Probing long COVID through a proteomic lens: a comprehensive two-year longitudinal cohort study of hospitalised survivors

Xiaoying Gu,[a,b,r] Siyuan Wang,[c,r] Wanying Zhang,[d,r] Caihong Li,[e,f,r] Li Guo,[g,h,r] Zai Wang,[b,i] Haibo Li,[b,d] Hui Zhang,[c,j] Yuhan Zhou,[k] Weijian Liang,[k] Hui Li,[b,c] Yan Liu,[c,l] Yeming Wang,[b,c] Lixue Huang,[m] Tao Dong,[n,o] Dingyu Zhang,[e,f,p,***,s] Catherine C. L. Wong,[c,q,**,s] and Bin Cao[b,d,j,q,s,*]

[a]National Center for Respiratory Medicine; State Key Laboratory of Respiratory Health and Multimorbidity; National Clinical Research Center for Respiratory Diseases; Institute of Respiratory Medicine, Chinese Academy of Medical Sciences; Department of Clinical Research and Data Management, Center of Respiratory Medicine, China-Japan Friendship Hospital, Beijing, PR China
[b]Changping Laboratory, Beijing, PR China
[c]State Key Laboratory of Complex Severe and Rare Diseases, Clinical Research Institute, Peking Union Medical College Hospital, Chinese Academy of Medical Science & Peking Union Medical College, Beijing, PR China
[d]National Center for Respiratory Medicine; State Key Laboratory of Respiratory Health and Multimorbidity; National Clinical Research Center for Respiratory Diseases; Institute of Respiratory Medicine, Chinese Academy of Medical Sciences; Department of Pulmonary and Critical Care Medicine, Center of Respiratory Medicine, China-Japan Friendship Hospital, Beijing, PR China
[e]Wuhan Jinyintan Hospital, Tongji Medical College of Huazhong University of Science and Technology, Wuhan, Hubei Province, PR China
[f]Wuhan Research Center for Communicable Disease Diagnosis and Treatment, Chinese Academy of Medical Sciences, Wuhan, Hubei Province, PR China
[g]Key Laboratory of Respiratory Disease Pathogenomics, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, PR China
[h]NHC Key Laboratory of Systems Biology of Pathogens and Christophe Merieux Laboratory, Institute of Pathogen Biology, Chinese Academy of Medical Sciences, Beijing, PR China
[i]Institute of Clinical Medical Sciences, China-Japan Friendship Hospital, Beijing, PR China
[j]Department of Pulmonary and Critical Care Medicine, Capital Medical University, Beijing, PR China
[k]Foreseen Biotechnology, Beijing, PR China
[l]Department of Infectious Disease, The Affiliated Yantai Yuhuangding Hospital of Qingdao University, Yantai, PR China
[m]Department of Respiratory and Critical Care Medicine, Beijing Hospital, National Center of Gerontology, Institute of Geriatric Medicine, Chinese Academy of Medical Sciences, Beijing, PR China
[n]Chinese Academy of Medical Science (CAMS) Oxford Institute (COI), University of Oxford, Oxford, UK
[o]MRC Human Immunology Unit, MRC Weatherall Institute of Molecular Medicine, University of Oxford, Oxford, UK
[p]Hubei Clinical Research Center for Infectious Diseases, Wuhan, Hubei Province, PR China
[q]Tsinghua University-Peking University Joint Center for Life Sciences, Beijing, PR China

eBioMedicine
2023;98: 104851
Published Online xxx
https://doi.org/10.1016/j.ebiom.2023.104851

Summary

Background As a debilitating condition that can impact a whole spectrum of people and involve multi-organ systems, long COVID has aroused the most attention than ever. However, mechanisms of long COVID are not clearly understood, and underlying biomarkers that can affect the long-term consequences of COVID-19 are paramount to be identified.

Methods Participants for the current study were from a cohort study of COVID-19 survivors discharged from hospital between Jan 7, and May 29, 2020. We profiled the proteomic of plasma samples from hospitalised COVID-19 survivors at 6-month, 1-year, and 2-year after symptom onset and age and sex matched healthy controls. Fold-change of >2 or <0.5, and false-discovery rate adjusted *P* value of 0.05 were used to filter differentially expressed proteins (DEPs). In-genuity pathway analysis was performed to explore the down-stream effects in the dataset of

*Corresponding author. National Center for Respiratory Medicine; State Key Laboratory of Respiratory Health and Multimorbidity; National Clinical Research Center for Respiratory Diseases; Institute of Respiratory Medicine, Chinese Academy of Medical Sciences; Department of Pulmonary and Critical Care Medicine, Center of Respiratory Medicine, China-Japan Friendship Hospital, No. 2 East Yinghuayuan St, Beijing 100029, PR China.
**Corresponding author. Department of Medical Research Center, State Key Laboratory of Complex Severe and Rare Diseases, Peking Union Medical College Hospital (Dongdan Campus), Chinese Academy of Medical Sciences & Peking Union Medical College, No. 1 Shuaifuyuan Wangfujing, Beijing, 100730, PR China.
***Corresponding author. Wuhan Jinyintan Hospital, Tongji Medical College of Huazhong University of Science and Technology; Wuhan Research Center for Communicable Disease Diagnosis and Treatment, Chinese Academy of Medical Sciences, No. 1 Yintan St, Wuhan, 430023, PR China.
E-mail addresses: caobin_ben@163.com (B. Cao), catclw321@126.com (C.C.L. Wong), 1813886398@qq.com (D. Zhang).
[r]These authors contributed equally.
[s]These authors also contributed equally.

significantly up- or down-regulated proteins. Proteins were integrated with long-term consequences of COVID-19 survivors to explore potential biomarkers of long COVID.

Findings The proteomic of 709 plasma samples from 181 COVID-19 survivors and 181 matched healthy controls was profiled. In both COVID-19 and control group, 114 (63%) were male. The results indicated four major recovery modes of biological processes. Pathways related to cell–matrix interactions and cytoskeletal remodeling and hypertrophic cardiomyopathy and dilated cardiomyopathy pathways recovered relatively earlier which was before 1-year after infection. Majority of immune response pathways, complement and coagulation cascade, and cholesterol metabolism returned to similar status of matched healthy controls later but before 2-year after infection. Fc receptor signaling pathway still did not return to status similar to healthy controls at 2-year follow-up. Pathways related to neuron generation and differentiation showed persistent suppression across 2-year after infection. Among 98 DEPs from the above pathways, evidence was found for association of 11 proteins with lung function recovery, with the associations consistent at two consecutive or all three follow-ups. These proteins were mainly enriched in complement and coagulation (COMP, PLG, SERPINE1, SRGN, COL1A1, FLNA, and APOE) and hypertrophic/dilated cardiomyopathy (TPM2, TPM1, and AGT) pathways. Two DEPs (APOA4 and LRP1) involved in both neuron and cholesterol pathways showed associations with smell disorder.

Interpretation The study findings provided molecular insights into potential mechanism of long COVID, and put forward biomarkers for more precise intervention to reduce burden of long COVID.

Funding National Natural Science Foundation of China; Chinese Academy of Medical Sciences (CAMS) Innovation Fund for Medical Sciences; Clinical Research Operating Fund of Central High Level Hospitals; the Talent Program of the Chinese Academy of Medical Science; Training Program of the Big Science Strategy Plan; Ministry of Science and Technology of the People's Republic of China; New Cornerstone Science Foundation; Peking Union Medical College Education Foundation; Research Funds from Health@InnoHK Program.

Keywords: Long COVID; Proteomic; Recovery modes; Lung function

Research in context

Evidence before this study

We searched PubMed for COVID-19 proteomics studies up to June 30, 2023 without language restriction. The search terms were (COVID-19 OR SARS-CoV-2 OR Coronavirus disease 2019 OR 2019-ncov) AND (proteomic). Majority of previous proteomics studies focused on COVID-19 disease severity. Others have been conducted with relatively short follow-up duration, and integration with long-term consequence data was limited. Thus, it is paramount to investigate potential mechanisms of long COVID, and reveal the underlying important proteins that may affect the long-term consequences of COVID-19 in a longer time dimension.

Added value of this study

This study uncovered changes in proteomic landscape across 2 years after SARS-CoV-2 infection, the time frame of which is the longest to the best of our knowledge. Four major recovery modes of biological process were indicated. The pathways including focal adhesion, ECM-receptor interaction, and regulation of action cytoskeleton pathways involved in cellular biology and hypertrophic cardiomyopathy and dilated cardiomyopathy pathways involved in cardiovascular system recovered between 6-month and 1-year after symptom onset which was relatively earlier. Majority of immune response pathway, complement and coagulation cascade, and cholesterol metabolism returned to similar status of matched healthy controls later and before 2-year after infection. Fc receptor signaling pathway still did not recover at 2-year follow-up, which may be mainly related to the long lasting antibody response. Pathways related to neuron generation and differentiation showed persistent suppression across 2-year after infection, with recovery rate becoming slow after 1-year follow-up. Among 98 differentially expressed proteins from the above pathways, 11 showed consistent association with lung function at least two consecutive follow-ups. These key proteins were mainly enriched in complement and coagulation and hypertrophic/dilated cardiomyopathy pathways.

Implications of all the available evidence

The study identified four recovery modes of different biological processes among COVID-19 survivors across 2 years after infection, and thus provided molecular insights into the potential mechanism of long COVID. In a context that diagnosis and treatment of long COVID are both challenging, the findings put forward biomarkers for more precise intervention to reduce burden of long COVID.

Introduction

Since its emergence, COVID-19 has spread around the globe and had a profound impact on health and economy. As of July 7, 2023, over 767 million confirmed COVID-19 cases and over 6.9 million deaths have been reported worldwide.[1] Recent studies have shown that a wide range of persistent symptoms can remain long after acute SARS-CoV-2 episode, and this phenomenon is referred to as long COVID by World Health Organization.[2–5] According to our longitudinal cohort from Wuhan, China, 55% of hospitalised COVID-19 survivors still reported at least one symptom of long COVID at 2-year after symptom onset, even though the proportion has declined compared to that at 6-month.[6–8] After more than three years of combating with SARS-CoV-2 pandemic, the large number of patients with long COVID is noteworthy.[9]

Long COVID is a debilitating condition that can impact a whole spectrum of people and involve multi-organ systems.[10–12] Considering potentially strained health resources and large economic burden, long COVID has aroused the most attention than ever. Previous studies indicated disease severity and sex at acute phase as joint risk factors of long COVID among hospitalised and home-isolated COVID-19 patients.[8,13] The hypothesised pathophysiologic mechanisms of long COVID included residual virus in tissue, autoimmunity resulting from cross reactivity of SARS-CoV-2-specific antibodies with host proteins, and imbalance in renin-angiotensin system.[14,15] However, these hypotheses needs to be further demonstrated in mechanistic and large cohort studies.

Up to now, the diagnosis and treatment of long COVID are still challenging, not to mention biomarkers for early identification. C-reactive protein, D-dimer, leukocytes, interleukin-6 and lactate dehydrogenase were shown to be higher among people with long COVID.[16] However, these findings were driven by studies with less than 6-month follow-up and provided limited information for physiological mechanisms of long COVID. The more promising biomarkers of long COVID needs to be explored systematically, which could be achieved by proteomics studies. Identification and functional analysis of differentially expressed proteins (DEPs) could provide evidence for the recovery situation of biological process among COVID-19 survivors. Further exploring relationships between DEPs and long-term consequences can help more specifically recognise effect of unrecovered biological process on long COVID phenotypes. Some proteomics studies have been conducted with relatively small sample size and short follow-up duration, and integration with long-term consequence data was limited.[17–21] Thus, it is paramount to investigate potential mechanisms of long COVID, and reveal the underlying important proteins that may affect the long-term consequences of COVID-19 in a longer time dimension.

