# Plasma proteomic signature predicts who will get persistent symptoms following SARS-CoV-2 infection

Gabriella Captur,[a,b,c,1] James C. Moon,[b,d,1] Constantin-Cristian Topriceanu,[a,b] George Joy,[b,d] Leo Swadling,[e] Jenny Hallqvist,[f] Ivan Doykov,[f] Nina Patel,[f] Justyna Spiewak,[f] Tomas Baldwin,[f] Matt Hamblin,[d] Katia Menacho,[b,d] Marianna Fontana,[b,g] Thomas A. Treibel,[b,d] Charlotte Manisty,[b,d] Ben O'Brien,[h,i,j,k] Joseph M. Gibbons,[l] Corrina Pade,[l] Tim Brooks,[m] Daniel M. Altmann,[n] Rosemary J. Boyton,[o,p] Áine McKnight,[l] Mala K. Maini,[e] Mahdad Noursadeghi,[e] Kevin Mills,[f,2] and Wendy E. Heywood,[f,2]*
UK COVIDsortium Investigators

[a]UCL MRC Unit for Lifelong Health and Ageing, 33 Bedford Place, London WC1B 5JU, UK
[b]Institute of Cardiovascular Science, University College London, Gower Street, London WC1E 6BT, UK
[c]The Royal Free Hospital, Center for Inherited Heart Muscle Conditions, Cardiology Department, Pond Street, Hampstead, London NW3 2QG, UK
[d]Barts Heart Center, The Cardiovascular Magnetic Resonance Imaging Unit and The Inherited Cardiovascular Diseases Unit, St Bartholomew's Hospital, West Smithfield, London EC1A 7BE, UK
[e]Division of Infection and Immunity, University College London, London WC1E 6JF, UK
[f]Translational Mass Spectrometry Research Group, UCL Institute of Child Health and Great Ormond Street Hospital, 30 Guilford Street, London WC1N 1EH, UK
[g]The Royal Free Hospital, Cardiac MRI Unit, Pond Street, Hampstead, London NW3 2QG, UK
[h]Department of Perioperative Medicine, St. Bartholomew's Hospital, Barts Health NHS Trust, West Smithfield, London EC1A 7BE, UK
[i]Department of Cardiac Anesthesiology and Intensive Care Medicine, German Heart Center, Augustenburger Platz 1, 13353 Berlin, Germany
[j]Department of Cardiac Anesthesiology and Intensive Care Medicine, Charité Berlin, Augustenburger Platz 1, 13353 Berlin, Germany
[k]Outcomes Research Consortium, Department of Outcomes Research, The Cleveland Clinic, 9500 Euclid Ave P77, Cleveland, OH 44195, USA
[l]Blizard Institute, Barts and the London School of Medicine and Dentistry, Queen Mary University of London, London E1 2AT, UK
[m]National Infection Service, Public Health England, Porton Down, UK
[n]Department of Immunology and Inflammation, Imperial College London, London W12 0NN, UK
[o]Department of Infectious Disease, Imperial College London, London SW7 2AZ, UK
[p]Lung Division, Royal Brompton Harefield Hospitals, Guy's and St Thomas' NHS Foundation Trust, London SW3 6NP, UK

## Summary

**Background** The majority of those infected by ancestral Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) during the UK first wave (starting March 2020) did not require hospitalisation. Most had a short-lived mild or asymptomatic infection, while others had symptoms that persisted for weeks or months. We hypothesized that the plasma proteome at the time of first infection would reflect differences in the inflammatory response that linked to symptom severity and duration.

**Methods** We performed a nested longitudinal case-control study and targeted analysis of the plasma proteome of 156 healthcare workers (HCW) with and without lab confirmed SARS-CoV-2 infection. Targeted proteomic multiple-reaction monitoring analysis of 91 pre-selected proteins was undertaken in uninfected healthcare workers at baseline, and in infected healthcare workers serially, from 1 week prior to 6 weeks after their first confirmed SARS-CoV-2 infection. Symptom severity and antibody responses were also tracked. Questionnaires at 6 and 12 months collected data on persistent symptoms.

**Findings** Within this cohort (median age 39 years, interquartile range 30–47 years), 54 healthcare workers (44% male) had PCR or antibody confirmed infection, with the remaining 102 (38% male) serving as uninfected controls.

eBioMedicine 2022;00: 104293
Published online xxx
https://doi.org/10.1016/j.ebiom.2022.104293

*Corresponding author at: UCL Translational Mass Spectrometry Research Group, Center for Translational Omics, Genetics & Genomic Medicine, UCL Great Ormond Street Institute of Child Health, 30 Guilford Street, London WC1N 1EH, UK.
*E-mail address:* w.heywood@ucl.ac.uk (W.E. Heywood).

[1] GC & JM joint first authors.

[2] KM & WH joint last authors.

Following the first confirmed SARS-CoV-2 infection, perturbation of the plasma proteome persisted for up to 6 weeks, tracking symptom severity and antibody responses. Differentially abundant proteins were mostly coordinated around lipid, atherosclerosis and cholesterol metabolism pathways, complement and coagulation cascades, autophagy, and lysosomal function. The proteomic profile at the time of seroconversion associated with persistent symptoms out to 12 months. Data are available via ProteomeXchange with identifier PXD036590.

**Interpretation** Our findings show that non-severe SARS-CoV-2 infection perturbs the plasma proteome for at least 6 weeks. The plasma proteomic signature at the time of seroconversion has the potential to identify which individuals are more likely to suffer from persistent symptoms related to SARS-CoV-2 infection.

**Funding information** The COVIDsortium is supported by funding donated by individuals, charitable Trusts, and corporations including Goldman Sachs, Citadel and Citadel Securities, The Guy Foundation, GW Pharmaceuticals, Kusuma Trust, and Jagclif Charitable Trust, and enabled by Barts Charity with support from University College London Hospitals (UCLH) Charity. This work was additionally supported by the Translational Mass Spectrometry Research Group and the Biomedical Research Center (BRC) at Great Ormond Street Hospital.



**Keywords:** COVID-19; Proteomics; Post-acute sequelae of SARS-CoV-2 (PASC)

### Research in context

*Evidence before this study*

The majority of individuals infected with ancestral SARS-CoV-2 during the UK first wave (starting March 2020) did not require hospital admission yet there were considerable differences in terms of symptom burden and clinical course, with some individuals experiencing more symptoms that persisted for weeks or months resulting in post-acute sequelae of SARS-CoV-2 (PASC).

*Added value of this study*

We analysed the plasma proteome in healthcare workers every week during the first wave of COVID-19 to understand whether the plasma proteomic abnormalities caused by infection associated with the development of PASC. We found that perturbation of the proteome continues for at least 6 weeks after the first positive test for COVID-19, tracking symptom severity and antibody response. Additionally, we identified a plasma proteomics signature at the time of infection that has the potential to predict those more likely to suffer from PASC.

*Implications of all the available evidence*

COVID-19 infection perturbs the plasma proteome for at least 6 weeks following first positive PCR test even if the infection is apparently mild or asymptomatic. The plasma proteomics signature at the time of seroconversion has the potential to identify which individuals are more likely to suffer from PASC out to 1 year.

## Introduction

Most people infected with SARS-CoV-2 do not require hospitalisation yet there is considerable heterogeneity in terms of symptom severity, duration and the number that go on to experience persistent symptoms lasting several weeks/months after the original infection (i.e. post-acute sequelae of SARS-CoV-2[1] [PASC] or long COVID). According to the World Health Organization definition, PASC occurs in individuals with a history of probable or confirmed SARS CoV-2 infection, usually 3 months from the onset of infection with symptoms and that last for at least 2 months and cannot be explained by an alternative diagnosis.[2] Though PASC is more common in hospitalized patients, epidemiological studies suggest that over a third of non-hospitalized patients without clinically severe infection, may also get PASC.[3,4] Rough first estimates suggest that as much as 30% of the SARS-CoV-2 health burden could be due to infection-induced disability and PASC, not death, with the overall economic cost of PASC reaching half the cost of COVID-related deaths in the UK.[5]

The UK COVIDsortium recruited a cohort of 731 HCW at three London hospitals early in the pandemic starting 23rd March 2020[6] before the availability of vaccines. HCW were followed-up longitudinally when fit to attend work with weekly nasopharyngeal swabs for SARS-CoV-2 PCR, blood sample collections for antibody tests and omics, and self-reporting health questionnaires, including symptom questionnaire at 6 and 12 months. Using a targeted mass spectrometry (MS) proteomic panel consisting of proteins associated with the immune response[7] enriched for neuroinflammatory

biomarkers, we examined the plasma proteome of infected HCW at baseline and throughout the 6 weeks following their first positive PCR test, and compared this to the baseline proteome of uninfected controls to determine whether perturbations in plasma proteins were associated with symptom severity in the acute period and with persistent symptoms at 12 months.

## Methods

### Ethics

The COVIDsortium Healthcare Workers bioresource was approved by the UK Research Ethics Committee (South Central - Oxford A Research Ethics Committee, reference 20/SC/0149, approval date 17th March 2020). The study conformed to the principles of the Helsinki Declaration, and all subjects gave written informed consent. Participants were not compensated for participation.

