Abnormal Upregulation of Cardiovascular Disease Biomarker PLA2G7 Induced by Proinflammatory Macrophages in COVID-19 patients

Yang LI^{1#}, Yongzhong JIANG^{2#}, Yi ZHANG^{1#}, Naizhe LI^{1#}, Qiangling YIN¹, Linlin LIU², Xin LV³, Yan LIU⁴, Aqian LI¹, Bin FANG², Jiajia LI⁵, Hengping YE⁶, Gang YANG⁷, Xiaoxian CUI⁸, Yang LIU¹, Yuanyuan QU¹, Chuan LI¹, Jiandong LI¹, Dexin LI¹, Shiwen WANG^{1,9*}, Zhongtao GAI^{3*}, Faxian ZHAN^{2*}, Mifang LIANG^{1,9*}

¹ NHC Key Laboratory of Medical Virology and Viral Diseases, National Institute for Viral Disease Control and Prevention, Chinese Center for Disease Control and Prevention, Beijing 102206, China

² Hubei Provincial Center for Disease Control and Prevention, Wuhan 430065, China

³ Qilu Children's Hospital, Cheeloo College of Medicine, Shandong University & Jinan Children's Hospital, Jinan 250022, China

⁴ Department of Microbiology, School of Basic Medical Science, Anhui Medical University, Hefei 230032, China

⁵ The Center for Scientific Research of the First Affiliated Hospital of Anhui Medical University, Hefei 230022, China

⁶ Xiantao Center for Disease Control and Prevention, Xiantao 433000, China

⁷ Xiangyang Center for Disease Control and Prevention, Xiangyang 441000, China

⁸ Shanghai Municipal Center for Disease Control and Prevention, Shanghai 200336, China

⁹ CDC-WIV Joint Research Center for Emerging Diseases and Biosafety, Wuhan, 430071, China.

[#] Yang LI, Yongzhong JIANG, Yi ZHANG and Naizhe LI contributed equally to the article.

*Correspondence:

Prof. Dr. Shiwen WANG E-mail: wangsw@ivdc.chinacdc.cn

Prof. Dr. Zhongtao GAI E-mail: gaizhongtao@sina.com

Prof. Dr. Fa-Xian Zhan Email: zhanfx@163.com

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

Prof. Dr. Mifang LIANG E-mail: liangmf@ivdc.chinacdc.cn

Abstract

BACKGROUND. Coronavirus disease 2019 (COVID-19) triggers distinct patterns of pneumonia progression with multiorgan disease, calling for cell- and/or tissue-type specific host injury markers.

METHODS. An integrated hypothesis-free single biomarker analysis framework was performed on nasal swabs (n=484) from patients with COVID-19 in GSE152075. The origin of candidate biomarker was assessed in single-cell RNA data (GSE145926). The candidate biomarker was validated in a cross-sectional cohort (n=564) at both nucleotide and protein levels.

RESULTS. Phospholipase A2 group VII (PLA2G7) was identified as a candidate biomarker in COVID-19. PLA2G7 was predominantly expressed by proinflammatory macrophages in lungs emerging with progression of COVID-19. In the validation stage, PLA2G7 was found in patients with COVID-19 and pneumonia, especially in severe pneumonia, rather than patients suffered mild H1N1 influenza infection. Up to 100% positive rates of PLA2G7 were positively correlated with not only viral loads in patients with COVID-19 but also severity of pneumonia in non-COVID-19 patients. Although Ct values of PLA2G7 in severe pneumonia was significantly lower than that in moderate pneumonia ($P=7.2e-11$), no differences were observed in moderate pneumonia with COVID-19 between severe pneumonia without COVID-19 ($P=0.81$). Serum protein levels of PLA2G7, also known as lipoprotein-associated phospholipase A2 (Lp-PLA₂), were further found to be elevated and beyond the upper limit of normal in patients with COVID-19, especially among the re-positive patients.

CONCLUSIONS. We firstly identified and validated PLA2G7, a biomarker for cardiovascular diseases (CVDs), was abnormally enhanced in COVID-19 patients at both nucleotide and protein aspects. These findings provided indications into the prevalence of cardiovascular involvements seen in COVID-19 patients. PLA2G7 could be a hallmark of COVID-19 for monitoring disease progress and therapeutic response.

FUNDING. This study was supported by grants from China Mega-Projects for Infectious Disease (2018ZX10711001), National Natural Science Foundation of China (82041023).

66 Introduction

67 Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is a novel enveloped
68 RNA betacoronavirus that emerged in December 2019 in Wuhan, China, and is the
69 causative etiology of coronavirus disease 2019 (COVID-19)(1, 2). Typical clinical
70 presentation of COVID-19 was a lung involvement, as evidenced by image tests, with
71 fever, cough and dyspnea(3, 4). Severe cases often developed acute respiratory distress
72 syndrome (ARDS) or even death(3, 4). The risk factors for increased disease severity
73 in patients with COVID-19 has been reported as older age (e.g., over 50 years old) and
74 the presence of comorbidities, including hypertension, diabetes mellitus, cardiovascular
75 disease etc.(5, 6). However, it rapidly became obvious that severe COVID-19 can also
76 occur in younger patients with no pre-existing comorbidities(7).

