


RESEARCH

Open Access



Multi-omics landscapes reveal heterogeneity in long COVID patients characterized with enhanced neutrophil activity

Ke Lin^{1†}, Jianpeng Cai^{1†}, Jingxin Guo^{1,2,3†}, Haocheng Zhang^{1†}, Gangqiang Sun^{1†}, Xun Wang^{4†}, Kun Zhu¹, Quanlin Xue¹, Feng Zhu⁵, Pengfei Wang⁴, Guanmin Yuan¹, Yuhan Sun¹, Sen Wang^{1,2,3*}, Jingwen Ai^{1,2,3*}  and Wenhong Zhang^{1,2,3}

Abstract

Background Omicron variant impacts populations with its rapid contagiousness, and part of patients suffered from persistent symptoms termed as long COVID. The molecular and immune mechanisms of this currently dominant global variant leading to long COVID remain unclear, due to long COVID heterogeneity across populations.

Methods We recruited 66 participants in total, 22 out of 66 were healthy control without COVID-19 infection history, and 22 complaining about long COVID symptoms 6 months after first infection of Omicron, referred as long COVID (LC) Group. The left ones were defined as non-long COVID (NLC) Group. We profiled them via plasma neutralizing antibody titer, SARS-CoV-2 viral load, transcriptomic and proteomics screening, and machine learning.

Results No serum residual SARS-CoV-2 was observed in the participants 6 months post COVID-19 infection. No significant difference in neutralizing antibody titers was found between the long COVID (LC) Group and the non-long COVID (NLC) Group. Transcriptomic and proteomic profiling allow the stratification of long COVID into neutrophil function upregulated (NU-LC) and downregulated types (ND-LC). The NU-LC, identifiable through a refined set of 5 blood gene markers (ABCA13, CEACAM6, CRISP3, CTSG and BPI), displays evidence of relatively higher neutrophil counts and function of degranulation than the ND-LC at 6 months after infection, while recovered at 12 months post COVID-19.

Conclusion The transcriptomic and proteomic profiling revealed heterogeneity among long COVID patients. We discovered a subgroup of long COVID population characterized by neutrophil activation, which might associate with the development of psychiatric symptoms and indicate a higher inflammatory state. Meanwhile, a cluster of 5 genes was manually curated as the most potent discriminators of NU-LC from long COVID population. This study can serve as a foundational exploration of the heterogeneity in the pathogenesis of long COVID and assist in therapeutic targeting and detailed epidemiological investigation of long COVID.

Keywords Long COVID, Omicron, Neutrophil degranulation, Transcriptome, Proteome

[†]Ke Lin, Jianpeng Cai, Jingxin Guo, Haocheng Zhang, Gangqiang Sun, and Xun Wang contributed equally to this manuscript.

*Correspondence:

Sen Wang

senwang@fudan.edu.cn

Jingwen Ai

jingwenai1990@126.com

Full list of author information is available at the end of the article



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

Introduction

Although COVID-19 has gradually turned into regional endemics, the relatively high portion of patients complaining to experience persistent symptoms, commonly termed as post-acute sequelae of COVID-19 (PASC), or long COVID, remains a global issue. According to the WHO, COVID-19 has caused 776 million human infections and 7.1 million deaths [1], with approximately ten percent of the patients experiencing PASC [2], undisputedly causing enormous medical burden [3, 4]. Therefore, to explore the concept and the potential mechanism behind PASC, and furthermore, to identify various subtypes of PASC, are of essential importance for deepening our knowledge of the long-term impact of COVID-19 infection.

Long COVID has neurological, psychiatric, gastrointestinal, and respiratory symptoms such as fatigue, shortness of breath, coughing, muscle weakness, and diarrhea, and has been repeatedly demonstrated in studies in the U.S., U.K., and other countries in a variety of regions and in different ethnic groups across ages and genders [5–9]. Although they can be classified into subgroups based on clinical symptoms, the diversity of symptoms demonstrated by long COVID patients makes it challenging to categorize them using a single clinical criterion [10].

Therefore, to distinguish different subgroups of long COVID patients, researchers must look beyond the clinical symptoms and seek in molecular and immune mechanisms. Unfortunately, describing the molecular mechanisms underneath long COVID has been proven to be even more complex, and although there have been many relevant studies, it is difficult to come to a relatively unified conclusion because of the complexity of the phenotypes. Previous studies have suggested persistent elevation of inflammation among long COVID patients [11], and TNF, IL-6, or NF- κ B-related signaling pathways have been found to be consistently elevated and established as potential prediction targets [12, 13]. Other researchers have identified possible mechanisms including inefficient clearing of viral RNA due to persisting reservoirs [14, 15], persistent reprogramming of immune cells [16], and endothelial inflammation or damaged tissues causing persistent hyperinflammation [17, 18].

Recent clinical research has shown differences in the severity and long COVID risk associated with the alpha, delta, and omicron variants [19, 20]. However, differentiating and comparing the long-term immunological changes caused by different variants in long COVID patients remains a challenge, and a consistent conclusion remains to be drawn. One significant reason for this is that a substantial portion of the population has experienced repeated infections with different variants, making

it difficult to isolate the effects of individual variants on long COVID symptoms and immunological changes.

Nevertheless, given the rapid contagiousness of the omicron variant and its impact on populations, it is essential to understand the mechanisms by which this currently dominant global variant leads to long COVID, and China's dynamic zero policy towards COVID-19 during 2020–2022 gave us a rare opportunity to recruit first-time Omicron infected patients and healthy controls. Our research is unique in that it utilizes omics analysis specifically among first-time infected patients during the Omicron wave in Shanghai, China, 2022, minimizing analytical interference caused by different virus variants, and we have conducted the 6-month and 12-month follow-ups on these patients as well. We aim to further explore the heterogeneous immune profiles during Omicron infections and their correlation with different clinical phenotypes, at the same time providing potential future medication target in the symptomatic population [13].

Methods

Enrollment of patients

Participants in this study were Omicron infected patient who had positive RT-PCR test or antigen test during the Omicron BA.2 outbreak in Shanghai and were hospitalized in Huashan hospital, all of whom had not been previously infected with COVID-19 during 2020–2022. Inclusion and exclusion criteria have been described previously (Supplementary Table 1) [8]. Participants' physical and mental conditions measured by questionnaires, described in our previous study, such as the Generalized Anxiety Disorder-7 (GAD-7) questionnaire and the Patients Health Questionnaire (PHQ-9), were collected at the 6-month and 12-month outpatient follow-ups (Supplementary Tables 2–6). Clinical laboratory test results were also collected, including blood routine and inflammatory indexes.

Based on the questionnaire results, we identified 22 individuals with severe long COVID symptoms reported during the 6-month follow-up and referred to this group as the long COVID (LC) Group. Age-, gender-, COVID-19-severity- and vaccination status matched COVID-19 convalescents that reported without any long COVID symptoms were enrolled as the non-long COVID (NLC) Group. Besides, 22 healthy controls without history of COVID-19 infection were screened by matching for age, gender, and vaccination status, referred to as the HC Group afterwards. Twelve of 22 participants in the LC Group and 12 of 22 individuals in the NLC Group attended the 12-month outpatient follow-up, and their questionnaires and laboratory examinations results were also collected. Peripheral blood samples were collected

from participants in the LC Group, NLC Group and the HC Group. Sample collection and clinical laboratory examinations were approved by the Ethics Committee of Huashan hospital (ethical approval number: 2023-750), and written informed consent was obtained from all participants enrolled.

Construction and production of variant pseudoviruses

Plasmids encoding the SARS-CoV-2 spike were constructed and cloned into the pCMV3 vector. HEK293T cells were transfected with spike-expressing plasmid using Polyethylenimine (Polyscience). VSV-G pseudotyped Δ Gluciferase ($G^*\Delta$ G-luciferase, Kerafast) was used to infect the cells one day post-transfection; 3 h after infection, cells were washed three times with PBS and moved to a fresh medium. The next day, the transfection supernatant was collected and clarified by centrifugation at 3000g for 10 min. Each viral stock was then incubated with 20% I1 hybridoma (anti-VSV-G; ATCC, CRL-2700) supernatant for 1 h at 37 °C to neutralize the contaminating VSV-G pseudotyped Δ G-luciferase virus before measuring titers and making aliquots to be stored at -80 °C.

Pseudovirus neutralization assays

Neutralization assays were performed by incubating pseudoviruses with serial dilutions of sera and scored by the reduction in luciferase gene expression. In brief, Pseudoviruses were incubated with serial dilutions of heat-inactivated sera in triplicate for 30 min at 37 °C. During the co-culture, Vero-E6 cells were trypsinized, resuspended with fresh medium, and then added into virus-sample mixture at a density of 2×10^4 cells/well. The mixture was incubated for an additional 18 h. The luminescence was quantified by Luciferase Assay System (Beyotime). Neutralization ID₅₀ values for sera was defined as the dilution at which the relative light units were reduced by 50% compared with the virus control wells (virus+cells) after subtraction of the background in the control groups with cells only. The ID₅₀ values were calculated using a nonlinear five-parameter dose response curve to the data in GraphPad Prism 9.

Whole blood and plasma droplet digital RT-PCR test of SARS-CoV-2

The droplet digital RT-PCR (ddRT-PCR) was performed using a TargetingOne Digital PCR System (TargetingOne, Beijing, China, licensed by NMPA, registration numbers: 20200025 and 20192220517), following the manufacturer's instructions. Briefly, the PCR reaction mixture (30 μ l) contained 15 μ l of RT-PCR mix (TargetingOne, 23301) and 15 μ l of extracted RNA was prepared. Then, the 30 μ l PCR reaction mixture and 180 μ l of droplet generation

oil were added into a droplet generation chip and droplet generation was performed using a Drop Maker (TargetingOne). The resulting droplet emulsion was amplified on an A300 thermal cycler (LongGene, Zhejiang, China) using the following PCR conditions: 55 °C for 15 min, 95 °C for 10 min, 45 cycles of 94 °C for 30 s and 57 °C for 1 min. After PCR, the PCR tube containing the droplets was connected to a droplet detection chip, followed by fluorescent signals detection using a Chip Reader (TargetingOne). To avoid the risk of viral infection and false positive results potentially due to laboratory contamination, all the experiments were performed inside the biosafety cabinet in a negative pressure biosafety laboratory using filter tips.

