The persistence of SARS-CoV-2 in tissues and its association with long COVID symptoms: a cross-sectional cohort study in China

Wenting Zuo, Di He*, Chaoyang Liang*, Shiyu Du, Zhan Hua, Qiangqiang Nie, Xiaofeng Zhou, Meng Yang, Haidong Tan, Jiuyang Xu, Yanbing Yu, Yuliang Zhan, Ying Zhang, Xiaoying Gu, Weijie Zhu, Hui Zhang, Hongyan Li, Weiliang Sun, Mingzhi Sun, Xiaolei Liu, Liguo Liu, Chuanzhen Cao, Rui Li, Jing Li, Yun Zhang, Yuting Zhang, Jing Guo, Ling Zhao, Chuan-Peng Zhang, Hongyu Liu, Shiyao Wang, Fei Xiao, Yeming Wang, Zai Wang, Haibo Li, Bin Cao*

Summary

Background Growing evidence suggests that symptoms associated with post-COVID-19 condition (also known as long COVID) can affect multiple organs and systems in the human body, but their association with viral persistence is not clear. The aim of this study was to investigate the persistence of SARS-CoV-2 in diverse tissues at three timepoints following recovery from mild COVID-19, as well as its association with long COVID symptoms.

Methods This single-centre, cross-sectional cohort study was done at China–Japan Friendship Hospital in Beijing, China, following the omicron wave of COVID-19 in December, 2022. Individuals with mild COVID-19 confirmed by PCR or a lateral flow test scheduled to undergo gastroscopy, surgery, or chemotherapy, or scheduled for treatment in hospital for other reasons, at 1 month, 2 months, or 4 months after infection were enrolled in this study. Residual surgical samples, gastroscopy samples, and blood samples were collected approximately 1 month (18–33 days), 2 months (55–84 days), or 4 months (115–134 days) after infection. SARS-CoV-2 was detected by digital droplet PCR and further confirmed through RNA in-situ hybridisation, immunofluorescence, and immunohistochemistry. Telephone follow-up was done at 4 months post-infection to assess the association between the persistence of SARS-CoV-2 RNA and long COVID symptoms.

Findings Between Jan 3 and April 28, 2023, 317 tissue samples were collected from 225 patients, including 201 residual surgical specimens, 59 gastroscopy samples, and 57 blood component samples. Viral RNA was detected in 16 (30%) of 53 solid tissue samples collected at 1 month, 38 (27%) of 141 collected at 2 months, and seven (11%) of 66 collected at 4 months. Viral RNA was distributed across ten different types of solid tissues, including liver, kidney, stomach, intestine, brain, blood vessel, lung, breast, skin, and thyroid. Additionally, subgenomic RNA was detected in 26 (43%) of 61 solid tissue samples tested for subgenomic RNA that also tested positive for viral RNA. At 2 months after infection, viral RNA was detected in the plasma of three (33%), granulocytes of one (11%), and peripheral blood mononuclear cells of two (22%) of nine patients who were immunocompromised, but in none of these blood compartments in ten patients who were immunocompetent. Among 213 patients who completed the telephone questionnaire, 72 (34%) reported at least one long COVID symptom, with fatigue (21%, 44 of 213) being the most frequent symptom. Detection of viral RNA in recovered patients was significantly associated with the development of long COVID symptoms (odds ratio 5·17, 95% CI 2·64–10·13, p<0·0001). Patients with higher virus copy numbers had a higher likelihood of developing long COVID symptoms.

Interpretation Our findings suggest that residual SARS-CoV-2 can persist in patients who have recovered from mild COVID-19 and that there is a significant association between viral persistence and long COVID symptoms. Further research is needed to verify a mechanistic link and identify potential targets to improve long COVID symptoms.

Funding National Natural Science Foundation of China, National Key R&D Program of China, Chinese Academy of Medical Sciences Innovation Fund for Medical Sciences, and New Cornerstone Science Foundation.

Introduction

As of Aug 10, 2023, the COVID-19 pandemic had led to more than 769 million confirmed cases and 6·9 million deaths.[1] In addition to respiratory disease, SARS-CoV-2 infection can lead to various symptoms beyond the respiratory tract. Severe cases might progress to viral sepsis and even death. The virus has been detected in multiple organs during post-mortem examination,[2] suggesting that extrapulmonary dissemination of the virus might be associated with systemic diseases.

Lancet Infect Dis 2024

Published **Online** April 22, 2024 https://doi.org/10.1016/S1473-3099(24)00171-3

See **Online/Comment** https://doi.org/10.1016/S1473-3099(24)00216-0

*Contributed equally

For the Chinese translation of the abstract see **Online** for appendix 1

National Center for Respiratory Medicine, State Key Laboratory of Respiratory Health and Multimorbidity, National Clinical Research Center for Respiratory Diseases, Institute of Respiratory Medicine, Chinese Academy of Medical Sciences, Department of Pulmonary and Critical Care Medicine, Center of Respiratory Medicine (W Zuo MD, D He MD, Prof C Liang MD, J Xu MD, X Gu PhD, W Zhu MD, H Zhang MD, M Sun PhD, H Liu MD, S Wang MD, Y Wang MD, Ha Li PhD, Prof B Cao MD), **Department of General Thoracic Surgery** (Prof C Liang, W Zhu, Prof F Xiao MD), **Department of Gastroenterology** (Prof S Du MD), **Division of Gastrointestinal Surgery** (Prof Z Hua MD), **Division of Breast and Thyroid Surgery** (Prof M Yang MD), **and Second Division of Hepatopancreatobiliary Surgery** (Prof H Tan MD, Prof X Liu MD, Prof L Liu MD), **Department of General Surgery, Department of Cardiovascular Surgery** (Q Nie MD), **Department of Urology** (Prof X Zhou MD, C Cao MD), **Department of Neurosurgery** (Prof Y Yu MD, R Li MD, C-P Zhang MD), **Department of Clinical Laboratory** (Prof Y Zhan MD), **Department of Anesthesiology and Operating Theatre**

(Yi Zhang MD), **Institute of Clinical Medical Sciences** (Ho Li MD, W Sun PhD, Prof J Li PhD, Yun Zhang PhD, Yut Zhang PhD, J Guo PhD, Prof Z Wang PhD), **Department of Pathology** (Prof L Zhao MD), **New Cornerstone Science Laboratory** (Prof B Cao), **and Department of Clinical Research and Data Management, Center of Respiratory Medicine** (X Gu), **China–Japan Friendship Hospital, Chaoyang District, Beijing, China; Peking Union Medical College, Chinese Academy of Medical Sciences, Beijing, China** (W Zuo, Prof B Cao); **Department of Pulmonary and Critical Care Medicine, China–Japan Friendship Hospital, Capital Medical University, Beijing, China** (D He, H Liu, H Zhang, Prof B Cao) ; **Changping Laboratory, Beijing, China** (M Sun, Ha Li, Prof B Cao); **Department of Neurosurgery, Peking University China–Japan Friendship School of Clinical Medicine, Beijing, China** (C-P Zhang); **Tsinghua University–Peking University Joint Center for Life Sciences, Beijing, China** (Prof B Cao)

Correspondence to:
Prof Bin Cao, Department of Pulmonary and Critical Care Medicine, National Center for Respiratory Medicine, National Clinical Research Center for Respiratory Diseases, China–Japan Friendship Hospital, Beijing 100029, China
caobin_ben@163.com

or

Prof Zai Wang, Institute of Clinical Medical Sciences, China-Japan Friendship Hospital, Chaoyang District, Beijing 100029, China
wzai_163pass@163.com

or

Dr Haibo Li, Department of Pulmonary and Critical Care Medicine, National Center for Respiratory Medicine, National Clinical Research Center for Respiratory Diseases, China-Japan Friendship Hospital, Beijing 100029, China
shrineswe@vip.qq.com

See Online for appendix 2

Research in context

Evidence before this study

Previous studies have detected SARS-CoV-2 in multiple autopsy specimens of patients who died from COVID-19 and biopsy specimens taken from patients who have recovered from COVID-19. Additionally, SARS-CoV-2 has been shown to cause long-term illness, commonly referred to as post-COVID-19 condition (also known as long COVID). We searched PubMed, Google Scholar, and Web of Science using keywords such as "COVID-19" OR "SARS-CoV-2" OR "viral persistence" AND "long COVID" OR "post-acute sequelae of SARS-CoV-2 infection" for articles published in English between Jan 1, 2020, and March 1, 2024. We found several articles aimed at describing the association between long-term presence of SARS-CoV-2 and localised long COVID symptoms, but the association between the persistence of SARS-CoV-2 and systemic long COVID symptoms is still inconclusive.