Here we proposed a framework for depicting the proteomic landscape of hospitalised COVID-19 survivors across 2 years after symptom onset, and understand the role of altered proteins after SARS-CoV-2 infection for long COVID using proteomic and artificial intelligence methods. To address the knowledge gaps, we profiled the proteins levels of plasma samples from hospitalised COVID-19 survivors at 6-month, 1-year, and 2-year after symptom onset and matched healthy controls. The proteins were integrated with long-term consequences of COVID-19 survivors collected from our 2-year cohort study to identify potential biomarkers for long COVID.

Methods

Study design and participants

Participants for the current study were from a cohort study of COVID-19 survivors discharged from hospital between Jan 7, and May 29, 2020. Inclusion and exclusion criteria of the cohort study have been described previously.[22] Briefly, all patients with laboratory confirmed COVID-19 discharged from hospital between Jan 7 and May 29, 2020, were eligible for participation. Patients were excluded if they died after discharge and before first follow-up; were living in a nursing or welfare home; had psychotic disorder, dementia, or osteoarthropathy; or were immobile. Three follow-up surveys were conducted at 6 months, 1 year, and 2 years after symptom onset.

A stratified disproportional random sampling procedure according to severity scale was used to select COVID-19 survivors to undergo pulmonary function test at 6-month follow-up visit. Patients requiring HFNC, NIV or IMV (severity scale $\geq$5) were all invited to receive the pulmonary function test. The ratio used to select patients not requiring supplemental oxygen (severity scale as 3) and those requiring supplemental oxygen (severity scale as 4) was 1:2. Finally, 516 patients were ascertained as the eligible patients to receive the above tests. A total of 349 survivors had completed pulmonary function tests at the 6-month visit,[22] and they were all invited to perform this test again at 1-year and 2-year visit. Among the sampling participants, COVID-19 survivors who completed 6-month, 1-year, and 2-year follow-up, with data for lung and kidney function, and with plasma samples at three follow-up visits were all included in the current study (n = 181). None of these 181 COVID-19 survivors was reinfected by SARS-CoV-2 after discharge and before 2-year follow-up.

To determine the recovery of COVID-19 survivors, we recruited community-dwelling adults without SARS-CoV-2 infection (non-COVID-19 cohort) from two districts of Wuhan city between Dec 24, 2020, and Jan 16, 2021. The inclusion and exclusion criteria have also been described previously.[7,8] A subgroup of participants

from non-COVID-19 cohort were invited to perform pulmonary function tests during 2-year follow-up visit. The 181 COVID-19 survivors were further matched (1:1) by age and sex to healthy controls who received pulmonary function tests and had plasma samples. Coarsened exact matching, a nonparametric method that matches based on bins of characteristics were used. The matching variables included age in 10-year bins and sex. Age of COVID-19 survivors and healthy controls at 2-year follow-up was used for matching. Each of 181 COVID-19 survivors was successfully matched with one community control. The plasma samples of community controls used for proteomic profiling were collected during 2-year follow-up visit for COVID-19. The community controls were not infected by SARS-CoV-2 when plasma samples were collected.

The study was approved by the Research Ethics Commission of hospital which enrolled the participants (KY-2020-78.01, KY-2020-78.03). Written informed consent was obtained from COVID-19 survivors who attended the follow-up visit and healthy controls.

Data collection at acute phase of COVID-19 survivors

Acute phase was defined as the time between symptom onset and hospital discharge. Data at acute phase was retrieved from electronic medical records, including demographic characteristics (age, sex, and cigarette smoking); clinical characteristics (self-reported comorbidities and symptom onset time); disease severity characterised by highest seven–category scale during hospital stay (3, hospitalised, not requiring supplemental oxygen; 4, hospitalised, requiring supplemental oxygen; 5, hospitalised, requiring high-flow nasal cannula (HFNC), noninvasive mechanical ventilation (NIV), or both; 6, hospitalised, requiring extracorporeal membrane oxygenation (ECMO), invasive mechanical ventilation (IMV), or both); and treatment (corticosteroids, antivirals including lopinavir–ritonavir, arbidol, chloroquine phosphate, and hydroxychloroquine, antibiotics, thymosin, and intravenous immunoglobulin).

Follow-up assessment of COVID-19 survivors

Eligible study participants were invited to attend face-to-face follow-up visits at the outpatient clinic of hospital at 6 months, 1 year, and 2 years after symptom onset. The 6-month, 1-year, and 2-year follow-up visits were conducted from June 16 to Sept 3, 2020, from Dec 16, 2020 to Feb 7, 2021, and from Nov 16, 2021 to Jan 10, 2022, respectively. A telephone survey was available for COVID-19 survivors at the 2-year follow-up visit as an alternative to the face-to-face interview, which was conducted by trained clinicians using the same questionnaires. The detailed 6-month 1-year, and 2-year follow-up procedures have been described previously.[7,8,22] Briefly, at each visit, they underwent a detailed interview, a physical examination, and a 6-min walking distance (6MWD) test; completed a series of questionnaires, including a self-reported symptom questionnaire, the modified British Medical Research Council (mMRC) dyspnea scale, the EQ-5D-5L questionnaire to assess HRQoL, the EuroQol Visual Analogue Scale (EQ-VAS; scores range from 0 to 100, with a higher score indicating a better health status) et al.[23–26] Core symptoms more specifically related to long COVID included fatigue or muscle weakness, smell disorder, taste disorder, and dysponea defined as mMRC $\geq$1.[27,28] Additionally, at the 1-year and 2-year visits, healthcare use after discharge was also collected by a questionnaire. Venous blood samples were drawn for the measurement of laboratory indicators and plasma isolation. The pulmonary function test was conducted using the Master Screen PFT (vyaire Medical GmbH, Hoechberg, Germany) according to American Thoracic Society (ATS) guidelines.[29]

In light of the emergency state of hospital at the early stage of COVID-19 pandemic, demographic information and self-reported comorbidity collected at admission may not be that accurate. To confirm demographic information and self-reported comorbidity collected at baseline, a standard questionnaire was designed and administered to obtain information including age, sex, education, cigarette smoking, alcohol consumption, personal medical history, and family history face-to-face by trained staff at 1-year and 2-year follow-up visit. All data collected was checked for completeness by the staff who collected information unless study participants were unwilling to provide the information. The data was finally recoded to reflect baseline characteristics of study participants according to time variables collected for some variables such as comorbidity and admission date.

Plasma isolation

Venous blood was collected from participants and processed within 12 h to isolate plasma. Plasma was separated by centrifugation at 300×*g* for 10 min and stored at –80 °C until testing.

Plasma sample preparation for DIA analysis and spectral library generation

Plasma samples were prepared in accordance with serum samples as described in our previous studies.[19,20] Purified peptides from all samples were collected as a mixture of 100 μg and processed for spectral library generation.

High-pH reversed-phase fractionation

Peptide fractionation was performed at pH 10 on a chromatographic system (Waters Xevo ACQUITY UPLC, USA). The purified peptides were collected every 60 s within 80 min and 62 fractions were obtained. The fractions were then mixed in pairs as a total of 31 samples, which were dissolved in 10 μL Milli-Q water

with 0.1% formic acid (FA) after drying process and spiked with iRT peptides before DIA analysis.

Liquid chromatography

A nanoElute liquid chromatography system (Bruker Daltonics) was applied for peptide separation at a flow rate of 300 nL/min. Water and ACN with 0.1% formic acid were mobile phase A and B, respectively. A non-linear gradient separation process was set to firstly increase from B% at 2%–22% within the time range of 70 min, secondly increase from B% at 22%–37% within 8 min, thirdly increase from B% at 37%–100% within 5 min, and finally maintained at 100% for the last 7 min before re-equilibration.

Mass spectrometry

All of the 31 fraction samples above were analysed on timsTOF Pro2 (Bruker). A CaptiveSpray nano-electrospray ion source was applied to interface LC with mass spectrometer. Generation of the spectral library was operated in data-dependent mode, with the accumulation and ramp time ranging from m/z 100 to 1700 in positive electrospray mode. The sample analysis was performed in data-independent mode, with an MS1 scan and 64 MS2 windows in a diaPASEF acquisition scheme covering the mass range from m/z 400 to m/z 1200. During the scanning mode, the ion mobility was set from 0.6 to 1.6 Vs/cm^2, and the collision energy was descented linearly as a function of the mobility from 59 eV at 1/K0 = 1.6 Vs/cm^2 to 20 eV at 1/K0 = 0.6 Vs/cm^2.

Generation of spectral libraries and DIA data analysis

Spectral libraries were generated with Spectronaut version 14.2 (Biognosys) against a UNIPROT human database (only reviewed entries) and the SARS-CoV-2 UNIPROT database. All the parameters were default.

Statistical analysis

Demographic and clinical characteristics of study participants

Demographic and clinical characteristics of hospitalised COVID-19 survivors and matched healthy controls, and clinical outcomes of hospitalised COVID-19 survivors at 6-month, 1-year, and 2-year after symptom onset were presented as mean (SD) or median (IQR) for continuous variables and expressed as absolute values along with percentages for categorical variables. For the comparison of characteristics between COVID-19 survivors and healthy controls, we used standardised mean differences (SMDs). For the comparison of symptoms, exercise capacity, health-related quality of life, healthcare use after discharge, lung function, and laboratory tests between different follow-up visits, we used paired t test, Wilcoxon signed-rank test, or McNemar test where appropriate. The assumption of symmetrical distribution for continuous variables and for difference of continuous variables between different follow-up visits was tested to determine the use of paired t test or Wilcoxon signed-rank test.

PCA and DEP analysis

Omicsbean software was applied for data imputation. The imputation model applied for proteomics data imputation was in accordance with our previous study.[30] We first used locally weighted polynomial regression (lowess in R v.3.6.3) to compute the local polynomial fit for protein number and protein-detecting rate in each group. Two boundary thresholds, 0.15 and 0.5, were used to separate the data into three parts. When a protein-detecting rate is <0.15, it is probably because the detected value is due to a technical error. For these proteins, no imputation was applied. When a protein-detecting rate is > 0.5, the missing value was probably due to the detection accuracy limitation of the LC/MS. In this case, the missing value was replaced with a median value. When a protein-detecting rate is between 0.15 and 0.5, it is probably because the protein expression is unstable for detection. Where PBA is the group missing rate and PA the total missing rate of each protein.

Then, we determined the predicted imputation number (IN) of each protein in each group,

$$\text{missp} = \text{PA} \times \left(\frac{\text{PBA}}{(\text{PBA} \times \text{PA}) + 0.05 \times (1 - \text{PA})} \right)$$

where PBA is the group missing rate and PA the total missing rate of each protein.

Then we determined the predicted imputation number (IN) of each protein in each group,

$$\text{IN} = (1 - \text{missp}) \times \text{Mi}$$

where Mi is the number of undetected sample number of a protein in group i.

Finally, the random method was used to determine the samples to be imputed. The imputation value was then defined by,

$$\text{IV} = \min\left(\frac{\text{Mi}}{2}, \text{IN} \right)$$

Imputed data were then normalised using LogNorm algorithm. PCA (muma v1.4 package, https://www.rdocumentation.org/packages/muma) was used to conduct clustering analysis of the samples. R package Genefilter (https://www.rdocumentation.org/packages/genefilter/versions/1.54.2) was applied to calculate the fold-change values of proteins. In this study, a fold change >2 or <0.5, and false-discovery rate adjusted P value (t test) < 0.05 were set for differential expression proteins filtration.