RNA isolation and library preparation for transcriptome

Total RNA was extracted using the TRIzol reagent (Invitrogen, CA, USA) according to the manufacturer's protocol. RNA purity and quantification were evaluated using the NanoDrop 2000 spectrophotometer (Thermo Scientific, USA). RNA integrity was assessed using the Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). Then the libraries were constructed using VAHTS Universal V6 RNA-seq Library Prep Kit according to the manufacturer's instructions. The transcriptome sequencing and analysis were conducted by OE Biotech Co., Ltd. (Shanghai, China).

RNA sequencing and differentially expressed genes analysis

The libraries were sequenced on an Illumina Novaseq 6000 platform and 150 bp paired-end reads were generated. About 56.82 raw reads for each sample were generated. Raw reads of fastq format were firstly processed using fastp and the low-quality reads were removed to obtain the clean reads [21]. Then about 46.89 clean reads for each sample were retained for subsequent analyses. The clean reads were mapped to the reference genome using HISAT2 [22]. FPKM [23] of each gene was calculated and the read counts of each gene were obtained by HTSeq-count [24]. PCA analysis were performed using R (v 4.3.1) to evaluate the biological duplication of samples. After the hierarchical clustering analysis, three samples were removed as outliers. Differential expression analysis was performed using the DESeq2 [25]. P value < 0.01 and absolute value of foldchange > 1 were set as the threshold for significantly differential expression gene (DEGs). Hierarchical cluster analysis of DEGs was performed using R (v 4.3.1) to demonstrate the expression pattern of genes in different groups and samples. The volcano plot as drew showed the expression of up-regulated or down-regulated DEGs using R packet ggplot. Based on the hypergeometric distribution, Reactome enrichment

analysis of DEGs were performed to screen the significant enriched term using R (v 4.3.1). Gene Set Enrichment Analysis (GSEA) was performed using GSEA software [26, 27] (v 4.3.2). The analysis used a predefined gene set, and the genes were ranked according to the degree of differential expression in the two types of samples. Then it is tested whether the predefined gene set was enriched at the top or bottom of the ranking list.

Plasma sample preparation for proteomics

Take out the samples from -80°C refrigerator in advance and thaw. Add 25 μl of plasma sample and 150 μl of Buffer B, centrifuge at room temperature at 4000 rpm for 2 min, Load 150 μl of incubation material working fluid (20 mg/ml) into centrifugation tube, incubate at 32°C for 60 min at 220 rpm; after incubation, centrifuge at 2500 rpm for 2 min, remove the supernatant; add 750 μl Buffer C, centrifuge at 2500 rpm for 2 min, remove the supernatant. Add 50 μl lysis solution (6 M urea, 2 M sulfourea), incubate at 32°C for 30 min at 220 rpm for denaturing. Then reduced sample with 3 μl 200 mM tris(2-carboxymethyl) phosphine (TCEP), 31.5°C , 600 rpm for 30 min and alkylated with 5 μl 800 mM iodoacetamide (IAA), 32°C , 220 rpm, 30 min at dark room. For protein digestion, 150 μl of 100 mM TEAB, 3 μl of trypsin (0.2 $\mu\text{g}/\mu\text{l}$) was added and incubate for 4 h at 32°C , 400 rpm. Then add 2 μl trypsin (0.2 $\mu\text{g}/\mu\text{l}$) again and incubate for 16 h at 32°C , 220 rpm. Quench the reaction by adding 22 μl of 10% trifluoroacetic acid and confirm that the peptide solution pH is between 2 and 3. The desalting operation was then performed using a SOLA μ (Thermo Fisher Scientific, San Jose, USA) desalting cartridge according to the product instructions. After desalting, use TMT-pro 16plex labeling reagent to label the peptide according to the product instruction, and then mix all samples into one sample. Fractionation was performed with a Waters XBridge Peptide BEH C18 column (300 \AA , 5 $\mu\text{m}\times 4.6\text{ mm}\times 250\text{ mm}$) under a DIONEX UltiMate 3000 Liquid Chromatogram. Mobile phase A was 10 mM ammonium hydroxide (pH=10), and mobile phase B was 98% ACN, 10 mM ammonium hydroxide (pH=10). Peptides were collected every one minute from 8% ACN to 35% ACN with a flowrate of 0.5 ml/min in 60 min, and then combined into 30 fractions. After SpeedVac dried, the 30 fraction samples were resuspended with 2% ACN, 0.1% Formic Acid and then sent for LC-MS analysis.

Mass spectrometry analysis for proteome, database search and statistical analysis

LC-MS/MS with the nanoflow DIONEX UltiMate 3000 RSLCnano System were coupled to an Orbitrap Exploris 480 mass spectrometer (Thermo ScientificTM, San Jose, USA), which equipped with a FAIMS ProTM (Thermo

ScientificTM, San Jose, USA), in data dependent acquisition (DDA) mode. Buffer A. 2% ACN, 98% H₂O containing 0.1% FA; Buffer B. 80% ACN in water containing 0.1% FA. All reagents were MS grade. For each acquisition, peptides were loaded onto a precolumn (3 μm , 100 \AA , 20 mm*75 μm i.d.) at a flowrate of 6 $\mu\text{l}/\text{min}$ for 4 min and then injected using a 30 min LC gradient (from 7 to 30% buffer B) at a flowrate of 300 nl/min (analytical column, 1.9 μm , 120 \AA , 150 mm*75 μm i.d.). Buffer A was 2% ACN, 98% H₂O containing 0.1% FA, and buffer B was 98% ACN in water containing 0.1% FA. All reagents were MS grade. The m/z range of MS1 was 375–1800 with the resolution at 60,000, normalized AGC target of 200% with the intensity threshold of $2e4$, and maximum ion injection time (max IT) of 85 ms. MS/MS experiment were performed with a resolution at 30,000, normalized AGC target of 200%, and max IT of 86 ms. isolation window was set to 0.7 m/z and first mass was set to 110 m/z. MS data were analyzed using the Fragpipe (version 19.1) search engine against the human protein database downloaded from UniProtKB (version 09/10/2022; 20420), with a precursor ion mass tolerance of 20 ppm, and fragment ion mass tolerance of 0.02 Da. Briefly, TMT pro-plex labels at lysine residues and the N-terminus, and carbamidomethylation of cysteine residues were set as static modifications. A cut-off criterion of a q-value of 0.01, corresponding to a 1% FDR, was set for filtering-identified peptides with highly confident peptide hits. Reporter ion ratio values of protein groups were exported for subsequent analysis. After the hierarchical clustering analysis, two samples were removed as outliers. PCA analysis were performed using R (v 4.3.1) to evaluate the biological duplication of samples. Reactome enrichment analysis of differentially expressed proteins (DEP) were performed to screen the significant enriched term using R (v 4.3.1).

Software and analysis

Computational analysis was carried out in R (v 4.3.1; release 16 June 2023). R (v 4.3.1) was used for computational analysis and drawing the violin plots, the heatmaps, bubble plots and bar plots. Heat maps were generated using the “pheatmap” library (v1.0.12), with data pre-normalized using scale function before plotting. Custom plotting, such as biological pathway analysis, was performed using the “ggplot2” library for base analysis, and then post processed in Adobe Illustrator (v 25.3.1). Differentially expressed proteins or genes were input into the STRING database for protein-protein interaction analysis to screen hub genes, and the protein-protein interaction (PPI) network was plotted using Cytoscape software (v 3.9.1). Random forest models were realized using the random forest function from the

“randomForest” R package (v 4.7–1.1). Violin plots and boxplots of neutralizing antibody titer and violin plots of ORF gene and N gene copy values were created using GraphPad Prism 9.5.0 software (GraphPad Software, San Diego, CA, USA). Graphical abstract was created with Biorender.com.

Results

General characteristics

In total, 66 of the participants were selected from the previous cohort study [8]. As noted in Methods, among the enrolled participants, 22 were patients reported with most severe long COVID symptoms (LC Group), 22 were participants without long COVID (NLC Group), and the rest were healthy controls (HC Group) with no COVID-19 infection history (Fig. 1a). The first and second sampling time was 6 and 12 months after the COVID-19 patient’s discharge, respectively, and 12 of the LC Group and 12 of the NLC Group attended the second follow-up. Baseline characteristics of enrolled participants showed no significant difference among groups (Table 1).

Enrollees had a mean age of 46 years (range 22–69), and 18 (27%) were female (Table 1). Participants in the LC and NLC Groups had asymptomatic or mild COVID-19 during the hospitalization. Besides, in the LC Group, 7 (32%) had chest tightness/chest pain, 17 (77%) had fatigue, and 10 (45%) had cough, and the average scores of the questionnaires were listed in Table 1. Detailed results of the questionnaires were listed in Supplementary Table 7.

Blood virus-neutralizing antibody titers and viral load in long COVID populations

Around 6 months after SARS-CoV-2 infection, neutralizing antibody titers of the LC and NLC Groups were significantly higher than that of the healthy donors, excluding antibody against BQ1.1. Although there was no evidence of a clear difference in these neutralizing antibody titers between the LC and NLC Groups, long COVID populations tended to have higher antibody levels (Fig. 1b).

