

Residual lung abnormality following COVID-19 hospitalisation is characterised by biomarkers of epithelial injury



Iain Stewart,^{a,b,*} Joseph Jacob,^{c,d,e} Joanna C. Porter,^{d,e} Bin Liu,^a Amanda L. Tatler,^{l,m} Nancy Gomez,^{l,m} Matthew R. Pugh,^{ac} Alison E. John,^a Richard J. Allen,^{f,g} John F. Blaikley,^{h,i} Nazia Chaudhuri,^j Emma Denny,^d Laura Fabbri,^{a,k} Peter M. George,^{a,k} Beatriz Guillen-Guio,^g Bibek Gooptu,^g Ian P. Hall,^{l,m} Ling Pei Ho,^{n,o} Ian Jarrold,^p Simon Johnson,^l Mark G. Jones,^q Fasihul Khan,^r Puja Mehta,^d Jane Mitchell,^{a,b} Philip L. Molyneaux,^{a,b} John E. Pearl,^s Karen Piper Hanley,^t Manuela Platé,^d Valerie Quinn,^a Pilar Rivera-Ortega,^h Laura C. Saunders,^u David J. F. Smith,^{a,b} Mark Spears,^v Lisa G. Spencer,^w Stefan C. Stanel,^{h,i} A. A. Roger Thompson,^{u,x} Simon Walsh,^k Jim M. Wild,^u Dan G. Wootton,^w Annemarie B. Docherty,^y Fergus Gleeson,ⁿ William Greenhalf,^{w,z} Ewen M. Harrison,^y Nazir Lone,^y Jennifer Quint,^a Anastasia Maslova,^{aa} Moritz Pohl,^{aa} Adam Stephens,^{aa} Simon Young,^{aa} Amisha Singapuri,^g Aarti Shikotra,^g Marco Sereno,^g Ruth M. Saunders,^g Matthew Richardson,^g Betty Raman,ⁿ Krishn Poinasamy,^p Hamish J. C. McAuley,^g Michael Marks,^{ab} Olivia C. Leavy,^f Linzy Houchen-Wolloff,^g Alex Horsley,^h Victoria C. Harris,^g Neil Greening,^g Rachael A. Evans,^g Omer Elneima,^g James D. Chalmers,^v Christopher E. Brightling,^{g,i,r} Rachel C. Chambers,^{d,ad} Louise V. Wain,^{f,g,ad} and R. Gisli Jenkins,^{a,b,k,ad} on behalf of the PHOSP-COVID Collaborative Group and the UKILD Consortium



^aNational Heart and Lung Institute, Imperial College London, UK

^bNIHR Imperial Biomedical Research Centre, Imperial College London, UK

^cSatsuma Lab, Centre for Medical Imaging Computing, University College London, UK

^dUCL Respiratory, University College London, UK

^eUniversity College London Hospitals NHS Foundation Trust, UK

^fDepartment of Health Sciences, University of Leicester, UK

^gNIHR Leicester Biomedical Research Centre, University of Leicester, UK

^hManchester University NHS Foundation Trust, UK

ⁱSchool of Biological Sciences, University of Manchester, UK

^jUlster University, UK

^kRoyal Brompton and Harefield Clinical Group, Guy's and St Thomas' NHS Foundation Trust, UK

^lSchool of Medicine, University of Nottingham, UK

^mNIHR Nottingham Biomedical Research Centre, University of Nottingham, UK

ⁿOxford University Hospitals NHS Foundation Trust, UK

^oRespiratory Medicine Unit, Nuffield Department of Medicine, University of Oxford, UK

^pAsthma + Lung UK, UK

^qUniversity Hospital Southampton NHS Foundation Trust, UK

^rUniversity Hospitals of Leicester NHS Trust, UK

^sDepartment of Respiratory Sciences, University of Leicester, UK

^tSchool of Medical Sciences, University of Manchester, UK

^uSchool of Medicine & Population Health, University of Sheffield, UK

^vNHS Tayside & University of Dundee, UK

^wLiverpool University Hospitals NHS Foundation Trust, UK

^xNIHR Sheffield Biomedical Research Centre, University of Sheffield, UK

^yCentre for Medical Informatics, The Usher Institute, University of Edinburgh, UK

^zUniversity of Liverpool, UK

^{aa}Sysmex R&D Centre UK, UK

^{ab}London School of Hygiene & Tropical Medicine, UK

^{ac}Institute of Immunology and Immunotherapy, University of Birmingham, UK

Summary

Background Long term respiratory symptoms are reported following recovery of acute COVID-19 infection and residual lung abnormalities (RLA) on follow-up thoracic computed tomography (CT) after COVID-19 hospitalisation have been observed. It is unknown whether RLA are associated with epithelial lung injury.

eBioMedicine
2026;124: 106134

Published Online xxx
<https://doi.org/10.1016/j.ebiom.2026.106134>

*Corresponding author. National Heart and Lung Institute, Imperial College London, UK.

E-mail address: iain.stewart@imperial.ac.uk (I. Stewart).

^{ad}Authors contributed equally.

Methods Plasma was sampled from the observational Post HOSPitalisation-COVID cohort at five months post-hospitalisation. Epithelial injury biomarkers Krebs von den Lungen-6 (KL-6), matrix metalloproteinase 7 (MMP-7), surfactant protein-D (SP-D) and surfactant protein-A (SP-A) were assayed. In those without follow-up CT, RLA at-risk was defined by percent predicted DL_{CO} <80% and/or abnormal chest X-ray, otherwise they were considered low-risk. Follow-up CT RLA was defined as combined involvement of ground glass opacity and reticulation ≥10%.

Findings A total of 957 people were included, 846 people with no CT (at-risk n = 103; 12.2%), 111 people with follow-up CT (RLA ≥10% n = 85; 76.6%). All epithelial injury biomarkers were significantly elevated in people at-risk of RLA compared with low-risk. KL-6 and MMP-7 were significantly higher in people with ≥10% RLA than those with <10%, SP-D and SP-A did not reach significance. SP-D and SP-A were associated with percent involvement of reticulation (3.22%, 95% CI 1.19–5.24; 3.03%, 95% CI 0.76–5.30, respectively).

Interpretation RLA after acute COVID-19 infection were consistent with elevated epithelial injury biomarkers and pro-fibrotic signalling. Future studies should address the temporal association between fibrotic biomarkers and resolution or progression of radiological involvement.

Funding MRC-UK Research and Innovation and National Institute for Health Research (NIHR) rapid response panel to tackle COVID-19 (MR/V027859/1; COV0319; MR/W006111/1).

Copyright © 2026 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Keywords: COVID-19; Hospitalisation; Lung injury; Biomarker; Epithelial

Research in context

Evidence before this study

Some survivors of acute COVID-19 infection have long-term symptoms that could suggest ongoing lung impairment. Searches performed in MEDLINE and Embase for SARS-COV-2 studies with radiological lung follow-up estimated that 50% of participants had inflammatory patterns and 29% had fibrotic patterns at a median of 3 months post infection. Analysis of the UK nationwide Post-hospitalisation COVID-19 Study at 5-months follow-up suggested that up to 11% of people discharged from hospital following COVID-19 infection were at-risk of radiological residual lung abnormalities, such as ground glass opacity and reticulation. In people with pulmonary fibrosis, these radiological patterns are often consistent with persistent epithelial lung injury. Biomarker studies have identified associations with COVID-19 severity, however there are few studies that explore the relationship between biomarkers of epithelial injury and parenchymal lung abnormalities post-hospitalisation.

Added value of this study

Using the Post-hospitalisation COVID-19 Study at five months follow-up, we present analysis of four biomarkers of epithelial lung injury that have been validated in pulmonary fibrosis. We compare levels according to the risk of residual lung abnormality in 846 people defined according to dysfunctional gas exchange and abnormal chest X-ray. Greater levels of matrix metalloproteinase-7 and Krebs von den Lungen-6 were observed in participants at-risk of

residual lung abnormality, which was consistent in an internal replication cohort of 111 people with evidence from thoracic CT. We report epithelial injury biomarker associations according to the extent of radiological abnormalities and demonstrate that surfactant proteins were associated with reticulation and not ground glass opacity. Associations of epithelial injury biomarkers with radiological abnormalities were independent of age, sex and admission severity. We observed greater gene and protein expression levels of epithelial injury biomarkers in epithelial cells from COVID-19 lung tissue.

Implications of all the available evidence

These findings provide evidence that raised, circulating levels of epithelial lung injury biomarkers were observed five months following hospitalisation due to COVID-19, were associated with the extent of radiological abnormality, and may be useful for monitoring people at-risk of parenchymal lung changes. The results suggest that pro-fibrotic signalling cascades contribute to persistent epithelial lung injury observed in some people recovering from COVID-19, which may contribute to ongoing symptoms and restrictive lung function. Further research and surveillance are required to address the longitudinal association of epithelial lung injury biomarkers and radiological patterns following COVID-19 in order to understand whether fibrotic features are stable, resolving or progressive in the long-term.

Introduction

Following acute COVID-19 infection, long-term symptoms have been frequently reported by survivors.¹ Meta-analysis of post-COVID-19 lung sequelae has highlighted dyspnoea at a prevalence estimate of 37%,² impaired gas transfer at 38%, restrictive impairment at 17%, thoracic computed tomography (CT) evidence of fibrotic involvement at 29% and inflammatory involvement at 50%.³ These sequelae are consistent with lung injury that for some is chronic and can be associated with a poor recovery following acute COVID-19 infection.

The Post-Hospitalisation COVID-19 (PHOSP-COVID) study was designed to identify associations with good or poor recovery,⁴ with less than 30% of participants reporting feeling fully recovered at one year.⁵ Risk factors for poor recovery of breathlessness included older age, obesity, major comorbidities, and length of acute admission.⁶ The UK interstitial lung disease (ILD) Post-COVID study previously leveraged the PHOSP-COVID dataset to assess the prevalence and risk factors of residual lung abnormalities (RLA) on thoracic CT in people without evidence of prior clinical management for ILD.^{7,8} Important clinical risk factors for RLA included the need for ventilation during hospitalisation, abnormal follow-up chest X-ray, and impaired gas transfer at five months post-hospitalisation, with up to 11% of survivors estimated to be at-risk of RLA.