### Study design and population

We undertook an unmatched case control study nested within our COVIDsortium health care worker cohort. Participant screening, study design, sample collection, and sample processing have been described in detail previously[6,8–10] and the study is registered at ClinicalTrials.gov (NCT04318314, first posted 23rd March 2020: the date of first study member recruitment). Briefly, healthcare workers were recruited at St Bartholomew's Hospital, London, UK in the first week of lockdown in the United Kingdom (between 23rd and 31st March 2020). Participants underwent weekly evaluation using a questionnaire and biological sample collection (including PCR and serological assays) for up to 16 weeks when fit to attend work at each visit, with further follow up symptom questionnaires collected at 6 and 12 months. Participants with available plasma samples who had PCR-confirmed SARS-CoV-2 infection (Roche cobas® diagnostic test platform) at any time point from week 0 to week 6 were included as 'cases'. A subset of consecutively recruited participants at baseline without evidence of SARS-CoV-2 infection on nasopharyngeal swabs and who remained seronegative by both Euroimmun antiS1 spike protein and Roche anti-nucleocapsid protein throughout follow-up were included as uninfected HCW.

### Proteomic analysis

The proteomics panel selection was inspired by previous works[11–13] and was designed to capture novel findings/biomarkers beyond those that could be found by highly used conventional inflammatory panels. We included proteins associated with acute phase proteins, apolipoproteins, complement factors, coagulation cascade components, interleukin signalling pathway members (especially of interleukin-6), and modulators of inflammation. In addition, there is an increasing body of evidence in support of SARS-CoV-2 related neurological and cognitive dysfunction. These are associated to changes in central nervous system structure[14] and are postulated to occur secondary to either viral neurotropism or virally induced neuroinflammation.[15,16] Thus, approximatively 25% of our panel consists of neuroinflammatory biomarkers (Table S1).

Whole blood collected from individual participants was centrifuged on-site. Aliquoted plasma samples were stored in a freezer at −80°C until use. Ten microlitres (10 μl) of plasma were analysed from each sample. As previously reported[7] we developed a custom targeted MS based assay panel that looks at over 90 pro- and anti-inflammatory associated plasma proteins (see Table 1 on protocols.io[17]). Quantatypic peptides were selected based on previous proteomics data and from the online multiple-reaction monitoring (MRM) database (www.thegpm.org). All MS analyses were performed using a Xevo- TQ XS mass spectrometer coupled to a Acquity UPLC (LC) and quadrupole time of flight (QToF) Premier mass spectrometer (MS, Waters Corporation, Manchester, U.K.). Pooled plasma digest was used as a quality control (QC) which was run periodically between injections. QCs were monitored throughout the run and a coefficient of variation of +/-15% was considered acceptable for run. A standard curve 0−40 pmol was run at the start and end of the run. Two transitions were selected for the assay; one for quantitation and one for confirmation. The most abundant clean transitions without interfering non-specific peaks were selected using the synthetic peptides spiked into matrix. Chromatograms were analyzed using SkyLine software (MacCoss labs, USA). Semi-automated peak integration was processed by manual inspection to correct for false assignments. The reader was fully blinded to the clinical status of HCW throughout the analysis procedure. Only peptides that gave good quantitative data by assessment using a standard curve spiked in plasma, and good signal to noise ratio of the endogenous peptide, were used in the assay. Integrated peak areas were expressed as a ratio to internal standard and subsequently $z$-score normalised across analytes.

A mean standard curve linearity of $R^2$ 0.96 ± 0.04 was achieved across the calibration curves for the assayed proteins. Mean coefficients of variation for target peptides between replicate quality control samples was 9.1 ± 3.3%. The MS proteomics data have been deposited to the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) via the PRIDE partner repository[18] with the dataset identifier PXD036590.

### Calprotectin ELISA

Detection of calprotectin in human serum was assessed by enzyme-linked immunosorbent assay (ELISA; Hycult Biotech, Pennsylvania, USA, catalogue# HK214-01) according to the manufacturers protocol in all

| Characteristic | Overall $n = 156$ | PCR positive $n = 29$ | Seropositive $n = 52$ | Uninfected HCW $n = 102$ | $p$-value[#] PCR positive vs Uninfected \| Seropositive vs Uninfected |
|---|---|---|---|---|---|
| Age | 39 (30, 47) | 38 (30, 48) | 38 (30,47) | 37 (30, 47) | 0.802 \| 0.427 |
| Sex | | | | | |
| Female | 93 (60%) | 20 (69%) | 24 (46%) | 63 (62%) | |
| Male | 63 (40%) | 9 (31%) | 28 (54%) | 39 (38%) | 0.478 \| 0.646 |
| Ethnicity | | | | | |
| White | 83 (53·2%) | 17 (58·6%) | 30 (57·7%) | 52 (51%) | |
| Black | 15 (9·6%) | 3 (10·3%) | 7 (13·5%) | 8 (7·8%) | 0.614 \| 0.293 |
| Asian | 45 (28·8%) | 8 (27·6%) | 13 (25%) | 31 (30·4%) | |
| Other | 13 (8·3%) | 1 (3·4%) | 2 (3·8%) | 11 (10·8%) | |
| Participants with $n$ samples available | | | | | |
| $n = 1$ | 106 (67·9%) | 2 (6·9%) | 3 (5·8%) | 102 (100%) | |
| $n = 2$ | 11 (7.1%) | 3 (10.3%) | 10 (19.2%) | 0 (0%) | |
| $n = 3$ | 8 (5·1%) | 6 (20·7%) | 8 (15·4%) | 0 (0%) | |
| $n = 4$ | 9 (5.8%) | 6 (20.7%) | 9 (17.3%) | 0 (0%) | |
| $n = 5$ | 7 (4·5%) | 3 (10·3%) | 7 (13·5%) | 0 (0%) | |
| $n = 6$ | 8 (5·1%) | 6 (20·7%) | 8 (15·4%) | 0 (0%) | |
| $n = 7$ | 7 (4·5%) | 3 (10·3%) | 7 (13·5%) | 0 (0%) | |
| Case-definition symptoms* | 20 (12·8%) | 11 (39·9%) | 12 (23%) | 0 (0%) | |

***Table 1*: Baseline characteristics of the study cohort.**
Statistics presented as median (interquartile range) or *n* (%).
*For uninfected controls this applies to any case-definition symptoms reported at baseline or in the preceding 3 months. For PCR positive/Seropositive cases, any case-definition symptoms reported from baseline out to week-6 were counted as 1.
[#]Differences between PCR positive vs uninfected HCW and between seropositive vs uninfected HCW, were tested using Wilcoxon signed-rank test (for age) or Chi-squared test (for sex and ethnicity; Fisher's exact test was not used though some cells <5 as $n$>120).
*HCW, healthcare worker; PCR, polymerase chain reaction.*

uninfected controls at baseline and on the 52 seropositive HCWs at the time of first seroconversion. Serum was brought to room temperature (RT) and diluted 20x in provided dilution buffer prior to analysis. 100µL of standard and diluted sample were added to individual wells in the 96-well ELISA plate and incubated for 60 minutes at RT. Plates were washed 4 times with wash buffer using an automatic plate washer. 100 µL of diluted biotinylated tracer antibody was added to each well, the plate was then covered and incubated for 60 minutes at RT. The plate wash stage was repeated using the automatic plate washer. 100 µL of diluted streptavidin-peroxidase was added to each well. The plate was covered and incubated for 60 minutes at RT. Following a final round of plate washing, 100 µL of TMB substrate was added to each well. The plate was covered and incubated for 30 minutes at RT, following which, 100µL of stop solution was added. The plate was then read at 450nm using a Tecan Infinite/F200 plate reader.

### General statistical methods

All statistical analyses were performed in R[19] (version 4.1.2). Data were assumed to be nonparametric unless normality was demonstrated by visual assessment on histograms and using Shapiro-Wilk test and Q-Q plots. Continuous variables are expressed as mean ± 1 standard deviation (SD) or median (interquartile range [IQ]); categorical variables, as counts and percent. Correlations were performed with Pearson's correlation coefficient for parametric data. Kruskall–Wallis testing was performed to compare three or more groups, Wilcoxcon signed-rank test for non-parametric paired groups and Mann-Whitney U test for non-parametric unpaired groups. Contingency table difference tests were by Fisher's exact test if small numbers were available in each cell. Otherwise, Chi-square testing was used. Unless otherwise stated, significance was based on a false discovery rate (FDR) using Benjamini–Hochberg correction of less than 0.05. To compare global analyte perturbations between PCR positive cases and uninfected HCW, we employed a variation of the molecular degree of perturbation (MDP) method wherein we inputted normalized plasma concentrations of proteomics biomarkers as previously described,[20] instead of the standard gene expression values. Average plasma protein concentration levels and SD were first calculated in our baseline reference group (uninfected HCW), and biomarker-specific MDP scores calculated as the difference in concentration levels in PCR positive cases, from the reference average divided by the reference SD.[20] Associations between baseline symptom index, symptom duration, and anti-S1, anti-NP and total antibody levels with persistent symptoms were tested using logistic regression.

We applied random forest-based modelling using the Breiman and Cutler algorithm in R package 'randomForest', as previously described,[21] to rank the 91 plasma proteomics biomarkers according to their discriminative power to predict persistent symptoms at 12 months. All seropositive HCW were used as a training set. The model was trained with 1000 trees and then optimized for the number of variables selected for each tree. For variable selection, proteomics biomarkers were ranked based on mean decrease in accuracy and mean decrease in Gini scores. The final model and biomarker candidates were presented as receiver operator curve (ROC) curves. In the absence of a separate test dataset to validate the performance of the random forest multimarker panel in predicting persistent symptoms, we

undertook linear discriminant analysis (LDA) using R package 'MASS' in the same cohort of seropositive HCW to estimate the misclassification rate.