Previous studies revealed that viral persistence and reactivation was one of the potential mechanisms of

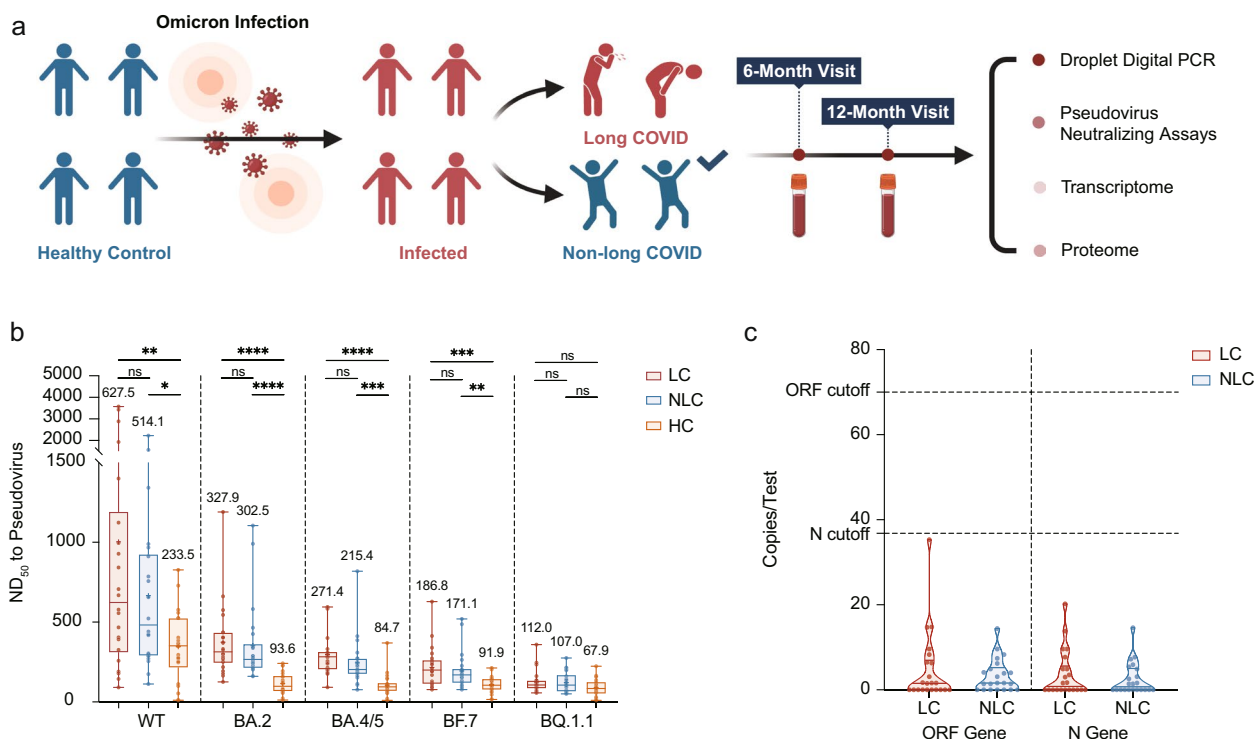


Fig. 1 Study design, and analysis of plasma neutralizing antibody titers and SARS-CoV-2 viral load. **a** The study population consisted of both healthy individuals and patients infected with Omicron. Among the COVID-19 convalescents, they were categorized into two groups based on the self-reported long COVID symptoms. Clinical data and blood samples were collected at the 6- and 12-month follow-ups. A comparative analysis was performed across the three groups by evaluating plasma viral load, neutralizing antibody titers, proteomics, and whole blood transcriptomics. **b** Pseudovirus neutralizing test was performed to measure the plasma-neutralizing activity against SARS-CoV-2 prototype and variants. The geometric mean neutralizing antibody titers are shown as box plots. P values were calculated from a paired two-tailed Student’s t-test. *** $P < 0.001$, ** $P < 0.01$, * $P < 0.05$. ns: no significant difference. **c** Violin plots show the copies of SARS-CoV-2 ORF gene and N gene per test detected via droplet digital PCR (ddPCR), indicating the viral load of SARS-CoV-2 in plasma

Table 1 Demographic and clinical characteristic of participants

Characteristics	LC Group (n = 22)	NLC Group (n = 22)	HC Group (n = 22)
Gender (n)			
Female (%)	16 (72.7)	16 (72.7)	16 (72.7)
Male (%)	6 (27.3)	6 (27.3)	6 (27.3)
Age, year (mean ± SD)	48.0 ± 13.1	45.7 ± 15.4	44.7 ± 12.8
20–39 years (%)	9 (40.9)	10 (45.5)	9 (40.9)
40–59 years (%)	7 (31.8)	5 (22.7)	10 (45.5)
≥ 60 years (%)	6 (27.3)	7 (31.8)	3 (13.6)
Vaccination status (n)			
Unvaccinated (%)	3 (13.6)	3 (13.6)	0 (0)
1 dose (%)	1 (4.5)	1 (4.5)	0 (0)
2 doses (%)	9 (40.9)	9 (40.9)	13 (59.1)
3 doses (%)	9 (40.9)	9 (40.9)	9 (40.9)
Acute COVID-19 severity			
Asymptomatic	5 (22.7)	5 (22.7)	–
Mild/moderate	17 (77.3)	17 (77.3)	–
Severe/critical	0 (0)	0 (0)	–
Long COVID symptoms (self-reported)			
Fatigue	17 (77.3)	–	–
Sleep difficulties	19 (86.4)	–	–
Hair loss	12 (54.5)	–	–
Cough	10 (45.5)	–	–
Sore throat	11 (50.0)	–	–
Chest pain	7 (31.8)	–	–
Smell disorder	9 (40.9)	–	–
Taste disorder	11 (50.0)	–	–
Headache	10 (45.5)	–	–
Palpitations	13 (59.1)	–	–
Skin rash	5 (22.7)	–	–
Dizziness	14 (63.6)	–	–
Muscle pain	10 (45.5)	–	–
Joint pain	14 (63.6)	–	–
Loss of appetite	7 (31.8)	–	–
Nausea/vomiting	6 (22.2)	–	–

long COVID [15]. In this study, droplet digital PCR was performed for quantification of residual plasma SARS-CoV-2 among the LC and the NLC Groups. Results turned out to be negative in both LC and NLC Groups, indicating no persistent SARS-CoV-2 remaining in the blood 6 months after infection (Fig. 1c).

Transcriptome atlas suggested suppressive chemokines and interleukins signaling functions 6 months post COVID-19, with upregulated neutrophil activity in the LC Group

To gain insight into the molecular characteristics of long COVID, the transcriptomic profiles were compared among LC, NLC, and HC Groups. A principal

component analysis (PCA) was performed with the host gene expression profiles to examine clustering of the global transcriptomic profiles of all samples. Nevertheless, the separation was not notable among the three groups (Fig. 2a), and the transcriptomic profiles were not significantly different between the LC and NLC Groups (Supplementary Fig. 1a). In the differential gene expression analysis, pairwise comparisons were made among these 3 groups, and differentially expressed genes (DEGs) were visualized by volcano plots (Figs. 2b, c, 4a). The findings revealed significant differential expression of genes, with 117 genes, 288 genes and 286 genes showing notable distinctions ($P < 0.05$, $|\log_2\text{FoldChange}| > 1$) between the LC and NLC Groups, the LC and HC Groups, and