Mechanisms of parenchymal lung changes include persistent epithelial lung injury that can result in aberrant wound healing and deposition of extracellular matrix components, with validated biomarkers that are elevated in fibrotic lung disease including matrix metalloproteinase-7 (MMP-7), Krebs von den Lungen-6 (KL-6), surfactant protein-A (SP-A), and surfactant protein-D (SP-D).^{9–15} Whilst the UK-ILD Post-COVID evidence of RLA is consistent with other studies,^{16,17} it is unknown whether RLA represents sequelae that will resolve over time or lead to persistent epithelial lung injury that may be consistent with interstitial lung changes.¹⁸

The aim of this UK ILD Post-COVID study was to assess whether biomarker levels of epithelial lung injury were greater in people at-risk of RLA and whether this effect was consistent in participants with evidence of RLA involvement on CT.

Methods

Study population

The study cohort included participants of the PHOSP-COVID study, a prospective longitudinal cohort study of adults discharged from National Health Service hospitals across the United Kingdom after admission for confirmed or clinically diagnosed COVID-19.^{19,20} Data are accessible to legitimate researchers using the [Data and Sample Access Request Form](#); access to all data and

samples collected by PHOSP-COVID are approved by the Core Management Group and Executive board. PHOSP-COVID recruitment was stratified into two tiers, Tier 1 included individuals consenting to information recorded during acute admission and Tier 2 included consent for two follow-up visits post-discharge. Observations were restricted to Tier 2 individuals and initial follow-up data at median 5 months post-discharge; participants were discharged by the end of March 2021 with follow-up available within 240 days of discharge. Individuals withdrawing consent from PHOSP-COVID-19 were excluded. Individuals being managed for an *a priori* diagnosed ILD or pulmonary fibrosis were excluded. Follow-up thoracic CT scans were acquired if the patient remained symptomatic at their clinic appointment and there was concern from the examining physician. Given restrictions in the ability to perform lung function tests during the pandemic (as aerosol generating procedures) and the insensitivity of CXRs in detecting subtle lung damage, there was an impetus to request CT scans based on purely clinical concern. CT scans were not included in the protocol, varied across UK hospital sites, and were identified through the PHOSP-COVID study via linkage to a radiological database. The PHOSP-COVID-19 study was used to collect demographic information from hospital records, including age and sex recorded at birth, as well as clinical characteristics including hospital admission severity based on the level of oxygen support needed, ranging from oxygen supplementation, to continuous positive airway pressure (CPAP) and invasive mechanical ventilation (IMV); chest X-ray records categorised as normal, abnormal (“suggestive of lung fibrosis”, “extensive, persistent changes greater than one-third of lung involvement” and “indeterminate”), or other; lung function measures where available.

Exposures and outcomes

The primary exposure for those without CT was classification within the RLA at-risk population, whilst for those with follow-up CT, the primary exposure was visually scored RLA equal to or greater than 10% lung involvement. RLA at-risk was defined by percent predicted DL_{CO} <80% and/or abnormal chest X-ray, as previously described.⁷ The first clinically indicated CT performed between 90 days and 240 days post discharge were scored. The presence of RLA on volumetric CTs was scored on a lobar basis by a radiologist blinded to clinical data. The percentage parenchymal involvement of ground-glass opacities (GGO) and reticulation were quantified visually. The sum of the two patterns was averaged across lobes to quantify RLA. Airway dilatation or traction bronchiectasis was not quantified alongside GGO and reticulation as they can be transitory in the acute setting. The outcome was the level of epithelial lung injury biomarkers MMP-7, KL-6, SP-D and SP-A. In secondary analyses, the exposure was specified as the biomarker level whilst RLA classification,

percentage lung involvement of GGO, or percentage lung involvement of reticulation, were included as outcomes.

Epithelial injury biomarker assay

All PHOSP-COVID Tier 2 participants were invited to provide a blood sample at the first follow-up visit, plasma processing and sample prioritisation for biomarker assays were performed as previously described.⁵ Briefly, blood samples were centrifuged and supernatant was aliquoted before being immediately frozen. Samples were provided to Sysmex for the measurement of KL-6, MMP-7, SP-A, and SP-D levels. The biomarker assays were run on a HISCL-800 analyser (Sysmex, Japan), which employs a two-step sandwich immunoassay system to detect antigens in plasma. Biotinylated monoclonal antibodies capture the analyte in the sample and bind to streptavidin-coated magnetic beads, before alkaline phosphatase-conjugated detection with a chemiluminescent readout. All samples were tested after being prediluted eight times and were within the assay's measurement range: 10–6000 U/mL for KL-6, 0.09 ng/mL to 100 ng/mL for MMP-7, 1–1000 ng/mL for SP-A, and 0.02–100 ng/mL for SP-D. Samples passing quality assurance and within limit of quantification were included in analysis.

Biomarker expression in lung epithelial cells

Gene expression levels of *MMP7* (MMP-7), *MUC1* (KL-6), *SFTPA1* (SP-A) and *SFTPD* (SP-D) were verified in lung epithelial cells from COVID-19 samples ($n = 9$) compared with non-COVID-19 controls ($n = 3$) using publicly available spatial transcriptomic data deposited in Gene Expression Omnibus under record [GSE190732](#).²¹ Post-mortem samples from COVID-19 related pneumonia donors were stratified into acute as 1–15 days of duration ($n = 3$), chronic as more than 15 days of duration ($n = 3$), and prolonged as 7–15 weeks of duration ($n = 3$). The Cell Ranger Software Suite (Version 3.1.0) was used to process raw sequencing data with the GRCh38 reference. Spatial RNA sequencing data analysis was performed in R (version 4.5.1) with Seurat (version 5.3.1), cluster annotation was performed using the Human Lung Cell Atlas reference. MMP-7 protein expression in epithelial cells in post-mortem lung tissue from fatal cases of COVID-19 was verified by immunohistochemistry. Briefly, 5 μ m thick sections of lung tissue from fatal COVID-19 cases ($n = 12$) and non-infected control samples ($n = 5$) collected pre-pandemic from normal adjacent regions of lung during resection surgery were dewaxed and rehydrated. Antigen retrieval was achieved by boiling the sections for 30 min in 10 mM Tris/0.5 M EDTA then stained with anti-MMP7 antibody (Abcam AB205525, 150 μ g/mL) overnight. Positive staining was visualised with 3',3'-diaminobenzidine and sections

scanned with a Hamamatsu™ NanoZoomer® S20 digital slide scanner.

Statistical analysis

To support multivariable modelling and comparable interpretation of non-normal distributions across different concentration units, biomarker levels were log transformed and z-standardised using all samples. The primary analysis tested the difference in mean biomarker levels according to RLA risk classification using unpaired t-tests, an internal replication analysis was performed according to evidence of RLA involvement. Mean values are reported with standard deviations (SD). Non-transformed, absolute biomarker concentrations were also compared using Wilcoxon rank sum tests. Logistic models were specified to calculate and compare the area under the receiver-operator characteristic (AUROC) of biomarkers and admission characteristics for RLA classification and RLA involvement, comparison of AUROC on the same outcome was performed using DeLong's test. Secondary analyses were specified using linear models to test the association of RLA $\geq 10\%$ on CT relative to $<10\%$, and RLA at-risk relative to low-risk, upon biomarker z-score in unadjusted and adjusted models. Covariates included in adjusted models were age and sex, severe admission (CPAP or IMV), and time difference between sampling and CT; 95% confidence intervals (95% CI) were derived using robust variance estimators. Normality of residuals were used to confirm linearity assumptions. Fractional regression was used to model the percentage involvement of reticulation or GGO according to biomarker z-score, specified as generalised linear models with binomial distribution and logit link. Coefficients are interpreted as the average difference in involvement per increment in z-score. A sub analysis in those with a second clinical CT scored was performed with Wilcoxon Rank Sum for difference in biomarker according to reticulation $>5\%$ involvement. Analyses were performed within the Scottish National Safe Haven Trusted Research Environment using Stata 17 MP. Wilcoxon rank-sum was used to test differences in gene expression with cells as the independent unit of analysis, using R.

Ethics

The PHOSP-COVID study received ethical approval from the Leeds West Research Ethics Committee (20/YH/0225) and is registered on the ISRCTN Registry (ISRCTN10980107), all included participants provided written informed consent. Lung tissue was obtained from postmortem examinations of fatal COVID-19 cases performed at two sites (South Wales and London). Postmortems conducted in South Wales, UK were performed between April and August 2020 (REC 19/NEC/0336). Postmortem tissue collected from London, UK were obtained from Imperial College London

Tissue Bank (Project number R20034) from fatal COVID-19 cases. COVID-19 was confirmed by positive SARS-CoV-2 swab either in-life during the patient's final illness or during postmortem in cases of suspected SARS-CoV-2 related death. All non-COVID-19 control tissue was collected from postmortem donors who had died of non-respiratory related causes. All postmortem material was obtained from either consent post-mortems or coronial postmortems in which consent for retention of tissue and use in research was obtained from the family or a person in a qualifying relationship.

Role of funders

PHOSP-COVID is jointly funded by a grant from the Medical Research Council (MRC)-UK Research and Innovation (UKRI) and the Department of Health and Social Care through the National Institute for Health Research (NIHR) rapid response panel to tackle COVID-19 (grant references: MR/V027859/1 and COV0319). The UKILD Consortium was funded by UKRI Ideas to Address COVID-19 (grant reference MR/W006111/1). The views expressed in the publication are those of the authors and not necessarily those of the National Health Service (NHS), the NIHR or the Department of Health and Social Care. The funders had no role in study design, data collection, data analyses, interpretation, or writing of this report.