Added value of this study

There is a dearth of research on SARS-CoV-2 persistence in diverse tissue samples collected from people who have recovered from COVID-19. Our study in patients who had recovered from mild COVID-19 includes not only blood samples but also surgical specimens from 13 types of solid tissues collected at 1 month, 2 months, or 4 months after infection. By using a large sample size, multiple timepoints of specimen collection, and highly sensitive digital droplet PCR, combined with other viral detection methods, we have provided preliminary insights into the persistence of SARS-CoV-2 in multiple organs after recovery from COVID-19, as well as its association with long COVID symptoms.

Implications of all the available evidence

SARS-CoV-2 can persist not only in patients critically ill with COVID-19 but also in multiple organs of patients who have recovered from mild COVID-19, with the proportion of tissues testing positive for viral nucleic acid gradually decreasing over time. We found a significant association between the persistent presence of residual SARS-CoV-2 and long COVID symptoms, suggesting that viral persistence is a risk factor for long COVID symptoms. Future studies are required to examine the potential mechanisms underlying the association between viral persistence and long COVID symptoms and identify potential targets for preventing the development of long COVID symptoms.

As well as in autopsy samples, residual SARS-CoV-2 RNA or protein have been detected in various sample types from patients with COVID-19, even after recovery. These include biopsy samples from the lungs, breasts, skin, appendix, intestine, adenoid, tonsils, taste buds, and olfactory neuroepithelium tissues, among others.[3–9] The virus has also been detected in stool samples[10] and plasma samples.[11,12]

It has been estimated that 6–68% of individuals who recover from COVID-19 have persistent symptoms, known as post-COVID-19 condition (also known as long COVID), affecting multiple organ systems, including fatigue, headache, shortness of breath, loss of smell or taste, and diarrhoea.[13,14] The mechanisms underlying long COVID are complex and uncertain.[15–18] Some studies have suggested that certain long COVID symptoms might be associated with viral persistence in relevant tissues, including the gut mucosa or epithelium, olfactory neuroepithelium, taste buds, and blood.[4,7,9,11]

However, there are, we believe, no large-scale studies investigating the persistence of SARS-CoV-2 in diverse tissues at different timepoints after recovery from COVID-19, and the association between viral persistence and long COVID symptoms remains unclear. In this study, we aimed to assess the persistence of viral nucleic acid and protein residues in diverse tissues collected from individuals who had recovered from mild COVID-19 using multiple techniques and explore the possible association between long COVID symptoms and viral persistence.

Methods

Study design and participants

This single-centre cross-sectional cohort study was done at China–Japan Friendship Hospital in Beijing, China. We enrolled patients who had mild COVID-19 (for definition see appendix 2 p 3) and confirmed positive through positive PCR or lateral flow tests during the omicron (BA.5·2 and BF.7) wave of SARS-CoV-2 around December, 2022. These patients were scheduled to undergo gastroscopy, surgery, chemotherapy, or immunotherapy, or were hospitalised for other reasons, at 1 month, 2 months, or 4 months after infection (as confirmed through PCR or lateral flow tests). Patient identifiers, including full name, date of birth, medical record number, and telephone number, were collected. We excluded patients who did not undergo treatments for various reasons, patients with repeat nasopharyngeal RT-PCR positivity at the time of the study, and patients who withdrew their consent. On the basis of previous reports on the organs in which SARS-CoV-2 persists,[3–11] we collected gastric mucosa samples, blood samples, and residual surgical samples from 13 types of solid tissue, including stomach, lung, skin, intestine, blood vessel, kidney, breast, thyroid, liver, brain, pancreas, gallbladder, and appendix, at approximately 1 month, 2 months, and 4 months after infection to assess the prolonged presence of the virus. For patients who underwent gastroscopy, we collected both oropharyngeal swab samples and gastric mucosa to account for potential contamination from the oral cavity.

According to WHO, long COVID symptoms are defined as new symptoms that occur beyond 3 months after infection or more severe symptoms than those experienced before COVID-19, and these symptoms should persist for at least 2 months with no other explanation.[19] To establish the association between viral persistence at 1 month, 2 months, 4 months, or any time after recovery from infection and long COVID symptoms as defined by WHO at 4 months, participants received follow-up telephone calls from trained physicians at around 4 months after infection and completed a symptom questionnaire over the telephone that was modified from our previous studies (appendix 2 pp 11–15).[20,21] We also collected data on variables including age, sex, BMI, history of cancer generally, current receipt of chemotherapy or immunotherapy, and drug interventions, and comorbidities including hypertension, diabetes, hyperlipidaemia, cardio-cerebrovascular disease, chronic respiratory diseases, anaemia, thyroid dysfunction, and number of COVID-19 vaccine doses received. The study adhered to the STROBE reporting guidelines for cohort studies. The Research Ethics Committee of the China–Japan Friendship Hospital approved this study (2023-KY-027), and written informed consent was obtained from all participants.

Procedures

One piece of gastric mucosa was collected and immediately frozen in liquid nitrogen for RNA extraction. Leftover samples from surgical pathology were collected with one portion fixed in 4% formalin for pathological sections, another portion frozen in liquid nitrogen for nucleic acid analysis, and the remaining portion fixed in 2·5% glutaraldehyde for electron microscopy observation.

4 mL of blood was collected by use of a vacuum blood collection tube containing edetic acid anticoagulant and centrifuged at 300×g for 10 min to obtain 140 µL plasma samples. The remaining blood was subjected to a density gradient centrifugation (700×g for 30 min) by use of lymphatic separation solution (Cytiva, Uppsala, Sweden) to obtain peripheral blood mononuclear cells (PBMCs). Granulocytes were obtained from the precipitates of density gradient centrifugation after erythrocyte lysis.

RNA from solid samples (surgical samples, gastroscopy samples, and blood cells) was extracted with TRIzol reagent (Invitrogen, Waltham, MA, USA) and from plasma samples by use of a QIAamp viral RNA mini kit (QIAGEN, Hilden, Germany) according to the manufacturer's instructions for virus detection.

To detect SARS-CoV-2 RNA in solid tissues, we did digital droplet PCR (ddPCR; QX200 AutoDG Droplet Digital PCR System, Bio-Rad, Hercules, CA, USA) using RNA extracted from the clinical samples. We selected two primer pairs, N1 and Orf1ab (appendix 2 p 16), which were reported to be 10–100 times more sensitive than quantitative PCR[22] to detect *N* and *ORF1ab* genes, respectively.[1] As a surrogate of replicating virus, we detected subgenomic RNA on the basis of the *E* gene,[23] which is generated from discontinuous transcription of the negative-strand, genome-length RNA that is created during replication of the virus and serves as the template for viral translation. We did quantitative real-time reverse-transcription PCR (qRT-PCR; QuantStudio Real-Time PCR Systems, ThermoFisher, Waltham, MA, USA) to measure the expression levels of *ACE2* and *TMPRSS2* in tumour or paratumour samples, and using the same primer sets as used for ddPCR on a subset of solid tissue samples to compare its sensitivity with ddPCR (appendix 2 p 3). We selected only one lung sample with the highest viral copy number through ddPCR detection.