Pathway and hierarchical clustering analysis

GO database (https://www.ebi.ac.uk/QuickGO/) and KEGG pathway database (https://www.kegg.jp/kegg/pathway.html) were applied for gene ontology and pathway enrichment analysis. Ggplot2 packages and Cytoscape v.3.5.1 implemented in the Omicsbean workbench were used to achieve Venn diagram, heatmap, and network visualization. Ingenuity pathway analysis was performed to explore the downstream effect in significantly up- or down-regulated protein datasets. The z-score algorithm was used to predict the regulatory mode (either activated or inhibited) of the biological process. Mfuzz v.2.46.0 (https://www.bioconductor.org/packages/release/bioc/html/Mfuzz.html) was used for exploring various sub-clustering models of gene expression among different groups. The type of hierarchical clustering method is agglomerative.

Machine learning

The random forest (RF) model is a widely-used machine learning approach that constructs multiple decision trees and combines them using standardised protein markers to achieve optimal classification performance. This method provides weight coefficients for each marker in the classification model, enabling interpretation of their relative importance. The majority voting method is used in RF to make a final decision. To construct the RF model, a training set $(x_i, y_i)_{i=1}^{n}$ (x_i represents the training data and y_i represents its label, n represents the number of samples in the training set), a set of m decision trees were built with individual weight functions W_j with the individual protein markers as each tree leaf j. The decision trees are constructed by selecting a random subset of the training set for each tree, thereby reducing overfitting. The splitting criterion for each tree node is chosen based on a random subset of the protein markers.

Once the RF model is constructed, it can be used to predict the label of a new testing set x' with proteomics data. Specifically, the model predicts the label of the time-based group (HC, 0.5Y, 1Y or 2Y) based on the outputs of the individual trees and the majority voting method. This prediction is made by aggregating the outputs of the individual trees, with each tree contributing a vote for the predicted label:

$$\hat{y} = \frac{1}{m}\sum_{j=1}^{m}\sum_{i=1}^{n} W_j(x_i, x')y_i = \sum_{i=1}^{n}\left(\frac{1}{m}\sum_{j=1}^{m} W_j(x_i, x')\right)y_i \qquad (1)$$

Random Forest is an ensemble learning algorithm based on decision trees and can be used for both classification and regression problems. The hyperparameters are as follows:

n_estimators: The number of trees in the forest is 100.

max_depth: The maximum depth of each tree is not limited (None).

max_features: All features are used for each split ("auto").

min_samples_split: Each node needs at least 2 samples for splitting.

min_samples_leaf: Each leaf node must have at least 1 sample.

bootstrap: Bootstrap sampling is used for building each tree (True).

Leave-one-out cross-validation (LOOCV) is a widely-used technique for assessing the performance of machine learning models. LOOCV splits the dataset into a training set of n-1 samples and a testing set of 1 sample. The model is trained on the training set and then used to predict the label of the held-out sample. This process is repeated n times, with each sample held out once. LOOCV provides an unbiased estimate of model performance and is particularly useful when the dataset is small.

Association of proteins with long-term consequences of COVID-19 patients

To identify biomarkers of long COVID and provide potential intervention at earlier time point, we evaluated association of 1370 protein at 6-month after symptom onset with long-term consequences of COVID-19 survivors at 6-month, 1-year, and 2-year follow-up. To find important biomarkers of long COVID, we further focused on 98 DEPs involved in immune system, complement and coagulation, neurological system, cellular activity, hypertrophic/dilated cardiomyopathy, and cholesterol metabolism. The imputed proteomics data was z-score standardised for further association analysis.

The associations of proteins at 6-month follow-up with lung function (DLCO of predicted value, TLC of predicted value, and RV of predicted value), percentage of predicted value for 6-min walking distance,[31] lymphocyte count, monocyte count, neutrophil count, neutrophil to lymphocyte count, eGFR, HbA1c, and triglyceride at 6-month, 1-year, and 2-year after symptom onset were assessed with multivariable adjusted generalised linear regression model. In the generalised linear regression model, normal response probability distribution with the identity link function was used. The linear relationship between continuous independent variables and dependent variables was assessed. Multivariable adjusted logistic regression models were used to explore association of proteins with core symptoms of long COVID (fatigue or muscle weakness, smell disorder, taste disorder, and dysponea).[27,28] The linearity assumption between continuous independent variables and logit probability of dependent variables was also assessed. OR (95% CI) or β (95% CI) for each long-term consequence were estimated with adjustment for age, sex, cigarette smoking (never-smoker, current or former smoker), education (college or higher vs high school or lower), body mass index (BMI), comorbidity (hypertension, diabetes, coronary heart diseases,

cerebrovascular diseases, malignant tumor, chronic obstructive pulmonary disease, and chronic kidney disease), and disease severity (scale 3, scale 4, scale 5–6). The minimally sufficient set of confounders were identified with directed acyclic graph.

All significance tests were two-sided, and false discovery rate (FDR) adjusted *P* values were employed to determine the statistical significance for the association of 1370 proteins with long-term consequences of COVID-19 patients. These statistical analyses were done with SAS, version 9.4 (SAS Institute, Inc, Cary, NC). Data visualization techniques utilized the ggplot package in R (version 4.1.2).

Role of funding source

The funder of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. All authors had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Results

Demographic and clinical characteristics of study participants

The demographic and clinical characteristics of 181 hospitalised COVID-19 survivors and 181 healthy controls are presented in Table 1. Healthy controls recruited from community-based cohort were matched (1:1) by age and sex with COVID-19 survivors. The mean age of COVID-19 survivors at 2-year follow-up and healthy controls were 57.7 (SD 11.8) and 57.6 (12.0) years, respectively, and 63% of study participants were male. There were minimal differences in characteristics between COVID-19 survivors and matched healthy controls, with SMDs below 0.10 except for smoking status, and history of hypertension and chronic kidney disease. Compared with COVID-19 survivors, matched healthy controls were more likely to be current or former smokers (18% vs 41%; SMD, 0.62). The proportions of hypertension (36% vs 30%; SMD, –0.22) and chronic kidney disease (4% vs 2%; SMD, –0.16) were slightly higher in COVID-19 survivors. Among 181 COVID-19 survivors, 93 (51%) required supplemental oxygen therapy during their hospital stay, and 49 (27%) required HFNC, non-IMV, IMV, or ECMO. 21 (12%) of 181 COVID-19 survivors were admitted to the intensive care unit (ICU). The mean duration from symptom onset to 6-month, 1-year, and 2-year follow-up visit was 193.7 (19.6) days, 356.2 (14.2) days, and 684.8 (14.5) days, respectively.

The long-term consequences of 181 COVID-19 survivors including symptoms, exercise capacity, health-related quality of life, healthcare use after discharge, lung function, and laboratory tests at 6-month, 1-year, and 2-year follow-up are shown in Table S1. The proportions of study participants with fatigue or muscle weakness, smell disorder, dyspnoea, 6-min walking distance less than the lower limit of the normal range (LLN), anxiety or depression, and mobility problem were lower at 2-year follow-up compared to the corresponding proportions at 6-month follow-up.

Plasma proteomic profiling of COVID-19 survivors and healthy controls

Plasma samples from 181 COVID-19 survivors at three time points (6-month, 1-year and 2-year after symptom onset) and 181 matched healthy controls were processed by data-independent acquisition (DIA) mass spectrometry. The proteomic data was successfully obtained for 171 COVID-19 survivors at 6-month follow-up, 178 COVID-19 survivors at 1-year follow-up, and 179 COVID-19 survivors at 2-year follow-up, and for all 181 healthy controls (Fig. 1a).

The proteomics data showed clear stratification of COVID-19 survivors at three follow-up visits and healthy controls according to principal component analysis (PCA) (Fig. 1b–d). The percentages of variance explained by three principle components are 51.5%, 6.3%, 4.2% for COVID-19 survivor at 6-month follow up and controls, 54.5%, 8.8%, 5.3% for COVID-19 survivor at 1-year follow up and controls, and 46.6%, 12%, 7% for COVID-19 survivor at 2-year follow up and controls, respectively. A total of 1370 proteins were identified from all samples. For the DEP analysis, 249 proteins levels were different (fold change >2 or <0.5 and FDR adjusted *P* value <0.05) between COVID-19 survivors and healthy controls (Fig. 1e). 172, 109, and 73 proteins were different at 6-month, 1-year, and 2-year after symptom onset among COVID-19 survivors compared with healthy controls, respectively (Fig. 1e, Figure S1). Of these protein, 70, 54, and 32 were up-regulated proteins, and 102, 55, and 41 were down-regulated at corresponding three follow-up visits. Among 249 DEPs, 176 were identified as recovered proteins at 2-year follow-up among COVID-19 survivors compared with healthy controls, whereas 73 were unrecovered (Fig. 1e).

To systematically understand dynamic change of proteins detected in plasma sample among COVID-19 survivor, clustering analysis was performed on 1370 proteins to analyse their variation among healthy controls, and COVID-19 survivors at 6-month, 1-year, and 2-year follow-up. Four clusters were identified, with first two clusters mainly containing down-regulated proteins, and the other two clusters mainly containing up-regulated proteins after COVID-19 (Figure S2, Figure S3). Cluster 1 was featured to contain proteins with levels significantly decreased among COVID-19 survivors at 1-year after symptom onset compared with healthy controls, whereas the levels at 6-month and 2-year were not substantially changed. According to the trajectory of protein in cluster 2, it mainly contained down-regulated proteins among COVID-19 survivors at 6-month after symptom onset, with the protein levels recovered to similar level as healthy controls at 2-year after symptom onset. Cluster 3 mainly contained

	COVID-19 survivors[a] (n = 181)	Matched healthy controls (n = 181)	SMD
Age, years	57.7 ± 11.8	57.6 ± 12.0	0.01
Sex			0.00
Male	114 (63%)	114 (63%)	
Female	67 (37%)	67 (37%)	
Cigarette smoking			0.62
Never-smoker	149 (82%)	107 (59%)	
Current or former smoker	32 (18%)	74 (41%)	
BMI, kg/m^2	24.7 ± 3.3	24.3 ± 2.8	−0.06
Comorbidity			
Hypertension	65 (36%)	55 (30%)	−0.22
Diabetes	18 (10%)	23 (13%)	−0.04
Coronary heart diseases	14 (8%)	13 (7%)	−0.07
Cerebrovascular diseases	5 (3%)	5 (3%)	−0.04
Chronic kidney disease	8 (4%)	4 (2%)	−0.16
Malignancy	2 (1%)	4 (2%)	0.10
COPD	0 (0%)	3 (2%)	NA
Autoimmune diseases	0 (0%)	1 (1%)	NA
Highest seven-category scale during hospital stay			
Scale 3: not requiring supplemental oxygen	39 (22%)	NA	NA
Scale 4: requiring supplemental oxygen	93 (51%)	NA	NA
Scale 5: requiring HFNC or non-IMV or both	45 (25%)	NA	NA
Scale 6: requiring ECMO or IMV, or both	4 (2%)	NA	NA
Treatment received during hospital stay			
Corticosteroids	65 (36%)	NA	NA
Antivirals	99 (55%)	NA	NA
Lopinavir-ritonavir	52 (29%)	NA	NA
Arbidol	63 (35%)	NA	NA
Antibiotics	154 (85%)	NA	NA
Thymosin	21 (12%)	NA	NA
Intravenous immunoglobulin	50 (28%)	NA	NA
Length of hospital stay, days	15.0 (12.0–25.0)	NA	NA
ICU admission	21 (12%)	NA	NA
Length of ICU stay, days	21.0 (8.0–46.0)	NA	NA
Time from symptom onset to 6-month follow-up, days	193.7 ± 19.6	NA	NA
Time from symptom onset to 12-month follow-up, days	356.2 ± 14.2	NA	NA
Time from symptom onset to 2-year follow-up, days	684.8 ± 14.5	NA	NA

Data are n (%), n/N (%), mean ± SD, or median (IQR). The differing denominators used indicate missing data. SMD = standardized mean difference. NA = not applicable. COPD = chronic obstructive pulmonary disease. HFNC = high-flow nasal cannula for oxygen therapy. IMV = invasive mechanical ventilation. ECMO = extracorporeal membrane oxygenation. ICU = intensive care unit. [a]Age at 2-year follow-up and sex of COVID-19 were matched with healthy controls. The baseline characteristics except for age at 2-year follow-up were shown for COVID-19 patients.