77
78 On-going data characterizing immunological features in patients with COVID-19 were
79 starting to emerge. Host response to SARS-CoV-2 infection was distinct in comparison
80 with other highly pathogenic coronaviruses and common respiratory viruses in cell
81 lines(8). Notably, reduced type I interferon activities were observed in both in vitro(8)
82 and in vivo(9) data. Delayed production of type I interferon resulted in boosted
83 cytopathic effects (CPE) and increased sensing of SARS-COV-2 threats promoted the
84 enhanced release of monocyte chemoattractants, such as CCL2(10), which contributed
85 an influx of monocytes into lungs(7). Autopsy studies have shown diffuse thickening
86 of the alveolar wall with mononuclear cells and macrophages infiltrating airspaces(11).
87 However, it is still mysterious why there were various complications of COVID-19,
88 including impaired function of the organs, especially heart(7). Among 1,216
89 hospitalized COVID-19 patients across 69 countries who did not have pre-existing
90 heart complications, almost half showed scan abnormalities that resemble the early
91 stages of heart failure(12).

92
93 Hypothesis-free biomarker studies could allow researchers to gain in depth insights into
94 patho-mechanisms underlying COVID-19. Several attempts have been made.
95 Expression of monocyte CD169 (mCD169), also known as sialic acid binding Ig like
96 lectin 1 (SIGLEC1), has been suggested as a biomarker in the early diagnosis of
97 COVID-19(13). But expression of SIGLEC1 could be enhanced by other viruses(14).

On the other hand, protein-level biomarkers that could predict the severity of COVID-19 patients were identified(15, 16). In particular, the level of lactate dehydrogenase (LDH) in the blood which was used to monitor the tissue damage was highly indicative of COVID-19 mortality with area under curve (AUC) > 0.9(15). These results suggested the development of additional cell- and/or tissue-type specific host injury markers were needed(17). Here, we started to employ the previously established integrated hypothesis-free single biomarker analysis framework(18) on bulk transcriptomic data (n=484) from nasal swabs retrieved from COVID-19 patients and controls. To further understand the contribution to COVID-19 of this gene, single cell RNA (scRNA) data from bronchoalveolar lavage fluids (BALFs) of moderate and severe COVID-19 patients were also explored. Thereafter, more than 500 clinical samples, including nasal swabs and serum samples, were collected to validate our findings at nucleotide and protein levels.

Results

In silico discovery of PLA2G7 as the biomarker of SARS-COV-2 infection

The clinical records in GSE152075 labelled as ‘not collected’ were removed. Batch-corrected count matrix was normalized by varianceStabilizingTransformation (vst) function in R package DEseq2(19). Following cluster analysis on the preprocessed expression matrix to remove outliers, in a total of 428 samples out of 484 were remaining (Supplementary Figure S1). As the scale-free topology fit index failed to reach values above 0.85 (Default Threshold), the soft-thresholding power of 12 was selected(20) (Supplementary Figure S2). In a total number of 8 modules were identified and constructed by WGCNA(21) analysis (Figure 1A; Supplementary Figure S3). To identify the SARS-COV-2 infection-related module, the Pearson correlation analysis, which involved calculating the Student asymptotic *P*-values for the correlations, between the module eigengenes (MEs) of each module and clinical traits was performed (Figure 1B). The red module was the module most relevant to SARS-COV-2 infection. Genes in red module (n = 60) were listed in Supplementary Table 1. Importantly, the disease ontology (DO)(22) analysis on these genes suggested not only upper respiratory tract disease, but also heart disease related terms (e.g., congestive heart failure and myocardial infarction) (Figure 1C). In addition, the network connections among the

most connected genes in the red module was displayed through Cytoscape(23) (Figure 1D).

The genes in the red module were considered as candidate hub genes in SARS-COV-2 infection. In a total of 6 genes were selected based on a gene significance threshold of 0.8 and a module membership significance of 0.25 (Figure 1E; Supplementary Figure S4). Next, hub gene selection based on XGBoost(24) was carried out. First of all, all the samples were randomly assigned at a 7:3 ratio to a training set (299 samples) and a test set (129 samples). The python packages “XGBoost”(24) and “scikit-learn”(25) were used for classification. Second, to obtain the best XGBoost model parameter combination (eta, gamma, max_depth, Min_child_weight, Subsample, and n_Colsample_bytree) with the highest classification accuracy, five-fold cross-validation and grid search were applied to the training set. Finally, the highest accuracy of classification was 0.876 which could be achieved through a single gene, phospholipase A2 group VII (PLA2G7) (Supplementary Figure S5). Moreover, the AUC score by R package “pROC”(26) in the training and test sets on this single gene was 0.850 and 0.820, respectively (Supplementary Figure S6). It should be noted that PLA2G7 was significantly upregulated in SARS-COV-2 positive group (logFC = 5.903, FDR P -value = $2.68e-40$) (Figure 1F).