Results

A total of 957 participants in the UKILD Post-COVID cohort had samples assayed for MMP-7 ($n = 955$), KL-6 ($n = 787$), SP-D ($n = 957$) and SP-A ($n = 875$), demographics were similar in those who did not have samples assayed (Supplementary Table S1). Of those included in biomarker sampling, 111 had a CT scored (11.6%) and 846 did not have a thoracic CT available (88.4%) (Fig. 1). Between those with CT scored and no thoracic CT available, representation of sex was comparable ($n = 70/111$; 63.1% and $n = 543/846$; 64.2%, respectively), age was similarly distributed (58.4 SD 11.2 and 57.0 SD 12.6, respectively) and participants were majority white ($n = 73/111$; 65.8% and $n = 645/846$; 76.2%, respectively) (Supplementary Table S2). Acute COVID-19 admission severity was similar between groups ($p = 0.059$), although there was greater representation of IMV in people with a thoracic CT ($n = 35/111$; 31.5%) than those without ($n = 185/846$; 21.9%).

Biomarker levels and RLA

A total of 103 people (12.2%) were classified as at-risk of RLA, whilst 742 participants were classified as low-risk (87.8%) (Supplementary Table S3a). Compared with people who were at low-risk, participants classified as at-risk of RLA had greater z-standardised levels of MMP-7 (0.39 SD 1.08 vs -0.07 SD 0.97, $p = 0.0001$),

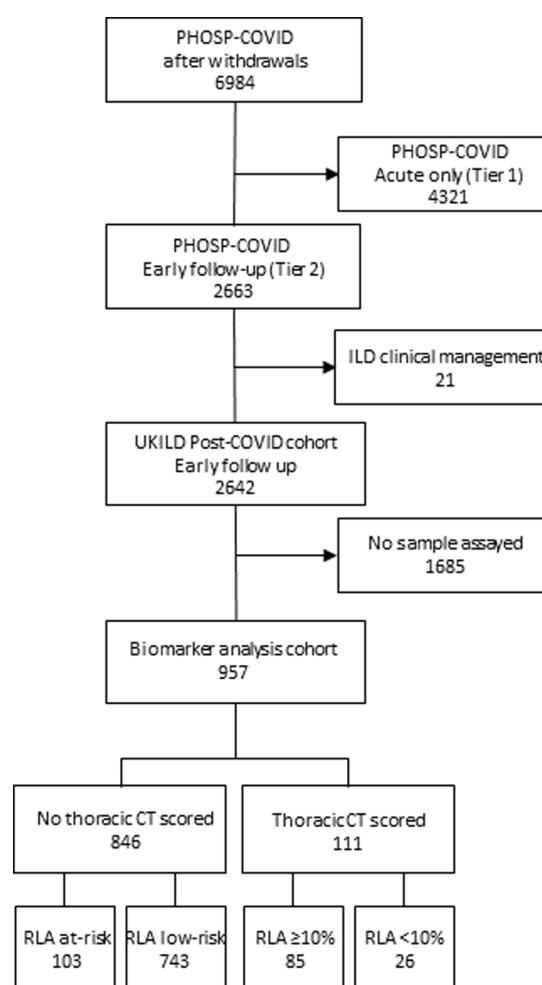


Fig. 1: Flow diagram of participant involvement in UK-ILD Post-COVID epithelial injury study. Inclusion of participants in the UK-ILD Post-COVID epithelial injury study from the PHOSP-COVID dataset based on early follow-up, no evidence of ILD clinical management, and epithelial injury biomarker assayed.

KL-6 (0.45 SD 1.19 vs -0.04 SD 0.97, $p = 0.0003$), SP-D (0.40 SD 1.08 vs -0.05 SD 0.99, $p = 0.0001$), and SP-A (0.41 SD 1.19 vs -0.07 SD 0.97, $p = 0.0004$) (Fig. 2a). Results were comparable for biomarker concentrations (Supplementary Table S4a).

In replication analysis performed in participants with a CT score, evidence of RLA $\geq 10\%$ was reported in 85 (76.6%), whilst RLA $< 10\%$ was observed in 26 (23.4%) (Supplementary Table S3b). Overall, the median time between thoracic CT and sample collection was 16 days (IQR -49 to 76), CT was a median 10 days prior to sampling in those with RLA $\geq 10\%$ and a median 21 days post sampling in those with RLA $< 10\%$ ($p = 0.090$). The mean z-standardised MMP-7 level was greater in those with RLA $\geq 10\%$ (0.32 SD 1.02 vs -0.54 SD 0.70, $p < 0.0001$), as well as KL-6 level (0.03 SD 0.86 vs -0.64 SD 0.92, $p = 0.0032$) (Fig. 2 b). There was no

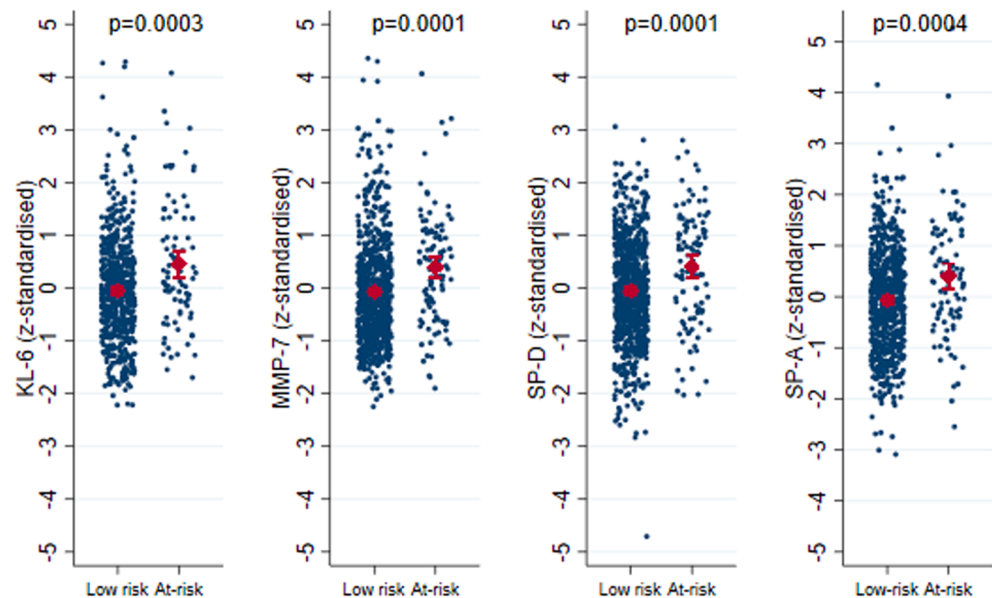
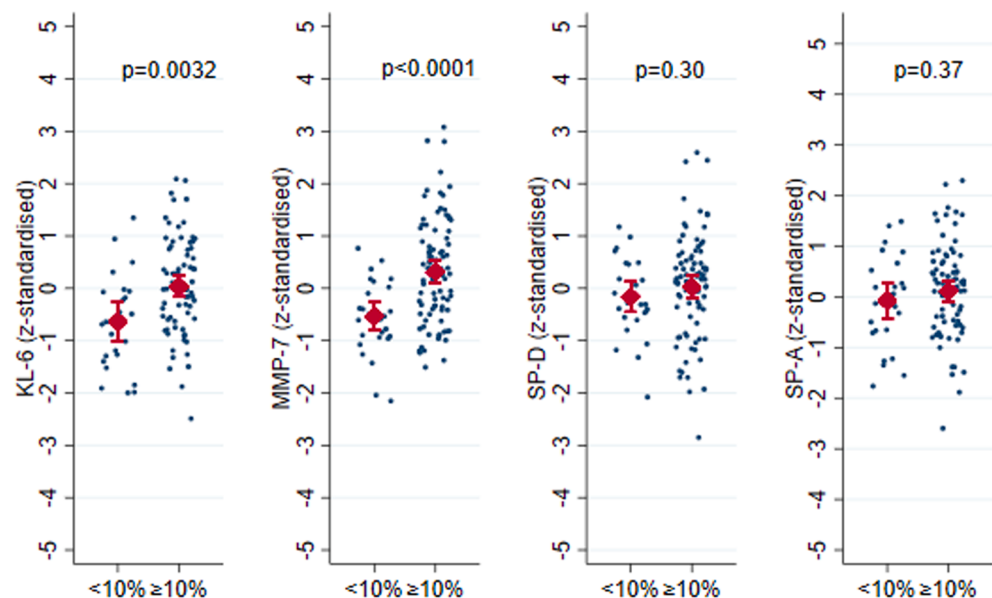
a RLA risk classification**b** RLA involvement on CT

Fig. 2: Comparison of epithelial injury biomarker levels according to residual lung abnormality risk classification and thoracic CT involvement (N = 957). a) Analysis of z-standardised KL-6, MMP-7, SP-D, SP-A levels between at-risk classification and low-risk classification in those without a thoracic CT (n = 846). b) Replication analysis of z-standardised KL-6, MMP-7, SP-D, SP-A levels between ≥10% residual lung abnormality involvement on thoracic CT and <10% (n = 111). p-values estimated using unpaired t-test, plots present mean values and standard deviation.

statistical difference in SP-D or SP-A levels between those with RLA ≥10% and those with RLA <10%. Results were comparable for biomarker concentrations (Supplementary Table S4b).