Table 1: **Characteristics of hospitalized COVID-19 survivors and matched healthy controls.**

proteins with levels significantly increased among COVID-19 survivors at 6-month after symptom onset, and recovered between 6-month and 1-year follow-up. Cluster 4 mainly contained up-regulated proteins after COVID-19 which recovered more slowly than those in cluster 3.

Dynamics of immune response and recovery of complement and blood coagulation system across 2-year after SARS-CoV-2 infection

Our data showed majority of DEPs enriched in immune response pathways were immunoglobulins (Fig. 2a). These immunoglobulins were involved in several pathways including regulation of B cell and lymphocyte activation, Fc receptor signaling pathway, and immunoglobulin mediated immune response. The COVID-19 survivors were not vaccinated at 6-month and 1-year follow-up and 81.2% of them were vaccinated at 2-year follow-up. 97.2% (176/181) of healthy controls were vaccinated. The level of most immunoglobulins among COVID-19 survivors returned to comparable level of vaccinated healthy controls before 2-year follow-up, with increased level of immunoglobulin heavy constant alpha 2 (IGHA2), immunoglobulin heavy constant gamma 3

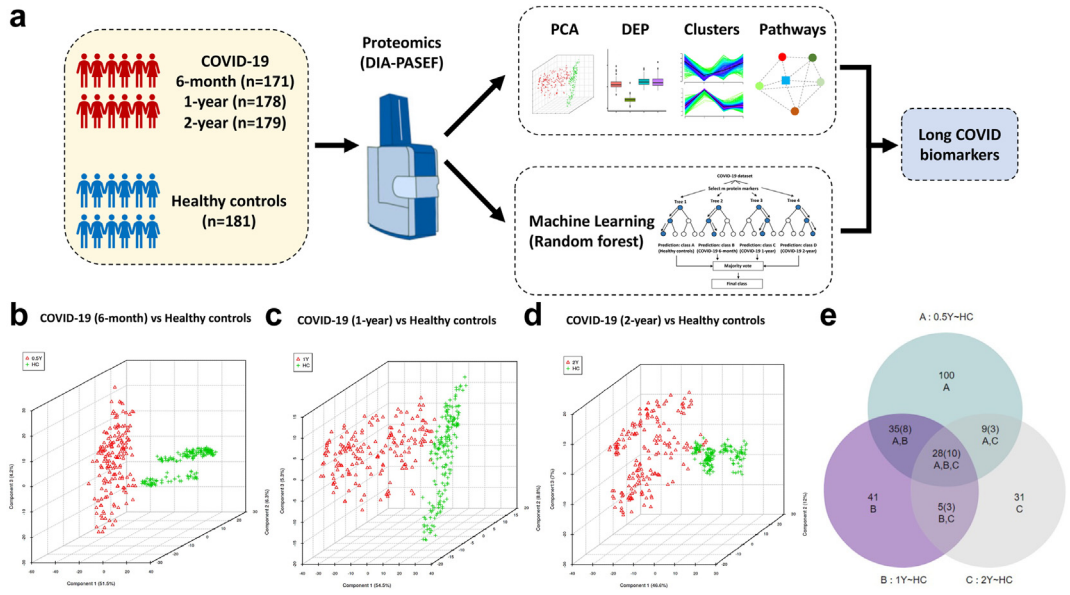

Fig. 1: **Summary of study design and differentially expressed proteins (DEPs) between COVID-19 and healthy controls**. a. An illustration of study design. 181 hospitalized COVID-19 survivors were matched with 181 healthy controls from communities according to age and sex. The plasma samples of COVID-19 survivors were collected at 6-month, 1-year, and 2-year after symptom onset. Proteomics data was successfully obtained for all 181 healthy controls and 171, 178, and 179 COVID-19 patients at 6-month, 1-year, and 2-year, respectively. b–d. Principal components analysis (PCA) showing the inter-group differences between COVID-19 patients at 6-month, 1-year and 2-year follow-up and healthy controls. e. Venn diagram for DEPs between COVID-19 patients at 6-month, 1-year and 2-year follow-up with healthy controls. DEPs met the criteria that fold change >2 or <0.5, two-tailed t-test, false discovery rate adjusted *P* value < 0.05.

(IGHG3), and IGHG4 involved in adaptive immune response, and immunoglobulin heavy variable 3–7 (IGHV3-7), IGHV3-74, IGHV3-21 and immunoglobulin lambda-like polypeptide 1 (IGLL1) involved in innate immune response before then. Other immunoglobulins with continuously increased level across 2 years after SARS-CoV-2 infection included immunoglobulin kappa variable 1D-16 (IGKV1D-16), IGKV2-29, and IGKV4-1 which are all V region of the variable domain of immunoglobulin light chains involved in adaptive immune response.[32] Except for immunoglobulins, proteins such as high mobility group protein B1 (HMGB1), galectin-1 (LGALS1) and CYFIP-related Rac1 interactor B (FAM49B) in regulation of lymphocyte activation pathway recovered before 2-year follow-up, whereas protein phosphatase 3 catalytic subunit alpha (PPP3CA) in Fc receptor signaling pathway was still down-regulated at 2-year after COVID-19. HMGB1 is a well-known damage associated molecular patterns (DAMP) that can induce strong inflammatory responses and correlate positively with the severity of COVID-19.[33,34]

GO enrichment analysis showed majority of immune response pathways, including regulation of B cell and lymphocyte activation, antigen receptor-mediated signaling pathway, and immunoglobulin mediated immune response recovered before 2-year after infection (Fig. 2b). Other pathways which were mainly Fc receptor related signaling pathways and immune response-activating cell surface receptor signaling pathway still did not recover at 2-year follow-up.

Hyperactivation of the complement and coagulation systems has been reported as part of the clinical syndrome of COVID-19.[35–39] Complement can induce platelet activation and activation of the coagulation cascade. Coagulation disorders among COVID-19 was found to be potentially associated with multi-organ injury or failure.[40] GO and KEGG enrichment analyses both showed complement and coagulation pathway recovered between 1-year and 2-year after COVID-19 (Fig. 2b, Table S2, Figure S4). Majority of DEPs involved in complement and coagulation cascades pathway, including down-regulated complement component C7, C9, complement factor D (CFD), CFB, talin-1 (TLN1), thrombospondin-1 (THBS1), integrin-linked protein kinase (ILK), plasminogen (PLG), and alpha-1-antitrypsin (SERPINA1), and up-regulated plasminogen activator inhibitor 1 (SERPINE1), alpha-actinin-4 (ACTN4), serglycin (SRGN), vinculin (VCL), and peptidyl-prolylcis-trans isomerase A (PPIA) have recovered between 6-month and 1-year follow-up before the whole pathway recovered, while few DEPs such as complement component C8 alpha chain (C8A) recovered between 1-year and 2-year follow-up compared with healthy controls (Fig. 2c). *C9, PLG, SERPINA1*, and *C8A*

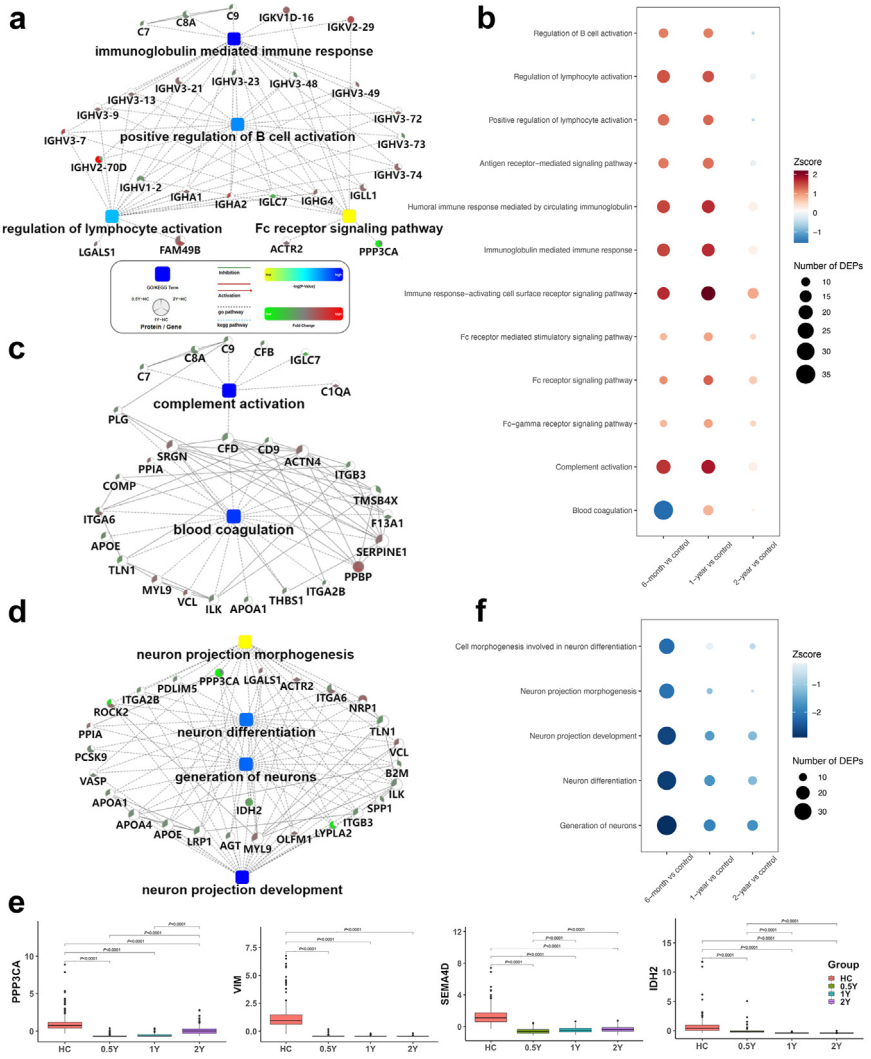

Fig. 2: **Dysregulation and recovery of immune, complement and coagulation, and neurological system among COVID-19 survivors across 2 years after symptom onset**. a, c, d. Interaction diagrams of the indicated pathways for differentially expressed proteins (DEPs) enriched in immune system (a), complement and coagulation (c), and neurological system (d) during 2-year follow-up when plasma proteins from COVID-19 patients at 6-month, 1-year, and 2-year after symptom onset were compared to those from healthy controls. The statistical significance of the pathways, represented by -log (*P* value) (Fisher's exact test), is shown by the color scale, with dark blue representing the highest statistical significance. Color bar from red to green represents the fold change of protein level from increasing to decreasing. Fold change indicates the protein level in COVID-19 patients compared to healthy controls. b, f. Bubble plots of GO pathways involved in immune, complement and coagulation (b) and neurological system (f). e. Boxplot of unrecovered DEPs in neurological system across 2 years after symptom onset among COVID-19 survivors compared with healthy controls. The box in the plot represents the interquartile range (IQR). The whiskers extend to a maximum of 1.5 times the IQR from the edges of the box. Any data points that fall outside the whiskers' range are considered outliers.

gene were specifically expressed in hepatocytes. SERPINE1, also known as plasminogen activator inhibitor-1 (PAI-1), is a proinflammatory cytokine that promote the coagulation process. The plasma level of SERPINE1 was strikingly elevated in severe COVID-19 patients, and even reached a similar level to that in bacterial septic patients.[41] The continuously upregulated proteins platelet basic protein (PPBP) and coagulation factor IX (F9) still did not recover at 2-year follow-up (Figure S5). PPBP is also involved in cytokine–cytokine receptor interaction pathway and chemokine signaling pathway. It is mainly released from activated platelets. Platelet level was previously found to be lower among severe COVID-19 and associated with mortality, and simultaneously PPBP was previously reported to be down-regulated in severe patients.[42,43] Furthermore, the expression of *PPBP* gene was enriched in NK cells and macrophages. *PPBP* gene was also indicated to be one of

the marker genes transcribed by the blood cells that can discriminate critically ill patients from severe patients.[44] The higher PPBP level among COVID-19 survivors compared with healthy controls and lower level among more severe patients indicate the protein may be activated at early stage of COVID-19 but was suppressed in severe patients.