PLA2G7 was induced by proinflammatory macrophages emerging along with progression of COVID-19

Clustering analysis on scRNA-seq data (GSE145926) showed 27 distinct clusters composed of macrophages, neutrophils, myeloid dendritic cells (mDCs), plasmacytoid dendritic cells (pDCs), natural killer (NK) cells, T cells, B cells, plasma cells and epithelial cells (Figure 2A, Supplementary Figure S7-8). The gene expression analysis showed PLA2G7 was expressed principally by macrophages (Figure 2A-B). To further understand the origin of PLA2G7, the 15 macrophages subclusters were re-integrated and re-clustered (Supplementary Figure S9). The expression patterns of PLA2G7, along with typical classification markers of macrophages (FCN1, SPP1 and FABP4)(27, 28), were then explored (Figure 2C). Thereafter, four heterogeneous subgroups of macrophages were identified, FCN1^{hi} group, FCN1^{lo}SPP1⁺ group, SPP1⁺ group and FABP4⁺ group, those were consistent with previous study(27)(Supplementary Figure

S10-11). It was worth of noting that the expression profile of PLA2G7 shared a close pattern to SPP1 (Figure 2D), which played an important role in idiopathic pulmonary fibrosis(28). Macrophages with FABP4, also known as the alveolar macrophages, were reduced with progression of COVID-19 while macrophages with FCN1 and SPP1, considered as the proinflammatory macrophages, were strikingly increased (Figure 2E). In addition, strong correlation between expression of PLA2G7 and progression of COVID-19 was discovered ($r = 0.927$, $P = 1.4e-4$) (Figure 2F). On the whole, PLA2G7 was predominantly expressed by these proinflammatory macrophages emerging along with progression of COVID-19.

PLA2G7 was detected in a retrospective collection of nasal swabs from both COVID-19 and pneumonia patients

In a total of 447 nasal swabs were collected and subject to SARS-COV-2 targeted qRT-PCR. The 447 nasal swabs were grouped as: Health controls ($n = 200$), COVID-19 group ($n = 134$), Influenza infection ($n = 52$), Severe pneumonia ($n = 41$) and Moderate pneumonia ($n = 20$). It should be noted the median (interquartile range [IQR]) of Ct of COVID-19 group in present study were relatively higher than that in public dataset GSE152075 (SARS-COV-2 Ct: 31.66 [28.62 – 34.00] vs. 21.31 [19.09 – 24.01]) (Figure 3A). To match the records in GSE152075, the COVID-19 group were divided into two groups with Ct value ≤ 25 and >25 . Thereafter, we next explored the relationship between positive perdition value (PPV) of PLA2G7 and Ct values of SARS-COV-2 in COVID-19 group. The PPV of PLA2G7 was correlated with Ct of SARS-COV-2 ($r = -0.96$, $p = 3.3e-12$) (Figure 3B), indicated that the PPV of PLA2G7 was positively correlated with SRAS-CoV-2 viral load. It was worth of noting more than 80% PLA2G7 positive rate were observed in COVID-19 patients with Ct of SARS-COV-2 ≤ 25 (Figure 3B-C), which showed comparable diagnostic performance in GSE152075 as shown in Figure 1E.

The positive rates of PLA2G7 in all different groups were shown in Figure 3C. There were no PLA2G7 positive cases in health controls. Among the 134 SARS-COV-2 positive cases, 47 (35.1%) were PLA2G7 positive. It was worth of noting more than 90% PLA2G7 positive rate were observed in COVID-19 patients with Ct of SARS-COV-2 ≤ 25 while 29.37 % were seen in the group with Ct of SARS-COV-2 >25 (Figure

3B-C). Surprisingly, a similar high detection rate of PLA2G7 was seen in patients with pneumonia (SRAS-CoV-2 negative). Notably, 100.00% of PLA2G7 positive rate were seen in group of severe pneumonia while 65.00% were noticed in moderate pneumonia patients. Conversely, only about 20.00% PLA2G7 positive rate were observed in patients with influenza infection. The Ct of PLA2G7 provided further insights (Figure 3D). The Ct values of PLA2G7 in group of SARS-COV-2 ($Ct \leq 25$) were significantly lower than that in other groups. In addition, records in group with severe pneumonia showed higher levels of PLA2G7 than that in group with moderate pneumonia group ($P=7.2e-11$) (Figure 3D; Supplementary Figure S12). It was worried to state that there were no significant differences between the Ct values of PLA2G7 in group of COVID-19 and that in severe pneumonia group (Figure 3D; Supplementary Figure S12).