### Clustering analysis

Principal component analysis (PCA) was performed using the factoextra package (https://cran.rproject.org/web/packages/factoextra/index.html) to extract and visualize PCA elements. Hierarchal clustering analysis was performed using function 'hclust', R/bioconductor package 'ComplexHeatmap' and k-means clustering.

### Pathway analysis

We completed a differential expression analysis of the 91 assayed proteins using a $\log_2$ fold change threshold of 1.5 and FDR threshold of 0.05. Proteins from the MS dataset were inputted in to PathfindR along with their respective $\log_2$ fold change and FDR value.[22] We used the KEGG pathway database (https://www.genome.jp/kegg/pathway.html). Only pathways that had a $p$-value of $\leq 0.05$ were considered. We implemented hierarchical clustering using a pairwise distance matrix based on the kappa statistics between the enriched terms. The active sub-network search requires a minimum of 10 imputed genes. The final table produced by PathfindR includes a table of significant pathways with an associated adjusted $p$-value, a fold enrichment value of the pathway, the lowest and highest $p$-values generated from each iteration of the pathways analysis, and the proteins with increased or decreased abundance from the input protein list for every pathway.[22]

### Role of funders

The funders had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

## Results

### Study participants

We undertook a nested cross-sectional case-control study of 156 HCW (Table 1) derived from a cohort of 400 HCW at one of the London hospitals (Barts Heart Center) recruited from 23rd March 2020 onwards. From within the original cohort, we obtained 218 blood proteomic profiles from 54 individuals with PCR or antibody confirmed infections of ancestral SARS-CoV-2 during the first UK wave (Figure 1). We followed 29 HCW with PCR positive SARS-CoV-2 infection and compared these with blood proteomic profiles obtained from baseline samples in 102 uninfected HCW who remained PCR negative and seronegative for SARS-CoV-2 spike (S1 RBD) and nucleocapsid (N) during 16 week follow-up. None of the SARS-CoV-2 infected individuals required hospitalization. A symptom severity index per study member was calculated as the sum of reported symptoms following the first positive PCR test (no symptoms = 0, atypical symptoms = 1, case-defining symptoms = 2), divided by the total symptom duration in days. Symptom questionaries were completed (Table S2) at each sampling visit and then again at 6 and 12 months to capture data on persistent symptoms. We used a simplified ontological approach capturing any HCW reporting 1 or more persistent symptom out to 12 months.

Twenty-five HCW seroconverted during follow-up. In our study seroconversion occurred about 2 weeks after positive PCR tests in up to 46% of HCW, with a median seroconversion lag of 1.8w (IQR 0 – 3w).

### Perturbation of the plasma proteome in the weeks following infection

Targeted proteomic multiple-reaction monitoring analysis of 91 proteins (Table S1) identified 12 (Figure S1) differentially expressed proteins (increased abundance) between PCR positive cases at the time of their first

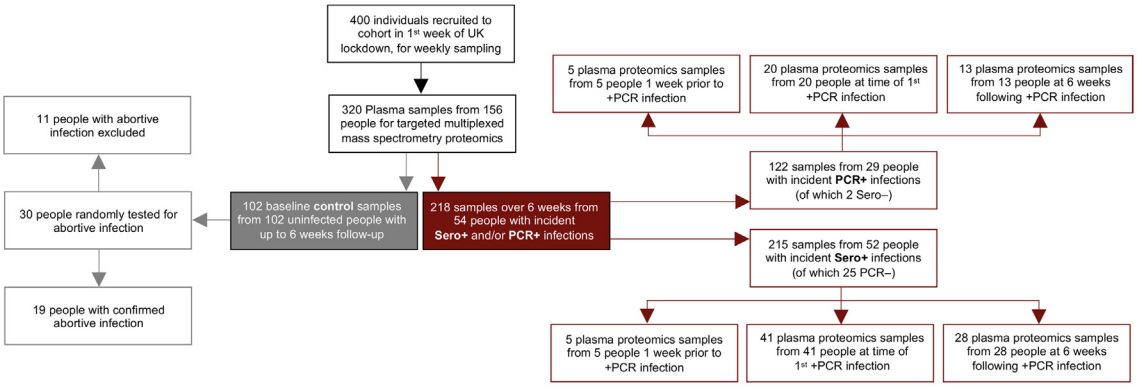


**Figure 1. Study consort diagram.**
*NIC, non-infected control; PCR, polymerase chain reaction; Sero+, seropositive.*

positive PCR test compared to uninfected HCW at baseline. The majority of these proteins were also found in increased abundance at the time of first seroconversion (Figure S2). The combined molecular degree of perturbation (MDP, see ***General Statistical Methods***) caused by this 12-biomarker panel was accentuated in those who had more symptoms during their acute SARS-CoV-2 infection (Figure 2a,b). Perturbed expression of these 12 proteins persisted for up to 6 weeks from the date of first positive PCR test (Figure 2c). Additional proteins were also found to be dysregulated at 6 weeks (Figure S3). Separate PCA for the combined plasma proteome (considering all 91 analytes) at each time point, confirmed sustained plasma proteomic perturbation out to week-6 (Figure 2d) and determined the main drivers of the observed variance. We observed overlapping plasma proteomics signatures between PCR positive and uninfected HCW at the time of first positive PCR test (week-0) and 1 week earlier, but progressive divergence thereafter and out to 6 weeks, largely determined by PC2 and

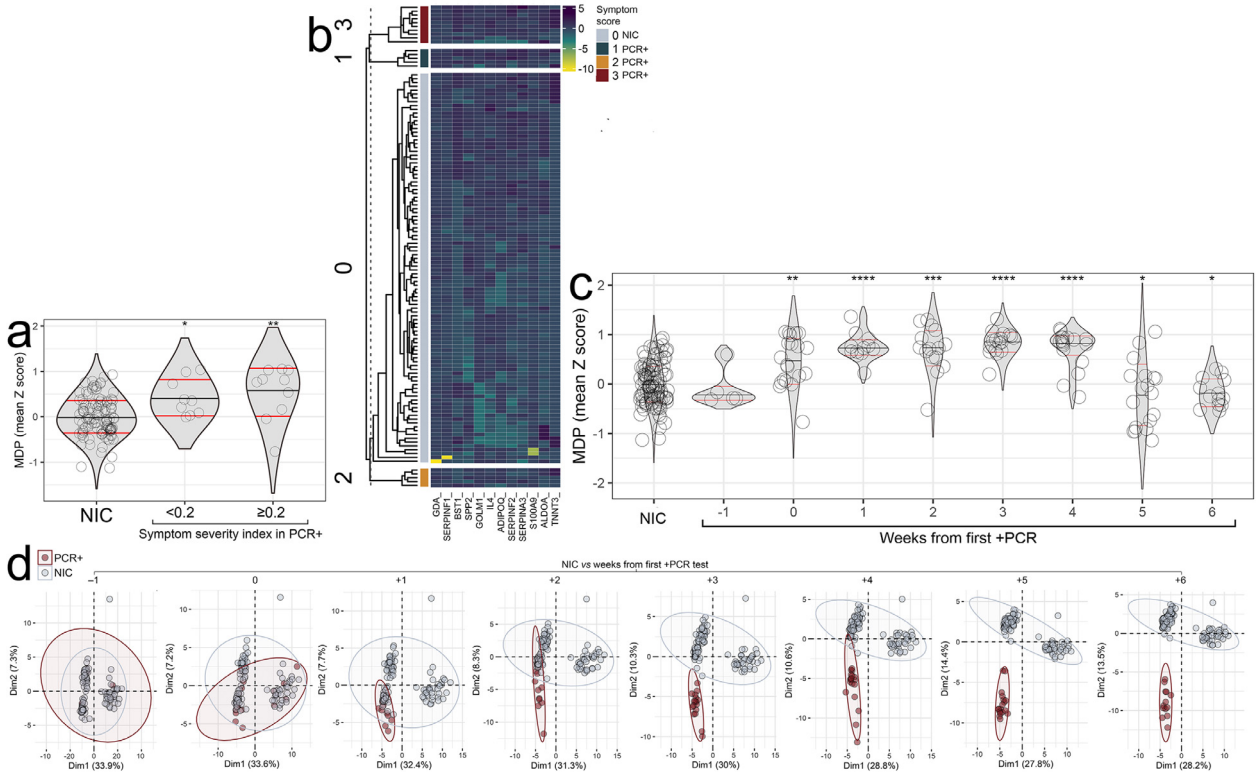


**Figure 2. Plasma proteome perturbation tracks symptom severity and lasts several weeks following SARS-CoV-2 infection**.

