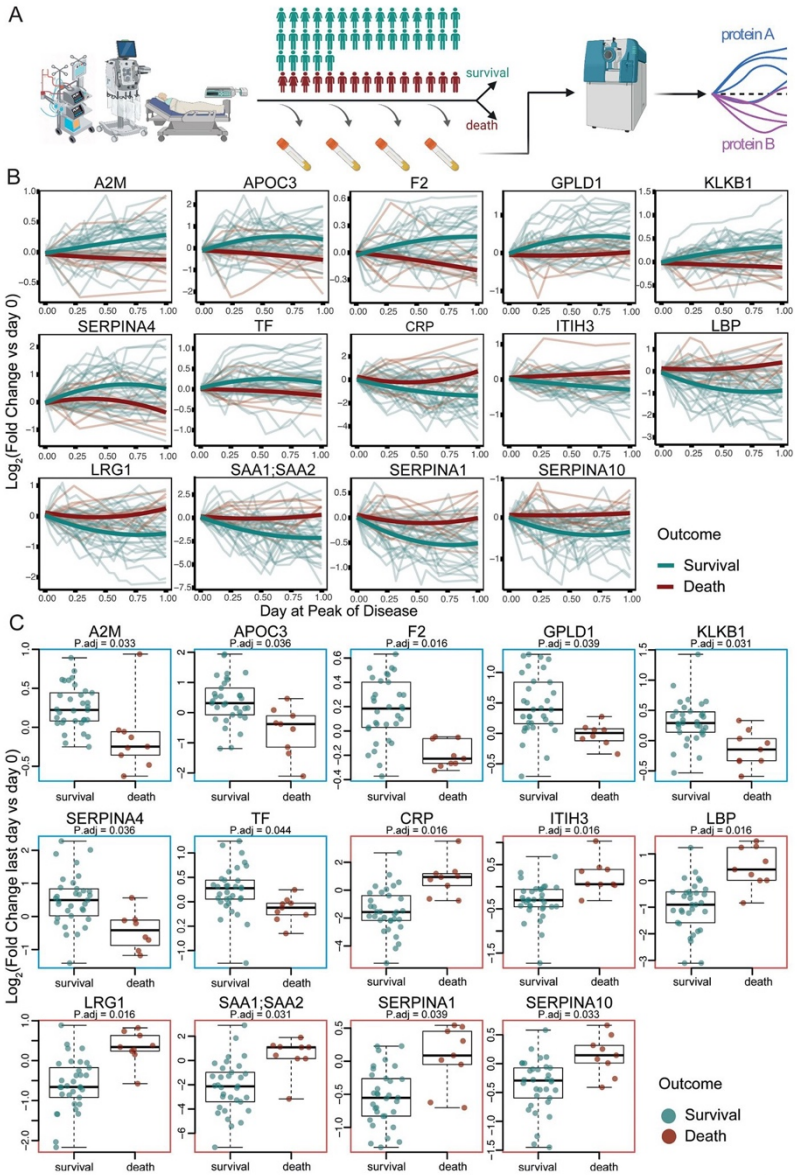


Fig 1. Protein concentration trajectories that differentiate survivors of critical COVID-19 from non-survivors. a) Fifty Patients with PCR-confirmed COVID-19 treated at Charité University Hospital Berlin, Germany, were sampled longitudinally, to generate high-resolution time series for 321 protein quantities. In parallel, precise clinical phenotyping was performed, including recording of intensive care and disease severity scores, treatment parameters, and outcome

(PA-COVID-19 data resource [14]). **b)** Protein level trajectories over time (FDR < 0.05), for which time-dependent concentration changes (y-axis: log2 fold change) during the peak of the disease differentiate survivors from non-survivors in critically ill patients (Methods). **c)** as **b)** but expressed as boxplots (log2 fold change last vs first day). Figure created with BioRender.com.

https://doi.org/10.1371/journal.pdig.0000007.g001

For diagnostic purposes and treatment decisions time series data is however impractical to obtain. We therefore explored the potential of using single time point samples to predict outcome. We chose the earliest sample obtained after the critical decision regarding escalation of treatment, i.e. the earliest sample obtained at the maximum treatment level (WHO grade 7), to generate an outcome predictor. The median time from sampling until the outcome was 39 (IQR 16–64) days in our cohort. Using 57 proteins for which targeted mass spectrometric assays (MRM assays) as listed in the MRMAssayDB [32] are available, indicating that they have been selected for a clinical or biomedical indication also in other context, we established a machine learning model based on parenclitic networks, a graph-based approach in which networks representing the deviation of an individual from the population are derived [33,34]. The networks are generated by considering every pair of analytes (proteins) individually and calculating the respective edge weight as the estimated probability of fatal outcome based on this pair of proteins. Predictive models are then generated by considering the topological differences between networks from individual cases (non-survivors vs. survivors) (Methods). We achieved high prediction accuracy on the test subjects, who were excluded when training the machine learning model (in a cross-validation fashion, see Methods), with AUC = 0.81 (95% CI 0.68–0.94) for the receiver-operating characteristic (ROC) curve (**Fig 2B**). Out of the 25 proteins with the highest relevance in the parenclitic model, 15 are components of the coagulation system and 8 proteins belong to the complement cascade (**S2 Table**). To further demonstrate that the proteomic data contains sufficient physiological information to allow outcome prediction, i.e. that the results are not restricted to a specific algorithm, we also tested a model based on a support vector machine (SVM). The SVM proved to be capable of survival prediction as well, albeit with inferior performance compared to the parenclitic network (**S1 Fig**).

To independently validate the potential of the plasma proteome to predict outcomes in critically ill COVID19 patients, we examined the performance of the parenclitic network trained on our prime cohort (Charité) on an independent cohort of 24 patients with critical COVID-19 from Austria (survival n = 19, death n = 5, median time between sampling and outcome 22 days, interquartile range 15–42 days) ('Innsbruck' cohort, Methods). Despite the validation cohort originating from a different hospital and health care system, the machine learning model demonstrated high predictive power on this independent cohort (AUROC = 1.0, P = 0.000047, **Fig 2C**). Using the cutoff value for survival prediction derived from the Charité cohort, the model correctly predicted the outcome for 18 out of 19 patients who survived and for 5 out of 5 patients who died in this independent 'Innsbruck' cohort.

Discussion

The prognostic value of several biomarkers (e.g. CRP, IL-6, ferritin) and clinical scores for predicting disease progression in COVID-19 at early disease stages, e.g. at hospital admission, is now well established [35,36]. For the comparatively homogeneous subgroup of severely ill patients already requiring mechanical ventilation and additional organ support, prediction of future disease trajectories and outcome (survival or death) is by far more challenging, and only limited data exist [17,37,38]. Moreover, clinical severity scores are often not validated for unconscious patients, and laboratory measurements are frequently confounded by intensive care treatment. Outcome of ICU patients may further be critically determined by resource

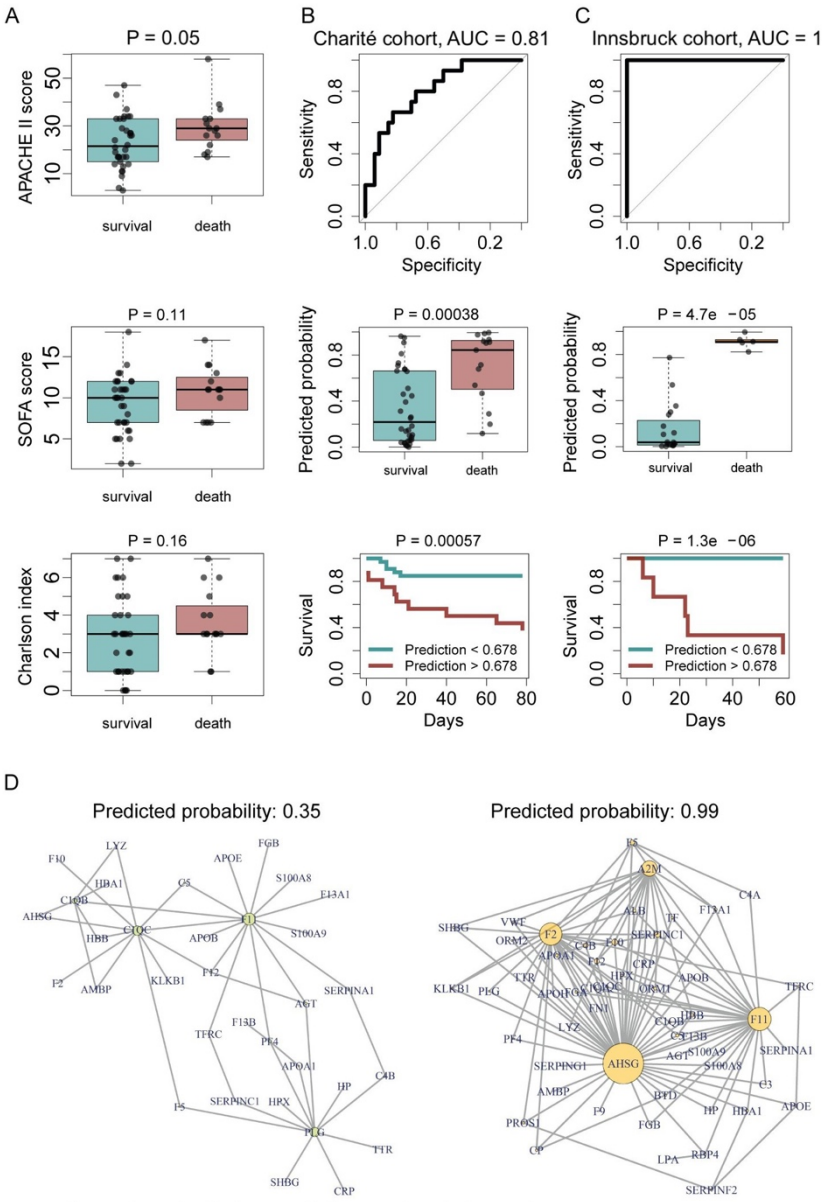


Fig 2. Prediction of survival or death in critically ill patients, from the first sampling time point at intensive care treatment level (WHO grade 7). a) Performance of established ICU risk assessment indices (APACHE II, SOFA and Charlson comorbidity

index) calculated at the time of ICU admission (APACHE II, Charlson comorbidity index) or at the first time point at WHO grade 7 (SOFA score) in predicting the outcome in critically ill patients. **b)** Prediction of survival or death in critically ill patients using proteomics. A machine learning model based on parenclitic networks (Methods) was trained on the samples of the Charité cohort closest to the time point of treatment escalation during intensive care (start of ECMO, RRT or vasopressors, i.e. WHO grade 7). The performance was assessed on the test samples, which were held out during training. **Upper panel:** The ROC curve indicates correct classification of survival vs non-survival with an AUROC of 0.81 (95% CI 0.68–0.94). **Middle panel**: The proteomic classifier was used to predict the probability of survival and non-survival, which is significantly different between the groups. **Lower panel**: Kaplan-Meier survival curves using a threshold of predicted probability (0.678) chosen to maximize Youden's J index (J = sensitivity + specificity—1). Log-rank test was used to compare survival rates between patients with predicted death risk < 0.678 (black) and > 0.678 (orange). **c) (upper, middle, and lower panels):** The model trained on the Charité cohort, was tested on an independent cohort (Innsbruck). **d)** Exemplary parenclitic networks from two patients in the independent Innsbruck cohort. Edges with weights > 0.5 are shown. Left panel: a network predicting low probability of death in a surviving patient. Right panel: a network predicting high probability of death in a non-survivor.

https://doi.org/10.1371/journal.pdig.0000007.g002

constraints, the varying level of experience with organ replacement therapies or the rates of superinfection, rendering prediction complex [38]. On the other hand, patients in intensive care units, and particularly those in need of special organ replacement therapies such as ECMO, require a disproportionately large share of resources compared to other patients, so decisions to initiate such therapies should be based on the best information and assessment possible. Prognostic tools in critically ill patients are hence of crucial importance to guide and tailor the treatment efforts. This is particularly true in a situation when health care systems are overstrained. Another key potential for the use of outcome predictors is clinical trial monitoring, where measurements of prognostic molecular signatures over time can be used to evaluate experimental therapies on an individual, time-resolved basis. Moreover, an accurate outcome predictor would allow us to test whether a given treatment changes the predicted trajectory of an individual patient.

Previously, we and others investigated plasma proteome alterations in COVID-19 [6–8,10,12,14], which show a remarkable ability to classify the severity of disease. For instance, our investigations showed that the host response in the early inflammatory phase creates a strong signature in the plasma proteome, and is critical as well as predictive about the future disease progression in severe COVID-19 [14]. New proteomic platform technologies have significantly gained precision and throughput compared to their predecessors, rendering the application of multivariate regression models more effective and bringing them increasingly close to routine clinical use [6]. Importantly, even without platform technologies, biomarkers identified in proteomic profiles can be translated into clinical use, e.g. by using standard techniques such as selective reaction monitoring (SRM) for the quantification of protein panels, or enzyme linked immunosorbent assays (ELISA) for the sensitive quantification of individual biomarkers.

Here, we show that an increase in specific inflammatory and acute phase proteins over time (e.g., SAA1;SAA2, CRP, ITIH3, LRG1, SERPINA1, and LBP) is associated with the risk of death from COVID-19, while an increase of kallikrein (KLKB1), kallistatin (SERPINA4), thrombin (F2), apolipoprotein C3 (APOC3), GPLD1, and the protease inhibitor A2M, is associated with survival. Interestingly, we and others have found all of these proteins to also be differentially expressed depending on disease severity in COVID-19 [6,7,10,12,14]. Moreover, there is substantial overlap with a panel of proteins predictive of mortality in COVID-19 identified by Völlmy et al. [39]. Hence, despite only a subset of proteins that are differentially concentrated depending on disease severity predict outcome, and the fact that typical single-centre ICU studies are conducted on small numbers of patients, this result indicates a high congruence and reproducibility of plasma proteome signatures across studies.

SAA1;SAA2, CRP, ITIH3, SERPINA1 are acute phase proteins that are also dysregulated in other inflammatory states including sepsis [40]. Increased LRG1 and LBP as well as decreased A2M [40,41] are indicators of an ongoing immune response, complementing

the general pro-inflammatory signature of these predictive proteins. APOC3 and GPLD1 are involved in lipid metabolism, which has been shown to be dysregulated in bacterial pneumonia, thereby associated with unfavorable outcomes [42]. Kallikrein is involved in the blood coagulation system, fibrinolysis, and the complement cascade, three systems known to be dysregulated in COVID-19 [43–45]. It mediates the cleavage of kininogen to bradykinin and des-Arg^9-bradykinin, a potent vasoactive peptide which is counter-regulated by ACE2, the cell entry receptor for SARS-CoV-2. Since the loss of ACE2 in COVID-19 supposedly leads to an imbalance of bradykinins, inhibition of the kallikrein-kinin system has been discussed as a treatment strategy in COVID-19 [46–48]. This hypothesis is not supported by our data, which indicate improved prognosis with increasing kallikrein levels. Kallikrein is counterbalanced by kallistatin, which equally increased over time in survivors in our study population, thereby potentially equilibrating the increase in the kinin-kallikrein system. Kallistatin is known for pleiotropic effects in vascular repair, endothelial function, and inflammation [49] and possesses protective properties in acute lung injury. According to our data kallistatin should be considered as a potential candidate for clinical testing in critical COVID-19 [50].

While prognostic assessments based on repeated measurements over time allow for treatment monitoring, including evaluation of experimental therapies in clinical trials, prognostic measurements from single time points are particularly valuable for timely patient management and resource allocation. We therefore employed a machine learning model to integrate proteomic measurements from the first time point at WHO grade 7, i.e. invasive mechanical ventilation and additional organ support therapy, in order to derive prognosis of outcome. We achieve high prognostic values, both in the exploratory cohort, as well as in a fully independent cohort.

The results are currently based on a comparatively small number of patients with adverse outcome. Given the naturally small sample sizes of ICU cohorts and the exploratory character of our study, findings will have to be validated in larger cohorts, before further steps can be undertaken to translate our findings into clinical practice in the future. The panel of proteins identified in our study should also be assessed for other conditions such as non-COVID-19 ARDS.

The majority of proteins with the highest relevance for the machine learning predictor were components of the coagulation system and the complement cascade (**S2 Table**). Both systems are known to be crucial for treatment and disease courses for severely ill COVID-19 patients [9,10]. This is particularly well illustrated by recent data from a multi-platform clinical trial indicating that a substantial proportion of patients with severe COVID-19 develop thromboembolic events despite therapeutic anticoagulation [51,52]. The protein with the highest relevance in our model is Fetuin-A (AHSG), which is known to be strongly downregulated in severe COVID-19 [10,14]. Of note, genetic polymorphisms associated with higher AHSG plasma concentrations were found to be protective in SARS-CoV-1 infection [53]. One important function of AHSG is regulation of inflammation through deactivation of macrophages [54], and there is emerging evidence that macrophages play a key role in pulmonary inflammation and dysfunction in COVID-19 [11,55–57]. A number of proteins identified as outcome predictors have also been shown to be differentially expressed in sepsis, including SAA1, CRP, SERPINA1, KLKB1, and A2M [40], indicating a general inflammatory signature rather than specific markers of COVID-19.

In summary, we have leveraged the power of the proteome to address a problematic diagnostic gap in the prognosis of the most critical form of COVID-19, that is not covered by established clinical assessments, such as the SOFA or APACHE II scores. We show that the proteome accurately predicts survival in critically ill patients with COVID-19, from samples

that were collected 39 days in median before the outcome. The findings warrant further prospective assessment of proteomic predictors and the described models in larger cohorts. The majority of proteins with high relevance in the model are components of the coagulation system and complement cascade, highlighting their critical role in progression and outcome of most severe COVID-19.

Methods

Charité patient cohort and clinical data

Patients included in this analysis are a sub-cohort of the PA-COVID-19 study conducted at Charité—Universitätsmedizin Berlin, a prospective observational cohort study on the pathophysiology of COVID-19 as described previously [14,19,29]. All patients with PCR-confirmed SARS-CoV-2 infection that progressed to critical disease (WHO grade 7, i.e. invasive mechanical ventilation and additional organ support), were eligible for inclusion. Exclusion criteria included refusal to provide informed consent by the patient or a legal representative, and any condition prohibiting serial biosampling. Patients were treated according to current clinical guidelines. Patients for whom limitation of therapy was decided according to the patient's wish were excluded from analysis. This includes three cases, for whom limitation of therapy was decided at a later time point according to the patient's presumed wish and predictably unfavorable outcome. All other patients received maximum intensive care treatment including organ replacement therapies at the discretion of the responsible physicians. One patient (ID 135), who was still hospitalized and clinically improving 5 months after admission, was classified as a survivor. One patient still in critical condition 5 months after admission was excluded due to uncertain outcome.