Persistent suppression of neuron generation and differentiation across 2-year after SARS-CoV-2 infection

The impact of SARS-CoV-2 on central nervous system and persistent neurological and neuropsychiatric symptoms post COVID have been widely reported.[45–47] A relatively large amount of DEPs that showed significant change at 6-month after symptom onset were involved in neuron generation and differentiation pathways. Nearly two thirds of neuron related DEPs recovered between 6-month and 1-year after COVID-19, and several proteins including fibronectin (FN1), proprotein convertase subtilisin/kexin type 9 (PCSK9), integrin alpha-6 (ITGA6), lysophospholipase II (LYPLA2), and rho-associated protein kinase 2 (ROCK2) recovered between 1-year and 2-year after COVID-19 (Fig. 2d). The expression of *FN1* and *PCSK9* gene was enriched in hepatocytes, whereas the others were not specific. Notably, 7 proteins were persistently up-regulated or down-regulated across 2-year after COVID-19, with quite slow recovery speed for some proteins (Fig. 2e, Figure S5). These proteins are down-regulated vimentin (VIM), semaphorin-4D (SEMA4D), PPP3CA, EH domain-containing protein 1 (EHD1), and isocitrate dehydrogenase [NADP], mitochondrial (IDH2) and upregulated superoxide dismutase (SOD1) and kunitz-type protease inhibitor 1 (SPINT1). This is quite a large proportion among all 24 proteins persistently up-regulated or down-regulated across 2-year after COVID-19. The gene expression of *VIM* and *SEMA4D* were enriched in Glial cells, and *PPP3CA* in excitatory neurons. SOD1 is a pivotal superoxide dismutase responsible for destroying free superoxide radicals, and is considered as a marker of oxidative stress.[48] SOD1 plasma level is associated with the severity of COVID-19.[49] The consistent elevated plasma level of this protein over 2 years may imply a not fully recovered status of cell damage.

The GO enrichment analysis indicated several neuron related pathways including generation of neurons, neuron differentiation, and neuron projection development were persistently suppressed across 2-year after COVID-19 even though number of DEPs involved in these pathways decreased (Fig. 2f).

Dysregulation and recovery of cellular biology, hypertrophic/dilated cardiomyopathy and cholesterol metabolism across 2-year after SARS-CoV-2 infection

Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis revealed several pathways recovered between 6-month and 1-year, mainly including focal adhesion, ECM-receptor interaction, regulation of actin cytoskeleton, dilated cardiomyopathy, and hypertrophic cardiomyopathy (Fig. 3a, Table S2, Figure S6).

Cell-substrate focal adhesion and cytoskeleton elements are important for maintaining structural integrity of endothelial barrier.[50] Endothelial dysfunction was indicated to be one of the major complications in COVID-19 patients. Disturbances to the actin cytoskeleton during infection of a cell by an RNA virus including SARS-CoV-2 is important to promote effective antiviral responses.[51] These indicated the perturbation of focal adhesion and actin cytoskeleton among COVID-19. In focal adhesion pathway, the levels of talin-1 (TLN1), thrombospondin-1 (THBS1), integrin beta-3 (ITGB3), integrin alpha-IIb (ITGA2B), integrin-linked protein kinase (ILK), cartilage oligomeric matrix protein (COMP), filamin-A (FLNA), and tenascin-N (TNN) were significantly down-regulated, whereas the levels of vinculin (VCL), alpha-actinin-4 (AGTN4), myosin regulatory light polypeptide 9 (MYL9), and laminin subunit gamma-1 (LAMC1) were significantly up-regulated at 6-month after symptom onset among COVID-19 survivors (Fig. 3b). Gene expressions of *TLN1*, *COMP*, *FLNA*, *VCL*, *MYL9*, and *LAMC1* were enriched in smooth muscle cells, with *COMP*, *FLNA*, and *LAMC1* also enriched in fibroblast. The downregulated level of integrin beta-3 (ITGB3) and integrin alpha-IIb (ITGA2B) were shared by regulation of actin cytoskeleton, ECM-receptor interaction, dilated cardiomyopathy, and hypertrophic cardiomyopathy pathways (Fig. 3b and c). The inflammatory cardiomyopathy including dilated cardiomyopathy have been reported to be predominantly mediated by viral infection.[52] A class of tropomyosin chain proteins including tropomyosin alpha-1 chain (TPM1), tropomyosin beta chain (TPM2), and TPM3 involved in dilated cardiomyopathy and hypertrophic cardiomyopathy pathways were all significantly upregulated at 6-month follow-up among COVID-19 survivors. The upregulated level of TPM2 was higher and recovery time was longer than that of TPM1 and TPM3. Although the above pathways have recovered at 1-year follow-up, up-regulated proteins such as collagen alpha-1(I) chain (COL1A1), and down-regulated proteins such as fibronectin (FN1), integrin alpha-6 (ITGA6), and rho-associated protein kinase 2 (ROCK2) involved in these pathways did not recover until 2-year follow-up.

The KEGG enrichment analysis revealed cholesterol metabolism pathway recovered between 1-year and 2-year follow-up (Fig. 3a, Table S2, Figure S7). In cholesterol metabolism pathway, apolipoprotein C–I (APOC1), APOC3, and proprotein convertase subtilisin/kexin type 9 (PCSK9) were down-regulated, and phosphatidylcholine-sterol acyltransferase (LCAT) was up-regulated at both 6-month and 1-year follow-up, and recovered at 2-year follow-up compared with healthy controls (Fig. 3d).

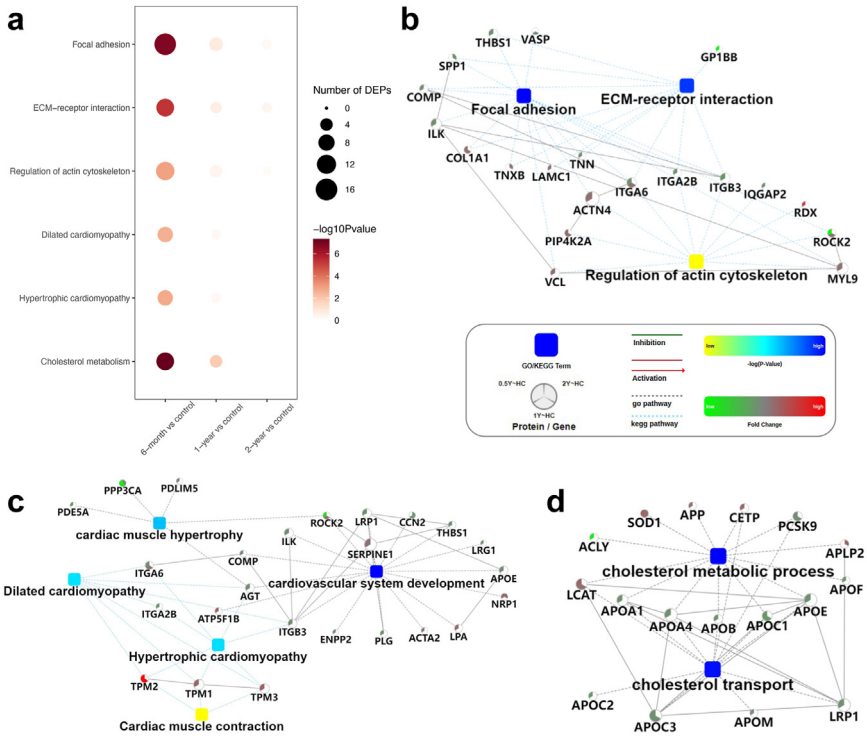

***Fig. 3:* Dysregulation and recovery of cellular activity, cardiovascular system, cholesterol metabolism, and hypoxia-related pathways among COVID-19 survivors across 2 years after symptom onset**. a. Bubble plots of KEGG pathways involved in cellular activity, cardiovascular system, cholesterol metabolism, and hypoxia-related pathways. b–e. Interaction diagrams of the indicated pathways for differentially expressed proteins (DEPs) enriched in cellular activity (b), cardiovascular system (c), and cholesterol metabolism (d) during 2-year follow-up when plasma proteins from COVID-19 patients at 6-month, 1-year, and 2-year after symptom onset are compared to those from healthy controls. The statistical significance of the pathways, represented by -log (*P* value) (Fisher's exact test), is shown by the color scale, with dark blue representing the highest statistical significance. Color bar from red to green represents the fold change of protein level from increasing to decreasing. Fold change indicates the protein level in COVID-19 patients compared to healthy controls.

Other DEPs involved in cholesterol metabolism pathway including APOA1, APOC2, APOE, APOB, prolow-density lipoprotein receptor-related protein 1 (LRP1) with decreased levels and cholesteryl ester transfer protein (CETP) with increased levels had recovered earlier before the whole pathway recovered. Gene expression except for *CETP* were all enriched in hepatocytes.

Differentially expressed proteins as biomarkers for long COVID

A random forest machine learning model based on 1370 proteins was built to differentiate COVID-19 survivors at 6-month, 1-year, and 2-year follow-up and healthy controls. The result showed only one COVID-19 survivor at 1-year after symptom onset was mistakenly classified as COVID-19 survivors at 6-month follow-up, with the accuracy of predicted model as 99.86% (Fig. 4a). To systematically understand the clinical indication of proteins detected in plasma sample, we evaluated associations of all 1370 proteins at 6-month after symptom onset with long-term consequences of COVID-19 survivors at 6-month, 1-year, and 2-year follow-up (Figure S8). There was evidence that T-complex protein 1 subunit alpha (TCP1), T-complex protein 1 subunit delta (CCT4), and Ras-related protein Rab-5B (RAB5B) were negatively associated with 6-min walking distance at 6-month follow-up with correction of multiple testing for 1370 proteins.