Admission characteristics of age, sex and admission severity provided an AUROC for RLA at-risk

classification of 0.63 (95% CI 0.57–0.69), addition of MMP-7 and KL-6 levels led to an AUROC of 0.70 (95% CI 0.64–0.76), with a significant improvement in performance (p = 0.008). Similarly, addition of MMP-7 and KL-6 improved the performance of admission characteristics in classifying evidence of RLA ≥10%

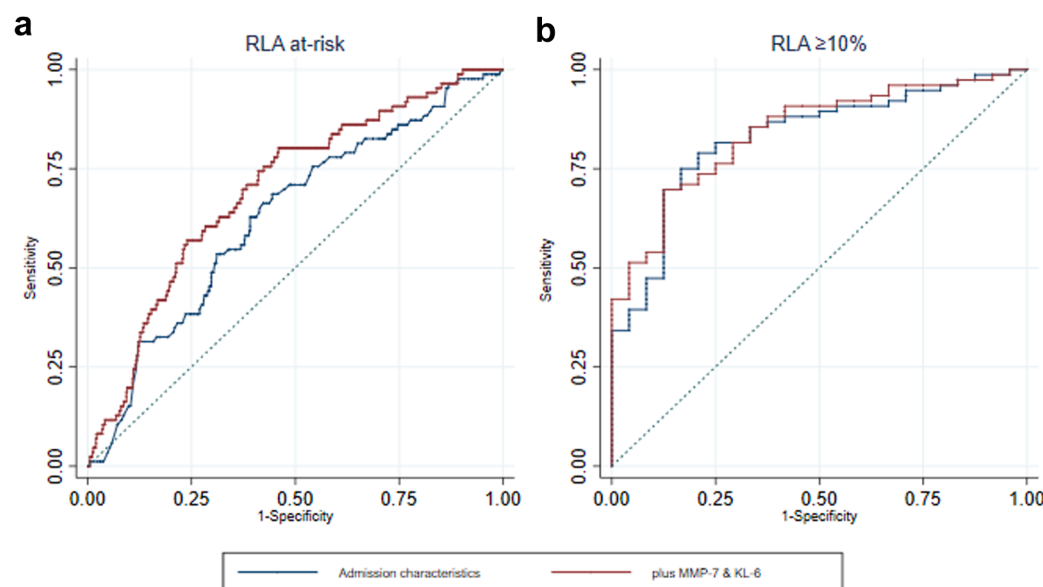


Fig. 3: Residual lung abnormality discrimination with epithelial injury biomarkers (N = 957). Area Under Receiver Operator Characteristic (AUROC) presented for ability to discriminate a) RLA at-risk participants from low-risk (n = 846) (blue AUROC 0.63 95% CI 0.57–0.69; red AUROC 0.70 95% CI 0.64–0.76; difference $p = 0.008$), b) RLA $\geq 10\%$ involvement on thoracic CT from RLA $< 10\%$ (n = 111) (blue AUROC 0.80 95% CI 0.69–0.91; red AUROC 0.84 95% CI 0.76–0.93; difference $p = 0.15$). Blue line represents logistic model specified with admission characteristics of age, sex and ventilation status. Red line represents logistic model specified with admission characteristics, plus MMP-7 levels and KL-6 levels.

involvement from 0.80 (95% CI 0.69–0.91) to 0.84 (95% CI 0.76–0.93), although this did not reach significance ($p = 0.15$) (Fig. 3).

Association between RLA and biomarker level

Participants with $\geq 10\%$ RLA involvement on CT had a 0.85 (95% CI 0.51–1.19, $p < 0.001$) higher MMP-7 z-score and a 0.67 (95% CI 0.26–1.08, $p = 0.001$) higher KL-6 z-score, compared with participants with RLA $< 10\%$ involvement on CT (Fig. 4). Associations were also observed in participants classified as at-risk of RLA relative to low-risk, the MMP-7 z-score was 0.47 (95% CI 0.25–0.69, $p < 0.001$) higher and the KL-6 z-score was 0.50 (95% CI 0.24–0.76, $p < 0.001$) higher. No statistically significant association was observed between RLA $\geq 10\%$ on CT and SP-D ($p = 0.30$) or SP-A ($p = 0.37$); whilst RLA at-risk was associated with a 0.46 (95% CI 0.24–0.67) and 0.47 (95% CI 0.22–0.73) higher z-score, respectively, when compared with low-risk participants. Effects were similar when adjusted for age, sex and admission severity.

Association between biomarker level and RLA involvement

In participants with CT scores, a unit increase in MMP-7 z-score was associated with a 3.58% (95% CI 2.04–5.13; $p < 0.001$) greater involvement of reticulation on average and a 6.25% (95% CI 3.46–9.03; $p < 0.001$) greater involvement of GGO on average (Fig. 5a, b).

A greater KL-6 z-score was similarly associated with percent reticulation (4.80%, 95% CI 2.65–6.96; $p < 0.001$) and GGO (5.81%, 95% CI 2.66–8.96; $p < 0.001$). Unit increments in SP-D z-score were associated with a 3.22% (95% CI 1.19–5.24; $p = 0.002$) greater involvement of reticulation, whilst no statistically significant association was observed with GGO (1.53%, 95% CI –1.39 to 4.46; $p = 0.30$) (Fig. 5c, d). Similarly, SP-A z-score was associated with more reticulation (3.03%, 95% CI 0.76–5.30; $p = 0.009$) but not GGO involvement (1.30%, 95% CI –2.01 to 4.61; $p = 0.44$). Effects were independent of age, sex, admission severity and time between CT and biomarker sampling (Supplementary Table S5).

In a sub analysis of 15 participants who underwent a second CT, median 163 days (IQR) from the initial CT, 301 days (IQR) from discharge and 149 days (IQR) from sampling date, SP-D levels were significantly higher in those with persistent reticulation involvement $> 5\%$ (Supplementary Table S6).

Biomarker expression in epithelial cells from COVID-19 lung tissue

Spatial RNA sequencing analysis demonstrated substantially greater gene expression of *MMP7*, *MUC1*, *SFTPA1* and *SFTPD* in epithelial cells from COVID-19 lung tissue than non-COVID-19 controls (Fig. 6a and d). In stratified analysis, greater gene expression levels were robustly observed in chronic and prolonged

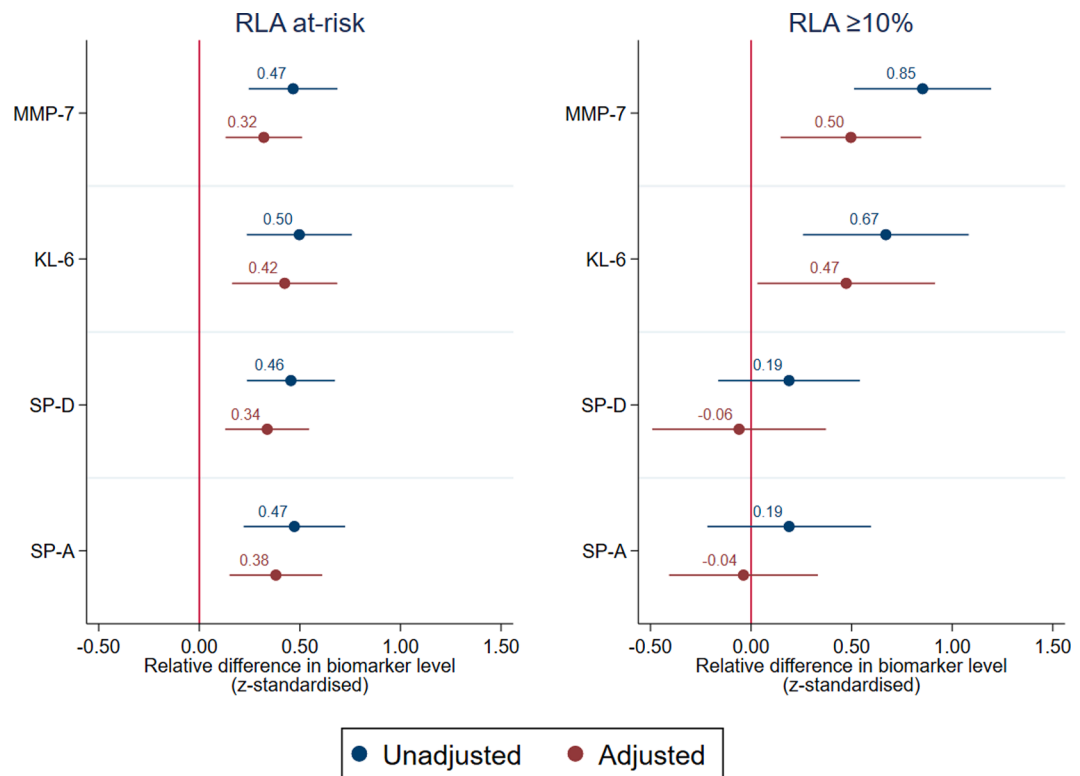


Fig. 4: Difference in biomarker levels according to residual lung abnormality (N = 957). Unadjusted and adjusted associations modelled with generalised linear model for the difference in z-standardised MMP-7, KL-6, SP-D, SP-A levels for the RLA at-risk classification relative to low-risk classification (n = 846), and RLA $\geq 10\%$ on thoracic CT relative to RLA $<10\%$ (n = 111). Covariates in adjusted models include age, sex and severe admission requiring ventilation, presented with 95% confidence interval.

COVID-19 pneumonia compared with control tissue (Supplementary Figure S1). MMP-7 protein expression was not observed in epithelial cells from non-COVID-19 archive lung tissue (Fig. 6e and g), however levels were clearly observed in alveolar epithelial cells within lung tissue from post-mortem COVID-19 cases (Fig. 6f & h).