Biosampling of EDTA plasma for proteome measurement was performed up to 3 times per week after inclusion. Disease severity was assessed according to the WHO ordinal scale for clinical improvement (World Health Organisation 2020). Clinical data were captured in secuTrial (interActive Systems GmbH, Berlin, Germany). Pseudonymized data exported from secuTrial were processed using JMP Pro 15 (SAS Institute Inc., Cary, NC, USA).

Innsbruck Patient cohort and clinical data

Serum samples from patients admitted to the intensive care unit at the Department of Medicine, University Hospital of Innsbruck with PCR-confirmed severe COVID-19 were collected within the first days (median 7.5, IQR 5–12) after admission, and written informed consent was obtained. Patients were treated according to national guidelines. The study was approved by the local ethics research committee EK-Nr. 1107/2020, and EK-Nr. 1103/2020 for follow-up.

Statistical analysis and multiple-testing correction

Statistical testing on proteomic and diagnostic data was performed in the R environment for statistical computing, version 3.6.0 [58], as described previously [14]. Briefly, all protein measurements were first log2-transformed and only protein groups matched to at least three different peptides were considered. Quantities of gene products corresponding to open reading frames IGxx (i.e. different types of immunoglobulin chains) were summed together to generate quantities representative of the overall levels of immunoglobulin classes (IGHVs, IGLVs, etc). Imputation of missing data was not performed. Significance testing for equal medians was performed using the Mann-Whitney U test, as implemented in the "wilcox.test" function of the "stats" R package. A non-parametric test was chosen here to minimise the influence of outliers on the calculated p-values. Multiple-testing correction

was performed using the Benjamini-Hochberg false discovery rate controlling procedure [59], implemented in the "p.adjust" function of the "stats" R package. Adjusted p-values below 0.05 were considered significant.

Identifying omics trajectories that are predictive of survival at the peak period of the disease

For each omics feature, the difference between its log2-levels at the last and the first sampling timepoints during the peak period of the disease was considered. This period was defined as the time when the patient was receiving the most intensive treatment during their stay in hospital, that is the time when the patient was at WHO grade 6 or 7. The distribution of this difference between survivors and non-survivors was compared using the Mann-Whitney U test. Only non-DNI patients with known outcome were included.

Prediction of survival

The first time point measured at the WHO grade 7 was selected per patient, to train the survival predictor. This ensured that 'future' information, encoded in the later time points, was not used for predictor training. To reduce the feature space used as input for the machine learning model, we limited it to the quantities of 57 proteins which are FDA-approved biomarkers with MRM assays available [32] and which were quantified with at least three different peptides in this study. Missing values were imputed using minimal value imputation, and the data were standardized.

Machine learning was carried out using the parenclitic networks approach [33,34]. Briefly, during training, for each pair of features, a radial SVM classifier is trained (using the svm() function from the "e1071" R package with default settings). For each sample, a network is then built, wherein vertices correspond to features and the edge weight is the death probability as predicted by the SVM classifier. Maximum, mean and standard deviation of the edge weights, as well as the numbers of edges with weights greater than 0.5 (i.e. fatal outcome is predicted) and nodes with at least one such edge are calculated. A LASSO classification model (alpha = 0.01) is then constructed on these 5 features using the glmnet() function of the "glmnet" [60] R package with default settings.

For the assessment of the classifier performance (Charité cohort), a cross-validation method was applied in the following way: the prediction was made for each sample by excluding (withholding) it from the dataset along with two other samples (chosen randomly with the constraint that out of 3 samples one corresponds to a non-survivor and two to survivors), training the classifier on the remaining (independent) samples and then generating predictions for the withheld samples using the trained model. Such a leave-3-out partition was generated randomly 50 times and the predictions for each sample were averaged. The partitioning strategy ensured that the evaluation of the predictive performance would not be affected by any potential overfitting, no matter how significant. For the assessment of the performance on an independent dataset (Innsbruck cohort), the classifier was trained on all the Charité samples and used to estimate the probabilities of fatal outcome on the Innsbruck cohort. The source code is provided in supplementary materials.

The 'relevance' scores for proteins in the parenclitic model were calculated as Kleinberg's authority centrality scores for the respective vertices in the "generalizing network". This network was generated by (i) replacing edge weights greater than 0.5 with 1.0 and weights less than 0.5 with 0.0 in the networks corresponding to non-survivors and (ii) averaging the resulting networks.

For survival prediction using support vector machines (SVM), the same data and the same selection of proteins as for the parenclitic network model was applied. The SVM was built in Python 3.8.5 using the SVC() function with an rbf-kernel and a gamma value of 0.005 as implemented in scikit-learn 0.23.2 [61]. To circumvent class-imbalances, balanced class-weights were assumed. For benchmarking the model a stratified 10-fold cross-validation was performed. The data were scaled to zero mean and unit variance based on the training data. The reported results for the Charité-cohort are based on the data that were withheld when constructing the model in each cross-validation step. For validating the model, a model was trained on all samples of the Charité-cohort and validated using the independent Innsbruck-cohort. p-Values were calculated using the Mann-Whitney U test as implemented in SciPy 1.5.2 [62]. AUC values and confidence intervals were obtained using the roc() function of the pROC R package.

We followed the guidelines for transparent reporting of multivariable prediction models for individual prognosis or diagnosis (TRIPOD) as proposed by the EQUATOR network [63].

Study approval

The study was approved by the ethics committee of Charité—Universitätsmedizin Berlin (EA2/066/20) and conducted in accordance with the Declaration of Helsinki and guidelines of Good Clinical Practice (ICH 1996). Written informed consent was obtained from all patients or legal representatives according to regulations set by the ethics committee of Charité—Universitätsmedizin Berlin. The study is registered in the German and the WHO international registry for clinical studies (DRKS00021688).

Supporting information

S1 Tripod Checklist. Transparent reporting of multivariable prediction models for individual prognosis or diagnosis (TRIPOD) checklist as proposed by the EQUATOR network [63]. Checklist includes location of key aspects within the manuscript.
(PDF)

S1 Table. Baseline, treatment, and outcome characteristics of patient cohort with severe COVID-19 receiving maximum therapy at Charité—University hospital Berlin.
(DOCX)

S2 Table. Top 25 proteins included in the machine learning model, ordered by their estimated 'relevance' scores (Methods). Red writing indicates proteins involved in the complement system. Blue writing indicates proteins involved in coagulation.
(DOCX)

S1 Fig. Performance of an SVM model in predicting survival for critical (WHO grade 7) COVID-19 patients. Left panel: Boxplot of the decision function of the SVM for the Charité cohort. Displayed is the performance on the test data that were not used for model training. Middle panel: Boxplot of the decision function of the SVM for the Innsbruck cohort using a pre-trained model based on the Charité-cohort. Right panel: ROC-Curve and AUC corresponding to the boxplots for the Charité-cohort (black) and for the Innsbruck-cohort (red). AUC values of 0.66 (95% CI 0.49–0.84) and 0.88 (95% CI 0.67–1.0) were obtained for the Charité and Innsbruck cohorts, respectively.
(EPS)

S1 Data. Scripts used to train and assess machine learning models as well as the respective input proteomics data.
(ZIP)

Acknowledgments

We thank Jan-David Manntz (Beckman, Germany) for help with the Biomek i7, Robert Lane, Jean-Baptiste Vincedent and Nick Morrice (SCIEX) for help with the TripleTOF 6600.

PA-COVID-19 Study group, Charité–Universitätsmedizin Berlin

Malte Kleinschmidt, Katrin M. Heim, Belén Millet, Lil Meyer-Arndt, Nils B. Müller, Ralf H. Hübner, Tim Andermann, Jan M. Doehn, Bastian Opitz, Birgit Sawitzki, Daniel Grund, Peter Radünzel, Mariana Schürmann, Thomas Zoller, Fridolin Steinbeis, Florian Alius, Philipp Knape, Astrid Breitbart, Yaosi Li, Felix Bremer, Panagiotis Pergantis, Susanne Fieberg, Anne Wetzel, Moritz Müller-Plathe, Timur Özkan, Carola Misgeld, Dirk Schürmann, Bettina Temmesfeld-Wollbrück, Britta Stier, Martin Möckel, Jan A. Graaw, Victor Wegener, Marc Kastrup, Felix Balzer, Daniel Wendisch, Sophia Brumhard, Sascha S. Haenel, Philipp Georg, Claudia Conrad, Kai-Uwe Eckardt, Lukas Lehner, Jan M. Kruse, Carolin Ferse, Roland Körner, Andreas Edel, Steffen Weber-Carstens, Alexander Krannich, Saskia Zvorc, Linna Li, Uwe Behrens, Sein Schmidt, Maria Rönnefarth, Christina Pley, Claudia Fink, Chantip Dang-Heine, Robert Röhle, Emma Lieker, Christian Wollboldt, Yinan Wu, Georg Schwanitz, Constanze Lüttke, Denise Treue, Michael Hummel, Victor M. Corman, Christian Drosten, Christof von Kalle

Author Contributions

Conceptualization: Leif Erik Sander, Florian Kurth, Markus Ralser.

Data curation: Vadim Demichev, Pinkus Tober-Lau, Lukasz Szyrwiel, Lena J. Lippert, Elisa T. Helbig, Paula Stubbemann, Nadine Olk, Charlotte Thibeault, Matthew White, Christoph B. Messner, Michael Joannidis, Thomas Sonnweber, Sebastian J. Klein, Alex Pizzini, Yvonne Wohlfarter, Sabina Sahanic, Richard Hilbe, Benedikt Schaefer, Sonja Wagner, Felix Machleidt, Carmen Garcia, Christoph Ruwwe-Glösenkamp, Tilman Lingscheid, Laure Bosquillon de Jarcy, Miriam S. Stegemann, Moritz Pfeiffer, Linda Jürgens, Sophy Denker, Daniel Zickler, Claudia Spies, Philipp Enghard, Rosa Bellmann-Weiler, Günter Weiss, Alexander Uhrig, Heinz Zoller, Judith Löffler-Ragg, Markus A. Keller, Ivan Tancevski, Holger Müller-Redetzky, Martin Witzenrath, Norbert Suttorp, Michael Mülleder.

Formal analysis: Vadim Demichev, Pinkus Tober-Lau, Tatiana Nazarenko, Oliver Lemke, Simran Kaur Aulakh, Harry J. Whitwell, Annika Röhl, Mirja Mittermaier, Lukasz Szyrwiel, Charlotte Thibeault, Nana-Maria Grüning, Oleg Blyuss, Spyros Vernardis, Matthew White, Christoph B. Messner, Aleksej Zelezniak, Archie Campbell, Caroline Hayward, David J. Porteous, Riccardo E. Marioni, John F. Timms, Alexey Zaikin, Stefan Hippenstiel, Michael Ramharter, Kathryn Lilley, Michael Mülleder.

Investigation: Vadim Demichev, Pinkus Tober-Lau, Oliver Lemke, Harry J. Whitwell, Annika Röhl, Anja Freiwald, Mirja Mittermaier, Daniela Ludwig, Clara Correia-Melo, Lena J. Lippert, Elisa T. Helbig, Paula Stubbemann, Charlotte Thibeault, Christoph B. Messner, Tilman Lingscheid, Claudia Spies, Andreas Edel, Nils B. Müller, Michael Ramharter, Kathryn Lilley, Michael Mülleder, Leif Erik Sander, Florian Kurth, Markus Ralser.

Methodology: Leif Erik Sander, Florian Kurth, Markus Ralser.

Resources: Harry J. Whitwell, Oleg Blyuss, Archie Campbell, Caroline Hayward, Riccardo E. Marioni, Markus A. Keller, John F. Timms, Alexey Zaikin, Stefan Hippenstiel, Martin Witzenrath, Norbert Suttorp.

Supervision: Leif Erik Sander, Florian Kurth, Markus Ralser.

Validation: Vadim Demichev.

Visualization: Vadim Demichev, Pinkus Tober-Lau, Simran Kaur Aulakh, Michael Mülleder, Leif Erik Sander.

Writing – original draft: Vadim Demichev, Pinkus Tober-Lau, Leif Erik Sander, Florian Kurth, Markus Ralser.

Writing – review & editing: Vadim Demichev, Pinkus Tober-Lau, Oliver Lemke, Simran Kaur Aulakh, Harry J. Whitwell, Annika Röhl, Anja Freiwald, Mirja Mittermaier, Lukasz Szyrwiel, Daniela Ludwig, Clara Correia-Melo, Lena J. Lippert, Elisa T. Helbig, Paula Stubbemann, Nadine Olk, Charlotte Thibeault, Oleg Blyuss, Spyros Vernardis, Matthew White, Christoph B. Messner, Michael Joannidis, Thomas Sonnweber, Sebastian J. Klein, Alex Pizzini, Yvonne Wohlfarter, Sabina Sahanic, Richard Hilbe, Sonja Wagner, Felix Machleidt, Carmen Garcia, Christoph Ruwwe-Glösenkamp, Tilman Lingscheid, Laure Bosquillon de Jarcy, Miriam S. Stegemann, Moritz Pfeiffer, Linda Jürgens, Sophy Denker, Daniel Zickler, Claudia Spies, Andreas Edel, Nils B. Müller, Philipp Enghard, Aleksej Zelezniak, Rosa Bellmann-Weiler, Günter Weiss, Archie Campbell, Caroline Hayward, David J. Porteous, Riccardo E. Marioni, Alexander Uhrig, Heinz Zoller, Ivan Tancevski, John F. Timms, Alexey Zaikin, Stefan Hippenstiel, Michael Ramharter, Holger Müller-Redetzky, Martin Witzenrath, Norbert Suttorp, Kathryn Lilley, Michael Mülleder, Leif Erik Sander, Florian Kurth, Markus Ralser.

References

1. Frantzi M, Latosinska A, Mischak H. Proteomics in Drug Development: The Dawn of a New Era? Proteomics Clin Appl. 2019; 13: e1800087. https://doi.org/10.1002/prca.201800087 PMID: 30724014
2. Clift AK, Coupland CAC, Keogh RH, Diaz-Ordaz K, Williamson E, Harrison EM, et al. Living risk prediction algorithm (QCOVID) for risk of hospital admission and mortality from coronavirus 19 in adults: national derivation and validation cohort study. BMJ. 2020; 371: m3731. https://doi.org/10.1136/bmj.m3731 PMID: 33082154
3. Knight SR, Ho A, Pius R, Buchan I, Carson G, Drake TM, et al. Risk stratification of patients admitted to hospital with covid-19 using the ISARIC WHO Clinical Characterisation Protocol: development and validation of the 4C Mortality Score. BMJ. 2020; 370: m3339. https://doi.org/10.1136/bmj.m3339 PMID: 32907855
4. Chassagnon G, Vakalopoulou M, Battistella E, Christodoulidis S, Hoang-Thi T-N, Dangeard S, et al. AI-driven quantification, staging and outcome prediction of COVID-19 pneumonia. Med Image Anal. 2021; 67: 101860. https://doi.org/10.1016/j.media.2020.101860 PMID: 33171345
5. Wynants L, Van Calster B, Collins GS, Riley RD, Heinze G, Schuit E, et al. Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal. BMJ. 2020; 369: m1328. https://doi.org/10.1136/bmj.m1328 PMID: 32265220
6. Messner CB, Demichev V, Wendisch D, Michalick L, White M, Freiwald A, et al. Ultra-High-Throughput Clinical Proteomics Reveals Classifiers of COVID-19 Infection. Cell Systems. 2020. pp. 11–24.e4. https://doi.org/10.1016/j.cels.2020.05.012 PMID: 32619549
7. Shen B, Yi X, Sun Y, Bi X, Du J, Zhang C, et al. Proteomic and Metabolomic Characterization of COVID-19 Patient Sera. Cell. 2020; 182: 59–72.e15. https://doi.org/10.1016/j.cell.2020.05.032 PMID: 32492406
8. Laing AG, Lorenc A, Del Molino Del Barrio I, Das A, Fish M, Monin L, et al. A dynamic COVID-19 immune signature includes associations with poor prognosis. Nat Med. 2020; 26: 1623–1635. https://doi.org/10.1038/s41591-020-1038-6 PMID: 32807934

9. Liu Y, Gao W, Guo W, Guo Y, Shi M, Dong G, et al. Prominent coagulation disorder is closely related to inflammatory response and could be as a prognostic indicator for ICU patients with COVID-19. J Thromb Thrombolysis. 2020; 50: 825–832. https://doi.org/10.1007/s11239-020-02174-9 PMID: 32761495

10. D'Alessandro A, Thomas T, Dzieciatkowska M, Hill RC, Francis RO, Hudson KE, et al. Serum Proteomics in COVID-19 Patients: Altered Coagulation and Complement Status as a Function of IL-6 Level. J Proteome Res. 2020; 19: 4417–4427. https://doi.org/10.1021/acs.jproteome.0c00365 PMID: 32786691

11. Schulte-Schrepping J, Reusch N, Paclik D, Baßler K, Schlickeiser S, Zhang B, et al. Severe COVID-19 Is Marked by a Dysregulated Myeloid Cell Compartment. Cell. 2020; 182: 1419–1440.e23. https://doi.org/10.1016/j.cell.2020.08.001 PMID: 32810438

12. Overmyer KA, Shishkova E, Miller IJ, Balnis J, Bernstein MN, Peters-Clarke TM, et al. Large-Scale Multi-omic Analysis of COVID-19 Severity. Cell Syst. 2021; 12: 23–40.e7. https://doi.org/10.1016/j.cels.2020.10.003 PMID: 33096026

13. Shu T, Ning W, Wu D, Xu J, Han Q, Huang M, et al. Plasma Proteomics Identify Biomarkers and Pathogenesis of COVID-19. Immunity. 2020; 53: 1108–1122.e5. https://doi.org/10.1016/j.immuni.2020.10.008 PMID: 33128875

14. Demichev V, Tober-Lau P, Lemke O, Nazarenko T, Thibeault C, Whitwell H, et al. A time-resolved proteomic and prognostic map of COVID-19. Cell Syst. 2021; 12: 780–794.e7. https://doi.org/10.1016/j.cels.2021.05.005 PMID: 34139154

15. Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: a severity of disease classification system. Crit Care Med. 1985; 13: 818–829. PMID: 3928249