To identify potential biomarkers of long COVID and provide intervention at earlier time point, we focused on association between levels of 98 DEPs involved in immune system, complement and coagulation, neurological system, cellular activity, hypertrophic/dilated cardiomyopathy, and cholesterol metabolism at 6-month follow-up with long-term consequences. 22 of them were key proteins that ranked as top 10% by random forest analysis to differentiate COVID-19 at 6-month follow-up and healthy controls (Fig. 4b). Evidence was found for association of some proteins at 6-month follow-up with multiple long-term consequences after multivariable adjustment at nominal significance level 0.05. For example, angiotensinogen (AGT) was positively associated with DLCO of predicted value at 6-month and 2-year follow-up (Fig. 4c), RV of predicted

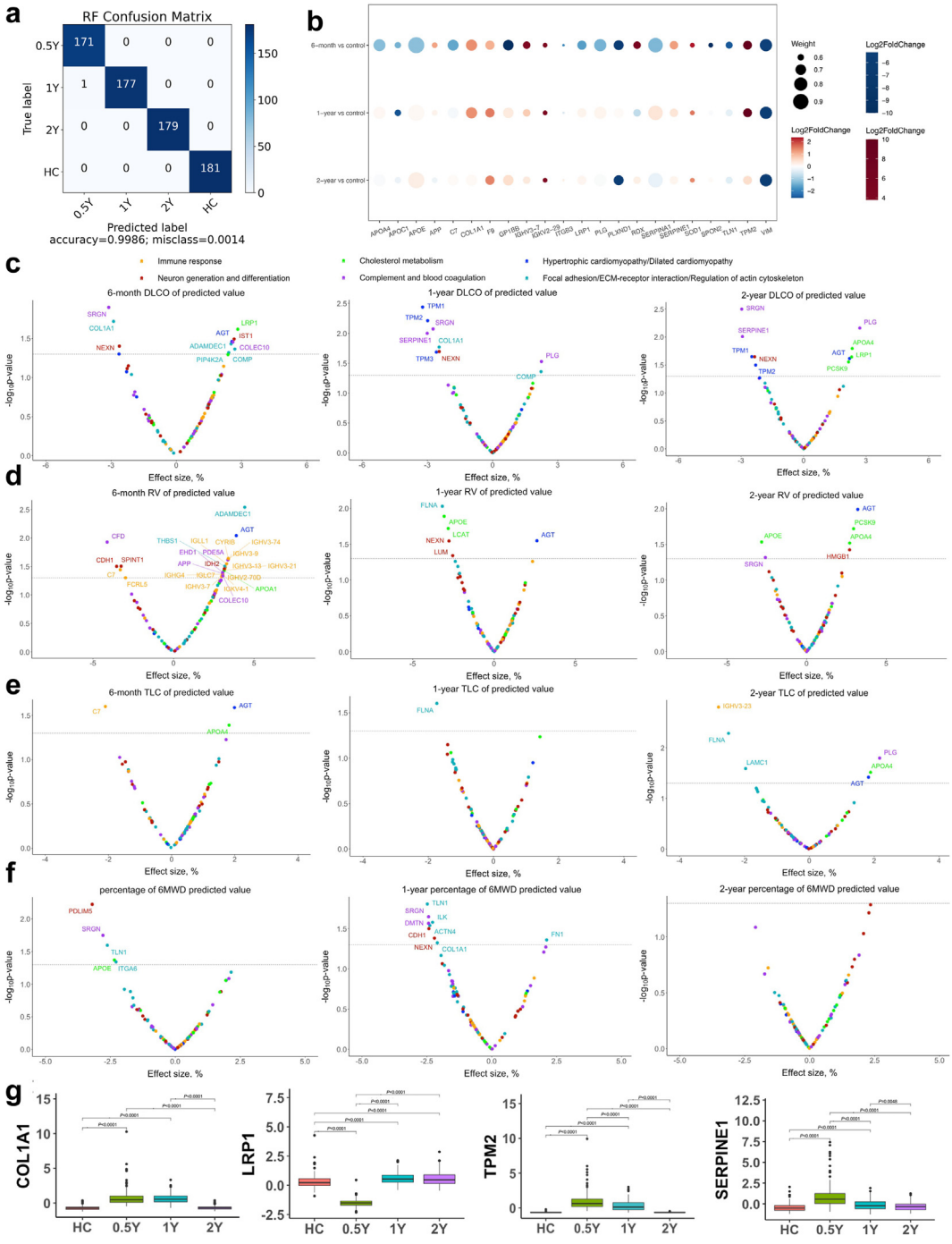

Fig. 4: **Key proteins identified for differentiation of COVID-19 from healthy controls and their association with long COVID**. a. Confusion matrix for random forest analysis to differentiate COVID-19 patients recovered after 6 month, 1 year, and 2 years, and healthy controls. b. Bubble plots of 22 key differentially expressed proteins (DEPs) belonging to immune system, complement and coagulation, neurological system, cellular activity, cardiovascular system, and cholesterol metabolism which are prioritized as top 10% of random forest analysis to differentiate COVID-19 at 6-month follow-up and healthy controls. c–f. Associations of 98 DEPs belonging to complement and coagulation, immune system, neuron generation and differentiation, cellular activity, dilated/hypertrophic cardiomyopathy, and cholesterol metabolism with DLCO of predicted value (c), RV of predicted value (d), TLC of predicted value (e), and 6-min walking distance of predicted value (f) at 6-month, 1-year and 2-year follow-up. DCLO = diffusion capacity for carbon monoxide. RV = residual volume. TLC = total lung capacity. g. Boxplot of DEPs from 22

value at three follow-ups (Fig. 4d), and TLC of predicted value at 6-month and 2-year follow-up (Fig. 4e). C7 showed negative association with RV and TLC of predicted value at 6-month follow-up. SRGN was negatively associated with DLCO of predicted value at three follow-ups and percentage of predicted value for 6-min walking distance at 6-month and 1-year follow-up (Fig. 4c–f).

Furthermore, the time point that protein can begin to have impact on long-term consequences among COVID-19 survivors varied among different proteins or same protein on different phenotypes. The evidence for association of COL1A1 with DLCO of predicted value was found at 6-month follow-up, whereas TPM2 and SERPINE1 showed association until 1-year after symptom onset. COL1A1 is highly expressed by myofibroblast and is a pivotal component in fibrotic ECM. The association of COL1A1 with lung function suggested that the pulmonary fibrosis as represented by COL1A1 level might persist for a long time. Evidence was found for the association of PLG with DLCO of predicted, smell disorder, and taste disorder at 1-year follow-up, but until 2-year follow-up with TLC of predicted value (Fig. 4c–e, Figure S9). Several drugs with PLG as target has been developed. The efficacy of these drugs were mainly antifibrinolytic and anti-inflammatory. ADAMDEC1 showed association with DLCO and RV of predicted value at 6-month follow-up, but with monocyte, lymphocyte, and neutrophil number until 2-year follow-up (Fig. 4c and d, Figure S10). The extrapulmonary manifestations caused by COVID-19 can be reflected from association of DEPs with extrapulmonary clinical indicators. CFD, lumican (LUM), and APOA4 were all negatively associated with eGFR at three follow-ups (Figure S11). ROCK2 and FCRL5 both showed positive associated with HbA1c at three follow-ups.

The relative intensity of key proteins which potentially have protective effect on long-term consequences such as LRP1 was lower, while others which potentially have deleterious effect such as COL1A1, TPM2, and SERPINE1 was higher among COVID-19 survivors at 6-month follow-up than that among healthy controls (Fig. 4g). Majority of these proteins returned to comparable level of healthy controls during 2-year recovery, whereas some of them such as SOD1 were remained at abnormal level after 2-year recovery compared to healthy controls (Figure S12).

Discussion

This study uncovered changes in proteomic landscape across 2 years after SARS-CoV-2 infection which is the longest one to the best of our knowledge. The results indicated four major recovery modes of different biological process. The pathways including focal adhesion, ECM-receptor interaction, and regulation of action cytoskeleton involved in cellular biology and hypertrophic cardiomyopathy and dilated cardiomyopathy pathways involved in cardiovascular system recovered between 6-month and 1-year after symptom onset which was relatively earlier (Fig. 5). Majority of immune response pathway, complement and coagulation cascade, and cholesterol metabolism returned to similar status of matched healthy controls later and before 2-year after infection. Fc receptor signaling pathway still did not recover at 2-year follow-up, which may be mainly related to the long lasting antibody response. Pathways related to neuron generation and differentiation showed persistent suppression across 2-year after infection, with recovery rate becoming much slower after 1-year follow-up. Furthermore, 11 DEPs from the above pathways showed nominally significant association with lung function recovery, with the associations consistent at two consecutive or all three follow-ups. These proteins were mainly enriched in complement and coagulation (PLG, SERPINE1, SRGN, COL1A1, FLNA, and APOE) and hypertrophic/dilated cardiomyopathy (TPM2, TPM1, and AGT) pathways. Two DEPs (APOA4 and LRP1) involved in both neuron and cholesterol pathways also showed similar associations with smell disorder.

The immune response during acute phase of COVID-19 has been widely elucidated, while the recovery pattern post COVID was not sufficiently explored.[53,54] Our study showed the dynamic change of immune response pathways, with majority pathways returning to similar status of vaccinated healthy controls, whereas Fc receptor related signaling pathways were still activated at 2-year after infection. Most of DEPs involved in these pathways were immunoglobulins. Except for preventing cells from being infected by SARS-CoV-2 which is mediated by neutralizing antibodies, antibodies can also interact with FcR-expressing immune cells including monocytes, macrophages, dendritic cells, neutrophils and natural killer cells or complement and mediate protection post-infection through their Fc domains.[55] The protective effect of antibody was found to be abrogated with neutralizing function but loss of Fc-mediated function in COVID-19 animal models.[56,57] Proper regulation of Fc-dependent immunoglobulins can contribute to disease resolution, whereas dysregulation can also play a role in immunopathology.[55] The Fc-dependent antibody effector functions of antibodies was reported to determine the outcome of SARS-CoV-2 infection

key DEPs. The box in the plot represents the interquartile range (IQR). The whiskers extend to a maximum of 1.5 times the IQR from the edges of the box. Any data points that fall outside the whiskers' range are considered outliers.

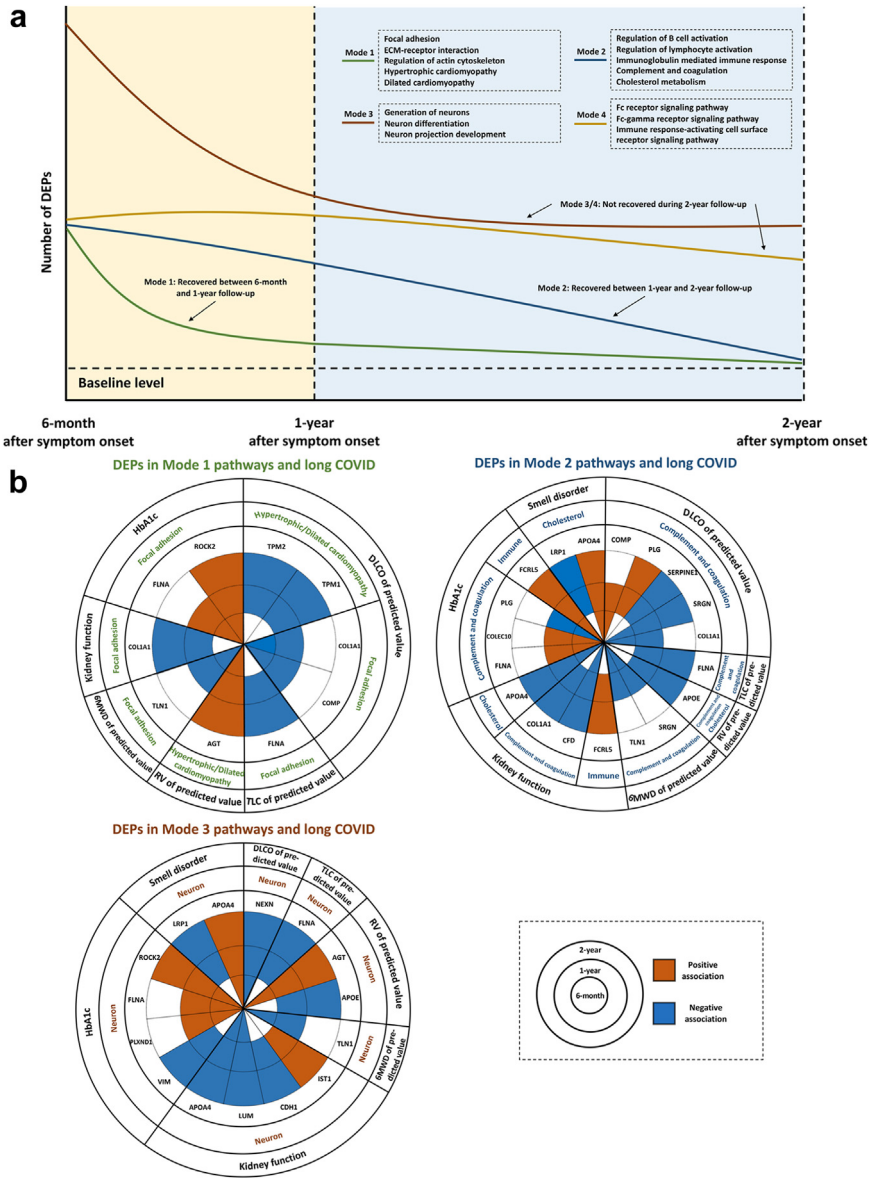

***Fig. 5*: Illustration of molecular modulation pattern across 2-years after symptom onset and association with long COVID among hospitalized COVID-19 survivors**. a. Illustration of molecular modulation pattern across 2-years after symptom onset among hospitalized COVID-19 survivors. The average numbers of differentially expressed proteins (DEPs) in the pathways of each mode were used to plot the curve. DEPs = differentially expressed proteins. b. Association of DEPs belonging to different recovery modes of biological process with long-term consequences of COVID-19 survivors. Proteins that showed consistent association with long-term consequences at two consecutive or all three follow-up were shown. Red shading indicate positive association and blue shading indicates negative association. Dashed circles indicate the association of protein levels at 6-month with consequences at 6-month, 1-year, and 2-year follow-up. DEPs = differentially expressed proteins. DLCO = diffusion capacity for carbon monoxide. RV = residual volume. TLC = total lung capacity. 6MWD = 6-min walking distance.