Discussion

People at-risk of RLA according to abnormal chest X-ray and/or impaired gas transfer at follow-up, and those with $\geq 10\%$ RLA on follow-up CT up-to eight months post-hospitalisation, had greater levels of circulating epithelial lung injury biomarkers. MMP-7 and KL-6, previously validated as pulmonary fibrosis biomarkers,^{12,13} were greater in those classified as at-risk of RLA compared with low-risk, a finding which was replicated in those with RLA $\geq 10\%$ on CT compared with $<10\%$. Similarly, MMP-7 and KL-6 levels improved the performance of major admission features in classifying RLA groups. SP-D and SP-A, which are produced predominantly in lung tissues,²² were positively associated with involvement of reticulation on CT, but not GGO, suggesting residual lung damage post-

COVID-19 that may be representative of fibrotic patterning. In a limited subsample, SP-D levels were greater in participants who had $>5\%$ reticulation involvement at a repeat CT. Circulating epithelial lung injury biomarkers may be useful in assessing risk of post-COVID-19 RLA, supporting clinical management in survivors of severe infection and those presenting with long term respiratory symptoms.

MMP-7 is a profibrotic metalloproteinase secreted by epithelial cells in numerous organs, whilst in the lungs it is specifically localised to activated alveolar and bronchiolar epithelial cells.²³ Localisation of MMP-7 to epithelial cells in COVID-19 lung tissue was verified by immunohistochemistry and spatial transcriptomics. MMP-7 degrades extracellular matrix components, activates cytokines and chemokines, and induces epithelial-mesenchymal transition.²⁴ Activity increases in response to tissue injury to enable repair and remodelling, whilst dysregulated activity is linked to increased profibrotic TGF- β signalling and irreversible tissue damage. High circulating MMP-7 levels have been associated with a greater risk of interstitial lung abnormality on CT at 10-years post-sampling and have been proposed as a biomarker of subclinical ILD.²⁵ An

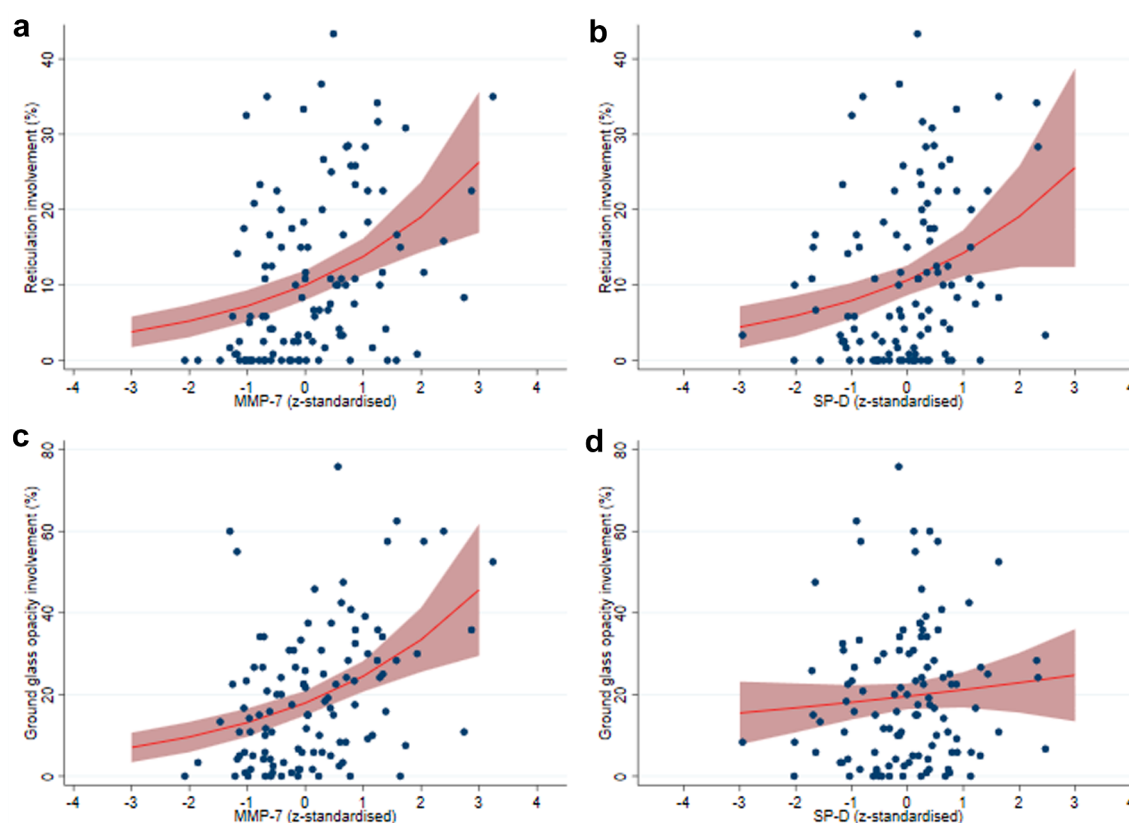


Fig. 5: Association between biomarker level and radiological pattern (N = 111). Unadjusted association of z-standardised biomarker level and percentage lung involvement, (a) MMP-7 and reticulation, (b) MMP-7 and ground glass opacity, (c) SP-D and reticulation, (d) SP-D and ground glass opacity. Estimates and marginal effects modelled with fractional regression, presented with 95% confidence interval.

individual participant data meta-analysis of 10 independent IPF cohorts demonstrated baseline MMP-7 levels were associated with a 23% greater risk of overall mortality and a 27% greater risk of 12-month disease progression, per standardised z-score (increment in standard deviation).²⁶ A previous study highlighted median concentrations of plasma MMP-7 of 12.1 ng/mL in IPF and 5.1 ng/mL in healthy controls,¹⁰ with high sensitivity and moderate specificity of 12.1 ng/mL as a predictive cut-off for all-cause mortality and transplant-free survival. Median concentrations of MMP-7 in those with RLA involvement $\geq 10\%$ and classified as at-risk were greater than this threshold, suggesting clinically relevant concentrations.

MMP-7 levels have been explored following COVID-19 infection in a number of small cohorts, with higher levels observed in COVID-19 patients compared with controls.²⁷ Within COVID-19 patients, higher levels have been observed in individuals who required invasive mechanical ventilation compared with those who did not,²⁸ MMP-7 was similarly higher at 2-months post-hospitalisation in severe compared to mild patients, although this discrepancy was not observed at

12-months.²⁹ Greater expression of MMP-7 has been observed within epithelial cells and macrophages from bronchoalveolar lavage fluid in severe COVID-19 infection.³⁰ By contrast, a plasma proteome study reported lower MMP-7 levels in Long COVID when compared with healthy controls, albeit with greater TGF- β levels and mediators of the integrin activation pathway implicated in profibrotic mechanisms.³¹ Upstream activators of the TGF- β pathway have also been implicated in diffuse alveolar damage following COVID-19.³² Although discrepancies in MMP-7 levels by severity have previously been reported, we found that associations of MMP-7 with radiological features in this study were independent of admission severity. Together, these findings implicate MMP-7 as a potentially important prognostic factor for poor recovery and long-term structural changes.

Higher levels of KL-6, SP-A and SP-D have been similarly associated with epithelial cell dysfunction and have all been demonstrated to be upregulated in IPF,^{13,15} conflicting findings have been reported regarding their association with IPF clinical outcomes of mortality, disease progression and change in forced vital