16. Ferreira FL. Serial Evaluation of the SOFA Score to Predict Outcome in Critically Ill Patients. JAMA. 2001. p. 1754. https://doi.org/10.1001/jama.286.14.1754 PMID: 11594901

17. Wang Z-H, Shu C, Ran X, Xie C-H, Zhang L. Critically ill patients with Coronavirus disease 2019 in a designated ICU: Clinical features and predictors for mortality. Risk Manag Healthc Policy. 2020; 13: 833–845. https://doi.org/10.2147/RMHP.S263095 PMID: 32765138

18. Zou X, Li S, Fang M, Hu M, Bian Y, Ling J, et al. Acute Physiology and Chronic Health Evaluation II Score as a Predictor of Hospital Mortality in Patients of Coronavirus Disease 2019. Crit Care Med. 2020; 48: e657–e665. https://doi.org/10.1097/CCM.0000000000004411 PMID: 32697506

19. Thibeault C, Mühlemann B, Helbig ET, Mittermaier M, Lingscheid T, Tober-Lau P, et al. Clinical and virological characteristics of hospitalised COVID-19 patients in a German tertiary care centre during the first wave of the SARS-CoV-2 pandemic: a prospective observational study. Infection. 2021. https://doi.org/10.1007/s15010-021-01594-w PMID: 33890243

20. Wang J, Li D, Dangott LJ, Wu G. Proteomics and its role in nutrition research. J Nutr. 2006; 136: 1759–1762. https://doi.org/10.1093/jn/136.7.1759 PMID: 16772433

21. Elhadad MA, Jonasson C, Huth C, Wilson R, Gieger C, Matias P, et al. Deciphering the plasma proteome of type 2 diabetes. Diabetes. 2020; 69: 2766–2778. https://doi.org/10.2337/db20-0296 PMID: 32928870

22. Hoogeveen RM, Pereira JPB, Nurmohamed NS, Zampoleri V, Bom MJ, Baragetti A, et al. Improved cardiovascular risk prediction using targeted plasma proteomics in primary prevention. Eur Heart J. 2020; 41: 3998–4007. https://doi.org/10.1093/eurheartj/ehaa648 PMID: 32808014

23. Suhre K, McCarthy MI, Schwenk JM. Genetics meets proteomics: perspectives for large population-based studies. Nat Rev Genet. 2021; 22: 19–37. https://doi.org/10.1038/s41576-020-0268-2 PMID: 32860016

24. Struwe W, Emmott E, Bailey M, Sharon M, Sinz A, Corrales FJ, et al. The COVID-19 MS Coalition-accelerating diagnostics, prognostics, and treatment. Lancet. 2020; 395: 1761–1762. https://doi.org/10.1016/S0140-6736(20)31211-3 PMID: 32473097

25. Geyer PE, Kulak NA, Pichler G, Holdt LM, Teupser D, Mann M. Plasma proteome profiling to assess human health and disease. Cell Syst. 2016; 2: 185–195. https://doi.org/10.1016/j.cels.2016.02.015 PMID: 27135364

26. Bruderer R, Muntel J, Müller S, Bernhardt OM, Gandhi T, Cominetti O, et al. Analysis of 1508 plasma samples by capillary-flow data-independent acquisition profiles proteomics of weight loss and maintenance. Mol Cell Proteomics. 2019; 18: 1242–1254. https://doi.org/10.1074/mcp.RA118.001288 PMID: 30948622

27. Ignjatovic V, Geyer PE, Palaniappan KK, Chaaban JE, Omenn GS, Baker MS, et al. Mass spectrometry-based plasma proteomics: Considerations from sample collection to achieving translational data. J Proteome Res. 2019; 18: 4085–4097. https://doi.org/10.1021/acs.jproteome.9b00503 PMID: 31573204

28. Messner CB, Demichev V, Bloomfield N, Yu JSL, White M, Kreidl M, et al. Ultra-fast proteomics with Scanning SWATH. Nat Biotechnol. 2021; 39: 846–854. https://doi.org/10.1038/s41587-021-00860-4 PMID: 33767396

29. Kurth F, Roennefarth M, Thibeault C, Corman VM, Müller-Redetzky H, Mittermaier M, et al. Studying the pathophysiology of coronavirus disease 2019: a protocol for the Berlin prospective COVID-19 patient cohort (Pa-COVID-19). Infection. 2020; 48: 619–626. https://doi.org/10.1007/s15010-020-01464-x PMID: 32535877

30. Quan H, Li B, Couris CM, Fushimi K, Graham P, Hider P, et al. Updating and validating the Charlson comorbidity index and score for risk adjustment in hospital discharge abstracts using data from 6 countries. Am J Epidemiol. 2011; 173: 676–682. https://doi.org/10.1093/aje/kwq433 PMID: 21330339

31. Varol Y, Hakoglu B, Kadri Cirak A, Polat G, Komurcuoglu B, Akkol B, et al. The impact of charlson comorbidity index on mortality from SARS-CoV-2 virus infection and A novel COVID-19 mortality index: CoLACD. Int J Clin Pract. 2020; e13858. https://doi.org/10.1111/ijcp.13858 PMID: 33237615

32. Bhowmick P, Mohammed Y, Borchers CH. MRMAssayDB: an integrated resource for validated targeted proteomics assays. Bioinformatics. 2018; 34: 3566–3571. https://doi.org/10.1093/bioinformatics/bty385 PMID: 29762640

33. Whitwell HJ, Blyuss O, Menon U, Timms JF, Zaikin A. Parenclitic networks for predicting ovarian cancer. Oncotarget. 2018; 9: 22717–22726. https://doi.org/10.18632/oncotarget.25216 PMID: 29854310

34. Krivonosov M, Nazarenko T, Bacalini MG, Franceschi C, Zaikin A, Ivanchenko M. DNA methylation changes with age as a complex system: a parenclitic network approach to a family-based cohort of patients with Down Syndrome. bioRxiv. bioRxiv; 2020. https://doi.org/10.1101/2020.03.10.986505

35. Danwang C, Endomba FT, Nkeck JR, Wouna DLA, Robert A, Noubiap JJ. A meta-analysis of potential biomarkers associated with severity of coronavirus disease 2019 (COVID-19). Biomark Res. 2020; 8: 37. https://doi.org/10.1186/s40364-020-00217-0 PMID: 32879731

36. Henry BM, de Oliveira MHS, Benoit S, Plebani M, Lippi G. Hematologic, biochemical and immune biomarker abnormalities associated with severe illness and mortality in coronavirus disease 2019 (COVID-19): a meta-analysis. Clin Chem Lab Med. 2020; 58: 1021–1028. https://doi.org/10.1515/cclm-2020-0369 PMID: 32286245

37. Huang C, Soleimani J, Herasevich S, Pinevich Y, Pennington KM, Dong Y, et al. Clinical Characteristics, Treatment, and Outcomes of Critically Ill Patients With COVID-19: A Scoping Review. Mayo Clin Proc. 2021; 96: 183–202. https://doi.org/10.1016/j.mayocp.2020.10.022 PMID: 33413817

38. Gupta S, Hayek SS, Wang W, Chan L, Mathews KS, Melamed ML, et al. Factors associated with death in critically ill patients with Coronavirus disease 2019 in the US. JAMA Intern Med. 2020; 180: 1436–1447. https://doi.org/10.1001/jamainternmed.2020.3596 PMID: 32667668

39. Völlmy F, van den Toorn H, Zenezini Chiozzi R, Zucchetti O, Papi A, Volta CA, et al. A serum proteome signature to predict mortality in severe COVID-19 patients. Life Sci Alliance. 2021; 4. https://doi.org/10.26508/lsa.202101099 PMID: 34226277

40. Thavarajah T, Dos Santos CC, Slutsky AS, Marshall JC, Bowden P, Romaschin A, et al. The plasma peptides of sepsis. Clin Proteomics. 2020; 17: 26. https://doi.org/10.1186/s12014-020-09288-5 PMID: 32636717

41. Rehman AA, Ahsan H, Khan FH. α-2-Macroglobulin: a physiological guardian. J Cell Physiol. 2013; 228: 1665–1675. https://doi.org/10.1002/jcp.24266 PMID: 23086799

42. Sharma NK, Tashima AK, Brunialti MKC, Ferreira ER, Torquato RJS, Mortara RA, et al. Proteomic study revealed cellular assembly and lipid metabolism dysregulation in sepsis secondary to community-acquired pneumonia. Sci Rep. 2017; 7: 15606. https://doi.org/10.1038/s41598-017-15755-1 PMID: 29142235

43. Risitano AM, Mastellos DC, Huber-Lang M, Yancopoulou D, Garlanda C, Ciceri F, et al. Complement as a target in COVID-19? Nat Rev Immunol. 2020; 20: 343–344. https://doi.org/10.1038/s41577-020-0320-7 PMID: 32327719

44. Abou-Ismail MY, Diamond A, Kapoor S, Arafah Y, Nayak L. The hypercoagulable state in COVID-19: Incidence, pathophysiology, and management. Thromb Res. 2020; 194: 101–115. https://doi.org/10.1016/j.thromres.2020.06.029 PMID: 32788101

45. Kashuba E, Bailey J, Allsup D, Cawkwell L. The kinin–kallikrein system: physiological roles, pathophysiology and its relationship to cancer biomarkers. Biomarkers. 2013; 18: 279–296. https://doi.org/10.3109/1354750X.2013.787544 PMID: 23672534

46. van de Veerdonk FL, Netea MG, van Deuren M, van der Meer JW, de Mast Q, Brüggemann RJ, et al. Kallikrein-kinin blockade in patients with COVID-19 to prevent acute respiratory distress syndrome. Elife. 2020;9. https://doi.org/10.7554/eLife.57555 PMID: 32338605

47. Colarusso C, Terlizzi M, Pinto A, Sorrentino R. A lesson from a saboteur: High-MW kininogen impact in coronavirus-induced disease 2019. Br J Pharmacol. 2020; 177: 4866–4872. https://doi.org/10.1111/bph.15154 PMID: 32497257

48. van de Veerdonk FL, Kouijzer IJE, de Nooijer AH, van der Hoeven HG, Maas C, Netea MG, et al. Outcomes associated with use of a kinin B2 receptor antagonist among patients with COVID-19. JAMA Netw Open. 2020; 3: e2017708. https://doi.org/10.1001/jamanetworkopen.2020.17708 PMID: 32789513

49. Chao J, Guo Y, Chao L. Protective role of endogenous kallistatin in vascular injury and senescence by inhibiting oxidative stress and inflammation. Oxid Med Cell Longev. 2018; 2018: 4138560. https://doi.org/10.1155/2018/4138560 PMID: 30622668

50. Lin W-C, Chen C-W, Huang Y-W, Chao L, Chao J, Lin Y-S, et al. Kallistatin protects against sepsis-related acute lung injury via inhibiting inflammation and apoptosis. Sci Rep. 2015; 5: 12463. https://doi.org/10.1038/srep12463 PMID: 26198099

51. NIH ACTIV Trial of blood thinners pauses enrollment of critically ill COVID-19 patients. 22 Dec 2020 [cited 29 Nov 2021]. Available: https://www.nih.gov/news-events/news-releases/nih-activ-trial-blood-thinners-pauses-enrollment-critically-ill-covid-19-patients

52. REMAP-CAP Investigators, ACTIV-4a Investigators, ATTACC Investigators, Goligher EC, Bradbury CA, McVerry BJ, et al. Therapeutic anticoagulation with heparin in critically ill patients with Covid-19. N Engl J Med. 2021; 385: 777–789. https://doi.org/10.1056/NEJMoa2103417 PMID: 34351722

53. Zhu X, Wang Y, Zhang H, Liu X, Chen T, Yang R, et al. Genetic variation of the human α-2-Heremans-Schmid glycoprotein (AHSG) gene associated with the risk of SARS-CoV infection. PLoS One. 2011; 6: e23730. https://doi.org/10.1371/journal.pone.0023730 PMID: 21904596

54. Ombrellino M, Wang H, Yang H, Zhang M, Vishnubhakat J, Frazier A, et al. Fetuin, a negative acute phase protein, attenuates TNF synthesis and the innate inflammatory response to carrageenan. Shock. 2001; 15: 181–185. https://doi.org/10.1097/00024382-200115030-00004 PMID: 11236900

55. Chua RL, Lukassen S, Trump S, Hennig BP, Wendisch D, Pott F, et al. COVID-19 severity correlates with airway epithelium-immune cell interactions identified by single-cell analysis. Nat Biotechnol. 2020; 38: 970–979. https://doi.org/10.1038/s41587-020-0602-4 PMID: 32591762

56. Shirato K, Kizaki T. SARS-CoV-2 spike protein S1 subunit induces pro-inflammatory responses via toll-like receptor 4 signaling in murine and human macrophages. Heliyon. 2021; 7: e06187. https://doi.org/10.1016/j.heliyon.2021.e06187 PMID: 33644468

57. Merad M, Martin JC. Pathological inflammation in patients with COVID-19: a key role for monocytes and macrophages. Nat Rev Immunol. 2020; 20: 355–362. https://doi.org/10.1038/s41577-020-0331-4 PMID: 32376901

58. R Core Team. R: A language and environment for statistical computing. https://www.R-project.org/. R Foundation for Statistical Computing, Vienna, Austria.; 2018.

59. Benjamini Y, Hochberg Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. J R Stat Soc. 1995; 57: 289–300.

60. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. J Stat Softw. 2010; 33: 1–22. PMID: 20808728

61. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine learning in Python. the Journal of machine Learning research. 2011; 12: 2825–2830.

62. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat Methods. 2020; 17: 261–272. https://doi.org/10.1038/s41592-019-0686-2 PMID: 32015543

63. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): the TRIPOD statement. Ann Intern Med. 2015; 162: 55–63. https://doi.org/10.7326/M14-0697 PMID: 25560714

Articles

A multiplex protein panel assay for severity prediction and outcome prognosis in patients with COVID-19: An observational multi-cohort study

Ziyue Wang,[a] Adam Cryar,[b] Oliver Lemke,[a] Pinkus Tober-Lau,[c] Daniela Ludwig,[a] Elisa Theresa Helbig,[c] Stefan Hippenstiel,[c] Leif-Erik Sander,[c,i] Daniel Blake,[d] Catherine S. Lane,[d] Rebekah L. Sayers,[d] Christoph Mueller,[e] Johannes Zeiser,[e] StJohn Townsend,[a] Vadim Demichev,[a] Michael Mülleder,[f] Florian Kurth,[c,g]* Ernestas Sirka,[b,#]** Johannes Hartl,[a,#]*** and Markus Ralser[a,h,#]****

[a]Department of Biochemistry, Charité – Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Am Chariteplatz 1, 10117 Berlin, Germany
[b]Inoviv, Mappin House, 4 Winsley St, London, United Kingdom
[c]Department of Infectious Diseases and Respiratory Medicine, Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Augustenburger Platz 1, 13353 Berlin, Germany
[d]SCIEX, Macclesfield, United Kingdom
[e]Agilent Technologies Sales & Services GmbH & Co. KG, Waldbronn, Germany
[f]Core Facility – High-Throughput Mass Spectrometry, Charité – Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Am Chariteplatz 1, 10117 Berlin, Germany
[g]Department of Tropical Medicine, Bernhard Nocht Institute for Tropical Medicine, and Department of Medicine I, University Medical Centre Hamburg-Eppendorf, Hamburg, Germany
[h]The Molecular Biology of Metabolism Laboratory, The Francis Crick Institute, London, UK
[i]Berlin Institute of Health at the Charité - Universitätsmedizin Berlin, Berlin, Germany

Summary

Background Global healthcare systems continue to be challenged by the COVID-19 pandemic, and there is a need for clinical assays that can help optimise resource allocation, support treatment decisions, and accelerate the development and evaluation of new therapies.

Methods We developed a multiplexed proteomics assay for determining disease severity and prognosis in COVID-19. The assay quantifies up to 50 peptides, derived from 30 known and newly introduced COVID-19-related protein markers, in a single measurement using routine-lab compatible analytical flow rate liquid chromatography and multiple reaction monitoring (LC-MRM). We conducted two observational studies in patients with COVID-19 hospitalised at Charité – Universitätsmedizin Berlin, Germany before (from March 1 to 26, 2020, n=30) and after (from April 4 to November 19, 2020, n=164) dexamethasone became standard of care. The study is registered in the German and the WHO International Clinical Trials Registry (DRKS00021688).

Findings The assay produces reproducible (median inter-batch CV of 10.9%) absolute quantification of 47 peptides with high sensitivity (median LLOQ of 143 ng/ml) and accuracy (median 96.8%). In both studies, the assay reproducibly captured hallmarks of COVID-19 infection and severity, as it distinguished healthy individuals, mild, moderate, and severe COVID-19. In the post-dexamethasone cohort, the assay predicted survival with an accuracy of 0.83 (108/130), and death with an accuracy of 0.76 (26/34) in the median 2.5 weeks before the outcome, thereby outperforming compound clinical risk assessments such as SOFA, APACHE II, and ABCS scores.

Interpretation Disease severity and clinical outcomes of patients with COVID-19 can be stratified and predicted by the routine-applicable panel assay that combines known and novel COVID-19 biomarkers. The prognostic value of

eClinicalMedicine 2022;49: 101495
Published online 9 June 2022
https://doi.org/10.1016/j.eclinm.2022.101495

*Corresponding author: Florian Kurth, Charité - Universitätsmedizin Berlin, Department of Infectious Diseases and Respiratory Medicine, Augustenburger Platz 1, 13353 Berlin, Germany. Tel.: +49 (0)30 450 553052.
**Corresponding author: Ernestas Sirka, Inoviv, Mappin House, 4 Winsley St, London W1W 8HF, United Kingdom, Tel.: +44 (0) 20 3239 0178.
***Corresponding author: Johannes Hartl, Charité – Universitätsmedizin Berlin, Department of Biochemistry, Charitéplatz 1, 10117 Berlin, Germany. Tel.: +49 (0)30 450 528317.
****Corresponding author: Markus Ralser, Charité – Universitätsmedizin Berlin, Department of Biochemistry, Charitéplatz 1, 10117 Berlin, Germany. Tel.: +49 (0)30 450 528141
E-mail addresses: florian.kurth@charite.de (F. Kurth), e.sirka@inoviv.com (E. Sirka), johannes.hartl@charite.de (J. Hartl), markus.ralser@charite.de (M. Ralser).
[#] These authors contributed equally to this work.

Articles

this assay should be prospectively assessed in larger patient cohorts for future support of clinical decisions, including evaluation of sample flow in routine setting. The possibility to objectively classify COVID-19 severity can be helpful for monitoring of novel therapies, especially in early clinical trials.