including risk for intubation or death among COVID-19 patients.[58,59] The dynamic change of these Fc-dependent immunoglobulin post COVID was not clearly elucidated. In our study, IGKV1D-16, IGKV2-29, and IGKV4-1 with continuously increased level across 2 years after SARS-CoV-2 infection were all involved in Fc receptor signaling pathway. The mechanism of persistently activated Fc receptor related signaling pathway and its potential effect for long-term consequences of COVID-19 survivors needs to be further explored.

Complement system was considered to play a crucial role in detecting and eliminating invading pathogens which could be protective during SARS-CoV-2 infection. However, as a 'double-edged sword' like many other components of the immune system, hyperactivation of the complement system was also implicated by previous studies in the pathophysiology of COVID-19.[35] Our study showed the activated complement pathways across 1-year post COVID and recovered at 2-year follow-up, whereas the down-regulation of complement proteins including C7 and C9 at 6-month follow-up and C8a at 1-year follow-up. Previous study reported the elevation of complement activation genes including *C7, C8A,* and *C9* among mild COVID-19 patients but not in asymptomatic ones.[60] Inhibition of protein involved in complement pathway such as C5a has been reported to alleviate viral infection-induced acute lung injury.[61,62] We thus speculated the down-regulation of complement protein post COVID was related to lung injury recovery. This is partially demonstrated by the negative association of C7 and RV of predicted and TLC of predicted value at 6-month follow-up found in this study. The above clues were also consistent with our previous finding of lung function trajectory, which lung function recovery detected between 6-month and 1-year post COVID.[63]

The syndromic feature such as loss of smell, taste disorder, and dizziness reported by COVID-19 patients suggested a potential for neural involvement in SARS-CoV-2 infection.[64] The proposed route for invasion of central nervous systems included hematogenous route with direct invasion of blood brain barrier, neuronal retrograde dissemination route through peripheral neurons, and transcribrial route through cerebrospinal fluid. Previous reports of the neuronal infection of SARS-CoV-2 were not consistent, with some studies finding neuroinvasive capacity and detecting SARS-CoV-2 in cortical neurons among patients who died of COVID-19,[46] while other studies reporting SARS-CoV-2 not to be a neurotropic virus.[65–67] The neurological effects of SARS-CoV-2 infection including white-matter-selective microglial reactivity among COVID-19 patients compared with individuals without SARS-CoV-2 infection was found.[68] However, attention should be paid that microglia did not have a neuronal origin but originated from the immune system.[69] Even if microglia are affected, this does not justify the assumption of neurotropism of SARS-CoV-2. In addition, circulating CCL11, a chemokine associated with cognitive impairment,[70] among people with mild SARS-CoV-2 infection and lasting cognitive symptoms post COVID was also reported to be higher than those without cognitive symptoms post COVID. Our study findings of persistent suppression of neuron generation and differentiation across 2-year after infection support the potential neurological effects of SARS-CoV-2 infection and elucidated the long-term recovery of neuron pathways. The pathways related to neuron generation and differentiation was dramatically inhibited at 6-month follow-up, with around one quarter of all DEPs enriched in these pathways. Although quite a large proportion of these DEPs have recovered at 1-year follow-up, several neuron pathways as a whole were still suppressed at 2-year follow-up. The neurological impact of these uncovered proteins can be reflected by their association with smell or taste disorder. Among proteins that showed nominally significant association with smell or taste disorder, APOA4, plexin-D1 (PLXND1), LRP1, LUM, TNN, and LGALS1 were involved in neuron pathways. Thus, when these neuron generation and differentiation pathways will recover to baseline level and prolonged neurological effect among COVID-19 survivors need to be better explored in large cohort studies with longer follow-up duration. However, the cellular and tissue origin of these potentially neuron related proteins needs to be further carefully demonstrated to better elucidate the neurological effect of SARS-CoV-2.

Uncovering the dynamic changes of proteomics across 2 years after infection provided clues for the potential mechanism of long COVID, and further analyzing association of DEPs with long-term consequences of COVID-19 survivors indicated the potential biomarkers of long COVID. The results elucidated some proteins could potentially have long-term impact on consequences, and association between some proteins and long COVID was shown later than the time point that we can identify the protein as DEP. Taking these points together with the diverse recovery of proteins in different pathways, the impact of DEPs on long-term consequences post COVID needs to be evaluated within a much longer time frame. Not only the unrecovered pathways and proteins at 2-year follow-up but also the lag effects of proteins even those recovered during 2-year after symptom onset on long-term consequence of COVID-19 survivors need to be paid attention.

Our study had several limitations. Firstly, our study only enrolled hospitalised COVID-19 survivors infected with original strain, thus was not able to depict the longitudinal proteomics landscape among outpatient with mild symptom or asymptomatic people and those infected with other strains. Secondly, proteome of study participants at acute phase was not profiled as the samples of study participants were not collected during hospitalisation considering the emergency state of hospitals at the early stage of COVID-19 pandemic. Thirdly, previous studies reported that amyloid fibrin microclots may exist in the plasma samples of long COVID people, which need a secondary trypsin digestion step for complete digestion.[71,72] It's possible that the current proteomic analysis may have missed some inflammatory molecules and antibodies due to the presence of these deposits that were not fully digested. In our study,

the inflammatory molecules SAA1, SAA4, α2AP and VWF have been identified, but no obvious difference was found between COVID-19 patients and healthy controls. Since these molecules are reported to be entrapped within the amyloid plasma deposits, the incomplete digestion process may possibly account for the undetected difference. Fourthly, the causal link between dynamic change of biological process uncovered by proteomics and long COVID needs to be further demonstrated with more rigorous mechanistic studies such as animal and cellular model studies. These mechanism studies will exclude potential perturbations of confounding factors that cannot be taken into consideration by current enrichment analyses and residual confounders such as smoking potentially existing in association analyses. Finally, the DEPs that showed consistent but nominally significant associations with long COVID also needs to be validated in cohort studies with larger sample size and longer follow-up duration.

In conclusion, our study identified four recovery modes of different biological processes among COVID-19 survivors across 2 years after infection, and thus provided molecular insights into the potential mechanism of long COVID. In a context that diagnosis and treatment of long COVID are both challenging, the findings put forward biomarkers for more precise intervention to reduce burden of long COVID.

Contributors

BC, CC.L.W, DZ and XG had the idea for and designed the study and all authors had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. XG, SW, WZ, YZ and ZW drafted the paper. XG, SW, and YZ did the analysis. HaiboL, HZ, WL, HuiL, YL, TD, CC.L.W, and BC critically revised the manuscript for important intellectual content and all authors agreed to submit the final version for publication. XG, CL, LG, HZ, YW, and LH completed the follow-up work, and collected and verified the sample and data. All authors agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Data sharing statement

Restrictions apply to the availability of these data and so are not publicly available. However, data are available from the authors upon reasonable request and with the permission of the institution.

Declaration of interests

We declare no competing interests.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (82241056, 82030002, 82200114, 32150005, and 92169107); the Chinese Academy of Medical Sciences (CAMS) Innovation Fund for Medical Sciences (CIFMS 2021-I2M-1–048, and 2022–12M-1-004); the Clinical Research Operating Fund of Central High Level Hospitals (2022-PUMCH-E-001); the Talent Program of the Chinese Academy of Medical Science (2022-RC310-10); the Training Program of the Big Science Strategy Plan (2020YFE0202200); the Ministry of Science and Technology of China (2022YFA0806000); and the Research Funds from Health@InnoHK Program launched by Innovation Technology Commission of the Hong Kong Special Administrative Region. This work was also supported by the New Cornerstone Science Foundation and Peking Union Medical College Education Foundation.

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.ebiom.2023.104851.

References

1 World Health Organization. WHO Coronavirus disease (COVID-19) dashboard. https://covid19.who.int/table. Accessed July 7, 2023.
2 Davis HE, McCorkell L, Vogel JM, Topol EJ. Long COVID: major findings, mechanisms and recommendations. *Nat Rev Microbiol*. 2023;21(3):133–146.
3 Crook H, Raza S, Nowell J, Young M, Edison P. Long covid-mechanisms, risk factors, and management. *BMJ*. 2021;374:n1648.
4 Levine RL. Addressing the long-term effects of COVID-19. *JAMA*. 2022;328(9):823–824.
5 Singh SJ, Baldwin MM, Daynes E, et al. Respiratory sequelae of COVID-19: pulmonary and extrapulmonary origins, and approaches to clinical care and rehabilitation. *Lancet Respir Med*. 2023;11(8):709–725.
6 Huang C, Huang L, Wang Y, et al. 6-month consequences of COVID-19 in patients discharged from hospital: a cohort study. *Lancet*. 2023;401(10393):e21–e33.
7 Huang L, Yao Q, Gu X, et al. 1-year outcomes in hospital survivors with COVID-19: a longitudinal cohort study. *Lancet*. 2021;398(10302):747–758.
8 Huang L, Li X, Gu X, et al. Health outcomes in people 2 years after surviving hospitalisation with COVID-19: a longitudinal cohort study. *Lancet Respir Med*. 2022;10(9):863–876.
9 Global Burden of Disease Long CC, Wulf Hanson S, Abbafati C, et al. Estimated global proportions of individuals with persistent fatigue, cognitive, and respiratory symptom clusters following symptomatic COVID-19 in 2020 and 2021. *JAMA*. 2022;328(16):1604–1615.
10 Nasserie T, Hittle M, Goodman SN. Assessment of the frequency and variety of persistent symptoms among patients with COVID-19: a systematic review. *JAMA Netw Open*. 2021;4(5):e2111417.
11 Almas T, Malik J, Alsubai AK, et al. Post-acute COVID-19 syndrome and its prolonged effects: an updated systematic review. *Ann Med Surg (Lond)*. 2022;80:103995.
12 Gu X, Huang L, Cui D, et al. Association of acute kidney injury with 1-year outcome of kidney function in hospital survivors with COVID-19: a cohort study. *eBioMedicine*. 2022;76:103817.
13 Blomberg B, Mohn KG, Brokstad KA, et al. Long COVID in a prospective cohort of home-isolated patients. *Nat Med*. 2021;27(9):1607–1613.
14 Ramakrishnan RK, Kashour T, Hamid Q, Halwani R, Tleyjeh IM. Unraveling the mystery surrounding post-acute sequelae of COVID-19. *Front Immunol*. 2021;12:686029.
15 Yao XH, He ZC, Li TY, et al. Pathological evidence for residual SARS-CoV-2 in pulmonary tissues of a ready-for-discharge patient. *Cell Res*. 2020;30(6):541–543.
16 Yong SJ, Halim A, Halim M, et al. Inflammatory and vascular biomarkers in post-COVID-19 syndrome: a systematic review and meta-analysis of over 20 biomarkers. *Rev Med Virol*. 2023;33(2):e2424.
17 Su Y, Yuan D, Chen DG, et al. Multiple early factors anticipate post-acute COVID-19 sequelae. *Cell*. 2022;185(5):881–895.e20.
18 Vijayakumar B, Boustani K, Ogger PP, et al. Immuno-proteomic profiling reveals aberrant immune cell regulation in the airways of individuals with ongoing post-COVID-19 respiratory disease. *Immunity*. 2022;55(3):542–556.e5.
19 Chen Y, Yao H, Zhang N, et al. Proteomic analysis identifies prolonged disturbances in pathways related to cholesterol metabolism and myocardium function in the COVID-19 recovery stage. *J Proteome Res*. 2021;20(7):3463–3474.
20 Chen Y, Zhang N, Zhang J, et al. Immune response pattern across the asymptomatic, symptomatic and convalescent periods of COVID-19. *Biochim Biophys Acta Proteins Proteom*. 2021;1870(2):140736.