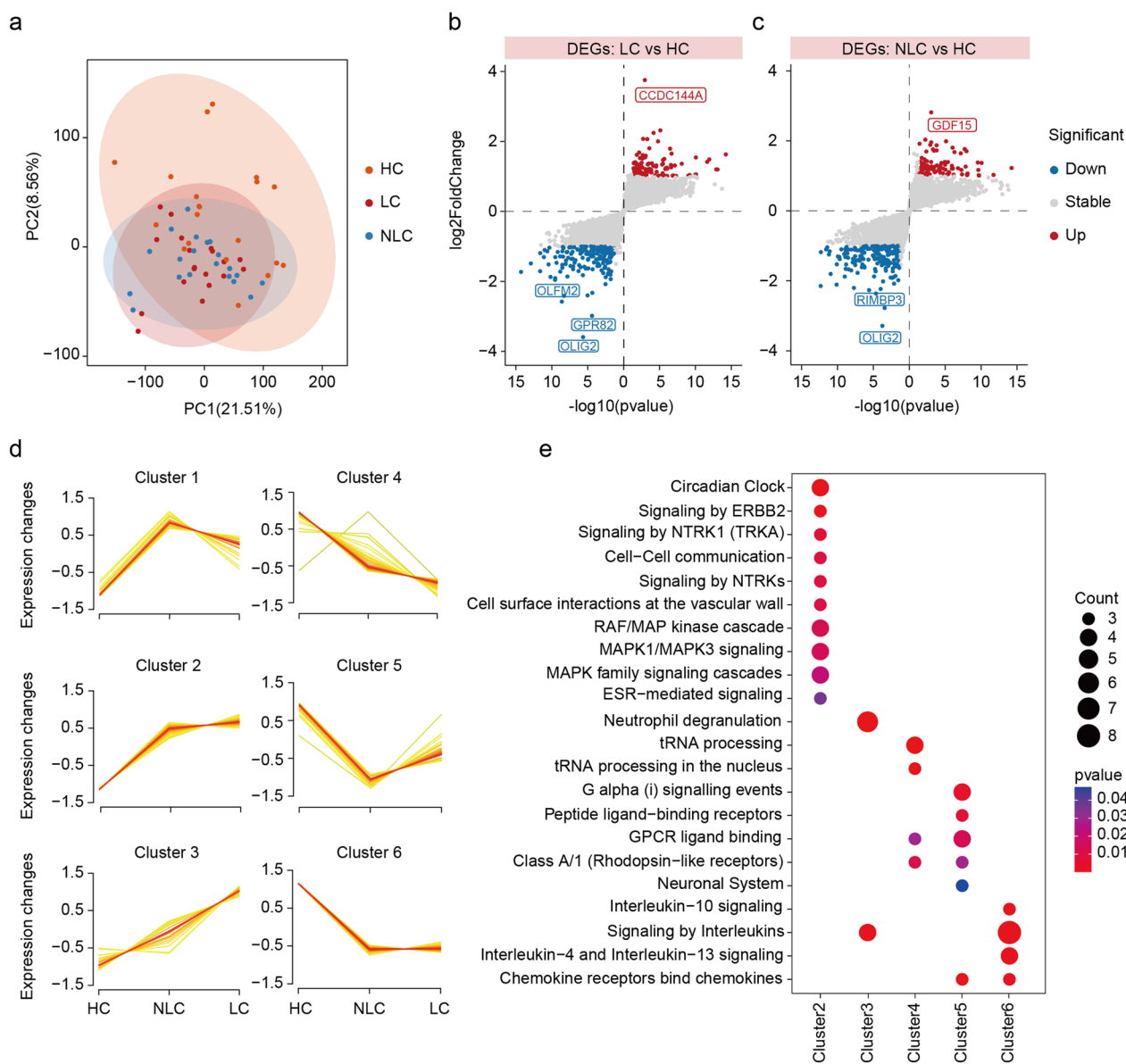


Fig. 2 Differential transcriptomic profiling across the LC, NLC and HC Groups. **a** The principal coordinate analysis plot showing the grouping of samples from the three groups based on global gene expression profiles. **b** The volcano plot of DEGs between the LC and HC Groups. The significant DEGs are highlighted in red (upregulated genes in LC) or blue (downregulated genes in LC). The insignificant DEGs are marked in grey. The most distinctly expressed genes are labeled. **c** The volcano plot of DEGs between the NLC and HC Groups. The significant DEGs are highlighted in red (upregulated genes in NLC) or blue (downregulated genes in NLC). The insignificant DEGs are marked in grey. The most distinctly expressed genes are labeled. **d** Expression patterns of both DEGs between the LC and HC Groups and DEGs between the NLC and HC Groups, across three groups. **e** Dot plot showing Reactome terms enriched in each cluster mentioned in Fig. 2d

the NLC and HC Groups, respectively. Furthermore, the results indicated that the transcriptomic differences before and after SARS-CoV-2 infection were greater than the differences between long COVID and non-long COVID.

To analyze distinct expression patterns of the LC, NLC, and HC groups, differential genes were classified into 6

clusters using Mfuzz (Fig. 2d). Genes in cluster 2 were enriched in MAPK-related signaling pathways. Genes in cluster 3, which were significantly higher expressed in the LC Group than the NLC Group and HC Group, were associated with neutrophil degranulation and interleukins (Fig. 2e). Cluster 4 genes were mainly involved in chemokine-related functions (Fig. 2e). Cluster 5 and

cluster 6 genes were highly expressed in the healthy individuals compared to those having had a COVID-19 infection, and the genes had an enrichment of chemokine and interleukins signaling pathways (Fig. 2e). Genes in cluster 5 and 6, including FOS, IL1B, CCR5, CCR2, CXCR3, and IL5RA, which represented functions of chemokines and interleukins, recovered to normal levels 12 months after infection (Supplementary Fig. 1b).

Our gene results suggested that, in contrast to the NLC Group, individuals in the LC Group had a stronger neutrophil function at 6 months post SARS-CoV-2 infection. Moreover, there was a decrease of chemokines and interleukins signaling functions in the participants at half a year post-infection, compared to the healthy individuals, and those functions recovered at the 12-month follow-up with the relief of long COVID symptoms.

Proteomic atlas revealed upregulated neutrophil activity in the LC Group

To analyze alterations in proteins associated with long COVID, proteomics profiles were compared among the LC, NLC, and HC Groups. Consistent with transcriptomic profiles, PCA results suggested that protein profiles were indistinguishable among the 3 groups (Fig. 3a). Pairwise comparisons were made among these 3 groups, in the differential protein expression analysis. The differentially expressed proteins (DEPs) were visualized by volcano diagrams (Figs. 3b, c, 4b), which revealed significant differential expression, with 76 proteins, 397 proteins, and 125 proteins showing notable distinctions ($P < 0.05$, $|\log_2\text{FoldChange}| > 0.25$) between the LC and NLC Groups, the LC and HC Groups, and the NLC and HC Groups, respectively. Similar to the transcriptomic profiles, the proteomic data also exhibited that the differences before and after infection were more significant in contrast to those between long COVID and non-long COVID individuals (Supplementary Fig. 2a).

To analyze different expression patterns of the LC, NLC, and HC groups, DEPs were classified into 6 clusters using Mfuzz (Fig. 3d). Proteins in cluster 1 were associated with biological oxidations pathway, which indicated that biological oxidations are more pronounced in the LC group compared to both the NLC and HC groups. This finding aligns with research that reports heightened oxidative stress following COVID-19 infection [28]. Specifically, the more intense oxidative response in the LC group compared to the NLC group suggested that oxidative stress may play a role in the development of long COVID symptoms [29, 30]. Paul and colleagues suggested that imbalances in redox reactions can lead to neurological manifestations and organ damage in long COVID [31]. Both clusters 2 and 3 showed enrichment in pathways related to neutrophil degranulation (Fig. 3e).

Proteins in cluster 3, which were significantly higher in the LC Group than the NLC Group and HC Group, were also enriched in pathways of extracellular matrix organization and diseases of metabolism (Fig. 3e). Cluster 4 and cluster 5 proteins were mainly involved in platelet function, suggesting that SARS-CoV-2 continued to have an impact on platelet function 6 months post infection (Fig. 3e). Proteins in cluster 6 had an enrichment of RHO GTPase cycle effectors pathways (Fig. 3e). Our proteins result resembled what we saw in transcriptomic data, which indicated that in contrast to the NLC Group, individuals in the LC Group had stronger neutrophil function at 6 months post SARS-CoV-2 infection.

Distinct neutrophil degranulation activation within long COVID patients

On the full-transcriptomic and proteomic evaluation, no distinct difference between LC group and NLC group was observed. Further PPI network analysis was performed using both genes and proteins significantly upregulated in the LC group compared to the NLC group (Fig. 4a, b). Remarkably, the largest PPI network cluster was correlated with the innate immune system, especially neutrophil and its degranulation function (Fig. 4c). In the LC group, expression of individual markers within the long COVID cohort was highly heterogeneous based on blood full transcriptome (Fig. 4d), which suggested potential subsets within the long COVID group with heterogeneous innate immunologic activity, especially neutrophil-related immune functions. In accordance with the hypothesis, unsupervised clustering of the long COVID group revealed a clear separation of the two subsets in the overall cohort (Fig. 4e, f). Pathway enrichment analysis of transcriptome profiling in reference to NLC group revealed that neutrophil degranulation biological pathway was positively associated with one of the subsets while negatively associated with the other one (Fig. 4g, h). The neutrophil degranulation upregulated group was hereafter referred as the neutrophil-function upregulated (NU-LC) subset, and the other one with downregulated neutrophil degranulation was named as the neutrophil-function downregulated (ND-LC) subset. Assessment of the neutrophil-related inflammation activity markers such as CXCL2, VWF, IL33, IL23R in the plasma protein sets showed significant abundances in the NU-LC Group compared to the ND-LC Group (Supplementary Fig. 2b). Participants in the NU-LC Group had higher level of neutrophil counts than the ND-LC Group and the NLC Group (Fig. 4i and Table 2). Further GSEA analysis revealed significant differences in the function of neutrophil degranulation between the two subsets (Fig. 4j). Furthermore, the enrichment of DEPs upregulated in the NU-LC compared to the ND-LC also demonstrated