Funding This research was funded in part by the European Research Council (ERC) under grant agreement ERC-SyG-2020 951475 (to M.R) and by the Wellcome Trust (IA 200829/Z/16/Z to M.R.). The work was further supported by the Ministry of Education and Research (BMBF) as part of the National Research Node 'Mass Spectrometry in Systems Medicine (MSCoresys)', under grant agreements 031L0220 and 161L0221. J.H. was supported by a Swiss National Science Foundation (SNSF) Postdoc Mobility fellowship (project number 191052). This study was further supported by the BMBF grant NaFoUniMedCOVID-19 – NUM-NAPKON, FKZ: 01KX2021. The study was co-funded by the UK's innovation agency, Innovate UK, under project numbers 75594 and 56328.

Keywords: COVID-19; SARS-CoV2; Biomarker; Clinical disease progression; Severity stratification; Disease prognosis; Machine learning; Targeted proteomics; LC-MS/MS

Research in context

Evidence before this study

We searched PubMed for articles published up to May 12, 2022. We used the search terms COVID-19 or SARS-CoV-2, and severity or outcome, and prognosis or prediction, and proteomics or protein panel or peptide panel, and plasma or serum, and mass spectrometry. The search returned 10 research articles, which mostly used explorative proteomics to identify putative protein markers associated with COVID-19 disease severity or outcome. However, no study translated identified biomarkers into a panel assay providing absolute quantification which can be deployed as targeted mass spectrometry platforms available to routine diagnostic laboratories.

Added value of this study

Proteomic panel assays hold the promise to outperform established intensive care unit outcome predictors such as APACHE II or SOFA in COVID-19, but so far their application in clinical routine is challenging for technical reasons. We select a panel of 50 peptides, derived from 30 proteins, whose functions have been associated with COVID-19 using discovery proteomics, and develop and analytically validate a scalable proteomic panel assay that is performed on instrumentation common in clinical laboratories. Applying the assay to two independent cohorts, we demonstrate accurate disease classification, and show that the marker panel is prognostic about outcome.

Implications of all the available evidence

The potential value of using the human plasma proteome in severity classification, risk assessment, and outcome prediction in COVID-19 has recently been uncovered in several studies. What was missing so far was translation of this research into a routine applicable assay. We present a protein marker panel which predicts survival in COVID-19 with high accuracy that can be implemented for routine laboratory testing. The described assay has the potential to improve clinical risk assessment for patients with COVID-19 by translating discovery proteomics findings to patient care.

Introduction

COVID-19 challenges healthcare systems worldwide, which is particularly apparent in areas with limited vaccine uptake. The outlook remains uncertain even in countries with high vaccination rates as the immunity conferred by the vaccines appears to diminish over time.[1–5] Moreover, SARS-COV-2 variants with capacity to evade immunity continue to emerge,[1,6–8] and may affect global medical care rapidly and unpredictably.

Biomarker tests that classify disease severity and are prognostic could help mitigate the impact of critical treatment choice by allowing to optimise resource allocation.[9–11] Indeed, clinical manifestation of COVID-19 is highly variable. For instance, 'happy hypoxia' describes situations where patients with COVID-19 present in a relatively well compensated clinical status, while molecular indicators indicate they are, in fact, severely ill.[12] Furthermore, in situations when healthcare systems reach maximum capacity, prognostic tests could support difficult clinical decisions, for instance to identify individuals that require maximum available support, irrespective of age and comorbidities.[13] Prognostic and severity-classifying tests could further help increase the likelihood of success and accelerate clinical trials by improving treatment efficacy assessments of COVID-19 therapies or stratifying patient populations for inclusion into the trials. They might also help detect clinically yet inapparent side effects and contribute to

patient safety. Moreover, during a pandemic, there is a need to conduct trials in a timely manner, and often in cohorts of limited size. When underpowered, clinical trials can lead to false positive and false negative assessments of a drug's efficacy.[14,15] Disease-severity and prognostic tests could hence help extrapolate more and patient-specific information. Unfortunately, the reliability of several risk-assessment scores conventionally used in ICU settings such as the Acute Physiology And Chronic Health Evaluation (APACHE II), Charlson Comorbidity Index (CCI), and Sequential Organ Failure Assessment (SOFA) scores appears to be limited in COVID-19.[16] Combinations of generic clinical readouts, e.g. blood oxygen saturation and interleukin-6 concentration, have been considered for outcome prediction at various disease-severity stages.[17] However, several predictive models that were reported early in the pandemic are now considered to be vulnerable to bias, can be obsolete due to therapeutic measures such as CRP-based models after anti-IL-6 treatment, and might not be suitable for the clinic.[18,19] Therefore, additional COVID-19-specific compound scores have been proposed recently (e.g. ABCS[20]), showing improved performance. A remaining limitation of compound scores is their reliance on various measurements of different nature, which makes them statistically challenging.

Proteomic datasets have repeatedly been successful at classifying and predicting COVID-19 severity and outcome,[9,10,16,21–23] and can quantify many proteins in parallel, from one sample and one measurement. Indeed, specifically in severe COVID-19 cases, proteomic predictors have outperformed APACHE II, CCI, and SOFA scores.[11,13,24,25] Proteomics also accelerated the characterisation of the antiviral host response, which improved our understanding of the COVID-19 disease by attributing the complement cascade, coagulation system, and apoprotein function to differences in COVID-19 pathology.[9,10,16,21–23,25] The application of discovery proteomics as a test for the clinical routine is, however, limited for technical, economic, and regulatory reasons.

The objective of this study was to develop a COVID-19 biomarker panel assay which runs on broadly available analytical instruments that could be deployed for clinical use within existing regulatory frameworks. Triple quadrupole mass spectrometers coupled to high-flow liquid chromatography are used in the clinic in other areas[26–29] and are widely available in large hospital laboratories, diagnostic laboratories, regulated (e.g. CLIA) laboratories, and contract research organisations. Biomarker tests developed on this platform can be accredited to existing regulatory standards in GCP, ISO:17025, ISO:15189, and CLIA environments, standardised and transferred across different instruments and laboratories, and thus deployed at scale rapidly. Triple-quadrupole-mass-spectrometry-based tests are cost effective to run at scale as sample preparation can be automated, consumables costs for the MS runs are typically <£10 per test, and the instrument uptime is typically >95%.

In order to establish a proteomic panel assay using analytical flow rate chromatography and multiple reaction monitoring on triple quadrupole instruments, we have mined discovery proteomics data from patients with COVID-19 and selected biomarkers that are informative about COVID-19 disease progression. The biomarkers were chosen for i) being prognostic of remaining duration of hospitalisation, disease aggravation, or being differentially concentrated in plasma depending on the treatment escalation level, used as a measure of disease severity, and ii) participating in biological processes that contribute to COVID-19 pathology, and iii) being technically and analytically suitable for the assay. Employing calibration curves with synthetic reference and stable-isotope-labelled (SIL) internal standards, the assay aims for the absolute quantification of up to 50 surrogate tryptic peptides corresponding to 30 plasma proteins. These function in inflammation (e.g. C-reactive protein), coagulation and vascular dysfunction (e.g. von Willebrand factor), complement cascade (e.g. Complement C1q subcomponent subunit C), and other biological processes altered by COVID-19 (e.g. Cystatin C).

We analytically validated the assay and implemented it in two analytical laboratories employing two different triple quadrupole LC-MS/MS platforms. The assay was then applied to two observational cohorts. We demonstrate that the assay captures host response to SARS-CoV-2 and thereby classifies and predicts COVID-19 disease severity. In one cohort, we tested the prognostic value of the panel. We found that the biomarker panel is predictive about survival weeks before outcome, and outperforms several commonly used risk-assessment scores.

Methods

Study design and participants

Patient samples were collected as part of an observational cohort.[13,24,25] The study protocol, patient characteristics, treatment and outcomes were described previously.[30–32] Briefly, all in-patients with PCR-confirmed SARS-CoV-2 infection treated at Charité – Universitätsmedizin Berlin, a tertiary care centre, were eligible for inclusion, regardless of age, gender, or disease severity, after written informed consent was obtained. We also included severely ill patients on invasive ventilation based on a deferred consent procedure in order to avoid bias towards mildly ill patients. The study size is based on availability of patients and feasibility, and all study patients were included in the analyses. An overview of the study design, cohorts, sampling, methodology and data analysis is provided in Supplementary Figure 1. Study date cutoffs were 1st to 26th of March 2020 (Cohort 2), and 4th April to 19th November 2020 (Cohort 3). The study is registered in the German and the WHO international registry for clinical studies (DRKS00021688). The study was approved by

Articles

N	164
Sex	
female, n (%)	39 (23.8)
male, n (%)	125 (76.2)
age, median [IQR]	60 [51-69]
body mass index, median [IQR]	29.4 [24.7-32.5]
maximum severity	
WHO3, n (%)	23 (14.0)
WHO4, n (%)	42 (25.6)
WHO5, n (%)	34 (20.7)
WHO6, n (%)	3 (1.8)
WHO7, n (%)	28 (17.1)
WHO8 (deceased), n (%)	34 (20.7)
treatment	
Dexamethasone, n (%)	112 (68.3)
Remdesivir	15 (9.1)
IMV, n (%)	61 (37.2)
ECMO, n (%)	33 (20.1)
RRT, n (%)	24 (14.6)
days hospitalized	16 [10-34]
days until discharge (w/o deceased)	12 [8-29]
days until death	32 [19-47]
sampling	
days since symptom onset, median [IQR]	13 [8-17]
days to outcome, median [IQR]	10 [5-24]

***Table 1*: Description of study cohort 3.**

the ethics committee of Charité - Universitatsmedizin Berlin (EA2/066/20). The cohorts are summarised in Table 1 and Supplementary Table 1.

Reagents and peptide standards

Reference peptide standards were custom synthesised where native peptides were obtained at ≥95% purity and stable isotope-labelled (SIL) internal standard peptides (ISTDs) - at ≥70% purity. Internal standards contained 4-6 amino acid tryptic tags mimicking the sequence in a corresponding human plasma protein and were labelled on C-terminal lysine (K) or arginine (R) with stable isotopes (K (U-$^{13}C_6$,$^{15}N_2$) or R(U-$^{13}C_6$,$^{15}N_4$)). All peptide stock solutions were prepared at 1 mg/ml in 50:50 v/v ddH_2O: acetonitrile mix, except for STDYGIFQINSR and VEGTAFVIFGIQDGEQR where 200 µl of DMSO were added to solubilise the peptides at 5 mg/ml which were then aliquoted and diluted to 1 mg/ml with 50:50 v/v ddH_2O: acetonitrile mix. Internal standard mix was prepared by pooling 20 µl of each SIL peptide, evaporating 200 µl of this mix to dryness and reconstituting in a denaturation buffer to the final concentration of 1.4 µg/ml for each peptide. Cassetted calibration curves were prepared by serial dilution of pooled native reference peptide standards as described below. After serial dilution, these samples were treated identically to respective clinical samples. Additional reagents employed are listed in the Supplementary Methods.

Sample preparation

Samples were prepared with minor modifications as described previously.[2,4] Briefly, samples were stored at -80°C for 11-12 months prior to preparation, and clinical samples and calibration lines were prepared as follows: 5 µl of citrate plasma were added to 55 µl of denaturation buffer, composed of 50 µl 8 M Urea, 100 mM ammonium bicarbonate, 5 µl 50 mM dithiothreitol (DTT) and internal standard mix. The samples were incubated for 1 h at room temperature (RT) before addition of 5 µl of 100 mM iodoacetamide (IAA). After a 30 min incubation at RT the samples were diluted with 340 µl of 100 mM ammonium bicarbonate and digested overnight with 22.5 µl of 0.1 µg/µl trypsin at 37 °C. The digestion was quenched by adding 50 µl of 10% v/v formic acid. The resulting tryptic peptides were purified on a 96-well C18-based solid phase extraction (SPE) plate (BioPureSPE Macro 96-well, 100mg PROTO C18, The Nest Group). The purified samples were resuspended in 120 µl of 0.1% formic acid and 20 µl or 0.2 µl were injected into two LC-MS/MS platforms (Agilent 6495C and SCIEX 7500 respectively).

Samples in Cohort 3 were prepared as described above, with the following modifications. Samples were stored at -80°C for 5-11 months prior to preparation EDTA plasma was used instead of citrate plasma, and internal standards were digested separately and added to pre-digested clinical and calibration samples before their injection into the LC-MS/MS system. Quality control (QC) samples consisted of pooled commercial control and COVID-19 human plasma (as described in a previous publication[2,4]), and were prepared alongside clinical and calibration curve samples in each cohort.

The COVID-19 sample pools used for the analytical validation were generated by pooling 5 µl of patient plasma from Cohort 3 according to their WHO treatment severity score. Only samples of patients that had not received dexamethasone at time of sampling were used.

Liquid chromatography - tandem mass spectrometry

Tryptic peptides were quantified on two LC-MRM platforms. All samples were analysed on the Agilent 6495C mass spectrometer, coupled to an Agilent 1290 Infinity II UHPLC system. Samples from Cohort 2 were additionally analysed on a SCIEX 7500 mass spectrometer coupled to an ExionLC AD UHPLC system (SCIEX, UK). Details on chromatography and mass spectrometry settings are described in the Supplementary Methods. MRM parameters are provided in the Supplementary Table 2 (Agilent 6495C) and Supplementary Table 3 (SCIEX 7500).

Establishment of the MRM based assay

The assay was first set up on the 6495C (Agilent) system. Preliminary transitions for the 50 selected peptides (consisting of several precursor ion charge states and respective product ions) were predicted by Skyline

v21.1.0.146.[33] The native peptide standard solution was then infused into LC-MS/MS system and 1 precursor ion per peptide with the highest relative intensity and 5 most abundant product ions were selected for collision energy optimisation using Skyline. From these 5 product ions, 2-5 experimentally optimised ion transitions per native peptide were ultimately selected for the panel based on the following criteria: i) highest relative signal intensity, ii) optimal chromatographic peak shape and iii) absence of interfering signals. Product ions of <300 m/z were excluded where possible to ensure specificity. Precursor and product ion-matched ISTD transitions were also included. Lastly, all selected transitions were combined into one scheduled MRM method, where the most abundant transition for each peptide was used for quantification, and 1-4 remaining transitions - for qualification. (Supplementary Table 2). For analytical cross-platform and cross-laboratory validation, the assay was set up on the 7500 (SCIEX) system in parallel following this approach (Supplementary Table 3).

Mass spectrometry data processing and calibration

LC-MRM data was processed using MassHunter Quantitative Analysis, v10.1 (Agilent platform) or SCIEX OS v2.0.1 (2020, Sciex platform). Peptide absolute concentration (expressed in ng/ml) was determined from calibration curves, constructed with native and SIL peptide standards, and manually validated. Linear regression analysis of each calibration curve was performed using custom R code or SCIEX OS (with 1/x weighting). The transitions used for quantification are shown in Supplementary Tables 2 and 3 respectively. Matching of native peptides and internal standards is detailed in the Supplementary Methods.

Analytical method validation

Method analytical validation was performed based on FDA Bioanalytical Method Validation criteria[34] where sensitivity, specificity, intra- and inter-batch precision, accuracy and matrix effects have been assessed. Five (5) independent calibration curves were prepared by serial dilution of native peptide standards in assay buffer (1), surrogate matrix (3) and pooled human plasma (1) across the final peptide concentration range of 0 - 146.7 µg/ml. Surrogate matrix (40 mg/ml bovine serum albumin (BSA)) calibration curves were prepared and analysed across 3 separate batches and all calibration curve samples were analysed in quintuplets. Linear 1/x weighted regression was used to test the linearity of the response of all calibration lines. To determine the intra- and inter-batch precision, the CV was calculated from the response ratios (native peptide peak area divided by ISTD area).

Lower limit of quantification (LLOQ) was defined as the lowest concentration point on the linear calibration curve where the inter-batch CV was ≤ 20%. Since analytical validation requirements for clinical assays are purpose and context dependent, and are influenced by the magnitude of change of target analyte levels in control versus disease samples, LLOQ CV cutoff was subsequently expanded to ≤40% for the remaining peptides. Upper limit of quantification (ULOQ) was defined as the highest calibration sample on the linear curve with a CV ≤ 20%.

Accuracy was assessed by treating 1 of the 5 replicates in each calibration curve in the surrogate matrix as pseudo-unknown samples, quantifying with the curve generated from the remaining 4 replicates, and then calculating a median accuracy of all replicates. Matrix effects were measured by comparing the slopes of calibration curve samples prepared in a BSA matrix and pooled human plasma. Here an Extra Sum of Square F test was used for statistical comparison with a p-value < 0.05 indicating potential matrix effects.

Statistical analysis

Details of statistical tests performed in this study are available in the Supplementary Methods; test results are provided in Supplementary Tables 4-6. (Adjusted) P values were considered significant when $P < 0.05$. In brief, significance testing of the trend between absolute peptide concentrations and the ordinal classification as provided by the WHO treatment escalation scale was performed using Kendall's tau (KT) statistics and where indicated, with multiple testing correction. Cross-laboratory/cross-instrument performance was evaluated by Pearson correlation coefficients. Statistical tests on shotgun plasma proteomics data were performed as described.[13]

Prediction of WHO grade and disease outcome

Clinical scores were extracted from the clinical information system or, where missing, manually calculated. CCI and APACHE II were determined at time of admission, SOFA (ICU patients only) at time of sampling, and ABCS at both admission and sampling. Note that due to imputation of the ABCS score memory leakage between training and test data for the ABCS score models can not be excluded for this particular comparison.

For the WHO grade and the outcome prediction a Support Vector Machine with rbf-kernel was constructed on the first sample measured for every patient (n = 164). The model was trained and validated using a shuffled stratified 10-fold cross-validation to avoid data leakage between training and validation data. For models trained on established risk assessments scores, only samples for which the respective score was determined were included in model construction and testing.

In addition, predictors based on logistic regression and the extra-trees algorithm were evaluated as well. Feature importances were extracted from a model trained on all data (n=164) without splitting the data

set. Detailed information on model construction, evaluation and metric calculations are reported in Supplementary methods.

Recommendations and guidelines

Mischak et al.[35] have described a set of practical recommendations for biomarker discovery in clinical proteomics. We have provided an assessment of this study with respect to therein stated reporting recommendations in Supplementary Table 7. This study further follows the *Strengthening the Reporting of Observational Studies in Epidemiology* (STROBE) reporting guideline for observational studies.[36]

Role of the funding source

The funding sources had no involvement in study design, data collection, or the manuscript. All authors reviewed the manuscript and had access to the data generated in the study. FK, ES, JH, and MR were responsible for the decision to submit the paper for publication.