To validate the accuracy of ddPCR, we did immunohistochemistry, which used enzyme-linked antibodies to detect SARS-CoV-2 S proteins in tissue sections (appendix 2 p 5). Additionally, immunofluorescence, which uses fluorescently labelled antibodies, was done to further visualise the presence of the virus in specific cell types (ie, epithelial cells and macrophages; appendix 2 pp 5–6). To further validate the accuracy of ddPCR, we used the RNAScope 2.5 HD assay kit (Advanced Cell Diagnostics, Newark, CA, USA) to perform RNA in-situ hybridisation to detect the *N* and *S* genes of SARS-CoV-2 on paraffin sections of tissue samples (appendix 2 p 6).

To investigate the effect of virus presence on host-gene transcription, transcriptome sequencing of vascular and lung tissues was done by Plastech Pharmaceutical Technology, Nanjing, China. Differentially expressed genes were mapped to biological pathways according to Kyoto Encyclopedia of Genes and Genomes pathway analysis (appendix 2 pp 7–8).

Statistical analysis

Baseline characteristics of patients are presented as mean (SD) for quantitative variables with reasonably symmetrical distributions and as absolute values with percentages for categorical variables. A McNemar test was done to investigate whether there was an association between the presence of SARS-CoV-2 in throat swabs and its presence in the gastric mucosa as a measure of contamination from nasal or oral sampling during gastroscopy. The exact two-sample Fisher–Pitman permutation test was used to assess whether there was a significant difference in the detection rates of the virus and the expression levels of virus receptors *ACE2* and *TMPRSS2* between tumour and paratumour tissues. To investigate whether individuals who are immunocompromised have difficulty in clearing the virus, we compared the virus detection ratios in individuals who were immunocompromised and immunocompetent, with the ratios reported as absolute values with percentages. To explore the association between viral persistence and long COVID symptoms, we used a multivariate logistic regression model to estimate the odds ratios (ORs) and 95% CIs accounting for

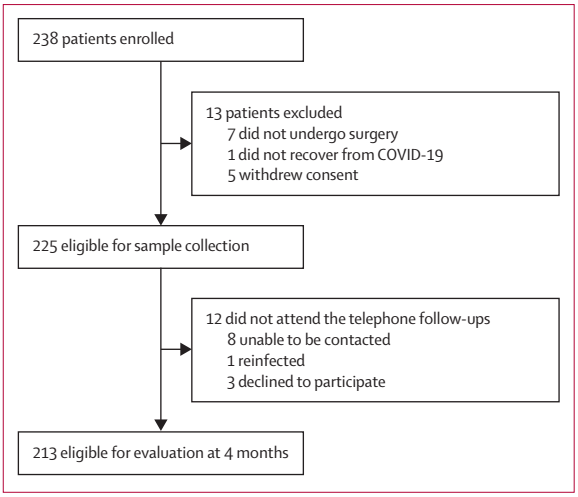

Figure 1: **Flow chart of the study**

confounders including age, sex, BMI, history of cancer generally, receipt of chemotherapy or immunotherapy, and comorbidities. We tested the linearity assumption underlying the logistic regression model for quantitative predictors (appendix 2 p 8). To investigate the association between viral copy number and long COVID symptoms, the tissue samples collected at 1 month, 2 months, 4 months, or at any time after infection were categorised into three groups (virus-free or low-virus group, medium-virus group, and high-virus group) on the basis of viral copy numbers of *N* or *ORF1ab* estimated with ddPCR by use of the classical K-Means algorithm, where the number of clusters was specified a priori. Markov distance was used as the distance function and excluded a sample with ultra-high viral RNA copy number. The symptom positive ratio was calculated as the number of tissues from patients with long COVID symptoms divided by the total number of tissues. All significance tests were two-sided, and a p value less than 0·05 was considered significant unless stated otherwise. All statistical analyses were done by use of SPSS version 26.0. The K-Means algorithm and the linearity-fit plot were done and generated in R, version 4.3.0. Other detailed materials and methods are described in the appendix 2 (pp 1–8).

Role of the funding source

The funders of the study had no role in study design, data collection, data analysis, data interpretation, writing of the report, or the decision to submit for publication.

Results

Of the 238 patients infected with SARS-CoV-2 around December, 2022, who were scheduled to undergo gastroscopy, surgery, chemotherapy, or immunotherapy, or who were hospitalised for other reasons, at China–Japan Friendship Hospital 1 month, 2 months, or 4 months after infection, 225 were eligible for sample collection. Of these, 213 (95%) participated in the 4-month follow-up and were included in the analysis of the association between SARS-CoV-2 persistence and long COVID symptoms (figure 1). 72 (34%) of 213 reported having at least one long COVID symptom at 4 months after infection (table). The mean age of patients with long COVID symptoms was 52·3 years (SD 13·5), and 35 (49%) were male. In comparison, the mean age of patients without long COVID symptoms was 55·0 years (SD 12·5), and 74 (52%) were male. Five (7%) patients with long COVID symptoms and 21 (15%) patients without received chemotherapy or immunotherapy. The majority of patients had received three doses of a COVID-19 vaccine before infection (56 [78%] of 72 patients with long COVID and 121 [86%] of 141 without; table).

We collected 317 tissue samples from 225 patients (appendix 2 p 16). These samples included 201 residual surgery specimens, 59 gastroscopy samples, and 57 blood component samples. Specifically, at 1 month post-infection, 53 samples from nine different tissue types were obtained from 38 patients. At 2 months post-infection, we collected 198 samples from 13 tissue types from 138 patients. Finally, at 4 months post-infection, we obtained 66 samples from nine tissue types from 49 patients (appendix 2 p 16). The tissue types included lung, skin, intestine, blood vessel, kidney, breast, thyroid, liver, stomach, brain, pancreas, gallbladder, and appendix, as well as blood components including plasma, granulocytes, and PBMCs. 63 samples each were from the lungs and stomach. 31 pairs of tumour and paratumour tissues were obtained (appendix 2 p 26), and 198 paratumour-only tissues were obtained to detect viral RNA or protein.

There was no association between the presence of SARS-CoV-2 in throat swabs and its presence in the gastric mucosa when detected by ddPCR (differential presence, p=0·0009 for *ORF1ab* and p=0·0008 for *N*), indicating that there was no contamination from nasal or oral sampling during gastroscopy (appendix 2 p 28). Over time, there was a gradual decrease in the proportion of solid tissues that tested positive for viral nucleic acid by ddPCR, with 16 (30%) of 53 solid tissue samples testing positive at 1 month after infection, 38 (27%) of 141 testing positive at 2 months, and seven (11%) of 66 testing positive at 4 months (figure 2A, B, D). Additionally, subgenomic RNA was detected, with four (8%) of 53 solid tissue samples testing positive for subgenomic RNA at 1 month, 21 (15%) of 141 at 2 months, and one (2%) of 66 at 4 months (figure 2C, E).

We also sorted tissue types into descending order on the basis of the proportion of tissue samples testing positive for SARS-CoV-2 *N*, *ORF1ab*, or subgenomic RNA. Viral *N* or *ORF1ab* RNA was found in liver, kidney, stomach, intestine, brain, blood vessel, lung, breast, skin, and thyroid tissues during the entire examined period (figure 3A–C). SARS-CoV-2 subgenomic RNA was also detected in 27 (44%) of 61 samples that were also positive

for *N1* or *ORF1ab*, with a tissue distribution including the liver, stomach, lungs, intestine, breasts, kidneys, and blood vessels (figure 3D).