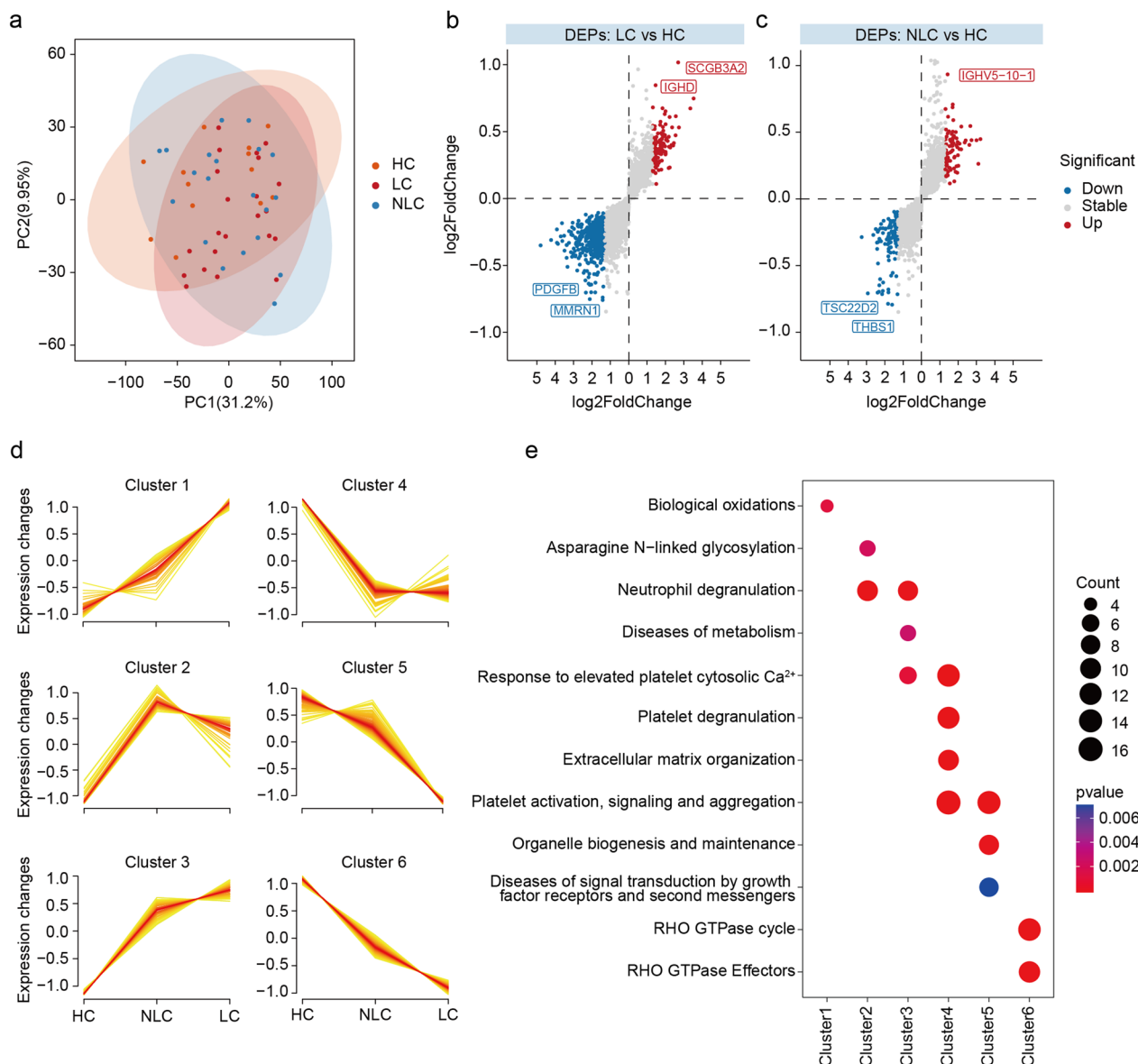


Fig. 3 Differential proteomic profiling across the LC, NLC and HC Groups. **a** The principal coordinate analysis plot showing the grouping of samples from the three groups based on global protein abundance profiles. **b** The volcano plot of DEPs between the LC and HC Groups. The significant DEPs are highlighted in red (upregulated genes in LC) or blue (downregulated genes in LC). The insignificant DEPs are marked in grey. The most distinctly upregulated or downregulated proteins are labeled. **c** The volcano plot of DEPs between the NLC and HC Groups. The significant DEPs are highlighted in red (upregulated genes in NLC) or blue (downregulated genes in NLC). The insignificant DEPs are marked in grey. The most distinctly upregulated or downregulated proteins are labeled. **d** Expression patterns of both DEPs between the LC and HC Groups and DEPs between the NLC and HC Groups, across three groups. **e** Dot plot showing Reactome terms enriched in each cluster mentioned in Fig. 3d

elevated neutrophil degranulation activity in the NU-LC Group (Fig. 4k).

Clinical distinctions in NU-LC subset

We further explored variations in neutrophil-related immune status within the NU-LC subset from 6 to 12 months post-COVID-19 infection. Analysis of gene expression scores associated with neutrophil

degranulation indicated that elevated neutrophil activity in the NU-LC group gradually returned to levels similar to the NLC group (Fig. 5a). The distinct neutrophil-related inflammatory signaling in the blood of the NU-LC and ND-LC subsets indicated clinical heterogeneity among long COVID cohorts. The NU-LC group had significantly elevated neutrophil counts compared to the ND-LC and NLC groups, though within normal

clinical ranges (Fig. 4i and Table 2). Increases in neutrophil counts were also connected with several broad markers of inflammation including C-reaction protein, procalcitonin (PCT) and erythrocyte sedimentation rate (ESR) (Fig. 5c). These increases also indicated different immune endoenvironments between NU-LC and ND-LC Groups, as patients in the NU-LC Group were potentially in a relatively inflammatory status. Moreover, elevated neutrophil counts were positively associated with psychological questionnaire scale score in the NU-LC group, such as depression and anxiety, which was opposite in the ND-LC group, establishing the implications in neutrophil-related long COVID symptoms (Fig. 5b, c).

Altogether, these findings indicate that while the long COVID questionnaires could help with identifying individuals with 'long COVID' symptoms, heterogeneity exists in the long COVID population. There was a group of people who had a relatively stronger neutrophil-related inflammatory response that was positively associated with a worse mental state, and they might be suffered from neutrophil-based immunopathological injury to some extent. The remaining group had a blood ecosystem status more similar to that of the non-long COVID population, suggesting their long COVID symptoms were more likely to be psychology-related without tissue or organ impairment.

Classifying NU-LC through machine learning

To accurately distinguish NU-LC individuals within the long COVID patients, we employed Random Forest (RF) modeling, which effectively classified NU-LC patients from all other long COVID individuals. The resulting predictive model exhibited exceptional robustness. Evaluation based on 10,000 models with randomized training

and test set splits yielded an average ROC AUC of 0.95 (SD+/-0.04). This indicates that, unlike the broader long COVID cohort, NU-LC patients could be efficiently identified regardless of the specific patient set used for model training. To this end, we manually curated a list of 5 genes from the most influential NU-LC discriminators, including ABCA13, CEACAM6, CRISP3, CTSG, and BPI (Fig. 6a). Those genes were then used as inputs for a new RF model (Fig. 6b). Despite the limited feature set, leveraging feature potency scores to guide parameter selection resulted in a model that remained effective in distinguishing the NU-LC Group. It achieved an average ROC AUC of 0.9596, underscoring that complete transcriptomic screening is not necessary for identifying these patients.

Significant differences in the expression of the five genes were observed between the ND-LC and NU-LC groups, with the NU-LC group showing significantly higher expression than the ND-LC group (Fig. 6c). Furthermore, the expression levels of these five genes are significantly higher in the NU-LC group compared to the NLC group, while the expression levels are similar between the HC and NLC groups (Supplementary Fig. 3). ABCA13 mediates transmembrane lipid transport, including cholesterol and phospholipids [32]. Its aberrant expression can cause lipid dysregulation and is linked to psychiatric disorders [33–35]. Research indicates that CEACAM6 facilitates neutrophil adhesion to endothelial cells, which in turn mediates their activation [36]. Moreover, upregulation of CRISP3 can promote leukocyte-mediated migration and neutrophil activation and degranulation and is closely associated with pulmonary fibrosis in COVID-19 patients [37, 38]. CTSG is mainly found in the azurophilic granules of neutrophils,

(See figure on next page.)

Fig. 4 Heterogeneity in long COVID patients with elevated neutrophil degranulation. **a** The volcano plot of DEGs between the LC and NLC Groups. The significant DEGs are highlighted in red (upregulated genes in LC) or blue (downregulated genes in LC). The insignificant DEGs are marked in grey. Distinctly expressed genes extracted from the core PPI cluster in Fig. 4c are labeled. **b** The volcano plot of DEPs between the LC and NLC Groups. The significant DEPs are highlighted in red (upregulated genes in LC) or blue (downregulated genes in LC). The insignificant DEPs are marked in grey. Distinctly upregulated proteins extracted from the core PPI cluster in Fig. 4c are labeled. **c** Networks generated by String database analysis using DEGs and upregulated DEPs upregulated in the LC Group compared to the NLC Group. The red and blue nodes indicate DEGs, and DEPs inputted for analyses, respectively. **d** Hierarchical clustering of the long COVID patients (LC Group). Major branches are differentially colored. **e** The principal coordinate analysis plot showing the grouping of samples from the NU-LC Group and ND-LC Group based on global gene expression profiles. **f** The principal coordinate analysis plot showing the grouping of samples from the NU-LC Group and ND-LC Group based on global protein abundance profiles. **g** Bar plots showing top5 enriched Reactome terms of DEGs between the ND-LC and NLC Group. The significant terms are highlighted in orange (terms enriched of upregulated genes in ND-LC) or green (terms enriched of downregulated genes in ND-LC). The insignificant terms are marked in grey (P value > 0.05 or Gene count < 4). **h** Bar plots showing top5 enriched Reactome terms of DEPs between the ND-LC and NLC Group. The significant terms are highlighted in orange (terms enriched of upregulated proteins in ND-LC) or green (terms enriched of downregulated proteins in ND-LC). The insignificant terms are marked in grey (P value > 0.05 or Gene count < 4). **i** Violin plots show the average of neutrophil counts of the NLC, ND-LC and NU-LC Groups. P values were calculated from a paired two-tailed Student's t-test. ***P < 0.001, **P < 0.01, *P < 0.05. ns: no significant difference. **j** Gene set enrichment analysis of DEGs between ND-LC and NU-LC Groups of Reactome Neutrophil Degranulation pathway dataset. Normalized enrichment score (NES) and P values are shown. **k** Dot plot showing enriched Reactome terms of DEPs between the ND-LC and NU-LC Group