Results

Peptide selection

50 peptides that corresponded to 30 plasma proteins (Supplementary Table 8) were selected from shotgun plasma proteomics data recorded on a deeply phenotyped cohort of patients with COVID-19 (Figure 1, PA-COVID-19 study cohort, N=139 inpatients, for which 687 plasma proteomes were measured in time series).[13,30] Target biomarkers were identified in a ranking exercise that focused on plasma proteins that are (i) prognostic for the remaining time in hospital for inpatients as a treatment-insensitive proxy for COVID-19 severity; ii) differ between treatment escalation levels, expressed on the WHO ordinal scale[37]; or (iii) are prognostic of future worsening, i.e. the progression to a higher WHO severity grade[13] of inpatients who were admitted with a milder disease which then deteriorated.[13,16,21,24,38] Part of the selected proteins are already monitored clinically in this or other indications, including SERPINC1 = Antithrombin-III,[13,39] C3 = Complement C3,[13] APOB = Apolipoprotein B,[13] SERPING1 = C1-inhibitor,[13,38] CST3 = cystatin-C,[13] VWF = von Willebrand factor,[9,13,38] CRP = C-reactive protein,[9,13,24,38] PLG = plasminogen,[9,13] KLKB1 = plasma kallikrein,[9,13] LYZ = lysozyme,[13] and APOA1 = Apolipoprotein A[13,24] (Figure 2a). For new markers, a selection criterion was that they were deposited in MRMAssayDB, as an indicator of being chosen as markers also in other settings.[40] Further, we evaluated technical parameters such as suitability for synthesis (as evaluated by the Thermo Peptide Analysing Tool), a mass range applicable for a MRM assay, and distribution over the chromatographic gradient (Figure 2b).

From the final selected panel, 18/30 proteins are associated with remaining time in hospital, 22/30 with disease severity, and 6/30 are prognostic of future worsening (Figure 2c). Six additional peptides included were prognostic for remaining time in hospital (PRG4, C3, EFEMP1, ORM2, FCGR3A, AFM, IGHVs). For data and statistics, see Supplementary Figures 2–4.

Lastly, we tested whether the selected surrogate peptides were unique in a Uniprot BLAST and manual human proteome FASTA text file search. This was true for 45/50 peptides. The remaining 5 peptides (CQSWSSMTPHR, EITALAPSTMK, WEMPFDPQDTHQSR, DSGSYFCR, ASDTAMYYCAR) were shared across closely related protein isoforms (Supplementary Table 8) but were retained in the panel composition as they fulfilled the selection criteria.

Establishment and analytical validation of a MRM-based, targeted COVID-19 biomarker assay

For each selected peptide, 2 standards were synthesised: 1) with a natural isotope distribution ('native'), and 2) with a C-terminal SIL amino acid to act as an ISTD, which contained a short tryptic tag to account for the digestion efficiency (Supplementary Table 8). In order to establish the assay for routine settings, we chose analytical flow rate reversed-phase chromatography. The native peptides were employed to optimise LC-MS/MS data acquisition method and quality of the Q1/Q3 (MRM) transitions (257 overall) on a 6495C (Agilent) system. The eluted peptides were well distributed along a 8.6-minute linear gradient and were quantified using a scheduled MRM method (Figure 2b).

We then tested intra- and inter-batch precision, linearity, LOQ, accuracy, and potential matrix effects. We calculated the coefficient of variation (CV) for the triplicate of independently prepared calibration curves. These were constructed from serial dilutions of native peptide standards in BSA (40 mg/ml), measured in technical pentuplicates (i.e. total of N=15) on the LC-MS/MS system. We used BSA as a surrogate matrix to test the analytical performance in the absence of the endogenous plasma peptides[41] and achieved a median intra-batch CV of 2.6% and median inter-batch CV of 10.9% across low (LLOQ), medium ((LLOQ+ULOQ)/2), and high (ULOQ) concentration points (Supplementary Table 9).

Additionally, we determined the limits of quantification (LOQ). 37 peptides exceeded the inter-batch CV criteria at LLOQ of ≤ 20%, and 10 additional peptides could be quantified with an expanded LLOQ CV cutoff of ≤40%. Calibration curves for 47 peptides revealed a median LLOQ of 143.26 ng/ml, and typically allowed quantification over 3–4 orders of magnitude on a linear dynamic range ($R^2 > 0.99$, Supplementary Table 9).

As analytical technologies are sensitive to matrix effects,[42] we evaluated parallelism in the surrogate BSA matrix compared to human plasma. We compared the

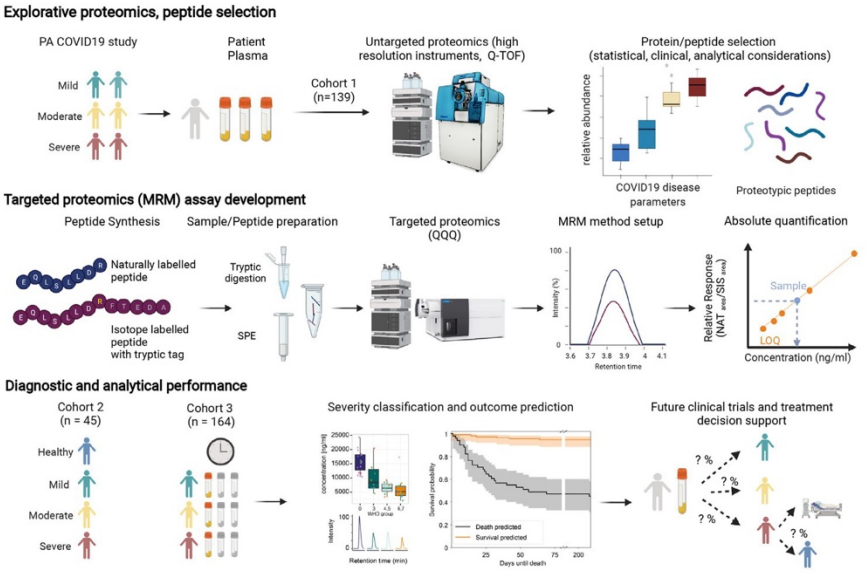


Figure 1. Schematic overview from peptide selection, to development and application of a COVID-19 biomarker panel. Top panel: Selection of 50 peptides derived from 30 plasma proteins as measured by discovery proteomics in a research setting[13,24] in the PA-COVID19 study cohort. Included proteins/peptides were selected accounting for their performance in the exploratory cohort, clinical, and analytical parameters. Middle panel: To generate an assay suitable for routine clinical laboratories, we established a targeted LC-MRM assay for conventional triple-quadrupole mass spectrometers, running reversed phase chromatography, at a high flow-rate. The assay was optimised using synthetic peptides, and allows for absolute quantification using stable isotope labelled internal standards ('AQUA peptides'[48]) with a short tryptic tag, to account for the sample preparation (tryptic digest) efficiency. Bottom panel: The assay was applied to two observational cohorts: a well balanced second ('1st wave') COVID-19 cohort,[25] with samples being measured in two laboratories, and a larger longitudinal cohort ('2nd wave'), with patients treated at the Charité Hospital, a national medical reference centre.

slopes obtained from calibration samples measured in a commercial human plasma sample with those measured in the surrogate matrix. 39/47 quantified peptide biomarkers showed no statistically significant matrix effect ($P > 0.05$). For the 8 peptides that differed significantly a matrix factor (slope plasma/slope BSA x 100%) was calculated and reported (Supplementary Table 9).

To test if peptide quantities from actual patient samples would be covered within the linear range of the calibration curves, we performed absolute quantification in plasma samples obtained from patients with COVID-19 (pooled COVID-19-patient samples of different WHO treatment escalation grades; see Methods). Peptide concentrations were covered within the determined linear range of the assay of pooled samples from WHO severity grade 3-7 (Supplementary Table 10).

The assay reports disease severity in an early pandemic cohort

Next, we assessed how the absolute concentration of the quantified biomarkers changed as a function of the COVID-19 treatment escalation level, a proxy for disease severity.[37] We applied the panel assay on plasma samples obtained from a deeply characterised, early-pandemic COVID-19 cohort, hospitalised between March 1 and 26, 2020 ('Cohort 2', n=45, Supplementary Table 1).[25] This cohort was suited for this validation step, as it was balanced, with patients with mild to severe COVID-19, and included healthy controls.[25,30] Furthermore, the cohort was sampled as citrate plasma, in difference to the exploratory cohorts in our previous studies to identify the biomarkers which were sampled as EDTA plasma. This test was hence also indicative if the biomarkers would allow stable conclusions across alternative sample matrices.

Samples were prepared using a semi-automated platform designed for precision.[24] 40/50 peptides were reliably quantified in the patient citrate plasma (Supplementary Figure 5, Supplementary Data 1). 32 peptides changed with disease severity, i.e. from uninfected (WHO 0) to mildly (WHO 3), moderately (WHO 4, 5) and severely (WHO 6, 7) affected patients (Figure 3a, Supplementary Figure 5, $P < 0.05$). Most of

Articles

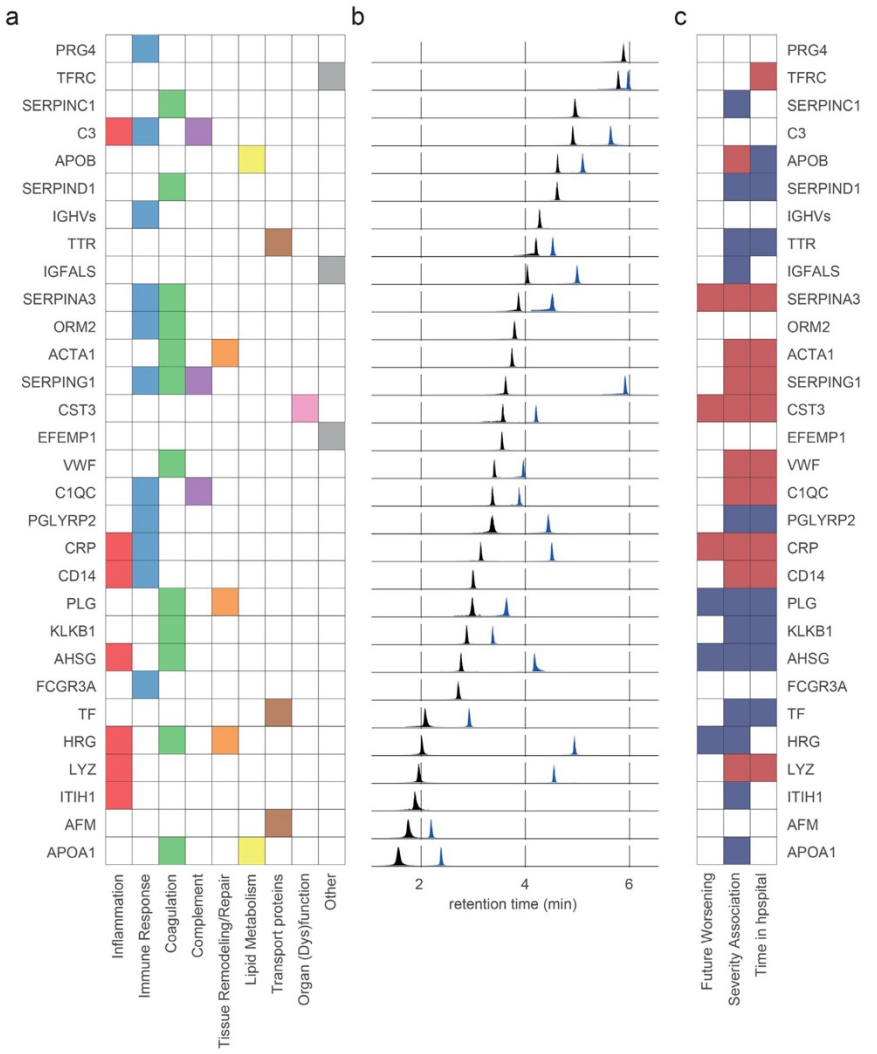


Figure 2. Selected peptides/proteins analysed by liquid chromatography, multiple reaction monitoring. a) Analysed proteins and their associated COVID-19-pathology-related processes as curated from literature. **b)** Extracted ion chromatograms (EIC) highlighting the chromatographic spread of the applied MRM transitions selected as quantifiers for the indicated marker peptides. Each EIC was normalised to the maximum intensity of the respective peptide. Note that the majority of proteins are captured by two peptides (arbitrarily coloured as blue, black). **c)** Proteins selected for the panel assay are associated with COVID-19 disease parameters, including severity and progression, and link to COVID-19 related processes. Coloured tiles indicate significant associations, with red/blue highlighting that the respective protein is up- or down-regulated in COVID-19 infected individuals in discovery proteomic data of cohort 1.[13] For data on individual peptides, see Supplementary Figures 2-4.

the chosen markers change in abundance between healthy and COVID-19-infected individuals, and further follow their respective trend with increasing treatment escalation level, such as peptides derived from the acute-phase proteins CRP and AHSG, or the innate-immune-response protein PGLYRP2 (Figure 3a, Figure 3b). Some of the peptides give a signal during specific disease state transitions, i.e. they differ between an infected and uninfected individual (such as peptides from the complement-related protein SERPING1 or the iron-binding protein TF (Figure 3a, Figure 3b)), or change the most during the most severe treatment escalations of COVID-19 (such as the kidney and inflammation marker CST3 (Figure 3a, Figure 3b)). As a profile, the protein marker quantities did classify the patients according to the treatment escalation score (Figure 3c).

Analytical cross-platform and cross-laboratory validation

The application of different instruments and sample matrices can lead to differences in the quantification of peptides. To evaluate the assay transferability, samples from Cohort 2 were measured on both the 6495C (Agilent) and the 7500 (SCIEX) LC-MS/MS platforms, in different laboratories. For 33/40 selected peptides, we obtained a clear cross-laboratory/cross-instrument correlation between the concentration measured in respective COVID-19-patient samples (Figure 3d, Supplementary Figure 6). One peptide (ESDTSYVSLK) suffered from one outlier, but otherwise had a good correlation on both platforms (Supplementary Figure 6). The remaining (six) peptides were close to the detection limit in the citrate plasma sample matrix, which caused limited correlation ($R^2 < 0.6$) between both platforms. Further, on a subset of peptides we observed correlation but different absolute values, pointing to differences in calibration.

Severity stratification and outcome prognosis in a longitudinal COVID-19 cohort

We next studied a larger, longitudinal cohort of the second wave of the pandemic, hospitalised between April 4 and November 19, 2020, ("Cohort 3"). Of 164 inpatients, 23 (14.0%) remained stable without the need for supplemental oxygen throughout their hospital stay, 80 (48.8%) required supplemental low- or high-flow oxygen, and 61 (37.2%) required invasive mechanical ventilation and, in all but 3 cases, additional organ support. Thirty-four (20.7%) patients died, including 5 with do not intubate/do not resuscitate orders in place (Table 1). This cohort was selected because the large number of samples (n=548 samples from 164 patients, Supplementary Data 2, Supplementary Table 11,) i) adds information about the technical stability of the assay and increases the statistical power to evaluate the results, ii) allowed us to assess a potential prognostic value of the assay and iii) to evaluate its applicability after dexamethasone became standard of care for patients requiring supplemental oxygen.

Reassuringly, despite the large number of samples acquired, split over three batches and measured over 10 days, and despite the different matrix (EDTA plasma), technical variation was low (Figure 4a). Peptides which were characteristic for disease severity in Cohort 2 (Figure 2b) also differentiated WHO grades in Cohort 3. (Figure 4b). Because they are of the highest practical value and because they are the furthest apart from outcome, we continued with the earliest sample of each patient. Of the 48 peptides quantified in the EDTA matrix, 34 had a significantly different trend between treatment escalation level (WHO3 to WHO7), with 12 peptides significantly increasing and 22 decreasing in concentration (Figure 4c, Supplementary Figure 7). As in Cohort 2, some markers indicated specific disease transitions, while others gradually changed with severity. For instance, CRP, CST3, or CD14 were increased in very severe forms of the disease (WHO7) (Figure 4c).

Next, we tested if the necessary treatment level could be predicted from the first available sample. We constructed a support vector machine (SVM) trained to differentiate between three different treatment groups on the basis of the severity markers: WHO3 (mild COVID-19, hospitalised, but no supplemental oxygen necessary), WHO4/5 (moderate COVID-19, hospitalised, supplemental low- or high-flow oxygen necessary), and WHO6/7 (severe COVID-19, hospitalised, intensive care and invasive mechanical ventilation necessary). The data was split in a training and a validation set in a cross-validated manner. The model predicted the WHO grades in the validation set from the peptide biomarker data. For most patients, the predicted WHO grade was in agreement with the actual treatment escalation (Figure 4d).