(a) MDP in plasma proteome for each individual expressed as the differences by number of standard deviations (*Z* scores) from the mean of uninfected HCW group, among samples from all uninfected HCW at baseline (n=102) and samples from PCR positive individuals at the time of their first positive PCR test result (n=29), the latter stratified according to symptom severity index. A symptom severity index per study member was calculated as the sum of reported symptoms following the first positive PCR test (no symptoms = 0, atypical symptoms = 1, case-defining symptoms = 2), divided by the total symptom duration in days. Violin plots depict median (black horizontal lines), IQR (red horizontal lines) and frequency distributions (FDR by pairwise Mann-Whitney tests with post hoc correction for each group compared to uninfected HCW, * $p = 0.016$; ** $p = 0.005$; non-significant $p$ value for differences between low and high symptom index PCR positive groups not shown). (**b**) Heatmap of 12 differentially abundant proteins between uninfected HCW (n=102) and PCR positive HCW at the time of first PCR positive test (n=29), grouped and then clustered according to symptom index. Colour bar represents range of normalized plasma concentration for each protein. Cluster '3' representing PCR positive cases with case-definition symptoms express the highest levels of these proteomics biomarkers compared to other PCR positive groups and uninfected HCW. Clusters '1' (PCR positive with no symptoms) and '2' (PCR positive with atypical symptoms) unsurprisingly cluster around '0' (uninfected controls with no symptoms). (**c**) MDP in plasma proteome comparing baseline uninfected HCW against serial samples from PCR positive individuals starting 1 week prior to first PCR positive test. (**d**) Principal component analysis plots comparing the 91-multimarker plasma proteome in all uninfected HCW (n=102) vs infected PCR positive participants (n=29) starting 1 week prior (left) to 6 weeks post (right) first PCR positive test. Individual points are coloured by group. Concentration ellipses of uninfected and infected HCW progressively diverge starting at week 0 (first PCR positive test), leading to complete separation of groups from week 3 onwards.

*FDR, false discovery rate; IQR, interquartile range; MDP, molecular degree of perturbation. Other abbreviations as in* Figure 1.

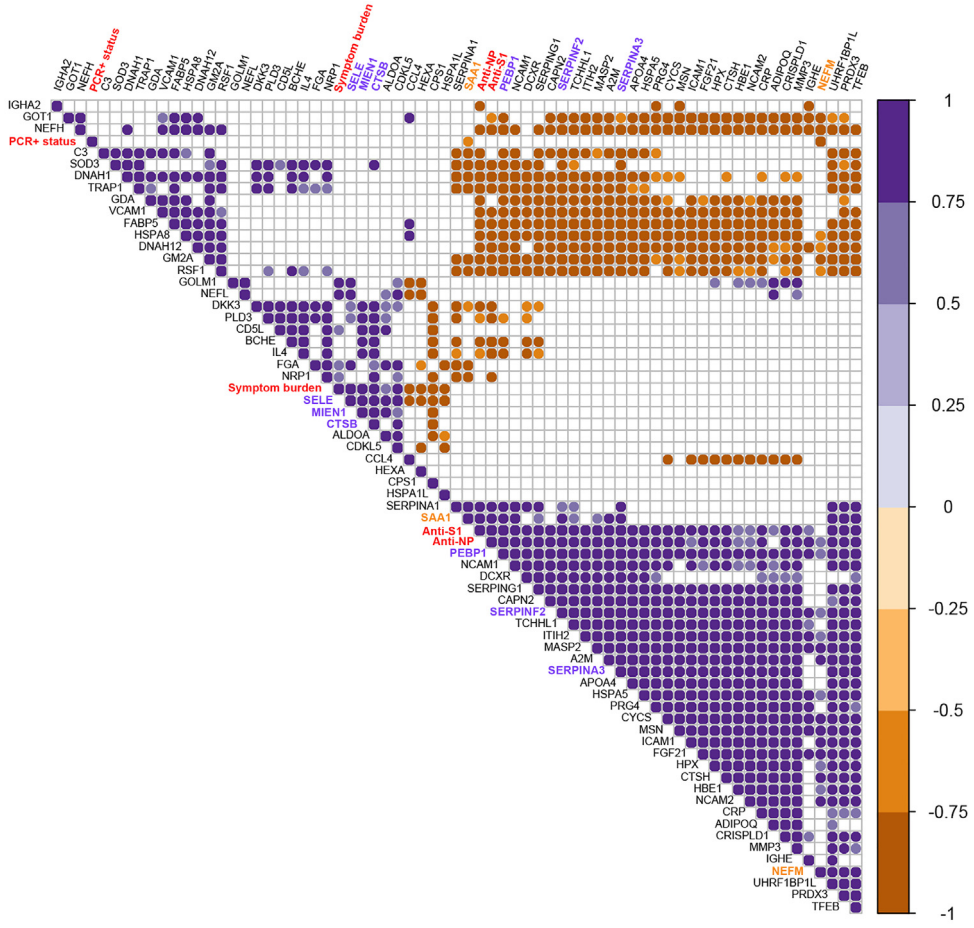


**Figure 3. Temporal correlogram of plasma proteomics trajectories with symptom burden, PCR-positivity and antibody titers during SARS-CoV-2 infection**.

Correlogram of individual normally-distributed plasma protein trajectories (i.e. protein levels during weeks -1 to 6) to Euroimmun anti-S1 spike protein (anti-S1) and Roche anti-nucleocapsid (anti-NP) titers, symptom-burden and PCR positive counts at these same timepoints, in infected participants ($n$=29) from 1 week prior to 6 weeks post, first PCR positive test. Only significant correlations ($p>0.05$) are shown. Circle colours reflect strength of the Pearson's correlation coefficient; interesting positive and negative correlations are highlighted in purple/orange and reported in the main text. *Other abbreviations as in* Figure 1.

to a lesser extent PC1. The top five drivers of this variance across all timepoints were Hemopexin (HPX, 3.777), Fibroblast growth factor 21 (FGF21, 3.754), Sulfhydryl oxidase 1 (QSOX1, 3.692), Neural cell adhesion molecule 2 (NCAM2, 3.665), and Mannan binding lectin serine peptidase 2 (MAPS2, 3.639). PCA also revealed two clusters of uninfected HCW, with minimal clinicodemographic differences between them (Table S3) and no differences stemming from their order in the MS runs (Figure S4).

### Plasma proteome trajectories track symptoms, antibody titer and PCR status during infection

Time-course trajectories of the 91 individual plasma proteins (Figure S5) starting 1 week prior, to 6 weeks post, first positive PCR test overlaid onto longitudinal plots of antibody titer, PCR status and symptom burden, revealed differential time-varying protein expression profiles during SARS-CoV-2 infection. Individual plasma protein trajectories that most strongly correlated (Figure 3) with symptom burden were: E-selectin (SELE, r = 0.97 $p$ = 0.00007, Pearson's correlation coefficient), Cathepsin B (CTSB r = 0.87 $p$ = 0.0005) and Migration and invasion enhancer 1 (MIEN1 r = 0.87 $p$ = 0.005); with PCR positive status: Neurofilament medium polypeptide (NEFM r = −0.90 $p$ = 0.003) and Serum amyloid A-1 protein (SAA1, r = −0.73 $p$ = 0.042); and with anti-S1/NP antibody titer: Alpha-2-antiplasmin (SERPINF2, r = 093 $p$ = 0.0008); Phosphatidylethanolamine-binding protein 1 (PEBP1, r = 0.93 $p$ = 0.0009); and Alpha-1-antichymotrypsin (SERPINA3, r = 0.81 $p$ = 0.0006).

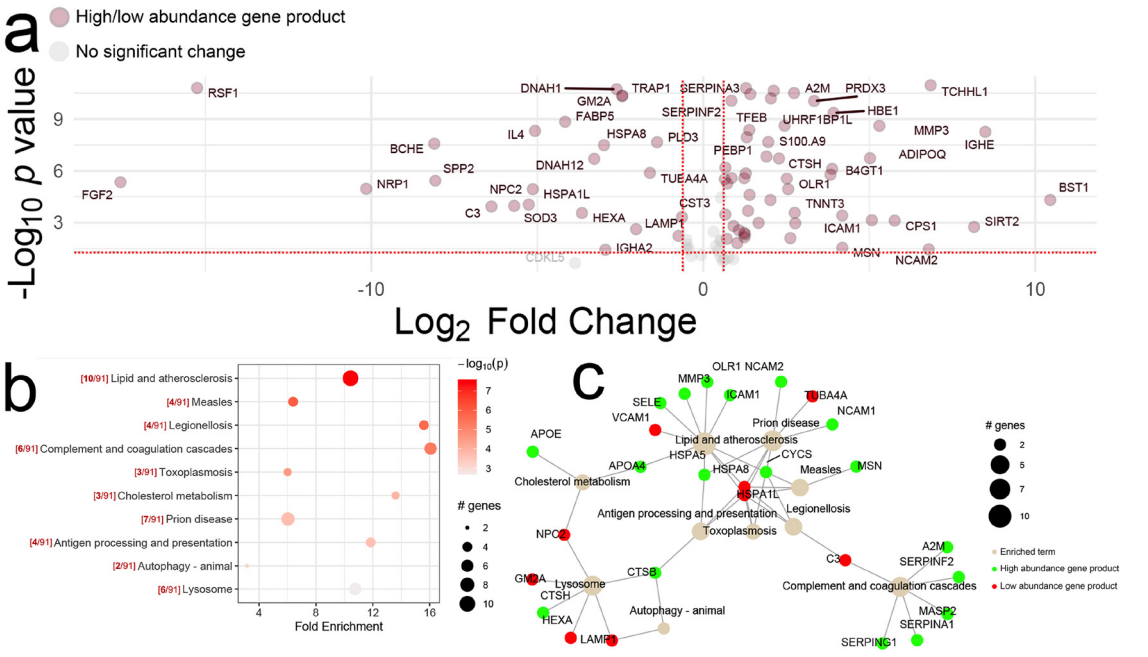


**Figure 4. Differentially detected plasma proteins and ranked pathway analysis for PCR positive state compared to uninfected controls.**

(**a**) Using this 91-protein multi-marker assay that was enriched from the outset for neuroinflammatory analytes, we observed 83 differentially detected proteins, log2 fold change threshold of 1.5 and FDR threshold of 0.05, between the plasma of infected participants (*n*=29) from 1 week prior to 6 weeks post first PCR positive test when compared to uninfected controls at baseline (*n*=52), along with ranked pathway analysis in **b** [also showing in red, the number of enriched genes per pathway relative to the total number of assayed genes]. Size of dots represents number of enriched genes within our neuroinflammatory enriched assayed set of 91 proteins, and intensity of colours represents *p*-value associated with enrichment analysis. (**c**) Network analyses for ranked pathway analysis (showing top 10 terms only) between these cases and controls, where abundantly present proteins are coloured in green and low abundance proteins are coloured in red. Node size of the pathway term represents number of input proteins implicated in the pathway.