Abnormal serum protein levels of Lp-PLA₂ in COVID-19 patients

PLA2G7 was also known as lipoprotein-associated phospholipase A₂ (Lp-PLA₂) in plasma(29). To understand the plasma levels of Lp-PLA₂, a total of 117 serum samples (36 health controls, 81 COVID-19 patients) were collected. Among the 81 COVID-19 patients, 47 serum samples were collected at various time points from patients during the hospitalization (Hospitalization Group), 19 were from the patients who had been discharged and then re-positive but asymptomatic (Re-positive Group), and 15 were from fully recovered individuals (Recovered Group). Compared with the health controls, serum levels of Lp-PLA₂ were higher in hospitalization and re-positive patients (Figure 4A). It was worth of noting that the hospitalization group showed significantly less levels of Lp-PLA₂ than that in re-positive group ($P=0.004$). Thereafter, we sought to explore the trend of Lp-PLA₂ along with time. The overall trend of Lp-PLA₂ in hospitalization group was reduced along with the days from initial symptoms onset (Figure 4B). To our surprise, higher plasma levels of Lp-PLA₂ was in moderate than in severe patients (Figure 4B). In addition, the serum levels of Lp-PLA₂ showed discriminating power between patients with COVID-19 and health individuals with AUC of 0.815 (95%CI 0.734 – 0.895) (Figure 4C).

Discussion

Given that the morbidity and mortality seen in COVID-19, a better understanding of the immunological underpinnings seen in patients infected with SARS-CoV-2 is

necessary to better identify biomarkers, especially these represented the therapeutic and diagnostic targets. It has been suggested that cytokine blood RNA level was not always correlated with the protein plasma level(9). We therefore applied our previous established hypothesis-free single biomarker analysis framework on nasal swabs from more than 400 donors with COVID-19 in GSE152075. Finally, upregulated PLA2G7, a biomarker for cardiovascular diseases (CVDs)(29), was identified as a hub gene in SARS-COV-2 infection (Figure 1). Coincidentally, phospholipase A2 Group IID (PLA2G2D), belonged to the same PLA₂ superfamily as PLA2G7 did, was considered as an increase factor to result in worse outcomes of mice infected by SARS-CoV(30). In addition, a 12-year follow-up survey of 25 patients who recovered from SARS-CoV infection found that 44% had cardiovascular system abnormalities(31) while almost half of hospitalized COVID-19 patients showed scan abnormalities that resemble the early stages of heart failure(12).

To understand the origin of PLA2G7, we re-analyzed scRNA-seq data from COVID-19 patients BALFs in GSE145926. We presented evidence-based findings that PLA2G7 was predominantly expressed by macrophages, especially the proinflammatory macrophages, including FCN1^{hi} group, FCN1^{lo}SPP1⁺ group and SPP1⁺ group (Figure 2). These proinflammatory macrophages were emerging with progression of COVID-19 suggesting PLA2G7 could be used to monitor the progress of COVID-19 (Figure 2E-F). Previous studies suggested there was no mRNA of PLA2G7 in monocytes, but expression was induced and maintained during differentiation into macrophages(32). Therefore, our results might provide insights to the hypothesis that dysregulated activation of the mononuclear phagocyte (MNP) compartment contributes to COVID-19-associated hyperinflammation(33). Hence, the PLA2G7 which was specifically induced by proinflammatory monocyte-driven macrophages in lungs was seen in patients suffered from COVID-19.

We next validated the expression of PLA2G7 in more than 400 nasal samples by qRT-PCR. Comparing to the diagnostic performance of PLA2G7 in public dataset (Figure 1E), the PPV of PLA2G7 in SARS-COV-2 positive samples was relatively low. This discrepancy might be due to the relatively low viral load in present cohort (Figure 3A). It was noted that PPV of PLA2G7 in SARS-COV-2 positive samples were over 0.8 when the Ct of SARS-COV-2 was lower than 25 (Figure 3B). As proinflammatory