An important clinical need for a COVID-19 assay is to be prognostic of outcomes.[43] An SVM was trained on data obtained from the earliest sample, in a cross-validated manner, to differentiate patients who later survived COVID-19 from patients with a fatal outcome (n=164, of which 130 survived (controls) and 34 died (cases)) (Figure 5). The trained outcome predictor correctly classified 81.7% of the patients (sensitivity = 0.765, (26/34) specificity = 0.831 (108/130), AUROC = 0.855) that were withheld while training the model (Figure 5a, b). Additionally, a decision curve analysis (Supplementary Figure 8) shows a higher net benefit over a long range of threshold probabilities for the trained SVM classifier compared to the reference strategies. To exclude that the predicting capabilities are limited to the method, two other predictors (logistic regression and extra-trees) using the same setup were evaluated (Supplementary Figures 9 and 10). Reassuringly, these predictors (logistic regression: sensitivity = 0.735, specificity = 0.854, AUROC = 0.848;

Articles

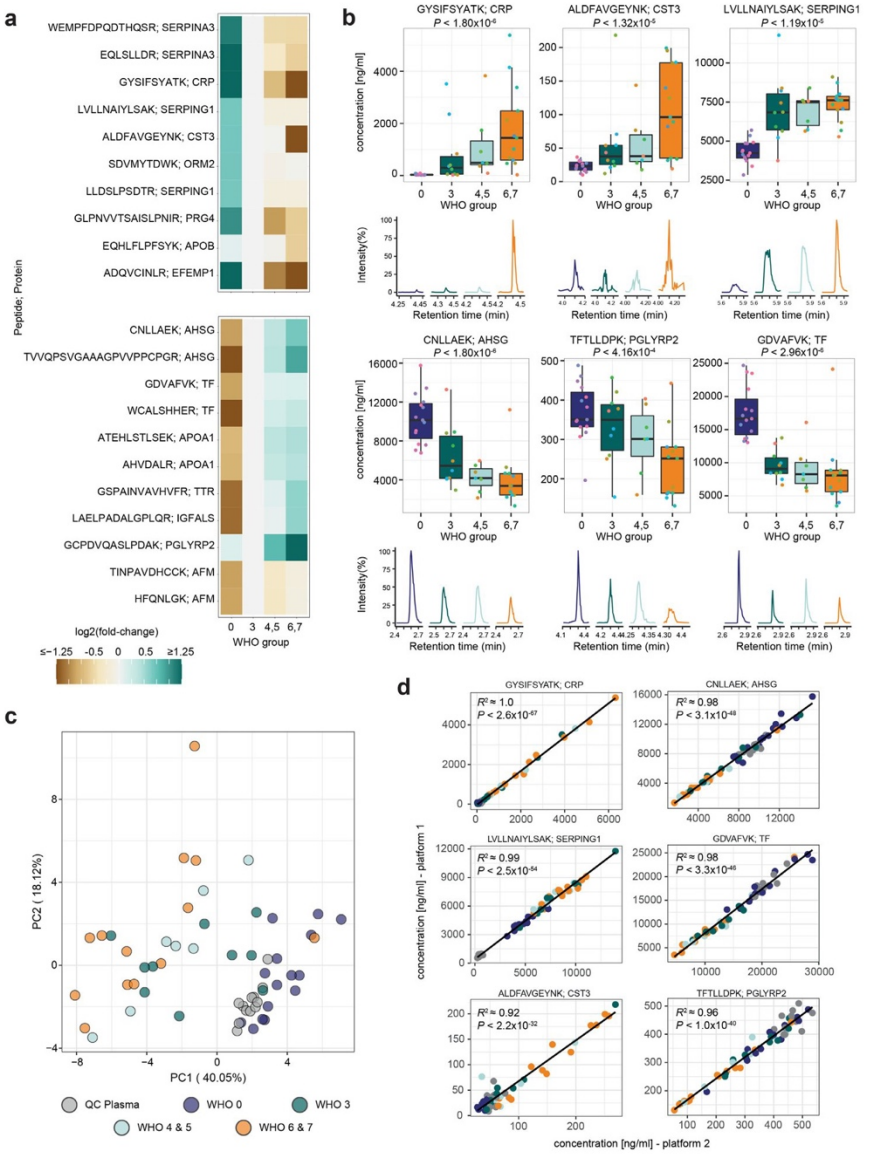


Figure 3. The protein biomarker assay reproducibly reports disease severity in a COVID-19 cohort. a) The established assay was applied to citrate plasma samples, collected for a balanced COVID-19 cohort studied during the first wave of the pandemic[24,25,30] ('Cohort 2') consisting of healthy volunteers (n=15, WHO 0), COVID-19 affected individuals requiring hospitalisation but no oxygen therapy (n=10 (WHO3), COVID-19 affected individuals requiring hospitalisation and non-invasive oxygen therapy (n=4, WHO4; n=3 WHO5), and severely affected hospitalised individuals requiring mechanical ventilation (n=3 (WHO6), n=10

extra-trees: sensitivity = 0.706, specificity = 0.815, AUROC = 0.836) show a comparable performance. To evaluate how well the SVM predictor performs compared to the clinical scores, we determined SOFA, APACHE II, and CCI scores as well as the COVID-19-specific ABCS score. SOFA, APACHE II, and ABCS, which are directly linked to the patient's disease severity, performed best among the four scores tested. Nonetheless, the MRM biomarker assay outperformed all other scores, as indicated by ROC analysis (Figure 5a, Supplementary Table 12.)

Discussion

The COVID-19 crisis has reminded us that novel infectious diseases can quickly challenge health systems on a global scale. Although COVID-19 has meanwhile become a well-studied disease, there remains unmet clinical need for personalised tests that can support clinical decision making and guide development of novel treatments. The most important decision points for the clinical management of COVID-19 are i) at admission, posing the question whether a patient needs inpatient care, ii) at regular inpatient wards, when decisions have to be made whether more intense monitoring and respiratory support (i.e. intensive care) are necessary during clinical deterioration, which is common during the second week of COVID-19 symptoms, and iii) at intensive care units, where decisions about additional organ replacement treatment such as extracorporeal membrane oxygenation need to be made.

Several investigations have highlighted the classification and prognostic value of plasma proteomes in COVID-19.[9,10,16,21–23] Discovery proteomic technologies are difficult to implement in a clinical routine. However, one can translate a proteomics result through the selection and validation of biomarker panels; and technologies, such as triple quadrupole mass spectrometers coupled to analytical flow rate chromatography are routinely used in clinical and regulated laboratories.[44,45] Indeed, their application in clinical proteomics is desirable as LC-MRM i) provides high sensitivity and specificity, ii) allows the inclusion of internal standards for higher precision and control over potential matrix effects,[46,47] iii) facilitates absolute quantification, enabling cross-platform transferability,[48,49] and iv) covers a large dynamic range.[50] This in turn enables the comparison of biomarkers with large abundance differences within one run, thus facilitating multiplexing of many biomarkers and downstream processing and statistical analysis.

To simplify the transition from discovery to applied proteomics, we have recently introduced a proteomics platform that uses analytical flow rate chromatography already at the discovery stage.[13,24] Herein, we used discovery proteomic data recorded with this platform, to design a multi-protein biomarker panel for severity stratification in COVID-19. The developed biomarker panel includes 50 peptides derived from 30 plasma proteins. These peptides were selected both for their association with COVID-19, for instance the innate immune response, the coagulation system, or the complement cascade,[13,24,25] and for technical reasons, for instance the distribution over the chromatographic gradient. The assay is robust and can be ported to different platforms and matrices, in some of which not all peptides will be quantifiable, as to be determined by a set of quality controls. In this study, we established the assay on two different routine-laboratory-compatible LC-MRM platforms. We demonstrate sensitivity, accuracy, precision, as well as overall reproducibility on both platforms.

We confirm in two temporally separated COVID-19-patient cohorts that the panel assay captures disease severity of SARS-CoV-2-infected individuals and discriminates the necessary treatment levels. This is particularly reassuring as the standard of care changed to include dexamethasone for treatment of respiratory failure in the outcome validation cohort (Cohort 3), following publication of the results of the recovery trial.[51] We also tested the prognostic value of the panel and found that it outperformed four clinical risk-assessment metrics, ABCS, CCI, SOFA, and APACHE II, in predicting

(WHO7) as well as QC plasma samples (n=12). Peptides with a significant concentration change (up- (top panel) and down-regulated (bottom panel)) distinguish healthy from infected individuals, as well as mild from severe forms of the disease. Heatmap displays the $\log_2$ fold-change of the indicated peptide to its median concentration in patients with a severity score of WHO3. Only significant peptides (adjusted $P < 0.05$) with an arbitrary fold-change difference <1/1.5 or >1.5 are shown, and $\log_2$ fold-changes <-1.25 or >1.25 are indicated by the same respective colour. For additional information, see Supplementary Figure 5. **b)** Visualisation of the response to COVID-19 based on selected peptides indicating different COVID-19 severity trends (changing with severity expressed according to the WHO ordinal scale (left, middle panel), and differentiating healthy from COVID-19 infected individuals (right panel)). Boxplots display the absolute concentration of selected peptides in patients in different severity groups as explained in (a). The extracted ion chromatograms (EIC) display the response of representative samples of individuals classified according to the treatment escalation WHO=0, WHO=3, WHO=5 and WHO=7. **c)** Unsupervised clustering by principal component analysis (PCA) based on the absolute concentration of 39 quantified peptides clusters patients with COVID-19 by severity. The peptide ADQVCINLR contained missing values and was omitted. **d)** Analytical reproducibility of the assay in two laboratories, running two different LC-MS/MS platforms, with independently optimised MRM transitions. Shown are linear correlations (pearson correlation) between absolute peptide concentrations. Selected peptides and colour code like in (c). For boxplots, the median is marked by a solid line, hinges mark the 25th and 75th percentiles, and whiskers show all values that, at maximum, fall within 1.5 times the interquartile range.

Articles

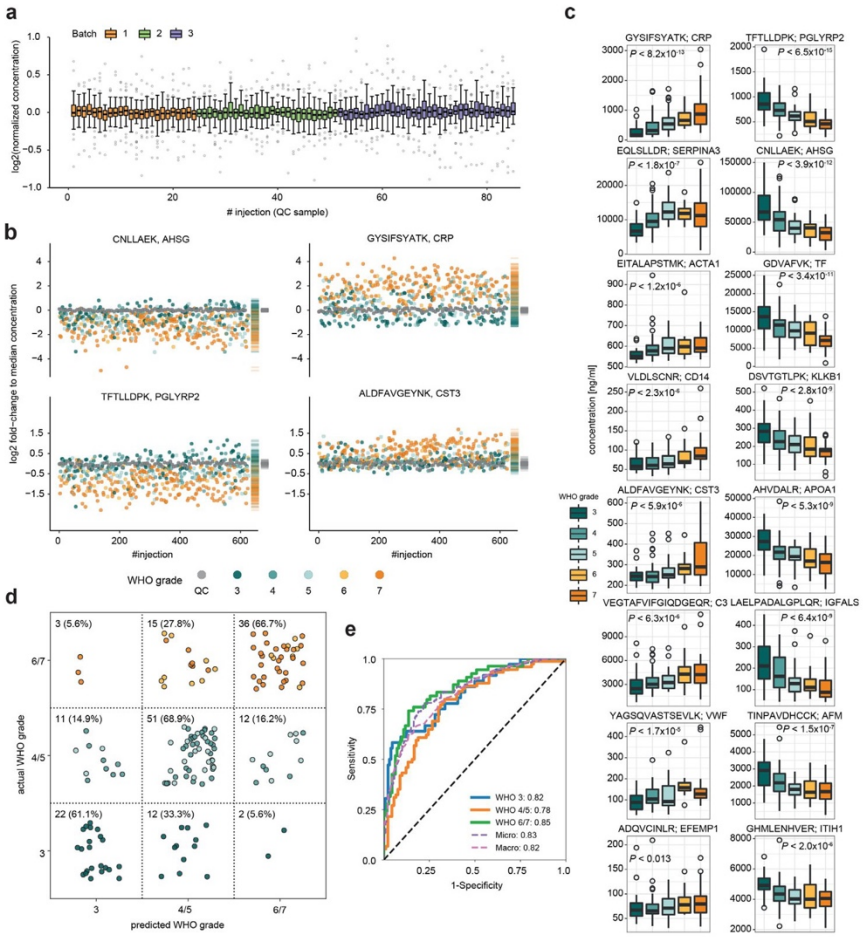


Figure 4. Diagnostic and analytical performance of the assay on a COVID-19 inpatient cohort treated during the 2nd wave of the pandemic. a) Quantitative performance (signal stability), during the measurement of 548 plasma proteome samples of patient Cohort 3 evaluated based on n=85 QC samples (coloured in grey, pool of COVID-19 samples as described in[24]) injected throughout the acquisition. Shown are the $\log_2$ fold-change of the absolute concentration for each of the 48 quantified peptides normalised to the median of the QC samples for the respective peptide. **b)** Peptide $\log_2$ absolute concentration fold change of two selected down- (AHSG, PGLYRP2) and two up-regulated (CRP, CST3) proteins for all samples acquired for the cohort described in (a). QC samples are shown in grey, all other samples are coloured according to the corresponding COVID-19 WHO treatment escalation score; rug plots on the right side of each peptide indicate the respective distributions. **c)** The 8 most significantly upregulated (left panel) and down-regulated (right panel) peptides indicative of COVID-19 disease severity, expressed as the treatment level according to the WHO scale. For illustration purposes, only one peptide per protein is displayed and one outlier sample in peptide EITALAPSMK was removed. The quantities of all quantified peptides are illustrated in Supplementary Figure 7. **d)** Confusion-matrix-like representation of the outcome of a multi-class classification model (SVM-based) trained to differentiate three WHO severity groups: grade 3, grades 4/5, and grades 6/7. Predictions were done on withheld samples that were not used for training the models (accuracy = 0.665, balanced accuracy = 0.656). The percentage denotes how many samples within each WHO severity group are assigned to each square. The positions of the points within each square were chosen randomly. Colour scheme according to (c). **e)** ROC-curves for the prediction of the WHO severity group from the first time point measured for every patient (grade 3, blue; grades 4/5,

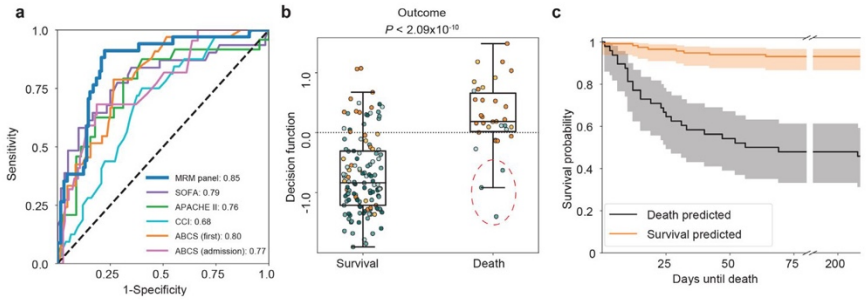


Figure 5. COVID-19 outcome prognosis. a) Receiver operating characteristic (ROC) curve for the prediction of survival and non-survival, from a single plasma sample (first sample measured for every patient, n=164) using an SVM-classifier. The blue curve denotes the model trained and benchmarked on measured proteomic data. The other curves denote models based on single severity scores (Sequential Organ Failure Assessment score (SOFA, purple, n=91), Acute Physiology And Chronic Health Evaluation (APACHE II, green, n=69), Charlson Comorbidity Index (CCI, cyan, n=157), and Age, Biomarkers, Clinical history, Sex score (ABCS, pink (admission, n=135) and orange (first sample, n=161)). **b)** Boxplot of the decision function of the SVM for every patient sorted according to the outcome and coloured with respect to the WHO grade at the day the sample was taken. Colour scheme according to Figure 4c. The red dashed circle indicates three patients with a high chance of predicted survival who died. These patients all were WHO4 patients with 'do not intubate (DNI)' orders in place due to other medical conditions. Therapy was therefore not escalated to invasive ventilation. The MRM assay classified those patients as milder COVID-19 cases, compared to other non-survivors. The cut point for the binary metrics is highlighted. Sensitivity = 0.765, (26/34); Specificity = 0.831 (108/130). Indicated *P* value reports differences between decision function distributions (survival vs. death) as calculated with a Mann-Whitney U rank test. **c)** Kaplan-Meier estimate of the survival function for survival predicted cases (orange) or non-survival predicted cases (black) with confidence interval (alpha=0.05). Patient survival data for each timepoint is provided in Supplementary Tables 13-15. All predictions were done on withheld samples that were not used for training the models. For boxplots, the median is marked by a solid line, hinges mark the 25th and 75th percentiles, and whiskers show all values that, at maximum, fall within 1.5 times the interquartile range.

the survival of COVID-19 inpatients, as revealed by a ROC analysis. Thus, the panel assay could be used to assess the current state of the patient, help monitor novel treatments, or stratify patients based on their responsiveness to novel therapeutic interventions. Furthermore, the assay can be employed to predict the future course of COVID-19, as exemplified by the prediction of disease outcome weeks into the future.

While the value of quantitative proteomic measurements for disease stratification was established in different cohorts, further validation of the outcome prediction will be required in an independent cohort. Similarly, not all eventualities of a complex disease such as COVID-19 are covered with this assay. Additional data to build improved models, and ideally prospective studies, will be of high value, especially to refine predictive models with regard to new SARS-CoV-2 variants and the additional treatment options such as JAK inhibitors. Similarly, it will be important to assess the performance of the assay in other populations, including ambulatory, asymptomatic, and patients affected by other infectious diseases. Another important criterion to be determined is the time point where the signature from this assay is most predictive of outcome. For instance, we recently described that the early spike of the inflammatory response and its gradual decrease is prognostic for patients surviving the disease.[13,43]

Proteomics can be prognostic of outcome in patients with similar disease severity, e.g. among severely affected patients[43] that are difficult to distinguish by clinical parameters. This means that survival prognosis using targeted proteomics could be improved beyond what was shown in this study for 'within-severity-group' prognosis, if biomarker panels were selected specifically for stratification within the respective COVID-19 severity group. Indeed, particularly within the group of severely affected individuals, some patients were predicted incorrectly, i.e. survived despite being predicted as non-survivors, or vice versa (Figure 5b). We assessed on a patient-by-patient basis whether there are medical reasons that could explain wrong predictions. Plotting outcome with respect to time until death (Kaplan–Meier survival analysis, Figure 5c) denotes no clear tendency for the correct and false predictions. However,

orange; grades 6/7, green). Dotted lines denote the micro (magenta) and macro average (pink) ROC-curves. Data shown and analysed in c)-d) are from the first time-point obtained for each individual; n=36 (WHO 3), n=47 (WHO 4), n=27 (WHO 5), n=16 (WHO 6), n=38 (WHO7). For boxplots, the median is marked by a solid line, hinges mark the 25th and 75th percentiles, and whiskers show all values that, at maximum, fall within 1.5 times the interquartile range.

we noted that the three samples with the smallest decision function across wrongly predicted patients with fatal outcomes belonged to WHO4 patients that had DNI ('do not intubate') orders in place (denoted with a red circle in Figure 5b). It is hence plausible that the assay correctly identified a milder form of COVID-19 in these three individuals; i.e. that without a strong comorbidity or a DNI order in place, these might have had a good chance to survive COVID-19. Similarly, we recently reported two cases where the proteomic signatures of patients correctly distinguished an influenza B from a SARS-CoV-2 infection, and that highlighted a patient that had to undergo chemotherapeutic cancer treatment just days before a SARS-CoV-2 infection.[24] Thus, protein signatures could in principle distinguish different (respiratory) infections or comorbidities, while additional research will be required to establish this for the presented protein panel.

COVID-19 will remain a central public health issue for the foreseeable future as new variants of concern with capacity to evade vaccine-induced immunity continue to emerge.[52] The underpinning targeted proteomics platform supports rapid iteration of the panel composition in case additional prognostic biomarkers are discovered. Taken together, this peptide panel and the underlying analytical platform hold potential to support a broader, continuous pandemic response in addition to their utility in hospitalised patient cohorts, which we demonstrate in the present study.