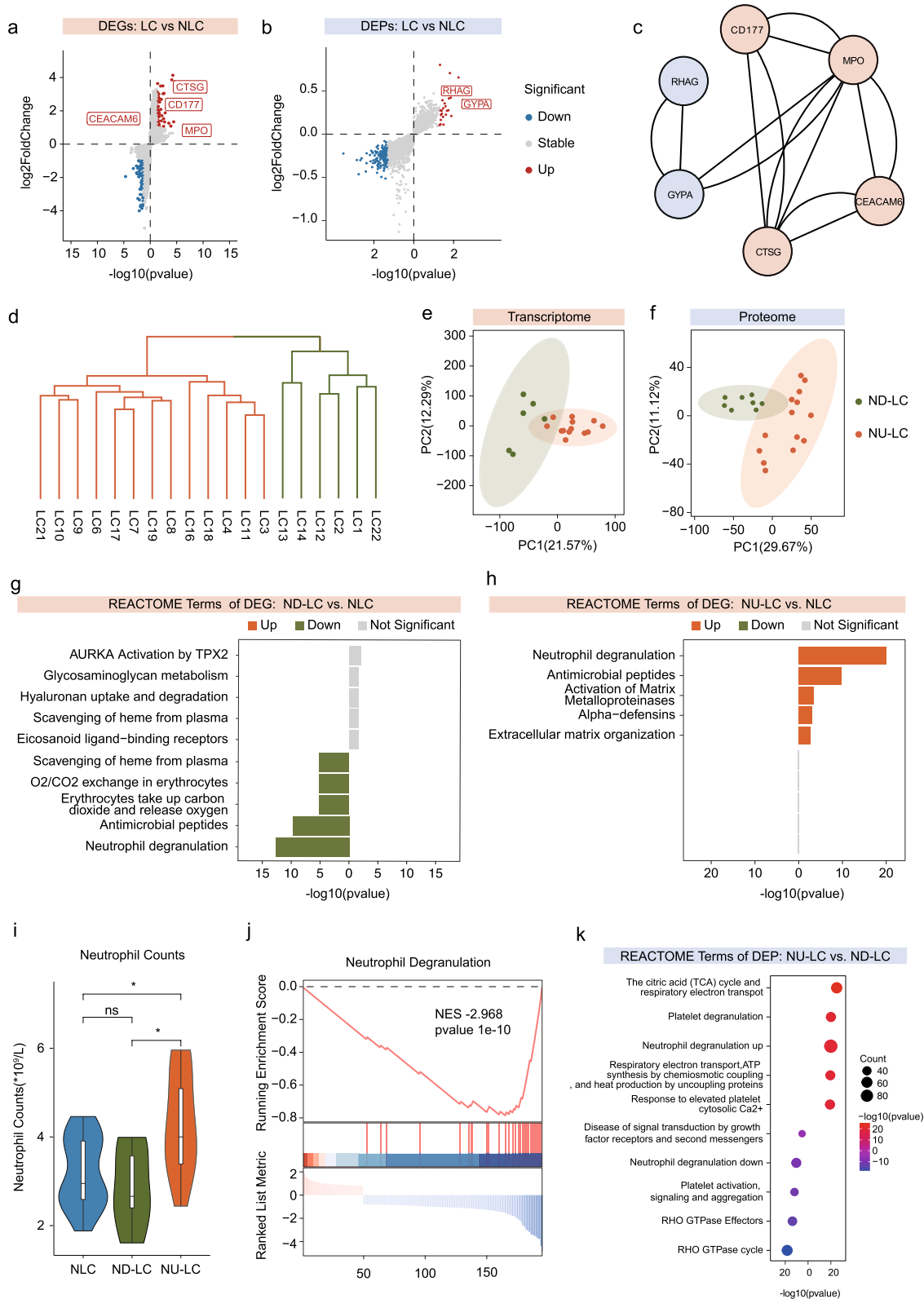


Fig. 4 (See legend on previous page.)

Table 2 Demographic and clinical characteristic of long COVID patients

Characteristics	ND-LC group (n = 6)	NU-LC group (n = 13)
Gender (n)		
Female (%)	2 (33.3)	3 (23.1)
Male (%)	4 (66.7)	10 (77.9)
Age, year (mean ± SD)	48 ± 13	48 ± 12
20–39 years (%)	2 (33.3)	5 (38.5)
40–59 years (%)	2 (33.3)	5 (38.5)
≥ 60 years (%)	2 (33.3)	3 (23.0)
Vaccination status (n)		
Unvaccinated (%)	1 (16.7)	1 (7.7)
1 dose (%)	0 (0)	1 (7.7)
2 doses (%)	3 (50.0)	5 (38.5)
3 doses (%)	2 (33.3)	6 (46.1)
Acute COVID-19 severity		
Asymptomatic	2 (33.3)	3 (23.1)
Mild/moderate	4 (66.7)	10 (77.9)
Severe/critical	0 (0)	0 (0)
Laboratory tests (mean ± SD)		
White blood cell, *10 ⁹ /l	5.1 ± 1.3	6.9 ± 1.4
Platelets, *10 ⁹ /l	276.0 ± 41.8	246.3 ± 49.5
Neutrophil, *10 ⁹ /l	2.7 ± 1.0	4.3 ± 1.1
C-reaction protein, mg/l	1.2 ± 0.4	1.8 ± 1.7
Procalcitonin, ng/ml	0.024 ± 0.009	0.032 ± 0.008
Erythrocyte sedimentation rate, mm/h	8.0 ± 4.6	5.6 ± 3.8
Anxiety symptom (GAD-7 score, mean ± SD)	7.8 ± 7.9	7.4 ± 6.9
Depression symptom (PHQ-9 score, mean ± SD)	10.8 ± 7.4	10.0 ± 8.2
PTSD symptom (PCL-C score, mean ± SD)	35.8 ± 13.5	35.5 ± 15.5
Fatigue symptom (Fatigue score, mean ± SD)	38.5 ± 11.3	46.5 ± 14.7

participating in the killing and digestion of phagocytosed pathogens [39], and increased levels of CTSG protein have been observed in the nasopharynx of COVID-19 patients [40]. Furthermore, BPI is primarily located in neutrophil granules and is a critical component of the innate immune system, playing an essential role in protecting the body from bacterial infections [41].

In the future, these genes could be set as specific targets and their expression levels would assist with identifying this specific patient subset among individuals with long COVID. However, the functional roles of these genes in the pathogenesis of long COVID require further investigation in future studies.

Discussion

Our study reached several conclusions. Here, we observed no serum residual SARS-CoV-2 six months after COVID-19 infection and no significant difference in neutralizing antibody titers between the LC Group and the NLC Group. Besides, chemokine and interleukin signaling was suppressed in our participants 6 months

post COVID-19 infection, which gradually returned to normal levels at 12 months. Finally, we found a subgroup of long COVID patients with upregulated neutrophil activity, revealing heterogeneity among the long COVID population. Those results unveiled long COVID development mechanisms underlying immunological changes caused by the Omicron variant among first-time infected individuals.

Previous studies have indicated the circulation of SARS-CoV-2 and spike proteins in the blood of long COVID patients [42, 43]. SARS-CoV-2 and its components might persist because of its reverse transcription ability, which alters subsequent gene expression [44, 45]. Other possible causes include the severity of the acute infection period and insufficient immune response [42]. However, our study did not detect residual SARS-CoV-2 in the bloodstream, likely because the Omicron variant, with its modified spike protein and mutated non-structural proteins, generates fewer infectious particles and has reduced fusogenicity and acute disease severity [46, 47]. Thus, it is plausible to contribute Omicron's

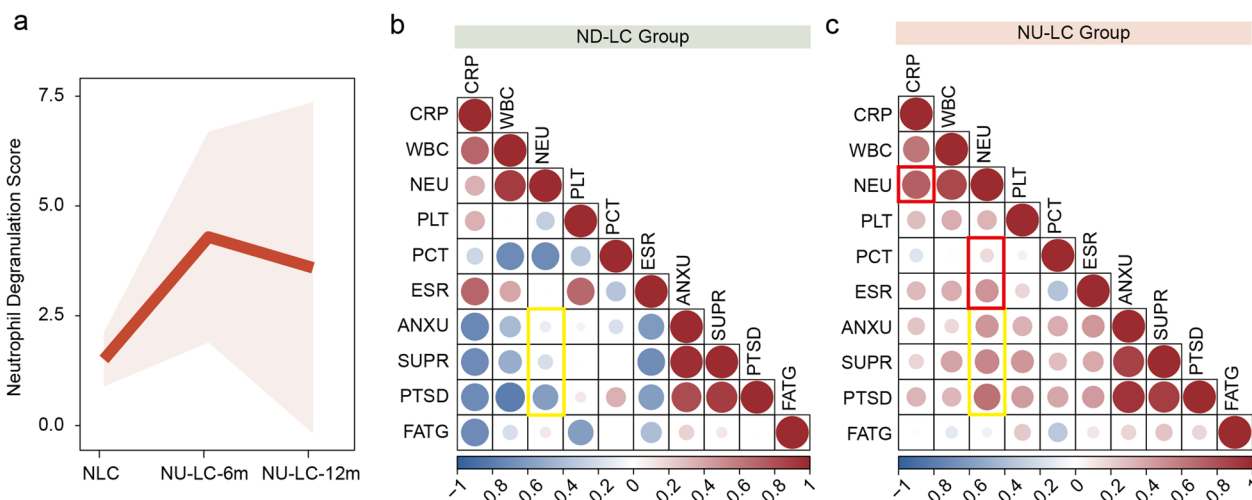


Fig. 5 Clinical features of NU-LC subset. **a** An interval plot with the mean and 95% confidence interval represents dynamic changes of Neutrophil Degranulation Score of NU-LC Group, from the 6-month visit to the 12-month visit, compared to the NLC Group. **b** Correlogram of ND-LC Group. Spearman rank order correlation values (r) are shown from blue (-1.0) to red (1.0); r values are indicated by color and circle size. Yellow rectangles denote correlation between neutrophil counts and scores of psychosocial questionnaires. **c** Correlogram of NU-LC Group. Spearman rank order correlation values (r) are shown from blue (-1.0) to red (1.0); r values are indicated by color and circle size. Yellow rectangles denote correlation between neutrophil counts and scores of psychosocial questionnaires. Red rectangles denote correlation between neutrophil counts and other inflammatory indexes

negligible viral persistence to its attenuated pathogenicity because of its decreased impact on host immune response and reduced genome reprogramming ability. Omicron characteristics above can also be observed in our previous studies, as our patients were mostly mild or asymptomatic during the acute onset of COVID [8]. Furthermore, population recruited had a history of only one COVID infection and most of them were vaccinated before the infection. Therefore, their immune systems were much less impacted, ruling out the long COVID mechanism of insufficient immune response.

Chemokine, which functions to recruit inflammatory cells, has been shown to increase in other long COVID studies, leading to various symptoms of long COVID, including fever, fatigue, or mild respiratory symptoms. Elevated cytokine levels trigger the CNS to perceive persistent inflammation, leading to exaggerated immune responses seen in long COVID [11, 48]. However, Scott and colleagues [49] revealed that specific chemokines, such as CXCR2, were decreased in the long COVID population and negatively correlated with the severity of illness during acute infection and fatigue symptoms afterwards, due to monocyte dysfunction. Our patients had downregulated chemokine-related functions, which could be due to the significant consumption of chemokines during acute infection or may indicate monocyte dysfunction. This finding may suggest a unique mechanism of Omicron that could be used to track the recovery of long COVID patients in clinical settings.