*Other Abbreviations as in* Figures 1 and 2.

### Pathways of activation in mild/asymptomatic SARS-CoV-2 infection

The volcano plot presented in Figure 4a explored the differential expression of our 91 assayed plasma proteins between infected and uninfected HCW. To identify pathways of relevance to the pathophysiology of non-severe SARS-CoV-2 infection amongst the assayed proteins, we performed active-subnetwork-oriented enrichment analysis using pathfindR.[22] Enrichments and ranked pathways based on our 91 assayed proteins are presented in Figure 4b,c. Lists of ranked pathways are available in Supplementary Data. Notably, differentially abundant proteins were mostly coordinated around lipid, atherosclerosis and cholesterol metabolism pathways as well as around complement and coagulation cascades, autophagy, and lysosomal function.

### Plasma proteome at seroconversion predicts persistent symptoms out to 12 months

Considering the plasma proteome of 52 HCW at the time of their first positive anti-S1/NP antibody test, 11 of whom reported persistent symptoms out to 12 months (21%), we identified a proteomic profile (Figure 5a) that predicted persistent symptoms with and area under the curve (AUC) = 1 (accuracy: 1; 95% confidence intervals [CI]: 0.9, 1; Figure 5b). Circulating plasma protein biomarkers at the time of seroconversion that most strongly associated with persistent symptoms included Iron-sulfur cluster co-chaperone protein HscB (HSCB), Heat shock protein HSP 90-beta (HSP90AB1), Amyloid-beta precursor protein (APP), Phospholipase D Family Member 3 (PLD3), Cystatin-C (CST3) and Calprotectin (S100-A9). We found that out of the top 20 proteomics biomarkers associated with persistent symptoms, 5 continued to be abnormally abundant at 6 weeks in the plasma of infected HCW (Figure S3). To internally validate the multi-marker's ability to discriminate between infected HCW who did and did not experience persistent symptoms out to 12 months, we performed LDA and estimated the probabilities of misclassification of observation (Figure 5c,d). From the confusion matrix (Table S4) the apparent error rate for the LDA is 0.06 and the percentage of correctly classified (PCC) for LDA was 94%.

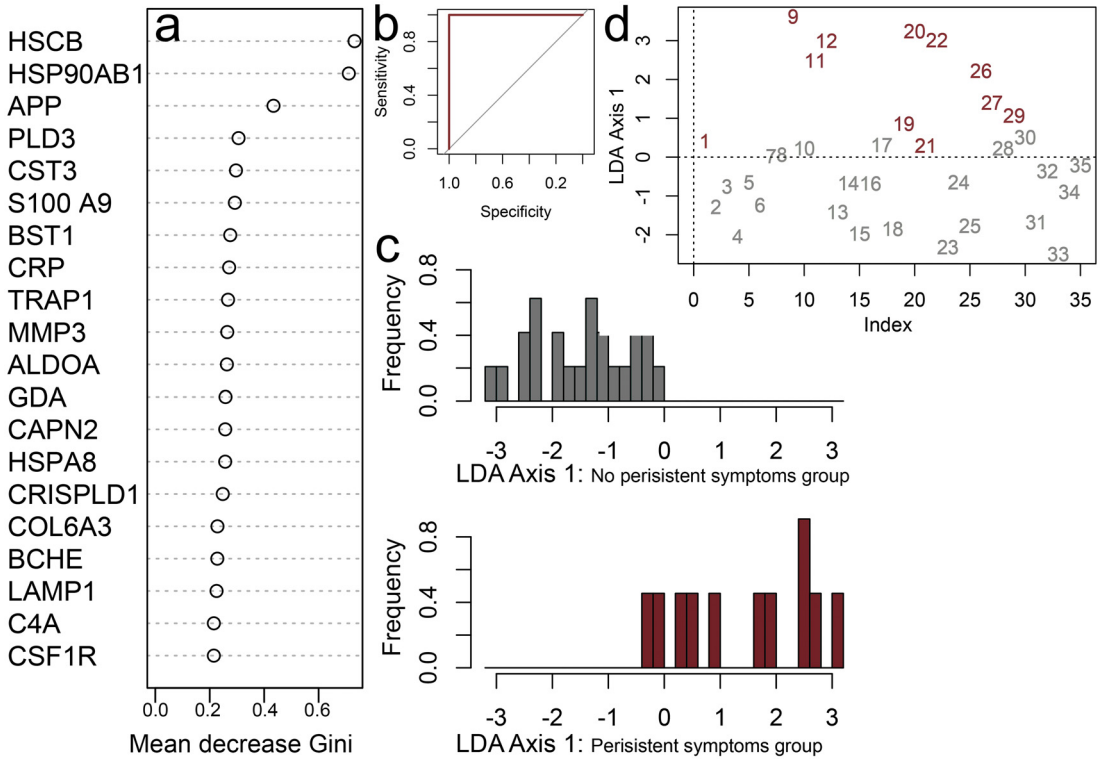


**Figure 5. Plasma proteome profile at time of seroconversion associates with persistent symptoms.**

Random forest multi-marker prediction of persistent symptoms at 12 months based on plasma proteomics profiles of healthcare workers at the time of first seroconversion for SARS-CoV-2. (**a**) Top 20 most predictive proteomics biomarkers are shown, ranked according to Gini coefficients. (**b**) In this cohort of *n*=156 healthcare workers, the multi-marker panel at the time of first seroconversion, predicts persistent symptoms at 12 months with an area under the curve (AUC) of 1.0. Applying linear discriminant analysis (LDA) to this same multi-marker panel (**c,d**) distinguishes between those with and without persistent symptoms at 12 months, with a percentage of correctly classified (PCC) = 94% (apparent error rate [AER] 0.06; two symptomatic cases misclassified "1" and "21").

For Calprotectin—one of the top predictors of persistent symptoms in our assay—we additionally undertook a separate ELISA on these plasma samples and noted a trend for higher levels in HCW with persistent symptoms vs those without (1072 ± 273 ng/mL vs 964 ± 327 ng/mL, $p$ = 0.150). This highlights the advantage of using a multi-marker panel instead of individual biomarkers.

Separately we also tested whether symptom index, symptom duration, and anti-S1, anti-NP or total antibody levels at baseline associated with persistent symptoms out to 12 months, but found no significant associations (respectively: β -0.03, $p$ = 0.680; β 0.18, $p$ = 0.244; β 0.01, $p$ = 0.334; β 1.76, $p$ = 0.255; β 0.03, $p$ = 0.096, logistic regression).

## Discussion

We report plasma proteomics responses to SARS-CoV-2 infection during the first UK wave, made possible due to high-frequency serial biosampling of HCW at risk of infection during the peak of the first epidemic wave in London. Non-severe SARS-CoV-2 infection markedly perturbs the plasma proteome from the time of first infection, and for the ensuing 6 weeks. This plasma proteomics perturbation is related to the extent of symptoms and associates with the presence of persistent symptoms (Figure 6).

The sustained perturbation of the plasma proteome following non-severe SARS-CoV-2 uncovered in the current study, fits with similar findings from a pilot using the same assay in individuals infected with SARS-CoV-2 showing mild or absent symptoms.[7] PCA of the plasma proteome in this independent cohort was most discriminant of infected cases (n = 10) compared to uninfected individuals (n = 10) at 6 weeks, in keeping with our PCA findings reported in Figure 2d. We go on to show that the plasma proteome of HCW in the week prior to PCR-confirmed SARS-CoV-2 infection is indistinguishable from that of uninfected HCW, that the earliest detectable perturbation coincides with the first PCR positive test result, and that it is progressively accentuated in subsequent weeks, resulting in complete separation of groups from week-five onwards.