Contributors

MM, ES, JH, MR conceptualised and supervised the study. ZW, AC, OL, ST, JH, ES formally analysed and curated the data. FK, MR, ES acquired funding. PTL, EH, LES and FK acquired clinical data, ZW, JH acquired LC-MS/MS data (on Agilent 6495), DB, CSL, RS acquired LC-MS/MS data (on SCIEX 7500). DL performed sample preparation. VD selected the peptides. ZW, AC, OL, ES, JH developed methodology. SH, LES, MM, FK, ES, JH and MR administered the project. CM, JZ provided resources. OL developed software. ZW, OL, AC, ES, JH validated the data and methodologies. ZW, OL, MM, JH visualised the results. ZW, OL, PTL, FK, ES, JH, MR wrote the original draft. ZW, OL, PTL, FK, ES, JH, MR revised and edited the manuscript. All authors reviewed the manuscript and had access to the data generated in the study. Authors involved in formal analysis and data curation, validation, and investigation directly worked with and verified the data. F.K., E.S., J.H. and M.R. decided to submit the manuscript.

Data sharing statement

Participant data of Cohort 3 that collected during the trial is available in (Supplementary Table 11). For other cohorts, respective information is available in previous publications.[13,24] Measurement data, as well as custom used code used for modelling / data analysis are available online.[53]

Declaration of interests

EM Scientific Limited (t/a Inoviv) and Charité – Universitätsmedizin Berlin (Ziyue Wang, Michael Mülleder, Vadim Demichev, Johannes Hartl and Markus Ralser) filed joint patent applications for the protein panel assay described herein - United States Application No: 63/156291, 63/283787 and 17/685756. Leif-Erik Sander received honoraria for lectures from Boehringer Ingelheim, Novartis, Berlin Chemie, GSK, Merck, Novartis, Sanofi. Ernestas Sirka and Adam Cryar are/were employees of EM Scientific Limited (t/a Inoviv). Daniel Blake, Rebekah L Sayers and Catherine S Lane are employees of SCIEX. Christoph Mueller and Johannes Zeiser are employees of Agilent Technologies.

Acknowledgments

Figure 1 and Supplementary Figure 1 was made using BioRender (Biorender.com). We thank Hezi Tenenboim for proofreading the manuscript. The Pa-COVID-19 study group[13,30] is acknowledged for study logistics and collection of biosamples and clinical data.

Funding

This research was funded in part by the European Research Council (ERC) under grant agreement ERC-SyG-2020 951475 (to M.R) and by the Wellcome Trust (IA 200829/Z/16/Z to M.R.). The work was further supported by the Ministry of Education and Research (BMBF) as part of the National Research Node 'Mass Spectrometry in Systems Medicine (MSCoresys)', under grant agreements 031L0220 and 161L0221. J.H. was supported by a Swiss National Science Foundation (SNSF) Postdoc Mobility fellowship (project number 191052). This study was further supported by the BMBF grant NaFoUniMedCOVID-19 – NUM-NAPKON, FKZ: 01KX2021. The study was co-funded by the UK's innovation agency, Innovate UK, under project numbers 75594 and 56328.

Supplementary materials

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.eclinm.2022.101495.

References

1 Pouwels KB, Pritchard E, Matthews PC, et al. Effect of Delta variant on viral burden and vaccine effectiveness against new SARS-CoV-2 infections in the UK. *Nat Med*. 2021.
2 Kearns P, Siebert S, Willicombe M, et al. Examining the Immunological Effects of COVID-19 Vaccination in Patients with

Conditions Potentially Leading to Diminished Immune Response Capacity – The OCTAVE Trial 2021.
3 Chemaitelly H, Tang P, Hasan MR, et al. Waning of BNT162b2 Vaccine Protection against SARS-CoV-2 Infection in Qatar. *N Engl J Med*. 2021.
4 Levin EG, Lustig Y, Cohen C, et al. Waning Immune Humoral Response to BNT162b2 Covid-19 Vaccine over 6 Months. *N Engl J Med*. 2021.
5 Tober-Lau P, Schwarz T, Vanshylla K, et al. Long-term immunogenicity of BNT162b2 vaccination in older people and younger health-care workers. *Lancet Respir Med*. 2021;9:e104–e105.
6 Bian L, Gao F, Zhang J, et al. Effects of SARS-CoV-2 variants on vaccine efficacy and response strategies. *Expert Rev Vaccines*. 2021;20:365–373.
7 Twohig KA, Nyberg T, Zaidi A, et al. Hospital admission and emergency care attendance risk for SARS-CoV-2 delta (B.1.617.2) compared with alpha (B.1.1.7) variants of concern: a cohort study. *Lancet Infect. Dis*. 2021.
8 Tao K, Tzou PL, Nouhin J, et al. The biological and clinical significance of emerging SARS-CoV-2 variants. *Nat Rev Genet*. 2021.
9 Overmyer KA, Shishkova E, Miller IJ, et al. Large-Scale Multi-omic Analysis of COVID-19 Severity. *Cell Syst*. 2021;12:23–40.e7.
10 D'Alessandro A, Thomas T, Dzieciatkowska M, et al. Serum Proteomics in COVID-19 Patients: Altered Coagulation and Complement Status as a Function of IL-6 Level. *J Proteome Res*. 2020;19:4417–4427.
11 Völlmy F, van den Toorn H, Zenezini Chiozzi R, et al. A serum proteome signature to predict mortality in severe COVID-19 patients. *Life Sci Alliance*. 2021;4.
12 Couzin-Frankel J. The mystery of the pandemic's "happy hypoxia. *Science*. 2020;368:455–456.
13 Demichev V, Tober-Lau P, Lemke O, et al. A time-resolved proteomic and prognostic map of COVID-19. *Cell Syst*. 2021.
14 Vegvari C, Cauët E, Hadjichrysanthou C, et al. Using Clinical Trial Simulators to Analyse the Sources of Variance in Clinical Trials of Novel Therapies for Acute Viral Infections. *PLoS One*. 2016;11:e0156622.
15 Shen C, Ferro EG, Xu H, et al. Underperformance of Contemporary Phase III Oncology Trials and Strategies for Improvement. *J Natl Compr Canc Netw*. 2021;19:1072–1078.
16 Gutmann C, Takov K, Burnap SA, et al. SARS-CoV-2 RNAemia and proteomic trajectories inform prognostication in COVID-19 patients admitted to intensive care. *Nat Commun*. 2021;12:3406.
17 Vatansever HS, Becer E. Relationship between IL-6 and COVID-19: to be considered during treatment. *Future Virol Future Medicine*. 2020;15:817–822.
18 Wynants L, Van Calster B, Collins GS, et al. Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal. *BMJ*. 2020;369:m1328.
19 Gupta RK, Marks M, Samuels THA, et al. Systematic evaluation and external validation of 22 prognostic models among hospitalised adults with COVID-19: an observational cohort study. *Eur. Respir J*. 2020;56.
20 Jiang M, Li C, Zheng L, et al. A biomarker-based age, biomarkers, clinical history, sex (ABCS)-mortality risk score for patients with coronavirus disease 2019. *Ann Transl Med*. 2021;9:230.
21 Park J, Kim H, Kim SY, et al. In-depth blood proteome profiling analysis revealed distinct functional characteristics of plasma proteins between severe and non-severe COVID-19 patients. *Sci Rep*. 2020;10:22418.
22 Ignjatovic V, Geyer PE, Palaniappan KK, et al. Mass Spectrometry-Based Plasma Proteomics: Considerations from Sample Collection to Achieving Translational Data. *J Proteome Res*. 2019;18:4085–4097.
23 Filbin MR, Mehta A, Schneider AM, et al. Plasma proteomics reveals tissue-specific cell death and mediators of cell-cell interactions in severe COVID-19 patients. *bioRxiv*. 2020.
24 Messner CB, Demichev V, Wendisch D, et al. Ultra-high-throughput clinical proteomics reveals classifiers of COVID-19 infection. *Cell Systems*. 2020.
25 Messner CB, Demichev V, Bloomfield N, et al. Ultra-fast proteomics with Scanning SWATH. *Nat Biotechnol*. 2021.
26 Keevil BG. LC-MS/MS analysis of steroids in the clinical laboratory. *Clin Biochem*. 2016;49:989–997.
27 Gaudl A, Kratzsch J, Ceglarek U. Advancement in steroid hormone analysis by LC-MS/MS in clinical routine diagnostics - A three year recap from serum cortisol to dried blood 17α-hydroxyprogesterone. *J Steroid Biochem Mol Biol*. 2019;192:105389.
28 Ma S, Guo Q, Zhang Z, et al. Expanded newborn screening for inborn errors of metabolism by tandem mass spectrometry in newborns from Xinxiang city in China. *J Clin Lab Anal*. 2020;34:e23159.
29 Zhu Z, Gu J, Genchev GZ, et al. Improving the Diagnosis of Phenylketonuria by Using a Machine Learning-Based Screening Model of Neonatal MRM Data. *Front Mol Biosci*. 2020;7:115.
30 Kurth F, Roennefarth M, Thibeault C, et al. Studying the pathophysiology of coronavirus disease 2019: a protocol for the Berlin prospective COVID-19 patient cohort (Pa-COVID-19). *Infection*. 2020;48:619–626.
31 Mühlemann B, Thibeault C, Hillus D, et al. Impact of dexamethasone on SARS-CoV-2 concentration kinetics and antibody response in hospitalized COVID-19 patients: results from a prospective observational study. *Clin Microbiol Infect*. 2021;27. 1520.e7–1520.e10.
32 Thibeault C, Mühlemann B, Helbig ET, et al. Clinical and virological characteristics of hospitalised COVID-19 patients in a German tertiary care centre during the first wave of the SARS-CoV-2 pandemic: a prospective observational study. *Infection*. 2021;49:703–714.
33 MacLean B, Tomazela DM, Shulman N, et al. Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics*. 2010;26:966–968.
34 Bioanalytical-Method-Validation-Guidance-for-Industry.pdf n.d.
35 Mischak H, Allmaier G, Apweiler R, et al. Recommendations for biomarker identification and qualification in clinical proteomics. *Sci Transl Med*. 2010;2:46ps42.
36 von Elm E, Altman DG, Egger M, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Ann Intern Med*. 2007;147:573–577.
37 World Health Organization. *R&D Blueprint Novel Coronavirus COVID-19 Therapeutic Trial Synopsis (WHO)*. 2020.
38 Shen B, Yi X, Sun Y, et al. Proteomic and Metabolomic Characterization of COVID-19 Patient Sera. *Cell*. 2020;182:59–72.e15.
39 Liu Y, Gao W, Guo W, et al. Prominent coagulation disorder is closely related to inflammatory response and could be as a prognostic indicator for ICU patients with COVID-19. *J Thromb Thrombolysis*. 2020;50:825–832.
40 Bhowmick P, Roome S, Borchers CH. An Update on MRMAssayDB: A Comprehensive Resource for Targeted Proteomics Assays in the Community. *J Proteome Res*. 2021;20:2105–2115.
41 Jones BR, Schultz GA, Eckstein JA, et al. Surrogate matrix and surrogate analyte approaches for definitive quantitation of endogenous biomolecules. *Bioanalysis*. 2012;4:2343–2356.
42 Pino LK, Searle BC, Yang H-Y, et al. Matrix-Matched Calibration Curves for Assessing Analytical Figures of Merit in Quantitative Proteomics. *J Proteome Res*. 2020;19:1147–1153.
43 Demichev V, Tober-Lau P, Nazarenko T, et al. A proteomic survival predictor for COVID-19 patients in intensive care. *PLOS Digital Health. Public Library of Science*. 2022;1:e0000007.
44 Jans JJM, Broeks MH, Verhoeven-Duif NM. Metabolomics in diagnostics of inborn metabolic disorders. *Current Opinion in Systems Biology*. 2021:100409.
45 van der Gugten JG. Tandem mass spectrometry in the clinical laboratory: A tutorial overview. *Clinical Mass Spectrometry*. 2020;15:36–43.
46 Xu RN, Fan L, Rieser MJ, et al. Recent advances in high-throughput quantitative bioanalysis by LC-MS/MS. *J Pharm Biomed Anal*. 2007;44:342–355.
47 Hall TG, Smukste I, Bresciano KR, et al. Identifying and overcoming matrix effects in drug discovery and development. Tandem Mass Spectrometry—Applications and Principles. Intech Open, London, UK. 2012;18:390–419.
48 Gerber SA, Rush J, Stemman O, et al. Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. *Proc Natl Acad Sci U S A*. 2003;100:6940–6945.
49 Yang JJ, Han Y, Mah CH, et al. Streamlined MRM method transfer between instruments assisted with HRMS matching and retention-time prediction. *Anal Chim Acta*. 2020;1100:88–96.
50 Anderson L, Hunter CL. Quantitative Mass Spectrometric Multiple Reaction Monitoring Assays for Major Plasma Proteins*. *Mol Cell Proteomics*. 2006;5:573–588.
51 Collaborative Group RECOVERY, Horby P, Lim WS, Emberson JR, et al. Dexamethasone in Hospitalized Patients with Covid-19. *N Engl J Med*. 2021;384:693–704.
52 Iftekhar EN, Priesemann V, Balling R, et al. A look into the future of the COVID-19 pandemic in Europe: an expert consultation. *Lancet Reg Health Eur*. 2021;8:100185.
53 Hartl J. Supplementary Data: A multiplex protein panel assay for severity prediction and outcome prognosis in patients with COVID-19: An observational multi-cohort study. *Mendeley Data. v1*. 2022.

10 Lebenslauf

Mein Lebenslauf wird aus datenschutzrechtlichen Gründen in der elektronischen Version meiner Arbeit nicht veröffentlicht.

11 Komplette Publikationsliste

Geteilte Erstautorenschaften sind durch * gekennzeichnet.

31. Thibeault C*, Bardtke L*, Vanshylla K*, di Cristanziano V, Eberhardt KA, Stubbemann P, Hillus D, **Tober-Lau P**, Mukherjee P, Münn F, Lippert LJ, Helbig ET, Lingscheid T, Steinbeis F, Mittermaier M, Witzenrath M, Zoller T, Pa-COVID study group, Klein F, Sander LE, Kurth F. Short- and long-term T cell and antibody responses following dexamethasone treatment in COVID-19. *JCI Insight*. 2023;8(8). doi:10.1172/jci.insight.166711

    IF: 9,484 (2021-2022)

30. Kurth F*, **Tober-Lau P***, Lingscheid T, Bardtke L, Kim J, Angheben A, Gobbi FG, Mbavu L, Stegemann MS, Heim KM, Pfäfflin F, Menner N, Schürmann M, Mikolajewska A, Witzenrath M, Sander LE, Mayer B, Zoller T. Post-treatment haemolysis is common following oral artemisinin combination therapy of uncomplicated malaria in travellers. *J Travel Med*. Published online January 5, 2023. doi:10.1093/jtm/taad001

    IF: 39.194 (2021-2022)

29. Frommert LM*, Arumahandi de Silva AN*, Zernicke J, Scholz V, Braun T, Jeworowski LM, Schwarz T, **Tober-Lau P**, Ten Hagen A, Habermann E, Kurth F, Sander LE, Corman VM, Burmester GR, Biesen R, Albach FN, Klotsche J. Type of vaccine and immunosuppressive therapy but not diagnosis critically influence antibody response after COVID-19 vaccination in patients with rheumatic disease. *RMD Open*. 2022;8(2):e002650.

    IF: 5,806 (2021-2022)

28. Lingscheid T, Kinzig M, Krüger A, Müller N, Bölke G, **Tober-Lau P**, Münn F, Kriedemann H, Witzenrath M, Sander LE, Sörgel F, Kurth F. Pharmacokinetics of Nirmatrelvir and Ritonavir in COVID-19 Patients with End-Stage Renal Disease on Intermittent Hemodialysis. *Antimicrob Agents Chemother*. 2022;66(11):e0122922.

    IF: 5,938 (2021-2022)

27. Habermann E*, Gieselmann L*, **Tober-Lau P**, Klotsche J, Albach FN, Ten Hagen A, Zernicke J, Ahmadov E, Arumahandi de Silva AN, Frommert LM, Kurth F, Sander LE, Burmester GR, Klein F, Biesen R. Pausing methotrexate prevents impairment of Omicron BA.1 and BA.2 neutralisation after COVID-19 booster vaccination. *RMD Open*. 2022;8(2):e002639.

    IF: 5,806 (2021-2022)

26. Wang Z*, **Tober-Lau P***, Farztdinov V*, Lemke O, Schwecke T, Steinbrecher S, Muenzner J, Kriedemann H, Sander LE, Hartl J, Mülleder M, Ralser M, Kurth F. The

human host response to monkeypox infection: a proteomic case series study. *EMBO Mol Med*. 2022;14(11):e16643.

IF: 14.260 (2021-2022)

25. Gruell H*, Vanshylla K*, **Tober-Lau P**, Hillus D, Sander LE, Kurth F, Klein F. Neutralisation sensitivity of the SARS-CoV-2 omicron BA.2.75 sublineage. *Lancet Infect Dis*. 2022;22(10):1422-1423.

IF: 71,421 (2021-2022)

24. Steinbeis F, Knape P, Mittermaier M, Helbig ET, **Tober-Lau P**, Thibeault C, Lippert LJ, Xiang W, Müller-Plathe M, Steinbrecher S, Meyer HJ, Ring RM, Ruwwe-Glösenkamp C, Alius F, Li Y, Müller-Redetzky H, Uhrig A, Lingscheid T, Grund D, Temmesfeld-Wollbrück B, Suttorp N, Sander LE, Kurth F, Witzenrath M, Zoller T. Functional limitations 12 months after SARS-CoV-2 infection correlate with initial disease severity: An observational study of cardiopulmonary exercise capacity testing in COVID-19 convalescents. *Respir Med*. 2022;202:106968.

IF: 4,582 (2021-2022)

23. Kleeman SO, Cordioli M, Timmers PRHJ, Khan A, **Tober-Lau P**, Kurth F, Demichev V, Meyer HV, Wilson JF, Ralser M, Kiryluk K, Ganna A, Baillie K, Janowitz T. Cystatin C is associated with adverse COVID-19 outcomes in diverse populations. *iScience*. 2022;25(10):105040.