To investigate the effect of the potential immunosuppressive microenvironment of tumours on SARS-CoV-2 persistence, we compared the viral load in tumour tissues and paratumour tissues. The results showed no significant difference in the rate of viral nucleic acid detection between tumour tissues (nine [29%] of 31) and paratumour tissues (12 [39%] of 31; p=0·59; appendix 2 p 26). Furthermore, to explore whether any difference in viral load was due to different concentrations of the SARS-CoV-2 receptors ACE2 and TMPRSS2, we compared the expression levels of *ACE2* and *TMPRSS2* in tumour tissues and paratumour tissues, and the results showed that the mRNA levels of *ACE2* (p=0·83) and *TMPRSS2* (p=0·49) were not significantly different (appendix 2 p 26).

We also did qRT-PCR for 74 solid tissue samples using the same primer sets as in ddPCR assay, but only one sample, which had the highest copy number in ddPCR (8000 copies of *N* per 20 µL), showed a positive result (appendix 2 p 17). Virus typing was done by sequencing the full genome of the virus in this sample and aligning it with known virus sequences to identify its specific subtype (appendix 2 p 4), and the result revealed that the virus was classified as BA.5.2, which was one of the two most prevalent strains at that time[24] (appendix 2 p 17).

To investigate any potential mechanism associated with the relatively high viral load in this sample, we subjected the sample to whole-exome sequencing to identify any relevant genetic variations (methodological detail in appendix 2 pp 4–5). The results revealed a mutation of c.1687G>T:p.A563S in the *DNAAF2* gene. Mutations in this specific site had not previously been reported, but it has been suggested that the mutation in the other site of *DNAAF2* is associated with ciliary dysmotility.[25] Electron microscopy (methodological detail in appendix 2 p 6) showed the absence of outer or inner dynein groups in three of 26 cilia, indicating the possibility of ciliary motility defects in the respiratory tract, which could contribute to the relatively poor viral clearance (appendix 2 p 29).

It has been suggested that individuals with haematological malignancies who are immunocompromised might face challenges in clearing the virus.[26] To further investigate this, we collected blood samples from nine patients who had recovered from mild COVID-19 with blood disorders, including two patients with diffuse large B-cell lymphoma, three patients with multiple myeloma, and one each with follicular lymphoma, non-Hodgkin lymphoma, idiopathic thrombocytopenic purpura, and primary mediastinal large B-cell lymphoma, at 2 months after infection (appendix 2 p 18). We examined the copy numbers of SARS-CoV-2 RNA in these samples by ddPCR. Viral RNA was detected in plasma in three (33%) of nine patients, in granulocytes in one (11%) patient, and in PBMCs in two (22%) patients. As controls, we simultaneously collected blood samples from ten volunteers who had recovered from mild COVID-19 2 months after infection and tested their viral RNA levels. However, no viral RNA was detected in this group of samples (appendix 2 p 30).

	With long COVID (n=72)	Without long COVID (n=141)
Days post-SARS-CoV-2 infection	77·6 (35·0)	73·9 (30·6)
Age, years	52·3 (13·5)	55·0 (12·5)
Sex		
Male	35 (49%)	74 (52%)
Female	37 (51%)	67 (48%)
BMI, kg/m^2	23·9 (3·0)	24·6 (4·1)
Comorbidities		
Hypertension	19 (26%)	44 (31%)
Diabetes	8 (11%)	13 (9%)
Cardio-cerebrovascular disease	9 (13%)	13 (9%)
Hyperlipidaemia	8 (11%)	16 (11%)
Thyroid dysfunction	2 (3%)	5 (4%)
Anaemia	2 (3%)	6 (4%)
Chronic respiratory diseases	4 (6%)	5 (4%)
Reason for surgery or hospitalisation		
Lung tumour	21 (29%)	37 (26%)
Intestinal tumour	1 (1%)	8 (6%)
Cholecystitis	1 (1%)	2 (1%)
Liver tumour	1 (1%)	2 (1%)
Thyroid tumour	3 (4%)	7 (5%)
Appendicitis	1 (1%)	0
Brain tumour	0	3 (2%)
Breast tumour	2 (3%)	7 (5%)
Renal tumour	4 (6%)	6 (4%)
Gastric tumour	3 (4%)	3 (2%)
Varicose veins	3 (4%)	17 (12%)
Blood cancer	2 (3%)	5 (4%)
Pancreatic tumour	0	2 (1%)
Other	30 (42%)	42 (30%)
Drug interventions*		
Antibiotics	33 (46%)	73 (52%)
Anti-inflammatory drugs	0	3 (2%)
Glucocorticoids	6 (8%)	13 (9%)
Antiviral drugs	0	3 (2%)
Chemotherapy or immunotherapy	5 (7%)	21 (15%)
Number of COVID-19 vaccine doses administered		
None	4 (6%)	9 (6%)
One dose	2 (3%)	1 (1%)
Two doses	10 (14%)	10 (7%)
Three doses	56 (78%)	121 (86%)

Data are n (%) or mean (SD). *Antiviral drugs included oseltamivir, baloxavir, nirmatrelvir–ritonavir, famciclovir, and ganciclovir. Anti-inflammatory drugs including loxoprofen and flurbiprofen were taken daily for a week. Glucocorticoids such as prednisone acetate, methylprednisolone, dexamethasone, and celestone were used for short periods. Chemotherapy or immunotherapy drugs including cyclophosphamide, etoposide, rituximab, thymalfasin, and basiliximab were administered over varying durations. Antibiotics including meropenem, vancomycin, posaconazole, and several others were prescribed for specific durations as well (appendix p 15).

Table: **Baseline characteristics of patients who had recovered from mild COVID-19**