macrophages in lungs played central roles in pneumonia(34), we postulated that PLA2G7 could be detected in not only patients with COVID-19, but also patients with pneumonia. The positive rate of PLA2G7 in nasal swabs from donors with pneumonia was higher than that in health controls and patients suffered seasonal influenza infection (Figure 3C). It was worth of noting that 100% positive rate of PLA2G7 were observed in severe pneumonia. In addition, the Ct values of PLA2G7 were significantly lower in severe pneumonia than in moderate pneumonia ($p = 7.2e-11$). This was consistent with our findings that expression of PLA2G7 was correlated with progression of COVID-19 (Figure 2F), suggesting PLA2G7 could be also used to monitor the progress of pneumonia. On-going epidemiological evidence substantiated the correlation between pneumonia and subsequent CVDs(35). Up to 30% of patients hospitalized for Community-acquired pneumonia (CAP) developed CVDs(36). Moreover, pneumonia has been suggested as a risk factor of CVDs(37). However, the patho-mechanisms by which pneumonia may trigger or promote subsequent CVDs still remain unclear. The expression profile of PLA2G7 was observed in patients with pneumonia could help to understand patho-mechanisms underlying pneumonia and CVDs. On the other hand, Ct values of PLA2G7 in group COVID-19 patients with high viral loads (SARS-COV-2 $Ct \leq 25$) were significantly lower than that in the rest groups (Figure 3D). Importantly, no significant differences between the Ct values of PLA2G7 in group of COVID-19 patients with SARS-COV-2 $Ct > 25$ and that in severe pneumonia group. These data might provide an indication that more than half of COVID-19 hospitalized patients developed conditions with CVDs(12).

The abnormal levels of plasma protein of Lp-PLA₂ have been seen in COVID-19 patients, especially the non-hospitalization patients (Figure 4). First of all, the aberrant plasma levels of Lp-PLA₂ were observed in COVID-19 patients from not only hospitalization group, but also re-positive group (Figure 4A). Strikingly, the plasma level of Lp-PLA₂ in hospitalization group was significantly lower than that in re-positive group ($p = 0.004$) (Figure 4A). It was necessary to point out all the patients in re-positive group were asymptomatic. Therefore, the re-positive patients who had great potential risks on CVDs might, however, receive limited medical attention on CVDs. In addition, although there was no significant difference ($p = 0.16$) in the plasma level of Lp-PLA₂ between fully recovered COVID-19 patients and health controls, it was

worried to state that the mass of Lp-PLA₂ in more than half (53.3%) of recovered patients (n = 15) was beyond the upper limit of normal (Figure 4A). Coincidentally, a recent study showed 78% of recovered COVID-19 patients, including illness ranging from asymptomatic to moderate symptoms, suffered issues association with the heart(38). Secondly, although the Lp-PLA₂ was reduced along with time in hospitalized COVID-19 patients, it should be noted that the plasma levels of Lp-PLA₂ in these patients were beyond the upper limit of normal, especially the moderate patients (Figure 4B). According to National Health Commission of China (NHC), some of the patients (later COVID-19 confirmed cases) first went to see a doctor because of cardiovascular symptoms, rather than respiratory symptoms(39). Even without symptoms and signs of interstitial pneumonia, cardiac involvement as a complication associated with COVID-19 has been reported⁽⁴⁰⁾. Compare to severe group, the moderate patients who always presented fewer and/or lighter symptoms of CVDs led to limited medical care on CVDs. This might initiate follow-up comorbidities with heart in these patients.

To our knowledge, this is the first study that uncover the relationship between PLA2G7 and COVID-19 from both nucleotide and protein aspects. The results of our study provide important indications into the prevalence of cardiovascular involvements seen in COVID-19 patients, including status ranging from hospitalization to recovery period. In addition, our findings reveal the aberrant profile of PLA2G7 persists beyond the period of acute manifestation, especially in the discharged and re-positive patients. Importantly, the expression profile of PLA2G7 in COVID-19 patients with moderate pneumonia shows similar patterns seen in severe pneumonia, suggesting COVID-19 might enhance the expression of PLA2G7 (Figure 3). Although the long-term health effects of role of PLA2G7 in COVID-19 cannot yet be determined, on-going epidemiological evidence substantiated heart issues were occurred after COVID-19(41). Thus, our findings may provide an indication of potentially considerable burden of CVDs in large and growing parts of the population and urgently require confirmation in a larger cohort.

Nevertheless, our study has certain limitations. The expression of PLA2G7 was not validated in nasal swabs from the patients with severe COVID-19. In addition, the plasma levels of Lp-PLA₂ in pneumonia patients were also not tested. Due to limited

information of clinical traits of samples, confounding factors, such as pre-existing CVDs, which could increase the likelihood of PLA2G7 positivities, were not controlled. The in-house qRT-PCR assay of PLA2G7 was not optimized, leading to inadequate diagnostic performance. Most of the samples were retrospectively collected in Wuhan, China. It has been almost several months passed since all the samples were collected. The freshness of samples could also undermine the performance of the in-house PLA2G7 targeted qRT-PCR and commercial Lp-PLA₂ ELISA toolkit.

Taken together, we firstly identified proinflammatory monocytes-driven macrophages specific PLA2G7 in lungs, a typically biomarker of CVDs, which led to abnormal levels of plasma Lp-PLA₂, was enhanced in COVID-19. Importantly, PLA2G7 was detected in nasal swabs from patients with pneumonia. Although levels of PLA2G7 was significantly elevated in severe pneumonia comparing to moderate pneumonia, there were no significant differences between that in COVID-19 and severe pneumonia. Moreover, increased levels of Lp-PLA₂ in plasma could provide insights to higher mortality was seen in patients underlying comorbidities (e.g. hypertension, diabetes mellitus, cardiovascular disease) and severe COVID-19 could occur in patients without pre-existing complications. Nevertheless, the role of PLA2G7 in the pathogenesis of COVID-19 worth cautious tests and studies.