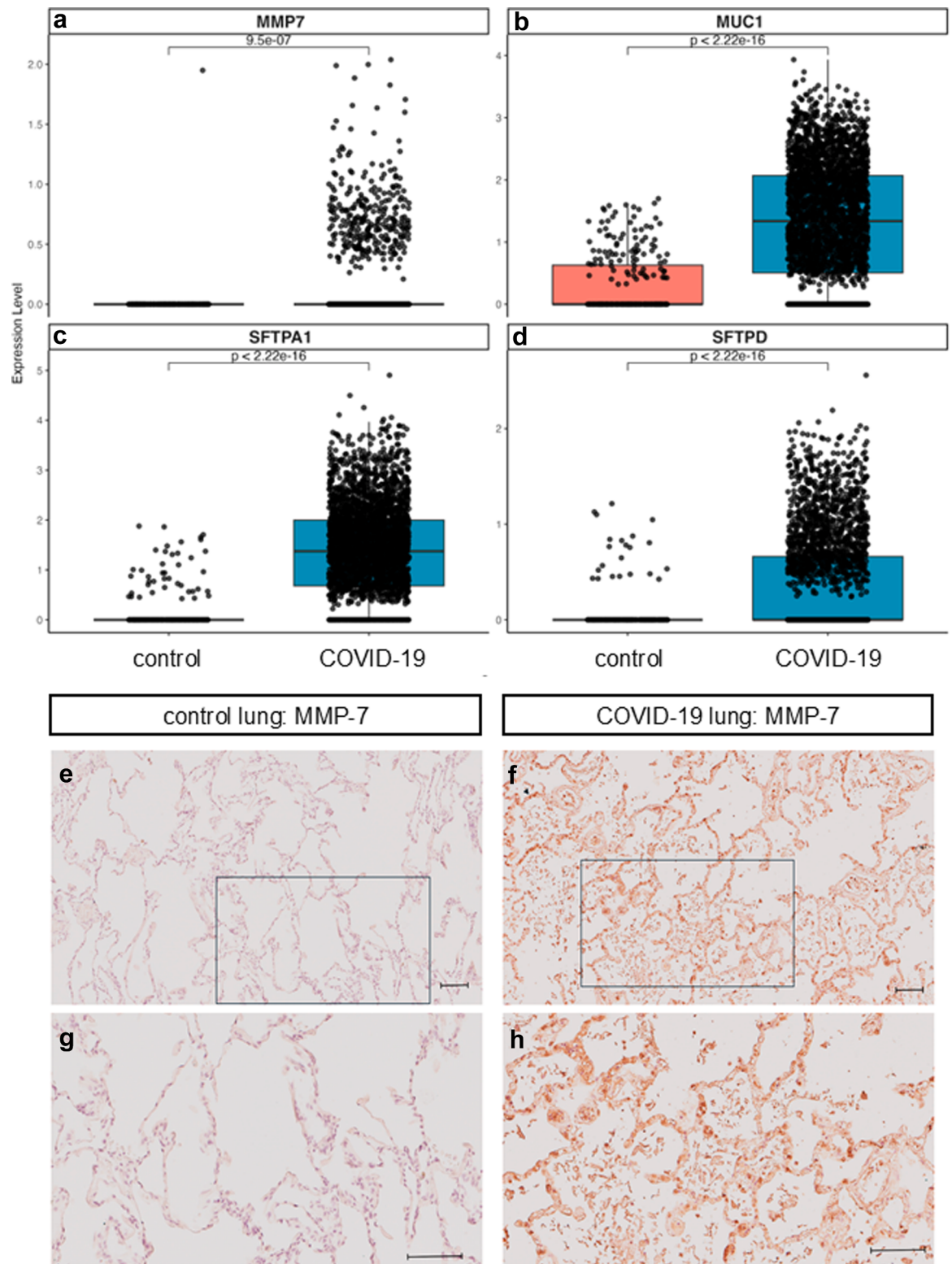


Fig. 6: Epithelial expression of epithelial injury biomarkers in COVID-19 lung tissue. Publicly available spatial RNA sequencing restricted to cells annotated as lung epithelium represents gene expression of a) MMP7, b) MUC1, c) SFTPA1, d) SFTPD in non-COVID-19 control (n = 3) and COVID-19 (n = 9) lung tissue (GSE190732). Box plots present median and interquartile range across conditions, difference tested by Wilcoxon rank-sum. Epithelial MMP-7 protein expression observed in representative images of control (e, g) and COVID-19 injured (f, h) lung tissue. Scale bars at 100 μ m, black box in panels e) and f) denote high power regions shown in g) and h), respectively.

capacity.²⁶ Spatial resolution of gene expression demonstrated that *MUC1* (KL-6), *SFPTA1* (SP-A) and *SFPTD* (SP-D) were expressed at greater levels in lung epithelial cells from COVID-19 donors compared with non-COVID-19 controls, which was consistent in chronic and prolonged disease strata.

KL-6 is a mucin-like glycoprotein expressed in regenerating type-II alveolar epithelial cells, released in response to epithelial injury to promote the migration and proliferation of fibroblasts.³³ Studies of the acute phase of COVID-19 infection demonstrated serum KL-6 as a biomarker of COVID-19 severity and poor clinical outcomes.^{34,35} In the present study, KL-6 levels were elevated after the acute infection in individuals with greater levels of RLA involvement on CT and in those who were considered at-risk of RLA. KL-6 and MMP-7 performed independently of admission features in classifying RLA risk, supporting these biomarkers as tools for prioritising radiological evaluation in people with ongoing respiratory symptoms after the acute COVID-19 infection.

Surfactant proteins are synthesised and secreted by type-II alveolar epithelial cells to facilitate transport and function of surfactant lipids, playing an important role in immunity and clearance of harmful exposures. Abnormal surfactant protein regulation can trigger apoptosis of regenerative alveolar cells and initiate fibrosis.³⁶ Surfactant proteins are produced predominantly within the lung, with leakage into circulation proposed to be indicative of a breakdown in the integrity of the alveolar air-blood barrier. Recent studies have demonstrated an upregulation of SP-D in the acute phase of COVID-19 infection.^{37,38} The observed relationship between surfactant proteins and reticulation after the acute infection suggests epithelial lung injury that may be consistent with fibrotic damage. An evaluation of thoracic CT changes in 57 severe participants between four and 15 months post-COVID-19 highlighted a decrease in GGO over time that was inversely correlated with reticulation involvement.³⁹ Future studies should determine whether longitudinal SP-D levels, and other epithelial lung injury biomarkers, are able to distinguish evidence of active fibrotic mechanisms from residual fibrotic remnants of infection.

A particular strength of this study is the utilisation of a large population of COVID-19 survivors hospitalised with varied severity of acute infection, applying mutually exclusive RLA definitions to address internal replication of biomarker levels. Findings were consistent after adjustment for demographics, severity of ventilation during admission, and time from sampling to CT. Whilst the PHOSP-COVID cohort is a nationwide study, the study was limited by internal replication alone and no healthy control group, people in comparator groups of low-risk of RLA or RLA<10% had evidence of elevated MMP-7 levels, however elevated levels were consistent with the highest levels of radiological involvement. We

would propose the measurement of serum MMP-7 to guide the performance of CT and justify risks associated with radiation exposure. Future studies should seek external validation of epithelial injury in individuals with non-resolving abnormalities on CT, including non-hospitalised individuals presenting to Long COVID clinics. Sequencing and immunohistochemistry of post-mortem lung samples from people with COVID-19 related pneumonia, including prolonged disease up to 15 weeks, supported epithelial cells as a source of epithelial injury biomarkers in COVID-19. This work can support reverse translation to preclinical models and in vitro systems for development of potential interventions and therapies targeting persistent mechanisms of epithelial injury. The study design was defined by an *a priori* hypothesis, exploratory studies may identify other important pathways beyond epithelial dysregulation.

There are additional important limitations to our study. Assays reflect circulating biomarker levels, it is possible that lung epithelial cells may not be the sole source of MMP7 production. With the exception of MMP-7, biomarker concentrations in the PHOSP-COVID cohort were largely comparable to healthy controls reported in previous studies,^{10,15,40} highlighting that the extent of epithelial injury is not consistent with clinical presentations of ILD. Whilst associations of biomarker levels with RLA risk and involvement were observed, there was considerable variability in effect with a correspondingly modest AUC. Limited evidence of persistent epithelial injury should not be considered evidence of progressive injury, future follow-up and surveillance is necessary to address the natural history of RLA. Mechanical ventilation has been reported to lead to increased levels of epithelial injury biomarkers and fibrotic injury,^{41,42} which may explain some associations with RLA involvement, however effects were independent of admission severity. The radiological focus of this study was on parenchymal damage, with ground glass opacities and reticulation as the major patterns scored. Traction bronchiectasis is often a transitory feature post the acute setting and was not included in our visual CT score as airway dilatation with infection/inflammation can be reversible. A further limitation was the small number of people with longitudinal data and the small number of samples with spatial transcriptomics available. Future long-term follow-up studies should comprehensively characterise radiological features in post-COVID settings with designs that facilitate longitudinal biobanking. There were discrepancies between the date of sampling and date of clinically indicated CT, which may limit interpretation of injury resolution, although effects were also independent of time differences.

Conclusion

RLA on thoracic CT, observed beyond three months following COVID-19 related hospital discharge, were

associated with circulating epithelial lung injury biomarker levels. Higher levels in individuals at-risk of RLA suggests a substantial number of individuals hospitalised with COVID-19 may be living with subclinical lung injury and profibrotic signalling cascades. Future studies should address the association between temporal changes in fibrotic biomarker levels and the resolution or persistence of radiological lung involvement.

Contributors

Conceptualization: I.S., J.J., J.C.P., J.M.W., P.L.M., P.M.G., R.J.A., J.F.B., N.C., E.D., L.F., B.G.-G., B.G., I.P.H., I.J., S.J., F.K., P.M., J.M., J.E.P., K.P.-H., P.R.-O., L.C.S., D.J.F.S., M. Sp., L.G.S., S.C.S., A.A.R.T., S.W., A.M., M.P., A.St., S.Y., C.E.B., J.D.C., V.C.H., A.H., M.M., K.P., B.R., R.A.E., R.C.C., L.P.H., R.C.C., L.V.W., R.G.J. Data curation: S.A., E.M.H., M. Se., O.C.L., H.J.C.M., I.S., J.J., N.L., J.K.Q., S.W., M.R.P., A.St., N.G., F.G., A.B.D. Formal analysis: I.S., J.J., J.K.Q., M.P., A.St., B.L., A.L.T., N.G., A.J. Funding acquisition: C.E.B., L.V.W., R.A.E., J.D.C., L.P.H., A.H., M.M., K.P., B.R., O.L., M.R., O.E., H.J.C.M., A.S., M.S., R.S., V.C.H., L.H.W., N.J.G., A.M., S.Y., R.C.C., R.G.J. Project administration: A. Sh., A. Si., A.M., V.Q., A.L.T., A.J., M.R.P. Writing of original draft: I.S., J.J., J.C.P., R.C.C., L.V.W., and R.G.J. Writing and review & editing: I.S., J.J., J.C.P., B.L., A.J., A.L.T., N.G., M.R.P., R.J.A., J.F.B., N.C., E.D., L.F., P.M.G., B.G.-G., B.G., I.P.H., L.P.H., I.J., S.J., M.G.J., F.K., P.M., J.M., P.L.M., J.E.P., K.P.-H., M.P., V.Q., P.R.-O., L.C.S., D.J.F.S., M. Sp., L.G.S., S.C.S., A.A.R.T., S.W., J.M.W., D.G.W., A.B.D., F.G., W.G., E.M.H., N.L., J.Q., A.M., M.P., A.St., S.Y., C.E.B., J.D.C., O.E., R.A.E., N.G., V.C.H., A.H., L.H.W., O.C.L., M.M., H.J.C.M., K.P., M.R., B.R., R.M.S., M. Se., A. Sh., A. Si., R.C.C., L.V.W., R.G.J.

All authors have read an approved the final version of the manuscript. I.S., J.J., and J.K.Q., have accessed and verified the underlying data.

The PHOSP-COVID Collaborative Group provided the underlying data for analysis.

Data sharing statement

The protocol, consent form, definition and derivation of clinical characteristics and outcomes, training materials, regulatory documents, information about requests for data access, and other relevant study materials are available from www.phosp.org. Spatial transcriptomic data used in this study is publicly available from the Gene Expression Omnibus under record GSE190732.

Declaration of interests

ASh; AS; AM; BR; DW; DJFS; ED; EH; FG; HM; IJ; JW; JP; KPH; KP; LF; LS; LPH; LHW; LSp; MP; MS; MSp; MRP; MR; MM; MPO; NG; NGR; OE; RA; RS; SW; VQ; VH; WG declare no competing interests.