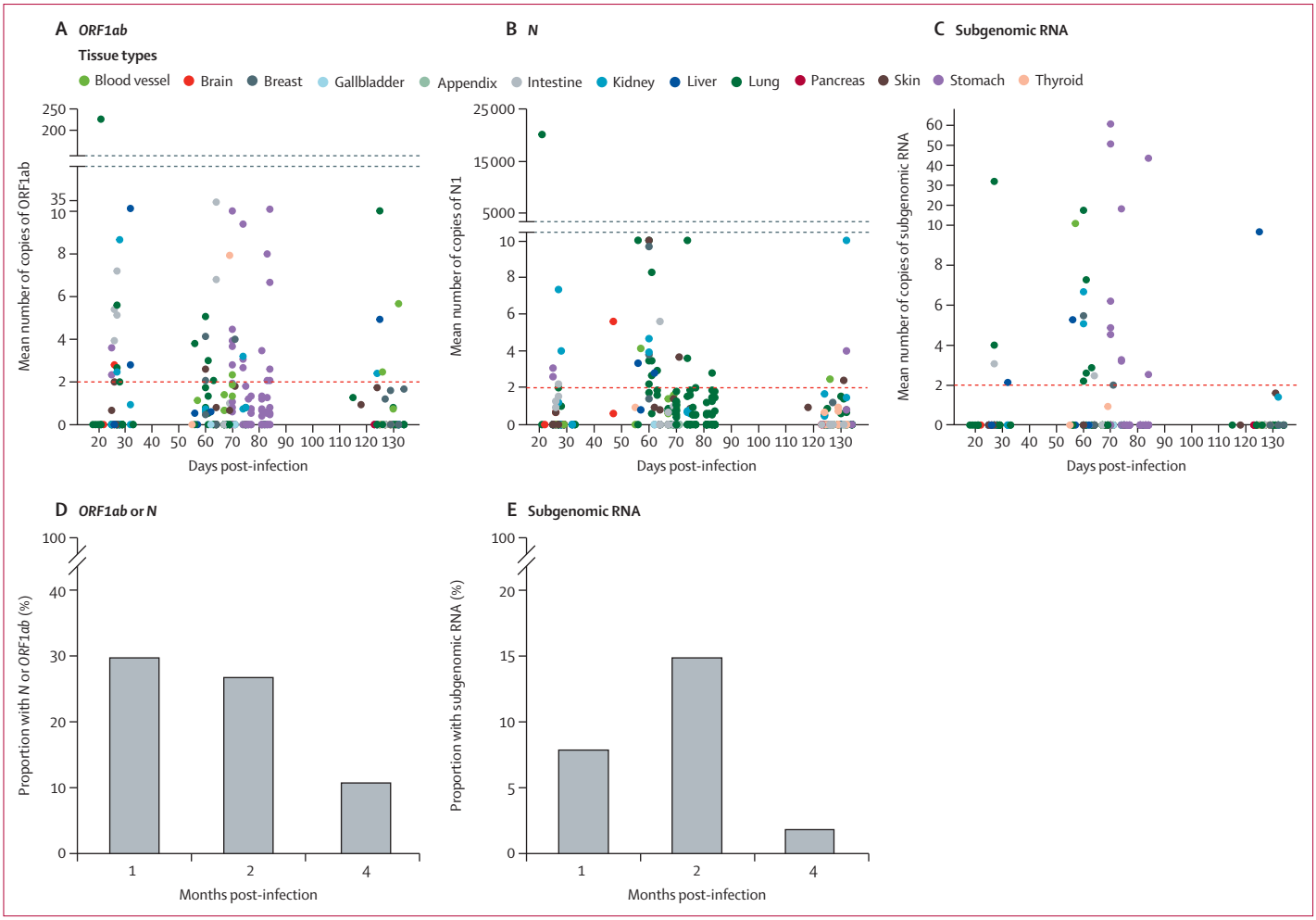

Figure 2: **Detection of SARS-CoV-2 RNA in various solid tissue types at 1 month, 2 months, and 4 months after infection**
Mean copy numbers of SARS-CoV-2 *ORF1ab* (A), *N* (B), and subgenomic RNA (C) in each tissue type at approximately 1 month, 2 months, and 4 months after infection, as measured by digital droplet PCR. The red dashed line represents the threshold (limit of detection; appendix p 3). Proportions of tissue samples collected at 1 month, 2 months, or 4 months after infection that tested positive for SARS-CoV-2 RNA (*ORF1ab* or *N*; D) or subgenomic RNA (E)

We confirmed that ddPCR was approximately 100 times more sensitive than qRT-PCR when using the N1 primer pair, which was consistent with a previous report[22] (appendix 2 p 19). Additionally, to further validate the accuracy of ddPCR, we did immunohistochemistry, RNA in-situ hybridisation, and immunofluorescence (appendix 2 pp 31–34). The results showed consistency among these assays (appendix 2 pp 20–21) and suggested that ddPCR has the highest sensitivity among the three methods and is the most reliable approach for qualifying and quantifying viral persistence. In addition, we observed the S protein in alveolar type I and type II epithelial cells, as well as in macrophages in lung samples, by immunofluorescence (appendix 2 p 31).

In the 213 patients who participated in the telephone follow-up questionnaire at 4 months after infection, the median follow-up after symptom onset was 127·0 days (IQR 126·0–129·0). In the 72 (34%) patients who reported at least one long COVID symptom, symptoms included fatigue, shortness of breath, palpitation, sleep difficulties, chest pain, diarrhoea, skin rash, cough, muscle weakness, hair loss, joint pain, abdominal pain, nausea or vomiting, myalgia, decreased appetite, headache, smell disorder, taste disorder, and dizziness, with fatigue (21%) being the most frequent symptom (appendix 2 p 22). After adjusting for potential confounders, we found a significant association between long COVID symptoms at 4 months and viral persistence at any timepoint post-infection (OR 5·17, 95% CI 2·64–10·13, p<0·0001). Long COVID symptoms at 4 months were significantly associated with viral persistence at 1 month and 2 months post-infection but not at 4 months. No association was found between long COVID symptoms and other variables (figure 4; appendix 2 p 22).

We further divided the tissue specimens collected around 1 month, 2 months, and 4 months and during the total period after infection into three groups depending

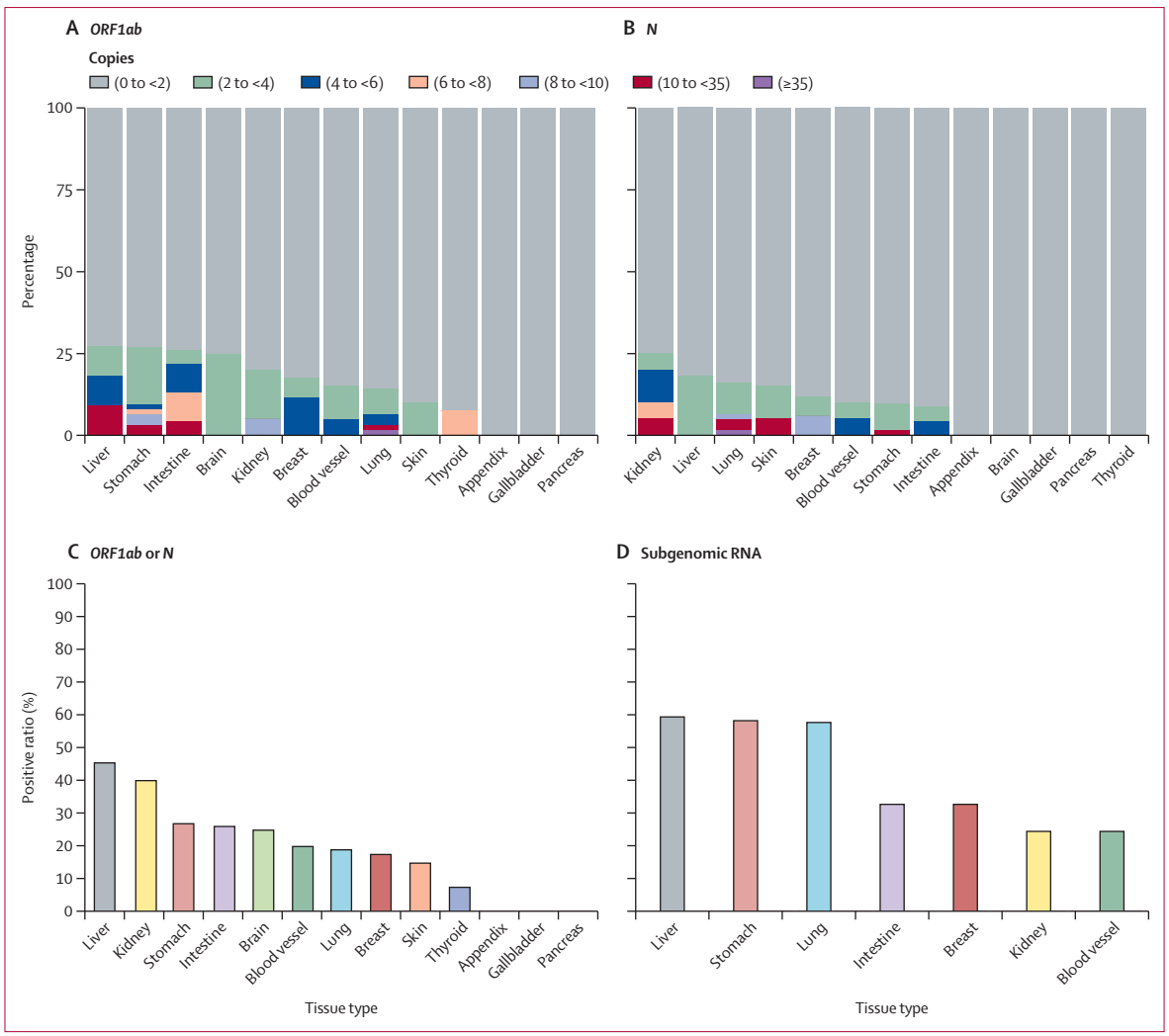

Figure 3: **Distribution of SARS-CoV-2 RNA in solid tissues at 1–4 months after infection**
Distribution of SARS-CoV-2 RNA in solid tissues, according to the RNA copy numbers of *ORF1ab* (A) and *N* (B) and the frequencies at which they were detected. Frequency of detection of SARS-CoV-2 RNA (*ORF1ab* or *N*; C) or subgenomic RNA in *N* or *ORF1ab* (D) positive solid tissues, in descending order. Samples were positive for SARS-CoV-2 RNA if they had at least two copies of *ORF1ab* or *N*. Samples were positive for subgenomic RNA if they had at least two copies of the subgenomic lead–*E* gene.

on virus copy number. The positive ratio of long COVID symptoms in each group was calculated (appendix 2 p 35). The results suggested that patients in the medium-virus group and high-virus group had a higher likelihood of developing long COVID symptoms than patients in the virus-free or low-virus group.