Materials and Methods

Data collection

In brief, data were obtained from the Gene Expression Omnibus (GEO) database (<http://www.ncbi.nlm.nih.gov/geo/>) in July 2020 using the keyword “SARS-COV-2”. The following exclusion criteria were applied to the expression profiling by high throughput sequencing: (1) concerned only cell model; (2) no or insufficient clinical data; and (3) used non-baseline (“healthy”) controls. After review, GSE152075, which contained 484 samples of nasal swabs, was selected for biomarker discovery. Among them, in a total of 430 subjects were positive with SARS-COV-2 infections while 54 were negative. According to batch information, the batch effect hidden in the raw count matrix of GSE152075 was removing by function `Combat_seq` in the R package `sva`(42).

Next, GSE145926 which included single-cell RNA sequencing (scRNA-Seq) data on bronchoalveolar lavage fluid (BALF) cells from three patients with moderate COVID-19, six patients with severe/critical infection and three healthy controls were also retrieved(27).

Differentially expressed genes screening

When the batch-corrected count matrix of GSE152075 was compared in SARS-COV-2 positive cases with negative cases, DESeq2(19) was applied for differential expressed genes (DEGs). Thereafter, correction for multiple testing was addressed by controlling the false discovery rate (FDR) using the Benjamini and Hochberg (B.H.) method for both above. Criteria for DEGs were an absolute $\log_2$ fold change ($\log_2FC$) of 0 and the FDR-adjusted P -value of < 0.05 .

Co-expression network construction

A co-expression network was constructed using the normalized GSE152075 data by the weighted correlation network analysis (WGCNA) in R (21). Briefly, quality assessment of GSE152075 samples was conducted using the cluster method. After removing outliers, the soft-thresholding power was then calculated, with the type of network set to signed. Network construction was then performed with correlation coefficient threshold was 0.85 (Default setting). In addition, the minimum number of genes in each module was 30 and the threshold for cut height was set to 0.25 to merge possible similar modules.

Identification of modules related to SARS-COV-2 infection

For a given module, the expression profile was summarized into a single characteristic expression profile, designated module eigengenes (MEs). MEs were considered as the first principal component in the principal component analysis (PCA). Thereafter, a Pearson correlation analysis, calculating the student asymptotic P -values for the correlations, between MEs and clinical traits (SARS-COV-2 negative, SARS-COV-2 positive, Gender and Age) was conducted.

Disease Ontology analyses

To understand the functions of enriched genes in interesting modules, Disease Ontology

(DO)(22) analyses were performed using clusterProfiler(43), identifying significant results based on a P -value < 0.05 and gene counts ≥ 2 .

Candidate single gene biomarker selection

The module that was most highly correlated with SARS-COV-2 infection was selected. Candidate genes in the module were determined by both gene significance and module membership. Thereafter, a single gene biomarker was selected using XGBoost with recursive feature elimination with cross-validation (RFECV) (24, 25).

Sample integration, dimensionality reduction and clustering on scRNA-data

Count matrixes from 10X CellRanger hdf5 in GSE145926 were obtained. The following criteria were then employed to filter low-quality cells: gene number between 200 and 6,000, unique molecular identifier (UMI) count $> 1,000$ and mitochondrial gene percentage < 0.1 . Thereafter, the gene-cell matrixes of all samples were integrated with Seurat (version: 3) to remove batch effects across different donors. Next, the integrated gene-cell matrix was normalized using 'LogNormalize' methods in Seurat(44) with default parameters. The top 2,000 variable genes were then identified using the 'vst' method in Seurat FindVariableFeatures function. Variables 'nCount_RNA' and 'percent.mito' were regressed out in the scaling step. PCA was performed using the top 2,000 variable genes. Then UMAP (Uniform Manifold Approximation and Projection) and tSNE (t-distributed Stochastic Neighbor Embedding) were performed on the top 50 principal components (PCs) for visualizing the cells. Graph-based clustering was performed on the PCA-reduced data for clustering analysis with Seurat v.3. The resolution was set to 1.2 to obtain a finer result(27). Briefly, the first 50 PCs of the integrated gene-cell matrix were used to construct a shared nearest-neighbor graph (SNN; FindNeighbors() in Seurat) and this SNN was used to cluster the dataset (FindClusters()) using a graph-based modularity-optimization algorithm of the Louvain method for community detection.

Macrophages re-integration

Macrophages of all samples were re-integrated using the first 50 dimensions of canonical correlation analysis and PCA with the parameter k.filter was set to 115. In the clustering step, parameter resolution was set to 0.8.