AART: grants from Heart Research UK, Janssen-Cilag Ltd, British Heart Foundation. Honoraria and support for attending meetings from Janssen-Cilag Ltd.

AH: support from MRC/UKRI/NIHR (MR/V027859/1), NIHR Manchester Biomedical Research Centre. Chair of NIHR Translational Research Collaboration.

AEJ: stock options and founder for Alevin Therapeutics.

ALT: research grants from Accession Therapeutics. Consulting fees from Accession Therapeutics. Honoraria from Universitas Hasanuddin, Australian Rare Lung Disease. Leadership as Treasurer of British Association for Lung Research. Scientific Advisory Board for Accessing Therapeutics.

ASi: support from MRC/UKRI/NIHR (MR/V027859/1; COV0319).

AD: support from Wellcome CDA 227856/Z/23/Z.

BGG: support from Wellcome fellowship (221680/Z/20/Z).

BG: grants from Alpha-1 Foundation, NIHR Leicester Biomedical Research Centre Respiratory Theme. Consulting fees from Third party anonymised consultancy discussions, GlaxoSmithKline. Honoraria from Limbic. Chair of Midlands ILD Research Alliance. Receipt of reagent from Galecto Biotech.

BL: grants from Action for Pulmonary Fibrosis Mike Bray Fellowship.

CB: support from MRC/UKRI/NIHR (MR/V027859/1; COV0319). Grants from Aretea, AstraZeneca, Chiesi, Genentech, GlaxoSmithKline, Regeneron Pharmaceuticals, Roche, Sanofi. Consulting fees from Aretea, AstraZeneca, Chiesi, Genentech, GlaxoSmithKline, Regeneron Pharmaceuticals, Roche, Sanofi.

FK: consulting fees from Guidepoint.

GJ: grants from Astra Zeneca, Galecto, GlaxoSmithKline, Nordic Biosciences, RedX, Pliant; honoraria from Chiesi, Roche, AstraZeneca, Gilead, GlaxoSmithKline, Boehringer Ingelheim. Consulting fees from AbbVie, AdAlta, Apollo Therapeutics, Arda Therapeutics, Astra Zeneca, Brainomix, Bristol Myers Squibb, Chiesi, Cohbar, Contineum Therapeutics, Galecto, GlaxoSmithKline, Media Therapeutics, RedX, Syndax, Pliant. Payment for expert testimony from Pinsent Masons LLP. Advisory board membership of Boehringer Ingelheim, Galapagos, Vicore. Leadership roles for NuMedii, Action for Pulmonary Fibrosis; Chair of Editorial Board at BMJ Open Respiratory Research.

IS: grants from the Rayne Foundation. Participation on the scientific advisory board for patientMpower Ltd. Honoraria from patientMpower. Support for attending meetings from ERS/ELF Travel grant.

IH: research grant from UKRI, senior investigator award from NIHR, grants from Wellcome Trust. Consulting fees from GlaxoSmithKline. Vice-chair of trustees at Asthma + Lung UK.

JC: grants from AstraZeneca, Genentech, Boehringer Ingelheim, Gilead, Chiesi, Insmed, Grifols, Trudell. Consulting fees from AstraZeneca, Chiesi, GlaxoSmithKline, Insmed, Grifols, Novartis, Boehringer Ingelheim, Pfizer, Janssen, Antabio and Zambon.

JM: Grants from British Heart Foundation Clinical Training Fellowship. Support for attending meetings from institute.

JQ: Grants from UKRI, NIHR, Health Data Research UK, Boehringer Ingelheim, AstraZeneca, Insmed, Sanofi. Consulting fees from GlaxoSmithKline, Sanofi, Chiesi, AstraZeneca.

JPo: Support from UKRI grant to UCL.

JB: support from NIHR BRC (NIHR 203308). Grants from North West Lung Centre.

JJ: grants from Gilead Sciences, GlaxoSmithKline, Microsoft Research, Wellcome Trust, Rosetrees Trus, Chan Zuckerberg initiative, Cystic Fibrosis Trust. Consulting fees from Boehringer Ingelheim, Roche, GlaxoSmithKline, NHSX. Honoraria from Boehringer Ingelheim, Roche, GlaxoSmithKline, Takeda. Patents (GB2113765.8; GB2211487.0). Advisory board for Boehringer Ingelheim, Roche, GlaxoSmithKline.

LVW: support from UKRI (MR/V027859/1), GSK/Asthma + Lung UK (C17-1), NIHR (COV0319), Sysmex OGT. Grants from GlaxoSmithKline, Genentech and Orion Pharma, Wellcome Trust, MRC. Consultancy fees from Galapagos, Boehringer Ingelheim and GlaxoSmithKline. Honorarium from ERJ, MRC Board.

MJ: grants from Wellcome Trust, Asthma + Lung UK, Medical Research Council, Boehringer Ingelheim, AAIR Charity, The Royal Society, GlaxoSmithKline R&D. Consulting fees from Skyhawk Therapeutics.

NC: Consulting fees from Boehringer Ingelheim. Honoraria from AstraZeneca, Boehringer Ingelheim. Payment for expert testimony from National Institute of Health and Clinical Excellence. Advisory board member for DSMB, Insilico, Fortea. Trustee of Pulmonary Fibrosis Northern Ireland, board member of Rare Disease Partnership NI.

NL: director of Research, Intensive Care Society and chair for the Scottish Intensive Care Society Audit Group, Public Health Scotland.

OL: grants from Medical Research Council.

PG: grants from Boehringer Ingelheim. Honoraria from Boehringer Ingelheim, Roche, Teva, Cipla, Brainomix, AstraZeneca, Daiichi-Sankyo, Avalyn. Support for attending meetings from Boehringer Ingelheim, Roche. Advisory board member for GlaxoSmithKline. Stock options for Brainomix, senior medical director of Brainomix.

PLM: grants from AstraZeneca, GlaxoSmithKline, Asthma + Lung UK, Action for Pulmonary Fibrosis. Consulting fees from Hoffman La Roche, Boehringer Ingelheim, AstraZeneca, Trevi, Qureight, Endeavour, Redx. Honoraria from Boehringer Ingelheim, Hoffman La Roche. Advisory board member for United Therapeutics.

PRO: Consulting fees from Boehringer Ingelheim, Hoffman La Roche, Chiesi. Honoraria from Boehringer Ingelheim, Hoffman La Roche, Respiratory Effectiveness Group. Support for attending meetings from Boehringer Ingelheim, Respiratory Effectiveness Group. Other fees from Boehringer Ingelheim, Hoffman La Roche, CSL Behring, FibroGen, Vicore Pharma AB, Gilead Sciences, Chiesi, Galecto, Endeavour BioMedicines. Chair of IPF/ILD working group, Respiratory Effectiveness Group. Member of REMAP-ILD committees.

PM: Support from UCL Hospital Biomedical Research Centre, Medical Research Council and GlaxoSmithKline Clinical Training Fellowship.

RE: grants from UKRI/MRC/NIHR, Wolfson Foundation, Genentech/Roche. Consulting fees from AstraZenica/Evidera. Personal speaker fee from Moderna April 2023. Chair of ERS Group 01.02, chair of ATS Pulmonary Rehabilitation Assembly.

RC: support from MRC/UKRI/NIHR (MR/V027859/1; COV0319). Grants from Chiesi. Consulting fees from Arda Therapeutics.

SRJ: support from the NIHR Nottingham Biomedical Research Centre. Grants from Medical Research Council, LAM Action, Fundacio La Marato De TV3, LifeArc. Personal travel award from Ferrer. Charity role as a Trustee of LAM Action.

SY: support from Sysmex Corporation, Oxford Gene Technology IP Ltd. Travel costs from Oxford Gene Technology IP Ltd. Equipment from Sysmex Corporation. Other, former employee of Oxford Gene Technology IP Ltd, which is a Sysmex Group Company.

SS: support from MRC/UKRI/NIHR (MR/V027859/1; COV0319).

PHOSP-COVID Collaborative Group was jointly funded by MRC-UKRI Research and NIHR rapid response panel to tackle COVID-19 (grant references: MR/V027859/1 and COV0319), the UKILD Consortium was supported by UKRI (grant reference MR/W006111/1). The funders had no role in the study design, interpretation or decision to submit for publication.

Acknowledgements

The authors would like to acknowledge the support of the electronic Data Research and Innovation Service team (Public Health Scotland) for their involvement in obtaining approvals, provisioning, and linking data, and the use of the secure analytical platform within the National Safe Haven. This study would not be possible without all the participants who have given their time and support. The authors thank all the participants and their families. The authors thank the many research administrators, healthcare and social care professionals who contributed to setting up and delivering the study at all of the 65 NHS trust health boards and 25 research institutions across the United Kingdom, as well as all the supporting staff at the NIHR Clinical Research Network, Health Research Authority, Research Ethics Committee, Department of Health and Social Care, Public Health Scotland, and the United Kingdom Health Security Agency, and support from the ISARIC Coronavirus Clinical Characterisation Consortium (ISARIC4C). The authors thank the NIHR Office for Clinical Research Infrastructure (NOCRI) for support in coordinating the charities group. The PHOSP-COVID-19 industry framework was formed to provide advice and support in commercial discussions, and we thank the Association of the British Pharmaceutical Industry and NOCRI for coordinating this. The authors are very grateful to all the charities that have provided insight into the study: Action Pulmonary Fibrosis, Alzheimer's Research UK, Asthma + Lung UK, British Heart Foundation, Diabetes UK, Cystic Fibrosis Trust, Kidney Research UK, MQ Mental Health, Muscular Dystrophy UK, Stroke Association Blood Cancer UK, McPin Foundations, and Versus Arthritis. The authors thank the NIHR Leicester Biomedical Research Centre patient and public involvement group and the Long Covid Support Group.

Appendix A. Supplementary data

Supplementary data related to this article can be found at <https://doi.org/10.1016/j.ebiom.2026.106134>.