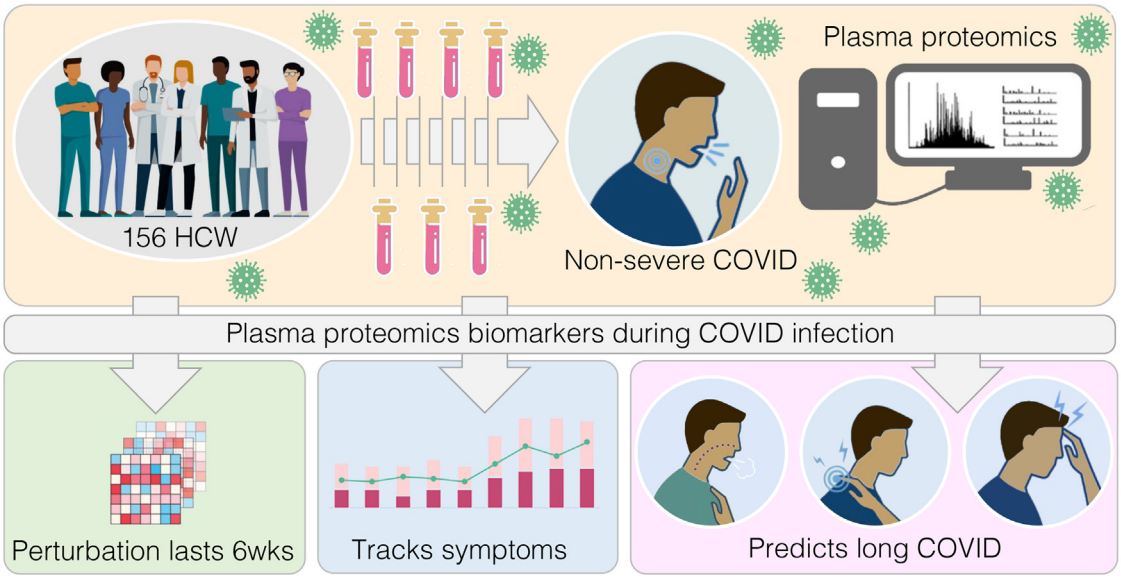


**Figure 6. Graphical abstract.**

A number of studies have previously undertaken a discovery proteomics approach to study the perturbation of the plasma proteome during acute infection,[23] including serial proteomics in mild or low symptom individuals,[24] and in hospitalized patients.[25,26] In line with previous works and using our assay that was enriched for neuroinflammatory biomarkers, we found that drivers of proteomic perturbation include oxidative stress markers[27] (e.g., QSOX1), metabolic reprogramming factors[28] (e.g., FGF21) and cell adhesion molecules[29] (e.g., NCAM2). However, plasma proteome signatures of severe hospitalized patients are expectedly different from ours as a result of acute inflammatory changes that would occur not just from initial infection but from the subsequent cascade of events that are not COVID-specific, such as mechanical ventilation or prolonged immobilisation. This can overshadow what may be the subtle changes occurring from infection. A previous study that also looked at non-hospitalized patients with SARS-CoV-2 found that certain symptoms such as fatigue and shortness of breath were still present up to 4-7 months after infection.[30] Only antibody serological analysis was performed and results suggested that lower baseline IgG was related to persistence of symptoms. We found no significant association between anti-S1 and anti-NP levels at baseline in those with persistent symptoms out to 12 months. Within our assay, one of the proteins that most strongly predicted persistent symptoms is APP, which in serum can act as an anti-coagulant. About 90% of soluble APP is thought to be derived by secretion from activated platelets and it acts as an inhibitor of coagulation factors IXa and Xia.[31] The iron-sulfur cluster biogenesis protein (HSCB also known as HSC20) which is a mitochondrial co-chaperone, has a key role in red blood cell erythropoiesis and hematopoiesis[32] and also predicted persistent symptoms. It is plausible that HSCB is being released into the blood secondary to mitochondrial disruption from COVID-mediated red blood cell destruction. Indeed, it has been amply shown[33] that acute COVID-19 infection causes multiple autoimmune heamolytic anaemia (AIHA) subtypes, beginning approximately 7 days after infectious symptoms. So what we may be capturing here by MS are milder subclinical forms of AIHA in non-severe SARS-CoV-2 infection which nevertheless cause measurable protein abnormalities in plasma. Extracellular HSP90 another heat shock protein which is functionally diverse but essential for maintaining healthy cells, could be indicative of vascular stress as it has been shown to be released by vascular smooth muscle cells under oxidative stress. Serum HSP90 is then thought to act as a cytokine stimulating IL-8[34] and it has a role as biomarker that can stratify systemic sclerosis in relation to pulmonary function.[35] PLD3 is an endo lysosomal protein with no detectable phospholipase activity. Along with PLD4 it has an anti-inflammatory function as it acts as an ssDNA exonuclease that breaks down Toll-like receptor 9 ligands mediating the TLR9 response.[36] We previously described Cystatin-C as being raised after SARS-CoV-2 infection[7] which we confirm in this larger cohort, however, it also appears to have utility to predict persistent symptoms. In mice, serum CST3 is controlled by the anti-inflammatory cytokine IL10 of which increasing levels suppress CST3 expression.[37] A longitudinal study looking at immune mediators show IL10 levels are significantly elevated at 4 weeks only in

severe cases of SARS-CoV-2 infection and not in milder cases.[38] This corroborates what we observe for CST3 as mild infection has increased cystatin C that is not being suppressed by higher IL-10 levels. S100A9 forms part of calprotectin which is a marker of neutrophil-related inflammatory processes. Silvin et al. first described that elevated calprotectin can indicate abnormal myeloid cells driving severe infection in patients infected with SARS-CoV-2. Since then many studies have now determined calprotectin as a biomarker of severe SARS-CoV-2 infection[12,39,40] detectable in the blood and lungs. Whilst the levels we observed were vastly lower than those expected in severely affected patients, our data suggests that higher calprotectin levels in mildly symptomatic patients at the time of first infection could be a risk marker of persistent symptoms. Recent studies have shown that impaired pulmonary function is detectable several months after SARS-CoV-2 infection[41–43] and this could be the underlying cause for persistent symptoms in some individuals. The perturbed proteins we describe here have anti-coagulation and anti-inflammatory functions in health, while other altered proteins have been implicated in erythropoiesis and hematopoiesis[32] and have links with lung disease so it is plausible that we are detecting surrogate indicators of possible mild pulmonary function impairment.

Pathway enrichment analysis of samples from our HCW with non-severe SARS-CoV-2 infection revealed pathways previously reported including lipid atherosclerosis and cholesterol metabolism pathways,[28] complement and coagulation cascades,[29,44,45] lysosomal function and autophagy.[46] Host lipids play a vital role in the life cycle of viruses as they are required for viral genome replication, assembly, maturation and transport through cell membrane.[44] For example, SARS-CoV-2 infection relies on its spike protein binding to angiotensin converting enzyme 2 (ACE2) receptor. As the later resides into a cholesterol-rich environment, alteration in membrane cholesterol composition would interfere with membrane dynamics which can potentially result in failure of endocytosis.[45] In addition, apolipoproteins also modulate the immune response, inflammation and even haematopoiesis meaning that they have a pivotal role in viral responses.[11] Thus, it is not surprising that downregulated apolipoproteins associated with worse outcomes (e.g., apolipoprotein A,[46] apolipoprotein E[11]) are more abundant in the non-severe infections. The interplay between SARS-CoV-2 and the coagulation cascade has become apparent in the early stages of the pandemic when increased rates of thromboembolic events were noted in acute infections.[47] However, little is known about the coagulation dysregulation in the convalescence period. We found greater abundance in plasma of molecules such as $\alpha$2-macroglobulin. These might protect against thrombo-inflammation in the context of SARS-CoV-2 by enhancing antithrombin activity, and binding coagulation inducing proteases and inflammatory mediators.[48]

We report that the plasma trajectory of E-selectin displayed the strongest correlation with symptom burden which fits with previous data linking endothelial markers such as E-selectin to SARS-CoV-2 infection severity.[49,50] Increased expression of selectins and other cell adhesion molecules in the pulmonary endothelium would recruit immune cells such as neutrophils and monocytes that then release pro-inflammatory cytokines which are associated with pro-coagulation processes and the generation of reactive oxygen species which can further damage the endothelium.[51] To infect target cells, SARS-CoV-2 relies on host serine or cysteine proteases to cleave its spike protein into an active conformation. Correlated with symptom severity in our study, the serine protease Cathepsin B has been previously identified as an efficient activator of the spike protein in-vivo and provides a novel therapeutic target for drug development.[52] Cathepsin inhibitors exist and have been in development for more than a decade,[53] however they have not been tested in COVID-19 clinical trials whether alone[54] or in synergistic combination with cysteine proteases.[55] The fact that Omicron relies mostly on the cathepsin pathway for infecting cells,[56] is likely to heighten the interest around Cathepsin inhibitors.

More than 30% of individuals have persistent symptoms following acute SARS-CoV-2 infection.[57] A clear cause has not yet been found but current theories[58,59] revolve around immune dysregulation,[60] clotting dysfunction or virus persistence.[61] We show that signature biomarkers at the time of seroconversion for SARS-CoV-2 associates with the persistence of symptoms at 12 months (Figure 5). The most predictive biomarker according to the Gini coefficient was HSCB which facilitates the insertion of iron-sulfur clusters in many cytoplasmic and mitochondrial proteins.[62] Its increased synthesis at seroconversion hints towards the need to handle higher iron levels as a hyperferritinimic state has been previously linked to tissue damage (e.g., through enhanced oxidative stress and lipid peroxidase) and impaired immunity[63] which could explain the noted association with persistent symptoms. The functional intersection between iron metabolism and SARS-CoV-2 was highlighted when high levels of ferritin were previously found to associate with admission to intensive care and mortality.[64] Since then, high levels of ferritin were also shown to be present in those with ongoing impaired physical performance after SARS-CoV-2 infection.[65] This iron metabolism dysregulation in COVID-19 has a pathophysiological basis as the similarity between hepcidin and the SARS-CoV-2 spike protein has been recently discovered.[66]

Some of our persistent symptoms signature biomarkers were persistently elevated at 6 weeks (Figure S3). These can be broadly divided into two categories:

proteins related to inflammation and those related to the stress response. An ongoing elevated C-reactive protein suggests a residual systematic inflammatory response.[67] However, the upregulation of collagen (e.g., COL6A3) and metalloproteases genes (e.g., MMP3) suggests concomitant activation of the repair pathway via fibrotic remodelling. Increased expression of S100A9 was previously related to neurodegenerative amyloid disease.[68] Continued mitochondrial PRDX3 expression at 6 weeks suggests a continued oxidative stress response, while persistent upregulation of glycolysis-related genes such as ALDOA indicate a catabolic stress response,[69] that is further supported by clinical (e.g., weight loss[57]) and biochemical low albumin[67]) findings from other studies.