To investigate the potential mechanisms underlying the association between viral persistence and long COVID symptoms, we did transcriptome sequencing of 11 blood vessels and 24 lung tissues (appendix 2 pp 23–25). In lung tissues, we observed downregulated genes involved in the innate and adaptive immune defence against pathogens in the viral persistence group, such as *KLRD1*, *FYB1*, *VAV2*, *LILRB4*, *LILRB5*, *TICAM1*, *BTK*, *CD8A*, and *CD8B* (figure 5A, B). We also noted a significant downregulation of zinc finger protein-related genes in the viral persistence group, which play a role in defence against SARS-CoV-2.[27] These findings suggest that dysfunction in host immune defence might contribute to poor virus clearance.

In the blood vessel samples with viral persistence, we identified dysregulation of genes related to the complement and coagulation cascades, such as *FGG*, *VTN*, *F12*, *FGB*, *SERPINA1*, *C5*, *C1QB*, *SERPINE2*, *SERPINA5*, and *VSIG4* (figure 5C, D). We also observed dysregulation of genes involved in cholesterol metabolism pathways, such as *APOC3*, *APOA1*, *APOH*, *APOA2*, *LIPG*, *APOC1*, *SCARB1*, *CD36*, and *PLTP*, which is consistent with our previous research findings on the plasma proteome of long COVID.[28] These findings suggest that viral persistence might affect host cell functions, which could be another contributing factor to the occurrence of long COVID symptoms.

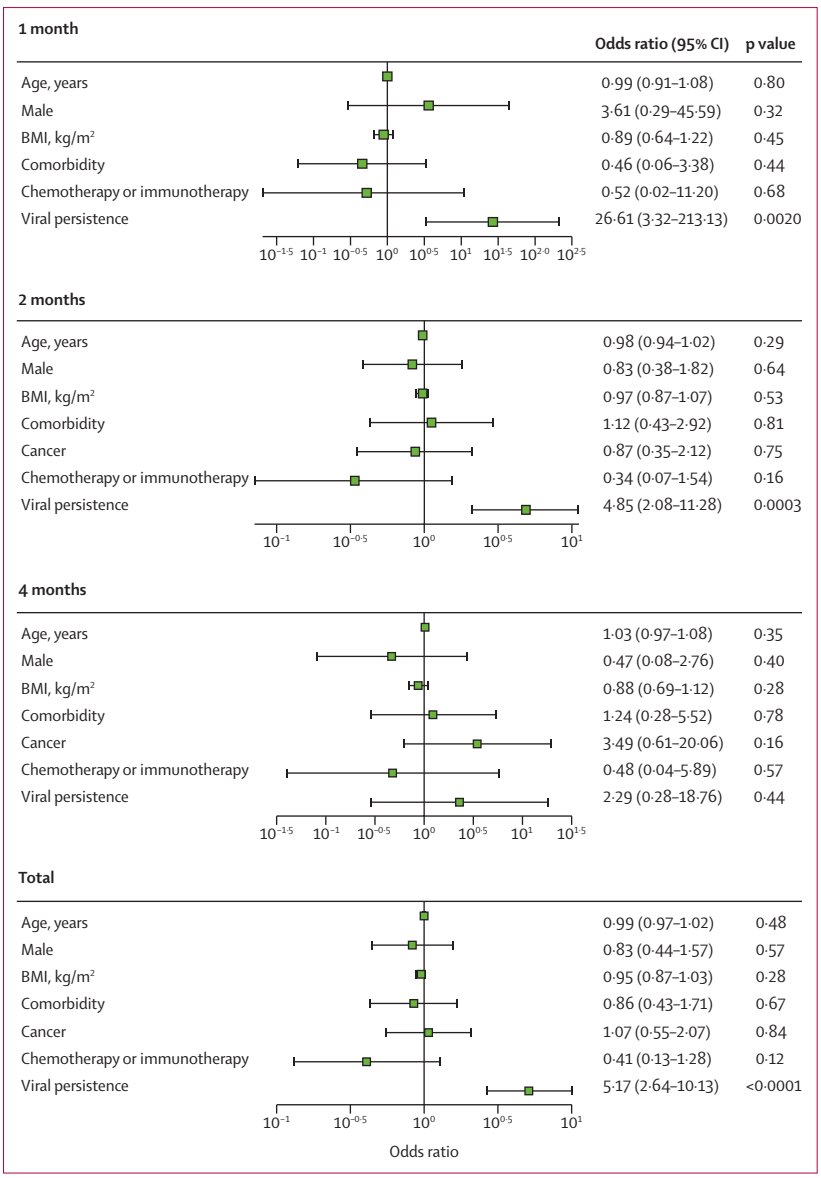

Figure 4: **Forest plots of associations between variables and long COVID symptoms**
The associations between different variables at 1 month, 2 months, 4 months, and any timepoint after infection and long COVID symptoms at 4 months after infection were analysed using a multivariable-adjusted logistic regression model to estimate odds ratios and 95% CIs for 213 patients who had recovered from mild COVID-19 (72 with and 141 without long COVID symptoms). Because of the presence of only four patients without tumours without any long COVID symptoms during the first month after infection, the cancer variable was excluded in the first month.

Discussion

In this study, we have observed the presence of viral nucleic acid in solid tissue samples from various organs, including the lung, liver, kidney, stomach, intestine, brain, breast, thyroid, blood vessels, and skin, in a proportion of patients who had recovered from mild COVID-19 at 1 month, 2 months, and 4 months after infection. The detection rate markedly decreased at 4 months post-infection, indicating a slow but ultimately effective viral clearance mechanism within the human body. In addition, we found that viral nucleic acids could be detected in a proportion of plasma samples, granulocytes, and PBMCs from patients who were immunocompromised 2 months after SARS-CoV-2 infection, but not in patients who were immunocompetent at a similar timepoint. Although the sample size for this analysis was small, it suggests impaired clearance of viruses in individuals who are immunocompromised. Most importantly, our research clearly showed an association between long COVID symptoms at 4 months after infection and the persistence of residual SARS-CoV-2 RNA.

Previous studies have reported viral persistence in diverse organs at 31–359 days after infection; however, these studies were carried out mainly on autopsy samples from people who died from COVID-19.[2,29] In recovered individuals, several studies have focused on detecting viral persistence in samples from the digestive and respiratory systems or blood samples.[4,7,11,30] Although the virus titre sharply decreased in the first month after infection, previous studies have not described the trend of virus clearance in diverse organs in individuals beyond 1 month after infection. Contrary to some studies,[6,8] we did not find viral RNA in appendix, gallbladder, or pancreas, but our small sample size might have contributed to this bias. The mechanism of long COVID is not completely understood, and a few studies have suggested that it might be related to the persistence of viruses.[9,10] However, the existing literature lacks sufficient evidence to support the association between viral persistence and systemic long COVID symptoms. The novelty of our research lies in studying viral persistence not only in blood and gastroscopy samples, but also in diverse surgical samples from patients who have recovered from mild COVID-19, and the comparison of virus copy numbers at different timepoints to show the trend of viral clearance. In addition, we clarified the association between viral persistence and long COVID symptoms and proposed possible mechanisms on the basis of transcriptomics data: deficient host antiviral response might lead to poor virus clearance and persistent virus might lead to dysfunction of host cells, although such a mechanism would need to be verified in future research.