References

- Davis HE, Assaf GS, McCorkell L, et al. Characterizing long COVID in an international cohort: 7 months of symptoms and their impact. *eClinicalMedicine*. 2021;38:101019.
- Cares-Marambio K, Montenegro-Jiménez Y, Torres-Castro R, et al. Prevalence of potential respiratory symptoms in survivors of hospital admission after coronavirus disease 2019 (COVID-19): a systematic review and meta-analysis. *Chron Respir Dis*. 2021;18:14799731211002240.
- Fabbri L, Moss S, Khan FA, et al. Parenchymal lung abnormalities following hospitalisation for COVID-19 and viral pneumonitis: a systematic review and meta-analysis. *Thorax*. 2023;78:191–201.
- Brightling CE, Evans RA, Singapuri A, Smith N, Wain LV, PHOSP-COVID Collaborative Group. Long COVID research: an update from the PHOSP-COVID scientific summit. *Lancet Respir Med*. 2023;11(11):e93–e94.
- Group P-CC. Clinical characteristics with inflammation profiling of long COVID and association with 1-year recovery following hospitalisation in the UK: a prospective observational study. *Lancet Respir Med*. 2022;10(8):761–775.
- Zheng B, Vivaldi G, Daines L, et al. Determinants of recovery from post-COVID-19 dyspnoea: analysis of UK prospective cohorts of hospitalised COVID-19 patients and community-based controls. *Lancet Reg Health Eur*. 2023;29:100635.
- Stewart I, Jacob J, George PM, et al. Residual lung abnormalities following COVID-19 hospitalization: interim analysis of the UKILD Post-COVID study. *Am J Respir Crit Care Med*. 2023;207(6):693–703.
- Wild JM, Porter JC, Molyneux PL, et al. Understanding the burden of interstitial lung disease post-COVID-19: the UK interstitial lung disease-long COVID study (UKILD-Long COVID). *BMJ Open Respir Res*. 2021;8(1):e001049.
- Rosas IO, Richards TJ, Konishi K, et al. MMP1 and MMP7 as potential peripheral blood biomarkers in idiopathic pulmonary fibrosis. *PLoS Med*. 2008;5(4):e93.
- Tzouveleakis A, Herazo-Maya JD, Slade M, et al. Validation of the prognostic value of MMP-7 in idiopathic pulmonary fibrosis. *Respirology*. 2017;22(3):486–493.
- Morais A, Beltrão M, Sokhatska O, et al. Serum metalloproteinases 1 and 7 in the diagnosis of idiopathic pulmonary fibrosis and other interstitial pneumonias. *Respir Med*. 2015;109(8):1063–1068.
- Song JW, Do KH, Jang SJ, Colby TV, Han S, Kim DS. Blood biomarkers MMP-7 and SP-A: predictors of outcome in idiopathic pulmonary fibrosis. *Chest*. 2013;143(5):1422–1429.
- Yokoyama A, Kohno N, Hamada H, et al. Circulating KL-6 predicts the outcome of rapidly progressive idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med*. 1998;158(5 Pt 1):1680–1684.
- Okamoto T, Fujii M, Furusawa H, Tsuchiya K, Miyazaki Y, Inase N. The usefulness of KL-6 and SP-D for the diagnosis and management of chronic hypersensitivity pneumonitis. *Respir Med*. 2015;109(12):1576–1581.
- Greene KE, King TE Jr, Kuroki Y, et al. Serum surfactant proteins-A and -D as biomarkers in idiopathic pulmonary fibrosis. *Eur Respir J*. 2002;19(3):439–446.
- McGroder CF, Zhang D, Choudhury MA, et al. Pulmonary fibrosis 4 months after COVID-19 is associated with severity of illness and blood leucocyte telomere length. *Thorax*. 2021;76(12):1242–1245.
- Cocconcelli E, Bernardinello N, Giraudo C, et al. Characteristics and prognostic factors of pulmonary fibrosis after COVID-19 pneumonia. *Front Med (Lausanne)*. 2021;8:823600.
- Houchen-Wolloff L, Poinasamy K, Holmes K, et al. Joint patient and clinician priority setting to identify 10 key research questions regarding the long-term sequelae of COVID-19. *Thorax*. 2022;77(7):717–720.
- Evans R, McAuley H, Harrison E, et al. Physical, cognitive, and mental health impacts of COVID-19 after hospitalisation (PHOSP-COVID): a UK multicentre, prospective cohort study. *Lancet Respir Med*. 2021;9(11):1275–1287.
- Elneima O, McAuley H, Leavy O, et al. Cohort profile: Post-Hospitalisation COVID-19 (PHOSP-COVID) study. *Int J Epidemiol*. 2024;53(1):dyad165.
- Mothes R, Pascual-Reguant A, Koehler R, et al. Distinct tissue niches direct lung immunopathology via CCL18 and CCL21 in severe COVID-19. *Nat Commun*. 2023;14(1):791.
- Hermans C, Dong P, Robin M, et al. Determinants of serum levels of surfactant proteins A and B and Clara cell protein CC16. *Biomarkers*. 2003;8(6):461–471.
- Fujishima S, Shiomi T, Yamashita S, et al. Production and activation of matrix metalloproteinase 7 (matrilysin 1) in the lungs of patients with idiopathic pulmonary fibrosis. *Arch Pathol Lab Med*. 2010;134(8):1136–1142.
- Ley B, Brown KK, Collard HR. Molecular biomarkers in idiopathic pulmonary fibrosis. *Am J Physiol Lung Cell Mol Physiol*. 2014;307(9):L681–L691.

- 25 Armstrong HF, Podolanczuk AJ, Barr RG, et al. Serum Matrix Metalloproteinase-7, respiratory symptoms, and mortality in community-dwelling adults. MESA (Multi-Ethnic Study of Atherosclerosis). *Am J Respir Crit Care Med.* 2017;196(10):1311–1317.
- 26 Khan FA, Stewart I, Saini G, Robinson KA, Jenkins RG. A systematic review of blood biomarkers with individual participant data meta-analysis of matrix-metalloproteinase-7 in IPF. *Eur Respir J.* 2021;59(4):2101612.
- 27 da Silva-Neto PV, do Valle VB, Fuzo CA, et al. Matrix metalloproteinases on severe COVID-19 lung disease pathogenesis: cooperative actions of MMP-8/MMP-2 axis on immune response through HLA-G shedding and oxidative stress. *Biomolecules.* 2022;12(5).
- 28 Chavez-Galan L, Ruiz A, Martinez-Espinosa K, et al. Circulating levels of PD-L1, TIM-3 and MMP-7 are promising biomarkers to differentiate COVID-19 patients that require invasive mechanical ventilation. *Biomolecules.* 2022;12(3):445.
- 29 Mulet A, Tarrasó J, Rodríguez-Borja E, et al. Biomarkers of fibrosis in patients with COVID-19 one year after Hospital discharge: a prospective cohort study. *Am J Respir Cell Mol Biol.* 2023;69(3):321–327.
- 30 Wu D, Yang XO. Dysregulation of pulmonary responses in severe COVID-19. *Viruses.* 2021;13(6):957.
- 31 Iosef C, Knauer MJ, Nicholson M, et al. Plasma proteome of Long-COVID patients indicates HIF-mediated vasculo-proliferative disease with impact on brain and heart function. *J Transl Med.* 2023;21(1):377.
- 32 Lee JTH, Barnett SN, Roberts K, et al. Integrated histopathology, spatial and single cell transcriptomics resolve cellular drivers of early and late alveolar damage in COVID-19. *Nat Commun.* 2025;16(1):1979.
- 33 Kohno N, Awaya Y, Oyama T, et al. KL-6, a mucin-like glycoprotein, in bronchoalveolar lavage fluid from patients with interstitial lung disease. *Am Rev Respir Dis.* 1993;148(3):637–642.
- 34 Cambier M, Henket M, Frix AN, et al. Increased KL-6 levels in moderate to severe COVID-19 infection. *PLoS One.* 2022;17(11):e0273107.
- 35 Azekawa S, Chubachi S, Asakura T, et al. Serum KL-6 levels predict clinical outcomes and are associated with MUC1 polymorphism in Japanese patients with COVID-19. *BMJ Open Respir Res.* 2023;10(1):e001625.
- 36 Gunther A, Schmidt R, Nix F, et al. Surfactant abnormalities in idiopathic pulmonary fibrosis, hypersensitivity pneumonitis and sarcoidosis. *Eur Respir J.* 1999;14(3):565–573.
- 37 Salvioni L, Testa F, Sulejmani A, et al. Surfactant protein D (SP-D) as a biomarker of SARS-CoV-2 infection. *Clin Chim Acta.* 2022;537:140–145.
- 38 Tong M, Xiong Y, Zhu C, et al. Serum surfactant protein D in COVID-19 is elevated and correlated with disease severity. *BMC Infect Dis.* 2021;21(1):737.
- 39 McGroder CF, Salvatore MM, D'Souza BM, Hoffman EA, Baldwin MR, Garcia CK. Improved pulmonary function and exercise tolerance despite persistent pulmonary fibrosis over 1 year after severe COVID-19 infection. *Thorax.* 2024;79(5):472–475.
- 40 Jiang D, Xiao H, Dong R, et al. Krebs von den Lungen-6 levels in untreated idiopathic pulmonary fibrosis. *Clin Respir J.* 2022;16(3):234–243.
- 41 Spadaro S, Fogagnolo A, Campo G, et al. Markers of endothelial and epithelial pulmonary injury in mechanically ventilated COVID-19 ICU patients. *Crit Care.* 2021;25(1):74.
- 42 Cabrera-Benitez NE, Laffey JG, Parotto M, et al. Mechanical ventilation-associated lung fibrosis in acute respiratory distress syndrome: a significant contributor to poor outcome. *Anesthesiology.* 2014;121(1):189–198.