Previous research has revealed heterogeneity among long COVID patients, and researchers have attempted to classify different long COVID subgroups [8, 50, 51]. Different patients show symptoms of long COVID involving various organs, which can be associated with unique markers indicating inefficient viral clearance and specific autoantibodies (such as anti-IFN- α 2 and anti-nuclear antibodies), in the blood [43, 52]. Multiple omics approaches were further performed to divide long COVID patients according to immunological features, including elevated inflammation marked by pro-inflammatory cytokines and differential expression of B and T cells, which can be used as clinical targets for the precise treatment of long COVID [11, 13, 43]. Nonetheless, there are still some immunologically unremarkable populations that cannot be categorized into any of the subpopulations of long COVID, which adds complexity to following omics studies and clinical treatments [11, 43].

Our study uniquely selected first-time COVID-19 infected patients during the Omicron wave in Shanghai, China, 2022, and matching was performed between individuals who later developed long COVID and those who did not to minimize analytical interference. Our research uncovered heterogeneity of long COVID post Omicron infections, with one subgroup exhibited upregulated neutrophil degranulation activity and elevated neutrophil counts, and another without these features. Similarly, elevated neutrophils have previously been proposed as a discriminatory feature among long

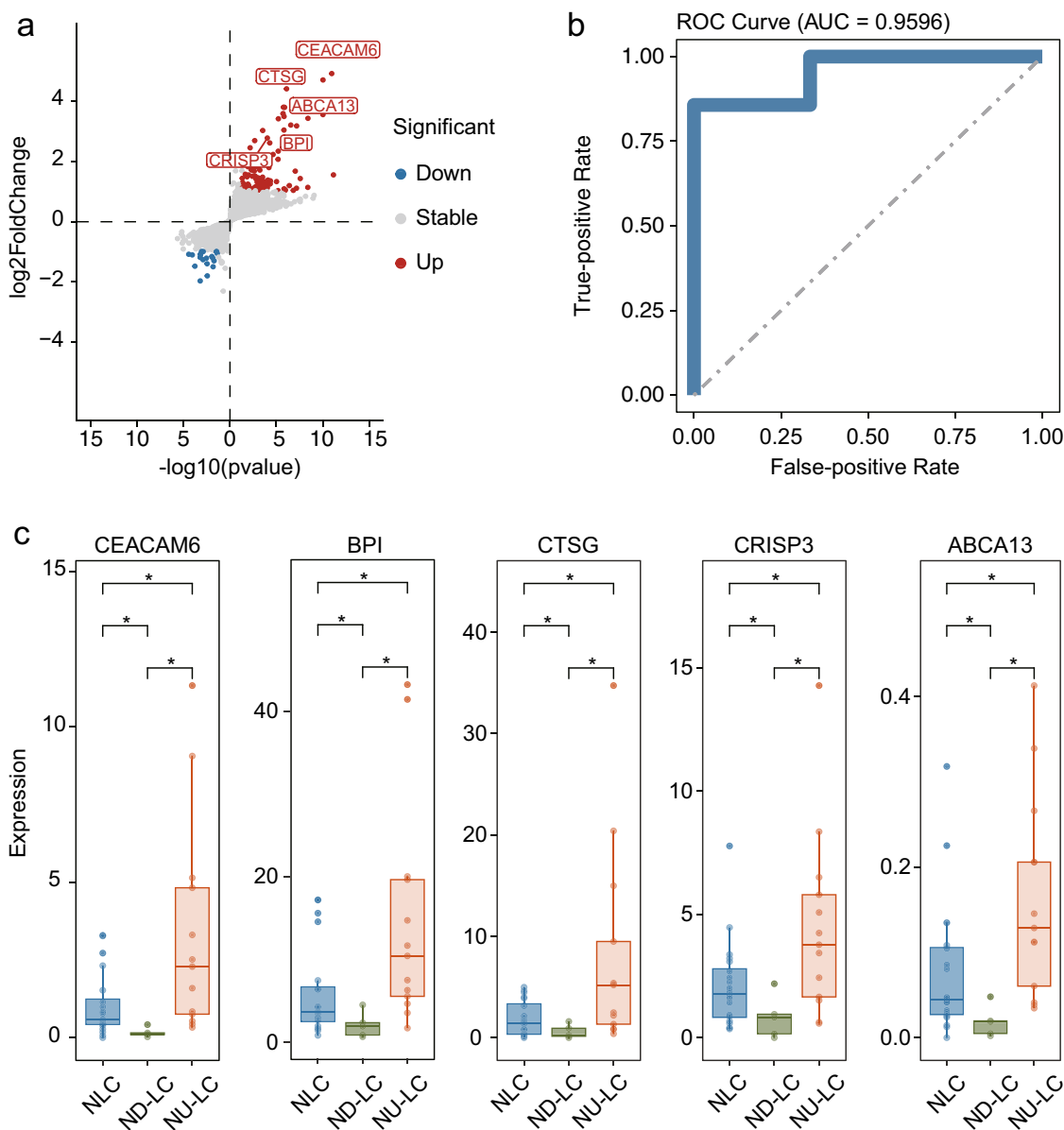


Fig. 6 Identifying NU-LC with small feature gene sets. **a** The volcano plot of DEGs between the NU-LC and ND-LC Groups. The significant DEGs are highlighted in red (upregulated genes in NU-LC) or blue (downregulated genes in ND-LC). The insignificant DEGs are marked in grey. The distinctly expressed genes selected as feature gene sets are labeled. **b** Receiver-operating characteristic (ROC) plot displaying one model with randomized train/test splits classifying NU-LC from other long COVID patients using the manually curated list of 5 high-potency genes identified in **a**. **c** The average expression level of the manually curated 5 genes, identified in **a**, of the NLC, ND-LC and NU-LC Groups, are shown as box plots. P values were calculated from a paired two-tailed Student's t-test. *** $P < 0.001$, ** $P < 0.01$, * $P < 0.05$. ns: no significant difference

COVID subgroups, suggesting different immunological profiles [11, 13]. In the study by Woodruff et al., the long COVID subgroup with neutrophil elevation is also identified with upregulation of other inflammation biomarkers, including IL-6, IL-8 and IL-1B. Furthermore, in another study, long COVID patients are further classified into two subgroups with different inflammatory profiles, as one of the subgroups have increased

neutrophil degranulation and displayed separate immunological features distinct from its counterpart, which has relatively lower cytokine and chemokine levels [11]. Researchers have reported that, abnormal elevation of neutrophils and their associated functions, as well as successive NETosis, indicating immunothrombosis and damaging of blood vessels inner linings, can be associated with severe COVID-19 [12, 53]. Besides, elevated

neutrophil counts and activity reflect an inflammatory immune status related to possible presence of inflammatory infiltration in some organs or tissues, thereby resulting in corresponding clinical symptoms. However, psychological factors could contribute to symptoms of the other group without neutrophil activation, given their normal indicators and mild clinical symptoms.

In this study, we employed machine learning techniques and identified a cluster of 5 genes classifiers for precise discrimination of long COVID population with different neutrophil activity status. It might be possible to categorize long COVID patients by rapidly assessing the expression levels of these 5 genes in future clinical applications without utilizing highly specialized technology, although their prediction abilities of the disease outcome were yet to be explored. Our research enriches current understanding of long COVID heterogeneity, serves as the basis for research of its mechanism and development, and provides insights for long COVID prevention, targeted therapeutics, and in-depth epidemiological investigations.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12967-024-05560-6>.

Supplementary Material 1. The principal coordinate analysis plot showing the grouping of samples from the two groups based on global gene expression profiles. The interval plots with the mean and 95% confidence interval represents dynamic changes of gene expression level of the LC Group, and the NLC Group from the 6-month visit to the 12-month visit, compared to the HC Group.

Supplementary Material 2. The principal coordinate analysis plot showing the grouping of samples from the two groups based on global protein abundance profiles. Violin plots show the average of protein abundance of the ND-LC and NU-LC Groups. P values were calculated from a paired two-tailed Student's t-test. *** $P < 0.001$, ** $P < 0.01$, * $P < 0.05$. ns: no significant difference.

Supplementary Material 3. Violin plots illustrating the gene expression levels of CEACAM6, BPI, CTSG, CRISP3, and ABCA13 across different cohorts: the HC, NLC, ND-LC, NU-LC and LC Group. P values were calculated from a paired two-tailed Student's t-test. *** $P < 0.001$, ** $P < 0.01$, * $P < 0.05$. ns: no significant difference.

Supplementary Material 4.

Supplementary Material 5.

Acknowledgements

We thank all participants in the study. We appreciate TargetingOne Corporation, Ltd for technical supports. We thank for the support from Shanghai Key Laboratory of Infectious Diseases and Biosafety Emergency Response, and Key Discipline Construction Plan from Shanghai Municipal Health Commission.

Author contributions

J.A. and K.L. conceived of the experiments. K.L., J.C., J.G., H.Z., G.S., X.W. analyzed the data and wrote the paper. Experiments were performed by K.L., J.C., X.W., K.Z., Q.X., F.Z., P.W., G.Y. and Y.S. J.A., S.W. and W.Z. supervised data collection and analysis, and paper draft.

Funding

This work is supported by National Natural Science Foundation of China [82341033, 82041010 to Wenhong Zhang], Autonomous deployment project of Shanghai Sci-Tech InnoCenter for Infection and Immunity (SSIII-202305), National Key Research and Development Program of China (2022YFC2009801, 2023YFC2043501 and 2024YFC3044600), Shanghai Municipal Health Commission (GWVI-11.1-07), and Shanghai Municipal Science and Technology Major Project (HS2021SHZX001).