### Limitations

Our study has some important limitations. This was a single-centre study and the generally low sample number as well as the low number of samples available from HCW in the week before their first PCR positive test, was small (n = 5). Though every attempt was made to minimise self-reporting bias, this could still have potentially influenced survey results. External independent cohort validation of the multi-marker proteomics profile predictive of persistent symptoms was not undertaken but will be a key focus of our future work. To predict persistent symptoms, we used proteomics signatures at the time of first seroconversion ($n$ = 52) due to smaller numbers available for first PCR positive test ($n$ = 29). Given all data collection occurred early during the first wave, differences by variant or those caused by vaccination status, were beyond the scope of this study.

## Conclusion

Non-severe SARS-CoV-2 infection perturbs the plasma proteome for at least 6 weeks following the first positive PCR test. Plasma proteomics signatures track symptom severity and antibody response, and have the potential to identify individuals who are more likely to suffer from persistent symptoms.

### Contributors

GC and JCM equally contributed to this work and are co-first authors. KM and WH equally contributed to this work and are co-senior authors. GC, JCM, WH and KM conceptualized the study reported. KM, WH and GC designed and supervised the proteomics experiments. GC, WH, NP, ID and TB have verified the underlying data. WH, ID, TB, JS and JH conducted the proteomics experiments. GC analysed the data and wrote the manuscript with input from all the authors. TB, JCM, CM, ÁM, TAT and MN conceptualized and established the HCW cohort. GJ, MH, KM, MF, KM, CM, TAT and JCM collected HCW samples. CT, GC, JCM, KM, WH, NP, CP, JMG, BOB, DMA, MKM, ÁM, TB, MN, AM, LS and RJB interpreted the data. All authors read and approved the final version of the manuscript.

### Data sharing statement

With publication all deidentified MS proteomics will be openly available via the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) and the PRIDE partner repository[18] with identifier PXD036590. All COVIDsortium HCW clinicodemographic data, including study protocols and templates of informed consent forms used for the study, are freely available following a data access request through the COVIDsortium data access portal. Source code and test data are available via GitHub (https://github.com/gcaptur/COVID-Proteomics).

### Declaration of interests

RJB and DMA are members of the Global T cell Expert Consortium and have consulted for Oxford Immunotec outside the submitted work.

### Acknowledgements

We gratefully acknowledge all the HCW participants for donating their samples and data for these analyses, and the research teams involved in consenting, recruitment and sampling of the HCW participants.

The COVIDsortium is supported by funding donated by individuals, charitable Trusts, and corporations including Goldman Sachs, Citadel and Citadel Securities, The Guy Foundation, GW Pharmaceuticals, Kusuma Trust, and Jagclif Charitable Trust, and enabled by Barts Charity with support from University College London Hospitals (UCLH) Charity. Wider support is acknowledged on the COVIDsortium website. Institutional support from Barts Health National Health Service (NHS) Trust and Royal Free NHS Foundation Trust facilitated study processes, in partnership with University College London and Queen Mary University of London. This work was additionally supported by the Translational Mass Spectrometry Research Group and the Biomedical Research Center (BRC) at Great Ormond Street Hospital.

GC is supported by the British Heart Foundation (BHF, SP/20/2/34841), the BHF Accelerator Award (AA/18/6/34223) and by the National Institute of Health Research (NIHR) UCL Hospitals BRC. GC, JCM and BOB are supported by the Barts Charity HeartOME1000 Grant (MGU0427/G-001411). KM and WH are supported by the NIHR BRC at Great Ormond Street Hospital for Children NHS Foundation Trust and UCL. RJB and DMA are supported by MRC (MR/S019553/1, MR/R02622X/1, MR/V036939/1, MR/W020610/1), NIHR Imperial Biomedical Research

Centre (BRC):ITMAT, Cystic Fibrosis Trust SRC (2019SRC015), NIHR EME Fast Track (NIHR134607), NIHR Long Covid (COV-LT2-0027), Innovate UK (SBRI 10008614) and Horizon 2020 Marie Skłodowska-Curie Innovative Training Network (ITN) European Training Network (No 860325). ÁM is supported by Rosetrees Trust, The John Black Charitable Foundation, and Medical College of St Bartholomew's Hospital Trust. JCM, CM and TAT are directly and indirectly supported by the UCLH and Barts NIHR BRCs and through the BHF Accelerator Award (AA/18/6/34223). TAT is funded by a BHF Intermediate Research Fellowship (FS/19/35/34374). MN is supported by the Wellcome Trust (207511/Z/17/Z) and by NIHR BRC Funding to UCL and UCLH. MKM is supported by UK Research and Innovation (UKRI)/NIHR UK-CIC, Wellcome Trust Investigator Award (214191/Z/18/Z) and Cancer Research UK Immunology grant (26603). ÁM, CM and JCM were supported by the UKRI/MRC Covid-19 Rapid response grant COV0331 MR/V027883/1.

## Supplementary materials

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.ebiom.2022.104293.

### References


1 Deer RR, Rock MA, Vasilevsky N, et al. Characterizing long COVID: deep phenotype of a complex condition. *EBioMedicine*. 2021;74:103722.
2 World Health Organization. A clinical case definition of post COVID-19 condition by a Delphi consensus, 6 October 2021. 2021.
3 Pérez-González A, Araújo-Ameijeiras A, Fernández-Villar A, Crespo M, Poveda E. Long COVID in hospitalized and non-hospitalized patients in a large cohort in Northwest Spain, a prospective cohort study. *Sci Rep*. 2022;12:3369.
4 Bliddal S, Banasik K, Pedersen OB, et al. Acute and persistent symptoms in non-hospitalized PCR-confirmed COVID-19 patients. *Sci Rep*. 2021;11:13153.
5 Briggs A, Vassall A. Count the cost of disability caused by COVID-19. *Nature*. 2021;593:502–505.
6 Augusto JB, Menacho K, Andiapen M, et al. Healthcare Workers Bioresource: Study outline and baseline characteristics of a prospective healthcare worker cohort to study immune protection and pathogenesis in COVID-19. *Wellcome Open Res*. 2020;5:179.
7 Doykov I, Hällqvist J, Gilmour KC, Grandjean L, Mills K, Heywood WE. "The long tail of Covid-19" - The detection of a prolonged inflammatory response after a SARS-CoV-2 infection in asymptomatic and mildly affected patients. *F1000Research*. 2020;9:1349.
8 Treibel TA, Manisty C, Burton M, et al. COVID-19: PCR screening of asymptomatic health-care workers at London hospital. *Lancet*. 2020;395:1608–1610.
9 Treibel TA, Manisty C, Andiapen M, et al. Asymptomatic health-care worker screening during the COVID-19 pandemic - authors' reply. *Lancet*. 2020;396:1394–1395.
10 Reynolds CJ, Swadling L, Gibbons JM, et al. Discordant neutralizing antibody and T cell responses in asymptomatic and mild SARS-CoV-2 infection. *Sci Immunol*. 2020;5.
11 Demichev V, Tober-Lau P, Lemke O, et al. A time-resolved proteomic and prognostic map of COVID-19. *Cell Syst*. 2021;12:780–794. .e7.
12 Wu M, Chen Y, Xia H, et al. Transcriptional and proteomic insights into the host response in fatal COVID-19 cases. *Proc Natl Acad Sci*. 2020;117:28336–28343.
13 Messner CB, Demichev V, Wendisch D, et al. Ultra-high-throughput clinical proteomics reveals classifiers of COVID-19 infection. *Cell Syst*. 2020;11:11–24.e4.
14 Douaud G, Lee S, Alfaro-Almagro F, et al. SARS-CoV-2 is associated with changes in brain structure in UK Biobank. *Nature*. 2022;604:697–707.
15 Yang AC, Kern F, Losada PM, et al. Dysregulation of brain and choroid plexus cell types in severe COVID-19. *Nature*. 2021;595:565–571.
16 Deleidi M, Isacson O. Viral and inflammatory triggers of neurodegenerative diseases. *Sci Transl Med*. 2012;4. 121ps3.
17 MRM-LC-MS/MS. Assay for inflammatrory associated proteins in serum. *protocols.io*. 2020. https://doi.org/10.17504/protocols.io.bm4vk8w6.
18 Vizcaíno JA, Côté RG, Csordas A, et al. The PRoteomics IDEntifications (PRIDE) database and associated tools: status in 2013. *Nucleic Acids Res*. 2013;41:D1063–D1069.
19 Core Team R. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing; 2013. URL; http://www.R-project.org/.
20 Oliveira-de-Souza D, Vinhaes CL, Arriaga MB, et al. Molecular degree of perturbation of plasma inflammatory markers associated with tuberculosis reveals distinct disease profiles between Indian and Chinese populations. *Sci Rep*. 2019;9:8002.
21 Toth R, Schiffmann H, Hube-Magg C, et al. Random forest-based modelling to detect biomarkers for prostate cancer progression. *Clin Epigenetics*. 2019;11:148.
22 Ulgen E, Ozisik O, Sezerman OU. pathfindR: An R package for comprehensive identification of enriched pathways in omics data through active subnetworks. *Front Genet*. 2019;10:858.
23 Filbin MR, Mehta A, Schneider AM, et al. Longitudinal proteomic analysis of severe COVID-19 reveals survival-associated signatures, tissue-specific cell death, and cell-cell interactions. *Cell Reports Med*. 2021;2:100287.
24 Gisby J, Clarke CL, Medjeral-Thomas N, et al. Longitudinal proteomic profiling of dialysis patients with COVID-19 reveals markers of severity and predictors of death. *Elife*. 2021;10.
25 Vedula P, Tang H-Y, Speicher DW, Kashina A. Protein posttranslational signatures identified in COVID-19 patient plasma. *Front Cell Dev Biol*. 2022;10:807149.
26 Al-Nesf MAY, Abdesselem HB, Bensmail I, et al. Prognostic tools and candidate drugs based on plasma proteomics of patients with severe COVID-19 complications. *Nat Commun*. 2022;13:946.
27 Saheb Sharif-Askari N, Saheb Sharif-Askari F, Mdkhana B, et al. Upregulation of oxidative stress gene markers during SARS-COV-2 viral infection. *Free Radic Biol Med*. 2021;172:688–698.
28 Shen T, Wang T. Metabolic Reprogramming in COVID-19. *Int J Mol Sci*. 2021;22.
29 Tong M, Jiang Y, Xia D, et al. Elevated expression of serum endothelial cell adhesion molecules in COVID-19 patients. *J Infect Dis*. 2020;222:894–898.
30 Augustin M, Schommers P, Stecher M, et al. Post-COVID syndrome in non-hospitalised patients with COVID-19: a longitudinal prospective cohort study. *Lancet Reg Heal – Eur*. 2021;6:100122.
31 d'Uscio L V, He T, Katusic ZS. Expression and processing of amyloid precursor protein in vascular endothelium. *Physiology*. 2016;32:20–32.
32 Crispin A, Guo C, Chen C, et al. Mutations in the iron-sulfur cluster biogenesis protein HSCB cause congenital sideroblastic anemia. *J Clin Invest*. 2020;130:5245–5256.
33 Jacobs JW, Booth GS. COVID-19 and immune-mediated RBC destruction. *Am J Clin Pathol*. 2022;157:844–851.
34 Chung S-W, Lee J-H, Choi K-H, et al. Extracellular heat shock protein 90 induces interleukin-8 in vascular smooth muscle cells. *Biochem Biophys Res Commun*. 2009;378:444–449.
35 Štorkánová H, Oreská S, Špiritović M, et al. Plasma Hsp90 levels in patients with systemic sclerosis and relation to lung and skin involvement: a cross-sectional and longitudinal study. *Sci Rep*. 2021;111.
36 Gavin AL, Huang D, Huber C, et al. PLD3 and PLD4 are single-stranded acid exonucleases that regulate endosomal nucleic-acid sensing. *Nat Immunol*. 2018;19:942–953.
37 Xu Y, Schnorrer P, Proietto A, et al. IL-10 controls cystatin C synthesis and blood concentration in response to inflammation through regulation of IFN regulatory factor 8 expression. *J Immunol*. 2011;186:3666–3673.
38 Zhao Y, Qin L, Zhang P, et al. Longitudinal COVID-19 profiling associates IL-1RA and IL-10 with disease severity and RANTES with mild disease. *JCI Insight*. 2020;5.