Retrospective collection of clinical samples and case definition

We enrolled nasal swabs from donors with health condition and diagnosis of COVID-19, influenza infection and pneumonia. In a total of 447 nasal swabs were collected. Among these samples, 200 were from health individuals; 134 moderate pneumonia cases of confirmed diagnosis of COVID-19 (SARS-COV-2 positive) and 20 suspected COVID-19 cases (SARS-COV-2 negative) with moderate pneumonia were collected from government authorized hospitals for COVID-19 in Wuhan during the outbreak of COVID-19 (January 2020 to April 2020); 52 cases with H1N1 infection were also included in Wuhan from October to November, 2019 while 41 cases from hospitalized children who were diagnosed as severe pneumonia (SARS-Cov2 negative) were collected from Qilu Children's Hospital, Jinan, Shandong in February, 2020. For the 117 serum samples, in a number of 81 serum samples from patients with COVID-19 and 36 from health donors were also collected from government authorized hospitals for COVID-19 in Wuhan during the outbreak of COVID-19. Among the 81 COVID-19 patients, 47 serum samples were collected at various time points from patients during the hospitalization, 19 were from the patients who had been discharged and then re-positive but asymptomatic, and 15 were from fully recovered individuals. There were 34 moderate cases, 8 severe cases and 5 critical cases in the 47 serum samples from hospitalized patients. Disease severity of the COVID-19 patients were defined as moderate, severe and critical based on the Interim Guidance for Novel Coronavirus Pneumonia (Trial Implementation of Seven Edition) by the National Health Commission of China issued on 3 March 2020 (<http://www.nhc.gov.cn/yzygj/s7653p/202003/46c9294a7dfe4cef80dc7f5912eb1989.shtml>).

Human studies

The research protocol was approved by the human bioethics committee of the Chinese Center for Disease Control and Prevention, and all participants provided written informed consent.

qRT-PCR assay for SARS-COV-2 and PLA2G7

All the nasal swabs were sent to BSL-3 lab in Hubei CDC with the standard operational

procedure (SOP) of the biosafety laboratory to re-test the SARS-COV-2 RNA before detection of PLA2G7. Total nucleic acid was extracted using the QIAamp viral RNA mini kit (Qiagen) and SARS-COV-2 targeted qRT-PCR was performed using a China Food and Drug Administration-approved commercial kit specific for SARS-CoV-2 detection (Da An Gene). This kit is based on one-step TaqMan qRT-PCR technique, and ORF1ab and N genes are selected as amplification target regions. Briefly, the 25-μL reaction mixture comprised 17 μL of PCR reaction Solution A, 3 μL of PCR reaction Solution B, and 5 μL of RNA extracts. The cycling conditions were as follows: 50 °C for 15 min, 95 °C for 15 min, followed by 45 cycles for 15 s at 94 °C and 45 s at 55 °C. Each qRT-PCR assay provided a cycle threshold (Ct) value, indicating the number of cycles surpassing the threshold for a positive test. The Ct value for a positive specimen was set at 40 cycles.

To study PLA2G7 transcripts expression, the primers and TaqMan probes selected from the Universal ProbeLibrary Assay Design Center (Roche Diagnostics) were applied. In briefly, the TaqMan qRT-PCR was carried out in AgPath-ID™ One-Step RT-PCR Reagents (Applied Biosystems) using a CFX96™ Real-Time PCR Detection System (Bio-Rad). Cycling conditions were 45°C for 10 min and 95°C for 10 min, followed by 40 cycles of 95°C for 15 s, 58°C for 45 s. At least three replicates were studied for each sample. The PLA2G7 expression levels were normalized to RHOG, and the relative expression was calculated as $2^{-\Delta\Delta C_t}$.

Measurement of Plasma levels of Lp-PLA2

The plasma levels of Lp-PLA₂ were detected by enzyme linked immunosorbent assay (ELISA) according to the instruction (Hotgen, registration no. 20192400366). In brief, serum was heat-inactivated at 56°C for 45 min. Each sample (20 μL) and standard were added to a 96-well plate. A total of 100 μl of enzyme-labeled reagent was added to all wells (except the blank well) and incubated at 37 °C for 1 h. The plate was washed 5 times before chromogenic reagents A and B (50 μl of each) were mixed and added to each well. The plate was incubated in the dark at 37°C for 10 min before 50 μl stop solution was added to each well to terminate the color reaction. The optical density (OD) 450 nm values in each well were determined by a microplate reader. The levels of Lp-PLA₂ were then calculated according to the standard curve. The upper limit of normal

level of Lp-PLA₂ were 175 ng/ml.

Statistical analysis

R (version 4.0.2) was used for most analyses, with hub gene selection being performed using XGBoost and Scikit in Python (version 3.6). Differences of median percentage among in group were compared using a Student's t-test (one-sided [less], unadjusted for multiple comparisons) with R ggpubr v.0.2.5. P-value ≤ 0.05 was considered statistically significant.