Availability of data and materials

The datasets during and/or analysed during the current study available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate

Sample collection and clinical laboratory examinations were approved by the Ethics Committee of Huashan hospital (ethical approval number: 2023-750), and written informed consent was obtained from all participants enrolled.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Department of Infectious Diseases, Shanghai Key Laboratory of Infectious Diseases and Biosafety Emergency Response, National Medical Center for Infectious Diseases, Huashan Hospital, Shanghai Medical College, Fudan University, Shanghai, China. ²Shanghai Sci-Tech Inno Center for Infection & Immunity, Shanghai, China. ³Institute of Infection and Health, Fudan University, Shanghai, China. ⁴Shanghai Pudong Hospital, Fudan University Pudong Medical Center, Shanghai Institute of Infectious Disease and Biosecurity, State Key Laboratory of Genetic Engineering, MOE Engineering Research Center of Gene Technology, School of Life Sciences, Fudan University, Shanghai, China. ⁵Department of Respiratory and Critical Care Medicine, Affiliated Wuxi Fifth Hospital of Jiangnan University, Wuxi Fifth People's Hospital, Wuxi, People's Republic of China.

Received: 22 March 2024 Accepted: 30 July 2024

Published online: 12 August 2024

References

- WHO. WHO COVID-19 Dashboard; 2024. <https://data.who.int/dashboards/covid19/>. Accessed 19 June 2024.
- Ballering AV, et al. Persistence of somatic symptoms after COVID-19 in the Netherlands: an observational cohort study. *Lancet* (London, England). 2022;400(10350):452–61.
- Altmann DM, Pagel C. Long covid: where are we, what does it say about our pandemic response, and where next? *BMJ*. 2023;383:2972.
- Parotto M, et al. Post-acute sequelae of COVID-19: understanding and addressing the burden of multisystem manifestations. *Lancet Respir Med*. 2023;11(8):739–54.
- National Academies of Sciences, E. and Medicine. A long COVID definition: a chronic, systemic disease state with profound consequences. Washington, DC: The National Academies Press; 2024. p. 186.
- Evans RA, Leavy OC, Richardson M, Elneima O, McAuley HJ, Shikotra A, et al. 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):761–75.
- Brannock MD, et al. Long COVID risk and pre-COVID vaccination in an EHR-based cohort study from the RECOVER program. *Nat Commun*. 2023;14(1):2914.
- Cai J, et al. A one-year follow-up study of systematic impact of long COVID symptoms among patients post SARS-CoV-2 omicron variants infection in Shanghai, China. *Emerg Microbes Infect*. 2023;12(2):2220578.

9. Zhang H, et al. 3-year outcomes of discharged survivors of COVID-19 following the SARS-CoV-2 omicron (B.1.1.529) wave in 2022 in China: a longitudinal cohort study. *Lancet Respir Med.* 2024;12(1):55–66.
10. Zhang H, et al. Data-driven identification of post-acute SARS-CoV-2 infection subphenotypes. *Nat Med.* 2023;29(1):226–35.
11. Talla A, et al. Persistent serum protein signatures define an inflammatory subcategory of long COVID. *Nat Commun.* 2023;14(1):3417.
12. Schultheiß C, et al. The IL-1 β , IL-6, and TNF cytokine triad is associated with post-acute sequelae of COVID-19. *Cell Rep Med.* 2022;3(6): 100663.
13. Woodruff MC, et al. Chronic inflammation, neutrophil activity, and autoreactivity splits long COVID. *Nat Commun.* 2023;14(1):4201.
14. Zuo W, et al. The persistence of SARS-CoV-2 in tissues and its association with long COVID symptoms: a cross-sectional cohort study in China. *Lancet Infect Dis.* 2024;24:845–55.
15. Chen B, et al. Viral persistence, reactivation, and mechanisms of long COVID. *Elife.* 2023;12: e86015.
16. Cervia-Hasler C, et al. Persistent complement dysregulation with signs of thromboinflammation in active Long Covid. *Science.* 2024;383(6680):eadg7942.
17. Pretorius E, et al. Persistent clotting protein pathology in Long COVID/ Post-Acute Sequelae of COVID-19 (PASC) is accompanied by increased levels of antiplasmin. *Cardiovasc Diabetol.* 2021;20(1):172.
18. Davis HE, et al. Long COVID: major findings, mechanisms and recommendations. *Nat Rev Microbiol.* 2023;21(3):133–46.
19. Antar AAR, et al. Long COVID brain fog and muscle pain are associated with longer time to clearance of SARS-CoV-2 RNA from the upper respiratory tract during acute infection. *Front Immunol.* 2023;14:1147549.
20. Canas LS, et al. Profiling post-COVID-19 condition across different variants of SARS-CoV-2: a prospective longitudinal study in unvaccinated wild-type, unvaccinated alpha-variant, and vaccinated delta-variant populations. *Lancet Digital Health.* 2023;5(7):e421–34.
21. Chen S, et al. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics.* 2018;34(17):i884–90.
22. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods.* 2015;12(4):357–60.
23. Roberts A, et al. Improving RNA-Seq expression estimates by correcting for fragment bias. *Genome Biol.* 2011;12(3):R22.
24. Anders S, Pyl PT, Huber W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics.* 2015;31(2):166–9.
25. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014;15(12):550.
26. Subramanian A, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A.* 2005;102(43):15545–50.
27. Mootha VK, et al. PGC-1 α -responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet.* 2003;34(3):267–73.
28. Georgieva E, et al. COVID-19 complications: oxidative stress, inflammation, and mitochondrial and endothelial dysfunction. *Int J Mol Sci.* 2023;24(19):14876.
29. Stufano A, et al. Oxidative damage and Post-COVID syndrome: a cross-sectional study in a cohort of Italian workers. *Int J Mol Sci.* 2023;24(8):7445.
30. Al-Hakeim HK, et al. Long-COVID post-viral chronic fatigue and affective symptoms are associated with oxidative damage, lowered antioxidant defenses and inflammation: a proof of concept and mechanism study. *Mol Psychiatry.* 2023;28(2):564–78.
31. Paul BD, et al. Redox imbalance links COVID-19 and myalgic encephalomyelitis/chronic fatigue syndrome. *Proc Natl Acad Sci U S A.* 2021;118(34): e2024358118.
32. Piehler AP, Ozcürümez M, Kaminski WE. A-subclass ATP-binding cassette proteins in brain lipid homeostasis and neurodegeneration. *Front Psychiatry.* 2012;3:17.
33. Knight HM, et al. A cytogenetic abnormality and rare coding variants identify ABCA13 as a candidate gene in schizophrenia, bipolar disorder, and depression. *Am J Hum Genet.* 2009;85(6):833–46.
34. Katzeff JS, et al. ATP-binding cassette transporter expression is widely dysregulated in frontotemporal dementia with TDP-43 inclusions. *Front Mol Neurosci.* 2022;15:1043127.
35. Dwyer S, et al. Investigation of rare non-synonymous variants at ABCA13 in schizophrenia and bipolar disorder. *Mol Psychiatry.* 2011;16(8):790–1.
36. Skubitz KM, Campbell KD, Skubitz AP. CD66a, CD66b, CD66c, and CD66d each independently stimulate neutrophils. *J Leukoc Biol.* 1996;60(1):106–17.
37. Banerjee A, et al. RNA-Seq analysis of peripheral blood mononuclear cells reveals unique transcriptional signatures associated with disease progression in dengue patients. *Transl Res.* 2017;186:62–78.e9.
38. Kooistra EJ, et al. Molecular mechanisms and treatment responses of pulmonary fibrosis in severe COVID-19. *Respir Res.* 2023;24(1):196.
39. Korkmaz B, Moreau T, Gauthier F. Neutrophil elastase, proteinase 3 and cathepsin G: physicochemical properties, activity and physiopathological functions. *Biochimie.* 2008;90(2):227–42.
40. Akgun E, et al. Proteins associated with neutrophil degranulation are upregulated in nasopharyngeal swabs from SARS-CoV-2 patients. *PLoS ONE.* 2020;15(10): e0240012.
41. Canny G, Levy O. Bactericidal/permeability-increasing protein (BPI) and BPI homologs at mucosal sites. *Trends Immunol.* 2008;29(11):541–7.
42. Craddock V, 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(2): e28568.
43. Su Y, et al. Multiple early factors anticipate post-acute COVID-19 sequelae. *Cell.* 2022;185(5):881–895.e20.
44. Zhang L, et al. Reverse-transcribed SARS-CoV-2 RNA can integrate into the genome of cultured human cells and can be expressed in patient-derived tissues. *Proc Natl Acad Sci U S A.* 2021;118(21): e2105968118.
45. Smits N, et al. No evidence of human genome integration of SARS-CoV-2 found by long-read DNA sequencing. *Cell Rep.* 2021;36(7): 109530.
46. Suzuki R, et al. Attenuated fusogenicity and pathogenicity of SARS-CoV-2 Omicron variant. *Nature.* 2022;603(7902):700–5.
47. Chen DY, et al. Spike and nsp6 are key determinants of SARS-CoV-2 Omicron BA.1 attenuation. *Nature.* 2023;615(7950):143–50.
48. Phetsouphanh C, et al. Immunological dysfunction persists for 8 months following initial mild-to-moderate SARS-CoV-2 infection. *Nat Immunol.* 2022;23(2):210–6.
49. Scott NA, et al. Monocyte migration profiles define disease severity in acute COVID-19 and unique features of long COVID. *Eur Respir J.* 2023;61(5):2202226.
50. Lam ICH, et al. Long-term post-acute sequelae of COVID-19 infection: a retrospective, multi-database cohort study in Hong Kong and the UK. *EClinicalMedicine.* 2023;60: 102000.
51. Huang C, et al. 6-month consequences of COVID-19 in patients discharged from hospital: a cohort study. *Lancet.* 2023;401(10393):e21–33.
52. Santa Cruz A, et al. Post-acute sequelae of COVID-19 is characterized by diminished peripheral CD8(+) β 7 integrin(+) β T cells and anti-SARS-CoV-2 IgA response. *Nat Commun.* 2023;14(1):1772.
53. Mohandas S, et al. Immune mechanisms underlying COVID-19 pathology and post-acute sequelae of SARS-CoV-2 infection (PASC). *Elife.* 2023;12: e86014.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.