21 Kovarik JJ, Bileck A, Hagn G, et al. A multi-omics based anti-inflammatory immune signature characterizes long COVID-19 syndrome. *iScience*. 2023;26(1):105717.
22 Huang C, Huang L, Wang Y, et al. 6-month consequences of COVID-19 in patients discharged from hospital: a cohort study. *Lancet*. 2021;397(10270):220–232.
23 Mahler DA, Wells CK. Evaluation of clinical methods for rating dyspnea. *Chest*. 1988;93(3):580–586.
24 Herdman M, Gudex C, Lloyd A, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Qual Life Res*. 2011;20(10):1727–1736.
25 Rabin R, de Charro F. EQ-5D: a measure of health status from the EuroQol Group. *Ann Med*. 2001;33(5):337–343.
26 Xie W, Wu Y, Wang W, et al. A longitudinal study of carotid plaque and risk of ischemic cardiovascular disease in the Chinese population. *J Am Soc Echocardiogr*. 2011;24(7):729–737.
27 Mizrahi B, Sudry T, Flaks-Manov N, et al. Long covid outcomes at one year after mild SARS-CoV-2 infection: nationwide cohort study. *BMJ*. 2023;380:e072529.
28 Ballering AV, van Zon SKR, olde Hartman TC, Rosmalen JGM. Persistence of somatic symptoms after COVID-19 in The Netherlands: an observational cohort study. *Lancet*. 2022;400(10350):452–461.
29 Standardization of spirometry, 1994 update. American thoracic society. *Am J Respir Crit Care Med*. 1995;152(3):1107–1136.
30 Tian W, Zhang N, Jin R, et al. Immune suppression in the early stage of COVID-19 disease. *Nat Commun*. 2020;11(1):5859.
31 Enright PL, Sherrill DL. Reference equations for the six-minute walk in healthy adults. *Am J Respir Crit Care Med*. 1998;158(5 Pt 1):1384–1387.
32 Lefranc MP. Immunoglobulin and T Cell receptor genes: IMGT(®) and the birth and rise of immunoinformatics. *Front Immunol*. 2014;5:22.
33 Stros M, Polanska EV, Hlavacova T, Skladal P. Progress in assays of HMGB1 levels in human plasma-the potential prognostic value in COVID-19. *Biomolecules*. 2022;12(4):544.
34 Wulandari S, Hartono, Wibawa T. The role of HMGB1 in COVID-19-induced cytokine storm and its potential therapeutic targets: a review. *Immunology*. 2023;169(2):117–131.
35 Afzali B, Noris M, Lambrecht BN, Kemper C. The state of complement in COVID-19. *Nat Rev Immunol*. 2022;22(2):77–84.
36 Georg P, Astaburuaga-Garcia R, Bonaguro L, et al. Complement activation induces excessive T cell cytotoxicity in severe COVID-19. *Cell*. 2022;185(3):493–512.e25.
37 Ali YM, Ferrari M, Lynch NJ, et al. Lectin pathway mediates complement activation by SARS-CoV-2 proteins. *Front Immunol*. 2021;12:714511.
38 Holter JC, Pischke SE, de Boer E, et al. Systemic complement activation is associated with respiratory failure in COVID-19 hospitalized patients. *Proc Natl Acad Sci U S A*. 2020;117(40):25018–25025.
39 Yu J, Gerber GF, Chen H, et al. Complement dysregulation is associated with severe COVID-19 illness. *Haematologica*. 2022;107(5):1095–1105.
40 Conway EM, Mackman N, Warren RQ, et al. Understanding COVID-19-associated coagulopathy. *Nat Rev Immunol*. 2022;22(10):639–649.
41 Kang S, Tanaka T, Inoue H, et al. IL-6 trans-signaling induces plasminogen activator inhibitor-1 from vascular endothelial cells in cytokine release syndrome. *Proc Natl Acad Sci U S A*. 2020;117(36):22351–22356.
42 Lippi G, Plebani M, Henry BM. Thrombocytopenia is associated with severe coronavirus disease 2019 (COVID-19) infections: a meta-analysis. *Clin Chim Acta*. 2020;506:145–148.
43 Shen B, Yi X, Sun Y, et al. Proteomic and metabolomic characterization of COVID-19 patient sera. *Cell*. 2020;182(1):59–72.e15.
44 Yatim N, Boussier J, Chocron R, et al. Platelet activation in critically ill COVID-19 patients. *Ann Intensive Care*. 2021;11(1):113.
45 Monje M, Iwasaki A. The neurobiology of long COVID. *Neuron*. 2022;110(21):3484–3496.
46 Song E, Zhang C, Israelow B, et al. Neuroinvasion of SARS-CoV-2 in human and mouse brain. *J Exp Med*. 2021;218(3):e20202135.
47 Solomon T. Neurological infection with SARS-CoV-2 - the story so far. *Nat Rev Neurol*. 2021;17(2):65–66.
48 Tavassolifar MJ, Aghdaei HA, Sadatpour O, et al. New insights into extracellular and intracellular redox status in COVID-19 patients. *Redox Biol*. 2023;59:102563.
49 Xu B, Lei Y, Ren X, et al. SOD1 is a possible predictor of COVID-19 progression as revealed by plasma proteomics. *ACS Omega*. 2021;6(26):16826–16836.
50 Karki P, Birukova AA. Microtubules as major regulators of endothelial function: implication for lung injury. *Front Physiol*. 2021;12:758313.
51 Acharya D, Reis R, Volcic M, et al. Actin cytoskeleton remodeling primes RIG-I-like receptor activation. *Cell*. 2022;185(19):3588–3602.e21.
52 Tschope C, Ammirati E, Bozkurt B, et al. Myocarditis and inflammatory cardiomyopathy: current evidence and future directions. *Nat Rev Cardiol*. 2021;18(3):169–193.
53 Sette A, Crotty S. Adaptive immunity to SARS-CoV-2 and COVID-19. *Cell*. 2021;184(4):861–880.
54 Brodin P. Immune determinants of COVID-19 disease presentation and severity. *Nat Med*. 2021;27(1):28–33.
55 Zhang A, Stacey HD, D'Agostino MR, Tugg Y, Marzok A, Miller MS. Beyond neutralization: Fc-dependent antibody effector functions in SARS-CoV-2 infection. *Nat Rev Immunol*. 2023;23(6):381–396.
56 Winkler ES, Gilchuk P, Yu J, et al. Human neutralizing antibodies against SARS-CoV-2 require intact Fc effector functions for optimal therapeutic protection. *Cell*. 2021;184(7):1804–1820.e16.
57 Suryadevara N, Shrihari S, Gilchuk P, et al. Neutralizing and protective human monoclonal antibodies recognizing the N-terminal domain of the SARS-CoV-2 spike protein. *Cell*. 2021;184(9):2316–2331.e15.
58 Begin P, Callum J, Jamula E, et al. Convalescent plasma for hospitalized patients with COVID-19: an open-label, randomized controlled trial. *Nat Med*. 2021;27(11):2012–2024.
59 Zohar T, Loos C, Fischinger S, et al. Compromised humoral functional evolution tracks with SARS-CoV-2 mortality. *Cell*. 2020;183(6):1508–1519.e12.
60 Faridl M, Mellyani K, Khoirunnisa K, et al. RNA sequence analysis of nasopharyngeal swabs from asymptomatic and mildly symptomatic patients with COVID-19. *Int J Infect Dis*. 2022;122:449–460.
61 Wang R, Xiao H, Guo R, Li Y, Shen B. The role of C5a in acute lung injury induced by highly pathogenic viral infections. *Emerg Microbes Infect*. 2015;4(5):e28.
62 Sun S, Zhao G, Liu C, et al. Inhibition of complement activation alleviates acute lung injury induced by highly pathogenic avian influenza H5N1 virus infection. *Am J Respir Cell Mol Biol*. 2013;49(2):221–230.
63 Zhang H, Li X, Huang L, et al. Lung-function trajectories in COVID-19 survivors after discharge: a two-year longitudinal cohort study. *eClinicalMedicine*. 2022;54:101668.
64 Baig AM, Sanders EC. Potential neuroinvasive pathways of SARS-CoV-2: deciphering the spectrum of neurological deficit seen in coronavirus disease-2019 (COVID-19). *J Med Virol*. 2020;92(10):1845–1857.
65 Khan M, Yoo SJ, Clijsters M, et al. Visualizing in deceased COVID-19 patients how SARS-CoV-2 attacks the respiratory and olfactory mucosae but spares the olfactory bulb. *Cell*. 2021;184(24):5932–5949.e15.
66 Pellegrini L, Albecka A, Mallery DL, et al. SARS-CoV-2 infects the brain choroid plexus and disrupts the blood-CSF barrier in human brain organoids. *Cell Stem Cell*. 2020;27(6):951–961.e5.
67 Khan M, Clijsters M, Choi S, et al. Anatomical barriers against SARS-CoV-2 neuroinvasion at vulnerable interfaces visualized in deceased COVID-19 patients. *Neuron*. 2022;110(23):3919–3935.e6.
68 Fernandez-Castaneda A, Lu P, Geraghty AC, et al. Mild respiratory COVID can cause multi-lineage neural cell and myelin dysregulation. *Cell*. 2022;185(14):2452–2468.e16.
69 Ginhoux F, Greter M, Leboeuf M, et al. Fate mapping analysis reveals that adult microglia derive from primitive macrophages. *Science*. 2010;330(6005):841–845.
70 Villeda SA, Luo J, Mosher KI, et al. The ageing systemic milieu negatively regulates neurogenesis and cognitive function. *Nature*. 2011;477(7362):90–94.
71 Pretorius E, Vlok M, Venter C, et al. Persistent clotting protein pathology in long COVID/Post-Acute Sequelae of COVID-19 (PASC) is accompanied by increased levels of antiplasmin. *Cardiovasc Diabetol*. 2021;20(1):172.
72 Kruger A, Vlok M, Turner S, et al. Proteomics of fibrin amyloid microclots in long COVID/post-acute sequelae of COVID-19 (PASC) shows many entrapped pro-inflammatory molecules that may also contribute to a failed fibrinolytic system. *Cardiovasc Diabetol*. 2022;21(1):190.