IF: 5,458 (2021-2022)

22. Wang Z*, Muecksch F*, Muenn F*, Cho A, Zong S, Raspe R, Ramos V, Johnson B, Ben Tanfous T, DaSilva J, Bednarski E, Guzman-Cardozo C, Turroja M, Millard KG, **Tober-Lau P**, Hillus D, Yao KH, Shimeliovich I, Dizon J, Kaczynska A, Jankovic M, Gazumyan A, Oliveira TY, Caskey M, Bieniasz PD, Hatziioannou T, Kurth F, Sander LE, Nussenzweig MC, Gaebler C. Humoral immunity to SARS-CoV-2 elicited by combination COVID-19 vaccination regimens. J Exp Med. 2022;219(10). doi:10.1084/jem.20220826

IF: 17,579 (2021-2022)

21. Pfäfflin F, Wendisch D, Scherer R, Jürgens L, Godzick-Njomgang G, Tranter E, **Tober-Lau P**, Stegemann MS, Corman VM, Kurth F, Schürmann D. Monkeypox in-patients with severe anal pain. *Infection*. Published online August 22, 2022. doi:10.1007/s15010-022-01896-7

IF: 7,455 (2021-2022)

20. Gruell H, Vanshylla K, Korenkov M, **Tober-Lau P**, Zehner M, Münn F, Janicki H, Augustin M, Schommers P, Sander LE, Kurth F, Kreer C, Klein F. SARS-CoV-2 Omicron sublineages exhibit distinct antibody escape patterns. *Cell Host Microbe*. Published online July 7, 2022. doi:10.1016/j.chom.2022.07.002

IF: 31,316 (2021-2022)

19. Wang Z*, Cryar A*, Lemke O, **Tober-Lau P**, Ludwig D, Helbig ET, Hippenstiel S, Sander LE, Blake D, Lane CS, Sayers RL, Mueller C, Zeiser J, Townsend S, Demichev V, Mülleder M, Kurth F, Sirka E, Hartl J, Ralser M. A multiplex protein panel assay for severity prediction and outcome prognosis in patients with COVID-19: An observational multi-cohort study. *EClinicalMedicine*. 2022;49(101495):101495.

    IF: 17,033 (2021-2022)

18. Lingscheid T*, Lippert LJ*, Hillus D, Kruis T, Thibeault C, Helbig ET, **Tober-Lau P**, Pfäfflin F, Müller-Redetzky H, Witzenrath M, Zoller T, Uhrig A, Opitz B, Suttorp N, Kramer TS, Sander LE, Stegemann MS, Kurth F. Characterization of antimicrobial use and co-infections among hospitalized patients with COVID-19: a prospective observational cohort study. *Infection*. Published online April 14, 2022. doi:10.1007/s15010-022-01796-w

    IF: 7,455 (2021-2022)

17. **Tober-Lau P***, Gruell H*, Vanshylla K*, Koch WM, Hillus D, Schommers P, Suárez I, Suttorp N, Sander LE, Klein F, Kurth F. Cross-variant neutralizing serum activity after SARS-CoV-2 breakthrough infections. *Emerg Infect Dis*. 2022;28(5):1050-1052.

    IF: 16,162 (2021-2022)

16. Vanshylla K*, **Tober-Lau P***, Gruell H*, Münn F, Eggeling R, Pfeifer N, Le NH, Landgraf I, Kurth F, Sander LE, Klein F. Durability of omicron-neutralising serum activity after mRNA booster immunisation in older adults. *Lancet Infect Dis*. 2022;22(4):445-446.

    IF: 71,421 (2021-2022)

15. Grishina A, Link F, Arend A, Kleemann F, **Tober-Lau P**, Andree D, Münn F, Gruendl M, Quante M, Lederhuber H, Albertsmeier M, Struller F, Grützmann R, Königsrainer A, Löffler MW. A survey among physicians in surgery and anesthesiology departments after the first surge of SARS-CoV-2 infections in Germany : Preparing for further challenges ahead. *Wien Klin Wochenschr*. Published online January 21, 2022. doi:10.1007/s00508-021-02000-z

    IF: 2,275 (2021-2022)

14. Gruell H*, Vanshylla K*, **Tober-Lau P**, Hillus D, Schommers P, Lehmann C, Kurth F, Sander LE, Klein F. mRNA booster immunization elicits potent neutralizing serum activity against the SARS-CoV-2 Omicron variant. *Nat Med*. Published online January 19, 2022. doi:10.1038/s41591-021-01676-0

    IF: 87,241 (2021-2022)

13. Demichev V*, **Tober-Lau P***, Nazarenko T, Lemke O, Kaur Aulakh S, Whitwell HJ, Röhl A, Freiwald A, Mittermaier M, Szyrwiel L, Ludwig D, Correia-Melo C, Lippert LJ, Helbig ET, Stubbemann P, Olk N, Thibeault C, Grüning NM, Blyuss O,

Vernardis S, White M, Messner CB, Joannidis M, Sonnweber T, Klein SJ, Pizzini A, Wohlfarter Y, Sahanic S, Hilbe R, Schaefer B, Wagner S, Machleidt F, Garcia C, Ruwwe-Glösenkamp C, Lingscheid T, Bosquillon de Jarcy L, Stegemann MS, Pfeiffer M, Jürgens L, Denker S, Zickler D, Spies C, Edel A, Müller NB, Enghard P, Zelezniak A, Bellmann-Weiler R, Weiss G, Campbell A, Hayward C, Porteous DJ, Marioni RE, Uhrig A, Zoller H, Löffler-Ragg J, Keller MA, Tancevski I, Timms JF, Zaikin A, Hippenstiel S, Ramharter M, Müller-Redetzky H, Witzenrath M, Suttorp N, Lilley K, Mülleder M, Sander LE, Kurth F, Ralser M, PA-COVID-19 Study group. A proteomic survival predictor for COVID-19 patients in intensive care. *PLOS Digit Health*. 2022;1(1):e0000007.

IF: noch nicht vergeben

12. Georg P*, Astaburuaga-García R*, Bonaguro L*, Brumhard S, Michalick L, Lippert LJ, Kostevc T, Gäbel C, Schneider M, Streitz M, Demichev V, Gemünd I, Barone M, **Tober-Lau P**, Helbig ET, Hillus D, Petrov L, Stein J, Dey HP, Paclik D, Iwert C, Mülleder M, Aulakh SK, Djudjaj S, Bülow RD, Mei HE, Schulz AR, Thiel A, Hippenstiel S, Saliba AE, Eils R, Lehmann I, Mall MA, Stricker S, Röhmel J, Corman VM, Beule D, Wyler E, Landthaler M, Obermayer B, von Stillfried S, Boor P, Demir M, Wesselmann H, Suttorp N, Uhrig A, Müller-Redetzky H, Nattermann J, Kuebler WM, Meisel C, Ralser M, Schultze JL, Aschenbrenner AC, Thibeault C, Kurth F, Sander LE, Blüthgen N, Sawitzki B, PA-COVID-19 Study Group. Complement activation induces excessive T cell cytotoxicity in severe COVID-19. *Cell*. 2022;185(3):493-512.e25.

IF: 66,850 (2021-2022)

11. **Tober-Lau P***, Schwarz T*, Vanshylla K*, Hillus D*, Gruell H, the EICOV/COVIM Study Group, Suttorp N, Landgraf I, Kappert K, Seybold J, Drosten C, Klein F, Kurth F, Corman VM, Sander LE. Long-term immunogenicity of BNT162b2 vaccination in older people and younger health-care workers. *Lancet Respir Med*. Published online October 20, 2021. doi:10.1016/S2213-2600(21)00456-2

IF: 102,642 (2021)

10. Hillus D*, Schwarz T*, **Tober-Lau P**, Vanshylla K, Hastor H, Thibeault C, Jentzsch S, Helbig ET, Lippert LJ, Tscheak P, Schmidt ML, Riege J, Solarek A, von Kalle C, Dang-Heine C, Gruell H, Kopankiewicz P, Suttorp N, Drosten C, Bias H, Seybold J, EICOV/COVIM Study Group, Klein F, Kurth F, Corman VM, Sander LE. Safety, reactogenicity, and immunogenicity of homologous and heterologous prime-boost immunisation with ChAdOx1 nCoV-19 and BNT162b2: a prospective cohort study. *Lancet Respir Med*. Published online August 12, 2021. doi:10.1016/S2213-2600(21)00357-X

IF: 102,642 (2021)

9. Schrezenmeier E*, Bergfeld L*, Hillus D*, Lippert JD, Weber U, **Tober-Lau P**, Landgraf I, Schwarz T, Kappert K, Stefanski AL, Sattler A, Kotsch K, Dörner T, Sander LE, Budde K, Halleck F, Kurth F, Corman VM, Choi M. Immunogenicity of

COVID-19 Tozinameran Vaccination in Patients on Chronic Dialysis. *Front Immunol*. 2021;12:690698.

IF: 8,768 (2021)

8. Völlmy F, van den Toorn H, Zenezini Chiozzi R, Zucchetti O, Papi A, Volta CA, Marracino L, Vieceli Dalla Sega F, Fortini F, Demichev V, **Tober-Lau P**, Campo G, Contoli M, Ralser M, Kurth F, Spadaro S, Rizzo P, Heck A Jr. A serum proteome signature to predict mortality in severe COVID-19 patients. *Life Sci Alliance*. 2021;4(9). doi:10.26508/lsa.202101099

IF: 5,781 (2021)

7. Mühlemann B*, Thibeault C*, Hillus D, Helbig ET, Lippert LJ, **Tober-Lau P**, Schwarz T, Müller MA, Witzenrath M, Suttorp N, Sander LE, Drosten C, Jones TC, Corman VM, Kurth F, Hippenstiel S, Haenel SS, Mittermaier M, Steinbeis F, Lingscheid T, Temmesfeld-Wollbrück B, Zoller T, Müller-Redetzky H, Uhrig A, Grund D, Ruwwe-Glösenkamp C, Stegemann MS, Heim KM, Hübner RH, Opitz B, Eckardt KU, Möckel M, Balzer F, Spies C, Weber-Carstens S, Tacke F, Dang-Heine C, Hummel M, Schwanitz G, Behrens UD, Rönnefarth M, Schmidt S, Krannich A, Jürgens L, Kleinschmidt M, Denker S, Pfeiffer M, Pascual-Leone BM, Mrziglod L, Machleidt F, Albus S, Bremer F, Doehn JM, Andermann T, Garcia C, Knape P, Krause PM, Lechtenberg L, Li Y, Pergantis P, Jacobi T, Ritter T, Yedikat B, Pfannkuch L, Zobel C, Kellermann U, Fieberg S, Bosquillon de Jarcy L, Wetzel A, Tabeling C, Brack MC, Müller-Plathe M, Kruse JM, Zickler D, Edel A, Stier B, Körner R, Müller NB, Stubbemann P, Olk N, Koch WM, Horn A, Stoyanova KK, Zvorc S, Kretzler L, Meyer-Arndt LA, Li L, Treue D, Briesemeister D, Schlesinger J, Sawitzki B, Bardtke L, Pohl K, Georg P, Wendisch D, Hiller AL, Brumhard S, Schmidt ML, et al. Impact of dexamethasone on SARS-CoV-2 concentration kinetics and antibody response in hospitalised COVID-19 patients - results from a prospective observational study. *Clin Microbiol Infect*. Published online June 15, 2021. doi:10.1016/j.cmi.2021.06.008

IF: 13,310 (2021)

6. Demichev V*, **Tober-Lau P***, Lemke O, Nazarenko T, Thibeault C, Whitwell H, Röhl A, Freiwald A, Szyrwiel L, Ludwig D, Correia-Melo C, Aulakh SK, Helbig ET, Stubbemann P, Lippert LJ, Grüning NM, Blyuss O, Vernardis S, White M, Messner CB, Joannidis M, Sonnweber T, Klein SJ, Pizzini A, Wohlfarter Y, Sahanic S, Hilbe R, Schaefer B, Wagner S, Mittermaier M, Machleidt F, Garcia C, Ruwwe-Glösenkamp C, Lingscheid T, Bosquillon de Jarcy L, Stegemann MS, Pfeiffer M, Jürgens L, Denker S, Zickler D, Enghard P, Zelezniak A, Campbell A, Hayward C, Porteous DJ, Marioni RE, Uhrig A, Müller-Redetzky H, Zoller H, Löffler-Ragg J, Keller MA, Tancevski I, Timms JF, Zaikin A, Hippenstiel S, Ramharter M, Witzenrath M, Suttorp N, Lilley K, Mülleder M, Sander LE, PA-COVID-19 Study group, Ralser M, Kurth F. A time-resolved proteomic and prognostic map of COVID-19. *Cell Syst*. 2021;12(8):780-794.e7.

IF: 11,091 (2021)

5. **Tober-Lau P***, Schwarz T*, Hillus D, Spieckermann J, Helbig ET, Lippert LJ, Thibeault C, Koch W, Bergfeld L, Niemeyer D, Mühlemann B, Conrad C, Kasper S, Münn F, Kunitz F, Jones TC, Suttorp N, Drosten C, Sander LE, Kurth F, Corman VM. Outbreak of SARS-CoV-2 B.1.1.7 Lineage after Vaccination in Long-Term Care Facility, Germany, February-March 2021. *Emerg Infect Dis*. 2021;27(8):2169-2173.

   IF: 16,126 (2021)

4. Schwarz T*, **Tober-Lau P***, Hillus D, Helbig ET, Lippert LJ, Thibeault C, Koch W, Landgraf I, Michel J, Bergfeld L, Niemeyer D, Mühlemann B, Conrad C, Dang-Heine C, Kasper S, Münn F, Kappert K, Nitsche A, Tauber R, Schmidt S, Kopankiewicz P, Bias H, Seybold J, von Kalle C, Jones TC, Suttorp N, Drosten C, Sander LE, Corman VM, Kurth F. Delayed Antibody and T-Cell Response to BNT162b2 Vaccination in the Elderly, Germany. *Emerg Infect Dis*. 2021;27(8). doi:10.3201/eid2708.211145

   IF: 16,126 (2021)

3. Thibeault C*, Mühlemann B*, Helbig ET, Mittermaier M, Lingscheid T, **Tober-Lau P**, Meyer-Arndt LA, Meiners L, Stubbemann P, Haenel SS, Bosquillon de Jarcy L, Lippert L, Pfeiffer M, Stegemann MS, Roehle R, Wiebach J, Hippenstiel S, Zoller T, Müller-Redetzky H, Uhrig A, Balzer F, von Kalle C, Suttorp N, Jones TC, Drosten C, Witzenrath M, Sander LE, Pa-COVID Study Group, Corman VM, Kurth F. Clinical and virological characteristics of hospitalised COVID-19 patients in a German tertiary care centre during the first wave of the SARS-CoV-2 pandemic: a prospective observational study. *Infection*. Published online April 22, 2021. doi:10.1007/s15010-021-01594-w

   IF: 7,455 (2021)

2. Azam TU, Shadid HR, Blakely P, O'Hayer P, Berlin H, Pan M, Zhao P, Zhao L, Pennathur S, Pop-Busui R, Altintas I, Tingleff J, Stauning MA, Andersen O, Adami ME, Solomonidi N, Tsilika M, **Tober-Lau P**, Arnaoutoglou E, Keitel V, Tacke F, Chalkias A, Loosen SH, Giamarellos-Bourboulis EJ, Eugen-Olsen J, Reiser J, Hayek SS, International Study of Inflammation in COVID-19. Soluble Urokinase Receptor (SuPAR) in COVID-19-Related AKI. *J Am Soc Nephrol*. 2020;31(11):2725-2735.

   IF: 10,121 (2020)

1. Frölich AM, **Tober-Lau P**, Schönfeld M, Brehm TT, Kurth F, Vinnemeier CD, Addo MM, Fiehler J, Rolling T. Brain magnetic resonance imaging in imported malaria. *Malar J*. 2019;18(1):74.

   IF: 1,780 (2018)

12 Danksagung

Allen voran möchte ich Florian Kurth danken, der mich nicht nur über Jahre unermüdlich betreut hat, sondern in mir eine Freude am wissenschaftlichen Arbeiten hervorgebracht hat, von der nicht einmal ich wusste, dass sie existiert. Die Möglichkeit, mich sowohl klinisch als auch im Labor auszutoben, Ideen und Gedanken zu (fast) jeder Tages- und Nachtzeit auszutauschen und den Blick über den Tellerrand zu wagen – sei es durch das Abtauchen in andere Fachrichtungen oder ferne Länder – kombiniert mit einer inspirierenden Ruhe und Fähigkeit, auch in den frustrierendsten Situationen noch etwas Gutes zu sehen, haben den wohl besten Grundstein für den weiteren Werdegang in der Forschung gelegt, den ich mir vorstellen kann.

Auch Leif Erik Sander möchte ich danken, zum einen für die offene und uneingeschränkte Einbindung in seine Arbeitsgruppe, zum anderen dafür, dass er stets im entscheidenden Moment mit neuen Ideen und einem frischen Blick zur Verfügung stand. Sein umfangreiches Fachwissen sowie die Eloquenz, mit der er dieses auch zu Papier bringt, ist ein Vorbild für junge Wissenschaftlerinnen und Wissenschaftler.

Mein besonderer Dank gilt den Kolleginnen und Kollegen am Institut für Biochemie – insbesondere Vadim Demichev, Johannes Hartl, Oliver Lemke und Markus Ralser sowie Michael Mülleder von der Core Facility High Throughput Mass Spectrometry – die mir mit der Plasma-Proteomik einen solch tiefen Einblick in ein mir vorher kaum bekanntes Forschungsfeld ermöglicht haben, und nie müde wurden, meine Fragen zu Workflows, bioinformatischer Aufbereitung und Statistik auf laienverständliche Art und Weise zu erklären, Metadaten zu hinterfragen und wissenschaftlichen Input zu geben.

Ich möchte Marcel Schmude dafür danken, dass er mich vor all diesen Jahren überhaupt erst zu Florian Kurths Arbeitsgruppe geführt hat; Friederike Münn, die mir stets mit den Erfahrungen ihrer eigenen Promotionsarbeit und allem Drumherum beistand; meinem Bruder Salo, der mit scharfem Blick auch noch den (hoffentlich) letzten Zahlendreher ausmerzte. Alle drei hatten stets ein offenes Ohr für die Hochs und Tiefs und waren zu jeglicher Unterstützung bereit. Auch meiner Mutter Kerstin möchte ich dafür danken, dass sie meine Gedanken hinterfragte, wenn ich den Wald vor lauter Bäumen nicht sah, und mir stets die Überzeugung mitgab, dass der von mir gewählte Weg schon der richtige sei.

Abschließend möchte ich allen meinen hier aus Platzgründen nicht namentlich erwähnten Kolleginnen und Kollegen, meiner Großmutter und allen weiteren Verwandten, Freundinnen und Freunden danken, die über all die Jahre nie aufgegeben haben zu fragen, wann ich denn nun endlich mit der Doktorarbeit fertig sei, und mir mit Rat und Tat beiseite standen, um dieses Ziel endlich zu erreichen.