Our study has several limitations. First, although it would be beneficial to explore the dynamic association between long COVID symptoms and viral load, ethical considerations prevent us from continuously observing virus clearance in the same patient. Second, variations in surgical numbers across different departments and time periods led to differences in the number of samples for different tissue types. Third, it is important to note that the prevalence of long COVID symptoms in our study is relatively high, possibly because the majority of patients included in our study had comorbidities or tumours, and thus underwent surgical treatments, chemotherapy, or immunotherapy. Fourth, several unmeasured confounders could not be accounted for, such as coagulation

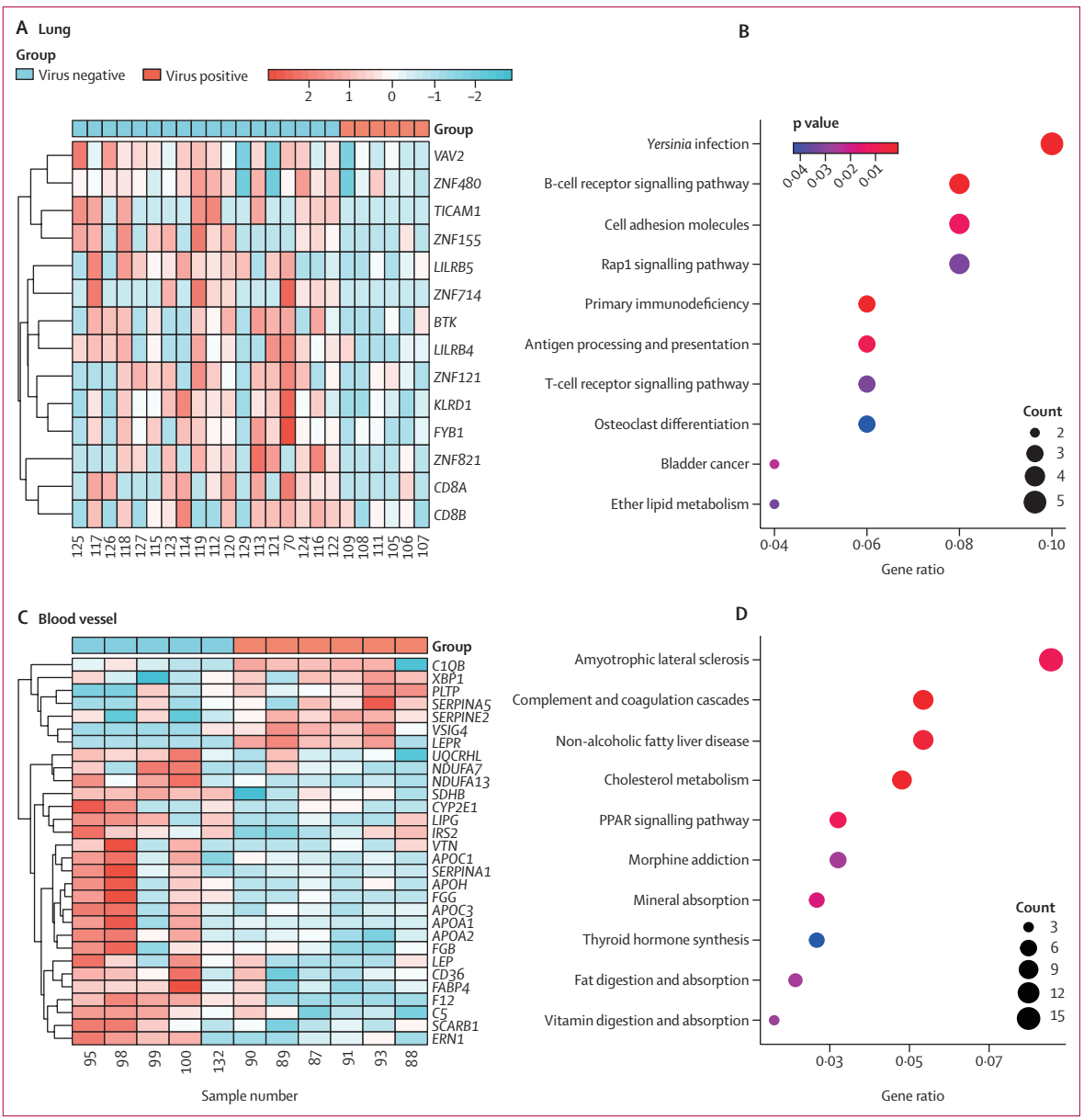

Figure 5: **Differential characteristics of lung and blood vessel tissues with or without viral persistence revealed by transcriptome sequencing**
Heat maps depict differentially expressed genes in the lungs (A) and blood vessels (C). Lung tissues include virus-negative samples collected at 1 month after infection (samples 125, 114, 112, 120, 70, and 116), 2 months after infection (samples 117, 126, 118, 115, 129, and 122), and 4 months after infection (samples 127, 119, 113, 121, 123, and 124), and virus-positive samples collected at 1 month after infection (sample 108) and 2 months after infection (samples 109, 111, 105, 106, and 107). Blood vessel tissues include virus-negative samples collected at 2 months after infection (samples 98, 100, and 132) and 4 months after infection (samples 95 and 99), and virus-positive samples collected at 2 months after infection (samples 90, 89, 87, 91, and 93) and 4 months after infection (sample 88). Bubble plots illustrate the enriched pathways of differentially expressed genes in lung tissue samples (B) and blood vessels (D) with or without viral persistence on the basis of Kyoto Encyclopedia of Genes and Genomes pathway analysis.

dysfunction, neurological dysfunction, and autoimmunity.[15] Fifth, we did not include any information on SARS-CoV-2-specific immune responses in the patients. Finally, the extremely wide CIs in the analyses of the association between viral persistence and long COVID symptoms at 1 month and 4 months after infection indicate sparse-data bias due to the low number of tissue samples available at those timepoints.[31]

In conclusion, we have identified an association between viral persistence in various tissues of the body and long COVID symptoms. Future studies should investigate the reasons for the persistence and its effect on the host, as research suggests that SARS-CoV-2 persistence in tissues might be associated with long-term immune dysregulation.[32] Screening methods should be developed to identify long COVID populations with viral persistence. In

addition, research including children should be considered in the future, since viral persistence not only exists in adults, but also in children, independent from disease severity, and can have an effect on their immune system.[33] As a new research direction, appropriately designed studies are needed to test the effects of antivirals or immunotherapies on long COVID symptom reduction. The host cell dysfunction caused by viral persistence might be a crucial aspect of long COVID pathogenesis. Biological processes such as protein translation and cell metabolism deserve further study as potential therapeutic targets.

Contributors

WZu, HZ, ZW, HaL, and BC conceived and designed the study. WZu and ZW drafted the paper. WZu, ZW, XG, HaL, and BC revised the paper. WZu, DH, and C-PZ did the analysis. WZu and DH collected the samples. WZu, QN, JX, WZh, XL, CC, HLiu, and SW collected and verified the data. CL, SD, ZH, QN, XZ, MY, HT, JX, YY, YulZ, YingZ, WZh, XL, LL, CC, RL, HLiu, SW, FX, and YW enrolled the patients. WZu, DH, ZW, HoLi, WS, JL, YunZ, YutZ, JG, and LZ participated in the experiments. WZu, DH, and MS did the the telephone questionnaires. All authors had full access to the data in the study, critically revised the manuscript for important intellectual content, and had final responsibility for the decision to submit for publication. All authors agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Declaration of interests

We declare no competing interests.