39 Mahler M, Meroni P-L, Infantino M, Buhler KA, Fritzler MJ. Circulating Calprotectin as a biomarker of COVID-19 severity. *Expert Rev Clin Immunol*. 2021;17:43143.
40 Shu T, Ning W, Wu D, et al. Plasma proteomics identify biomarkers and pathogenesis of COVID-19. *Immunity*. 2020;53:1108–1122. .e5.
41 Townsend L, Dowds J, O'Brien K, et al. Persistent poor health after COVID-19 Is not associated with respiratory complications or initial disease severity. *Ann Am Thorac Soc*. 2021;18:997–1003.
42 Cheon IS, Li C, Son YM, et al. Immune signatures underlying post-acute COVID-19 lung sequelae. *Sci Immunol*. 2021;6:eabk1741.
43 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;55:542–556.e5.
44 Kočar E, Režen T, Rozman D. Cholesterol, lipoproteins, and COVID-19: basic concepts and clinical applications. *Biochim Biophys Acta Mol Cell Biol Lipids*. 2021;1866:158849.
45 Proto MC, Fiore D, Piscopo C, et al. Lipid homeostasis and mevalonate pathway in COVID-19: basic concepts and potential therapeutic targets. *Prog Lipid Res*. 2021;82:101099.
46 Zhu Z, Yang Y, Fan L, et al. Low serum level of apolipoprotein A1 may predict the severity of COVID-19: a retrospective study. *J Clin Lab Anal*. 2021;35:e23911.
47 Lazzaroni MG, Piantoni S, Masneri S, et al. Coagulation dysfunction in COVID-19: the interplay between inflammation, viral infection and the coagulation system. *Blood Rev*. 2021;46:100745.
48 Seitz R, Gürtler L, Schramm W. Thromboinflammation in COVID-19: Can $\alpha(2)$ -macroglobulin help to control the fire? *J Thromb Haemost*. 2021;19:351–354.
49 Oliva A, Rando E, Al Ismail D, et al. Role of serum E-selectin as a biomarker of infection severity in coronavirus disease 2019. *J Clin Med*. 2021;10:4018–4028.
50 Birnhuber A, Fließer E, Gorkiewicz G, et al. Between inflammation and thrombosis: endothelial cells in COVID-19. *Eur Respir J*. 2021;58:2100377–2100382.
51 Violi F, Oliva A, Cangemi R, et al. Nox2 activation in Covid-19. *Redox Biol*. 2020;36:101655.
52 Zhao M-M, Yang W-L, Yang F-Y, et al. Cathepsin L plays a key role in SARS-CoV-2 infection in humans and humanized mice and is a promising target for new drug development. *Signal Transduct Target Ther*. 2021;6:134.
53 Li Y-Y, Fang J, Ao G-Z. Cathepsin B and L inhibitors: a patent review (2010 - present). *Expert Opin Ther Pat*. 2017;27:643–656.
54 Cannalire R, Stefanelli I, Cerchia C, Beccari AR, Pelliccia S, Summa V. SARS-CoV-2 entry inhibitors: small molecules and peptides targeting virus or host cells. *Int J Mol Sci*. 2020;21.
55 Padmanabhan P, Desikan R, Dixit NM. Targeting TMPRSS2 and Cathepsin B/L together may be synergistic against SARS-CoV-2 infection. *PLoS Comput Biol*. 2020;16:e1008461.
56 Padmanabhan P, Dixit NM. Evidence of increased Cathepsin B/L and decreased TMPRSS2 usage for cell entry by the SARS-CoV-2 Omicron variant. *bioRxiv*. 2022:1–9. https://doi.org/10.1101/2022.01.13.476267.
57 Groff D, Sun A, Ssentongo AE, et al. Short-term and long-term rates of postacute sequelae of SARS-CoV-2 infection: a systematic review. *JAMA Netw Open*. 2021;4:e2128568.
58 Su Y, Yuan D, Chen DG, et al. Multiple early factors anticipate post-acute COVID-19 sequelae. *Cell*. 2022;185:881–895.e20.
59 Klein J, Wood J, Jaycox J. et al. Distinguishing features of Long COVID identified through immune profiling. *medRxiv* 2022.
60 Galeotti C, Bayry J. Autoimmune and inflammatory diseases following COVID-19. *Nat Rev Rheumatol*. 2020;16:413–414.
61 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.
62 Maio N, Ghezzi D, Verrigni D, et al. Disease-causing SDHAF1 mutations impair transfer of Fe-S clusters to SDHB. *Cell Metab*. 2016;23:292–302.
63 Cherayil BJ. Iron and immunity: immunological consequences of iron deficiency and overload. *Arch Immunol Ther Exp*. 2010;58:407–415.
64 Cheng L, Li H, Li L, et al. Ferritin in the coronavirus disease 2019 (COVID-19): a systematic review and meta-analysis. *J Clin Lab Anal*. 2020;34:e23618.
65 Sonnweber T, Boehm A, Sahanic S, et al. Persisting alterations of iron homeostasis in COVID-19 are associated with non-resolving lung pathologies and poor patients' performance: a prospective observational cohort study. *Respir Res*. 2020;21:276.
66 Ehsani S. COVID-19 and iron dysregulation: distant sequence similarity between hepcidin and the novel coronavirus spike glycoprotein. *Biol Direct*. 2020;15:19.
67 Gameil MA, Marzouk RE, Elsebaie AH, Rozaik SE. Long-term clinical and biochemical residue after COVID-19 recovery. *Egypt Liver J*. 2021;11:74.
68 Horvath I, Iashchishyn IA, Moskalenko RA, et al. Co-aggregation of pro-inflammatory S100A9 with $\alpha$-synuclein in Parkinson's disease: ex vivo and in vitro studies. *J Neuroinflammation*. 2018;15:172.
69 Horiguchi H, Loftus TJ, Hawkins RB, et al. Innate immunity in the persistent inflammation, immunosuppression, and catabolism syndrome and its implications for therapy. *Front Immunol*. 2018;9:595.