Data sharing

Data including transcriptome sequencing and viral load information are available from the corresponding author on reasonable request.

Acknowledgments

We acknowledge all patients who participated in this study and their families. We thank all staff of this study team at China–Japan Friendship Hospital. We acknowledge the electron microscope research services in the Institute of Clinical Medical Sciences for assistance with the generation of electron microscope images, and the Morphological Technology Services for RNAscope experiments. This work was supported by the National Natural Science Foundation of China (82241056), National Key R&D Program of China (2023YFC2306300), National Natural Science Foundation of China (82170015, 82100009, 82030002), Chinese Academy of Medical Sciences Innovation Fund for Medical Sciences (2021-I2M-1-048), and New Cornerstone Science Foundation.

References

1 WHO COVID-19 dashboard (updated August 10, 2023). https://data.who.int/dashboards/covid19/cases?n=c (accessed Aug 10, 2023).
2 Stein SR, Ramelli SC, Grazioli A, et al. SARS-CoV-2 infection and persistence in the human body and brain at autopsy. *Nature* 2022; **612:** 758–63.
3 Xu Q, Milanez-Almeida P, Martins AJ, et al. Adaptive immune responses to SARS-CoV-2 persist in the pharyngeal lymphoid tissue of children. *Nat Immunol* 2023; **24:** 186–99.
4 de Melo GD, Lazarini F, Levallois S, et al. COVID-19-related anosmia is associated with viral persistence and inflammation in human olfactory epithelium and brain infection in hamsters. *Sci Transl Med* 2021; **13:** eabf8396.
5 Bussani R, Zentilin L, Correa R, et al. Persistent SARS-CoV-2 infection in patients seemingly recovered from COVID-19. *J Pathol* 2023; **259:** 254–63.
6 Goh D, Lim JCT, Fernaíndez SB, et al. Case report: persistence of residual antigen and RNA of the SARS-CoV-2 virus in tissues of two patients with long COVID. *Front Immunol* 2022; **13:** 939989.
7 Zollner A, Koch R, Jukic A, et al. Postacute COVID-19 is characterized by gut viral antigen persistence in inflammatory bowel diseases. *Gastroenterology* 2022; **163:** 495–506.e8.
8 Hany M, Zidan A, Gaballa M, et al. Lingering SARS-CoV-2 in gastric and gallbladder tissues of patients with previous COVID-19 infection undergoing bariatric surgery. *Obes Surg* 2023; **33:** 139–48.
9 Yao Q, Doyle ME, Liu Q-R, et al. Long-term dysfunction of taste papillae in SARS-CoV-2. *NEJM Evid* 2023; **2:** EVIDoa2300046.
10 Natarajan A, Zlitni S, Brooks EF, et al. Gastrointestinal symptoms and fecal shedding of SARS-CoV-2 RNA suggest prolonged gastrointestinal infection. *Med* 2022; **3:** 371–87.e9.
11 Swank Z, Senussi Y, Manickas-Hill Z, et al. Persistent circulating severe acute respiratory syndrome coronavirus 2 spike is associated with post-acute coronavirus disease 2019 sequelae. *Clin Infect Dis* 2023; **76:** e487–90.
12 Craddock V, Mahajan A, Spikes L, et al. Persistent circulation of soluble and extracellular vesicle-linked spike protein in individuals with postacute sequelae of COVID-19. *J Med Virol* 2023; **95:** e28568.
13 Proal AD, VanElzakker MB, Aleman S, et al. SARS-CoV-2 reservoir in post-acute sequelae of COVID-19 (PASC). *Nat Immunol* 2023; **24:** 1616–27.
14 National Center for Health Statistics–US Census Bureau. Long COVID 2022–2023 (updated Sept 17, 2023). Household pulse survey. https://www.cdc.gov/nchs/covid19/pulse/long-covid.htm (accessed Sept 17, 2023).
15 Davis HE, McCorkell L, Vogel JM, Topol EJ. Long COVID: major findings, mechanisms and recommendations. *Nat Rev Microbiol* 2023; **21:** 133–46.
16 Liu Y, Gu X, Li H, Zhang H, Xu J. Mechanisms of long COVID: an updated review. *Chin Med J Pulm Crit Care Med* 2023; **1:** 231–40.
17 Wong AC, Devason AS, Umana IC, et al. Serotonin reduction in post-acute sequelae of viral infection. *Cell* 2023; **186:** 4851–67.e20.
18 Constantinescu-Bercu A, Lobiuc A, Căliman-Sturdza OA, et al. Long COVID: molecular mechanisms and detection techniques. *Int J Mol Sci* 2023; **25:** 408.
19 Soriano JB, Murthy S, Marshall JC, Relan P, Diaz JV. A clinical case definition of post-COVID-19 condition by a Delphi consensus. *Lancet Infect Dis* 2022; **22:** e102–07.
20 Huang L, Li X, Gu X, et al. Health outcomes in people 2 years after surviving hospitalisation with COVID-19: a longitudinal cohort study. *Lancet Respir Med* 2022; **10:** 863–76.
21 Huang C, Huang L, Wang Y, et al. 6-month consequences of COVID-19 in patients discharged from hospital: a cohort study. *Lancet* 2023; **401:** e21–33.
22 Liu X, Feng J, Zhang Q, et al. Analytical comparisons of SARS-CoV-2 detection by qRT-PCR and ddPCR with multiple primer/probe sets. *Emerg Microbes Infect* 2020; **9:** 1175–79.
23 Dagotto G, Mercado NB, Martinez DR, et al. Comparison of subgenomic and total RNA in SARS-CoV-2 challenged rhesus macaques. *J Virol* 2021; **95:** e02370-20.
24 Pan Y, Wang L, Feng Z, et al. Characterisation of SARS-CoV-2 variants in Beijing during 2022: an epidemiological and phylogenetic analysis. *Lancet* 2023; **401:** 664–72.
25 Omran H, Kobayashi D, Olbrich H, et al. Ktu/PF13 is required for cytoplasmic pre-assembly of axonemal dyneins. *Nature* 2008; **456:** 611–16.
26 DeWolf S, Laracy JC, Perales M-A, Kamboj M, van den Brink MRM, Vardhana S. SARS-CoV-2 in immunocompromised individuals. *Immunity* 2022; **55:** 1779–98.
27 Qin S, Xu W, Wang C, et al. Analyzing master regulators and scRNA-seq of COVID-19 patients reveals an underlying anti-SARS-CoV-2 mechanism of ZNF proteins. *Brief Bioinform* 2021; **22:** bbab118.
28 Gu X, Wang S, Zhang W, et al. Probing long COVID through a proteomic lens: a comprehensive two-year longitudinal cohort study of hospitalised survivors. *EBioMedicine* 2023; **98:** 104851.
29 Roden AC, Boland JM, Johnson TF, et al. Late complications of COVID-19. *Arch Pathol Lab Med* 2022; **146:** 791–804.
30 Cao B, Wang Y, Wen D, et al. A trial of lopinavir-ritonavir in adults hospitalized with severe Covid-19. *N Engl J Med* 2020; 382: 1787–99.
31 Greenland S, Mansournia MA, Altman DG. Sparse data bias: a problem hiding in plain sight. *BMJ* 2016; **352:** i1981.
32 Peluso MJ, Ryder D, Flavell R, et al. Multimodal molecular imaging reveals tissue-based T cell activation and viral RNA persistence for up to 2 years following COVID-19. *medRxiv* 2023; published online July 31. https://www.medrxiv.org/content/10.1101/2023.07.27.23293177v1 (preprint).
33 Buonsenso D, Martino L, Morello R, Mariani F, Fearnley K, Valentini P. Viral persistence in children infected with SARS-CoV-2: current evidence and future research strategies. *Lancet Microbe* 2023; **4:** e745–56.