Conflict of Interests

The authors declared no competing interests.

Author Contributions

SW, ZG, FZ and ML conceived and designed the study and had full access to all of the data in the study. Yang L, YJ, YZ, NL, SW, ML drafted the paper. Yang L, YJ, YZ, NL, QY, Yan L, AL, JL, XC, JL, DL did the analysis. LL, XL, BF, HY, GY banked the samples. QY, Yang LIU, YQ, CL performed experiments. All authors critically revised the manuscript for important intellectual content and gave final approve for the version to be published. 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.

Funding

This study was supported by grants from China Mega-Projects for Infectious Disease (2018ZX10711001), National Natural Science Foundation of China (82041023). The funders had no role in the design, execution, or analysis of the study, nor in the preparation or approval of the manuscript.

Acknowledgments

We acknowledge all health-care workers involved in the diagnosis and treatment of patients in Wuhan and Jinan. We are grateful for all the patients for their supports and confidence in our study. We thank the researchers who obtained the sequencing data used in this study. We thank the helpful discussions with Xianbing YU from the

Chemistry Department at the University of Chicago and Chen CHEN from Beijing Ditan Hospital, Capital Medical University. We thank Yiming ZHOU and Yunke LI from Beijing Neoantigen Biotechnology Co. Ltd for their helps in single cell data analysis. We thank the valuable comments about PLA2G7 from clinical doctors, especially Guiqiang WANG from Peking University First Hospital and Jun ZHU from Fuwai Hospital, Chinese Academy of Medical Sciences.

Data Availability Statement

The datasets GSE152075 and GSE145926 for this study can be found in the Gene Expression Omnibus (GEO) database hosted by the National Center for Biotechnology Information of the US National Institutes of Health (<https://www.ncbi.nlm.nih.gov/geo/>). All data are available in the main text or the supplementary materials.

References

1. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, et al. A Novel Coronavirus from Patients with Pneumonia in China, 2019. *N Engl J Med*. 2020;382(8):727-33.
2. Lu R, Zhao X, Li J, Niu P, Yang B, Wu H, et al. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Lancet*. 2020;395(10224):565-74.
3. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet*. 2020;395(10223):497-506.
4. Chen N, Zhou M, Dong X, Qu J, Gong F, Han Y, et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *Lancet*. 2020;395(10223):507-13.
5. Richardson S, Hirsch JS, Narasimhan M, Crawford JM, McGinn T, Davidson KW, et al. Presenting Characteristics, Comorbidities, and Outcomes Among 5700 Patients Hospitalized With COVID-19 in the New York City Area. *Jama*. 2020;323(20):2052-9.
6. Docherty AB, Harrison EM, Green CA, Hardwick HE, Pius R, Norman L, et al. Features of 20 133 UK patients in hospital with covid-19 using the ISARIC

-
- 531 WHO Clinical Characterisation Protocol: prospective observational cohort
532 study. *Bmj*. 2020;369:m1985.
- 533 7. Wiersinga WJ, Rhodes A, Cheng AC, Peacock SJ, and Prescott HC.
534 Pathophysiology, Transmission, Diagnosis, and Treatment of Coronavirus
535 Disease 2019 (COVID-19): A Review. *JAMA*. 2020.
- 536 8. Blanco-Melo D, Nilsson-Payant BE, Liu WC, Uhl S, Hoagland D, Moller R, et
537 al. Imbalanced Host Response to SARS-CoV-2 Drives Development of
538 COVID-19. *Cell*. 2020;181(5):1036-45 e9.
- 539 9. Hadjadj J, Yatim N, Barnabei L, Corneau A, Boussier J, Smith N, et al. Impaired
540 type I interferon activity and inflammatory responses in severe COVID-19
541 patients. *Science*. 2020:eabc6027.
- 542 10. Channappanavar R, and Perlman S. Pathogenic human coronavirus infections:
543 causes and consequences of cytokine storm and immunopathology. *Seminars in*
544 *immunopathology*. 2017;39(5):529-39.
- 545 11. Xu Z, Shi L, Wang Y, Zhang J, Huang L, Zhang C, et al. Pathological findings
546 of COVID-19 associated with acute respiratory distress syndrome. *The Lancet*
547 *Respiratory Medicine*. 2020;8(4):420-2.
- 548 12. Dweck MR, Bularga A, Hahn RT, Bing R, Lee KK, Chapman AR, et al. Global
549 evaluation of echocardiography in patients with COVID-19. *Eur Heart J*
550 *Cardiovasc Imaging*. 2020.
- 551 13. Bedin A-S, Makinson A, Picot M-C, Mennechet F, Malergue F, Pisoni A, et al.
552 Monocyte CD169 expression as a biomarker in the early diagnosis of COVID-
553 19. *medRxiv*. 2020:2020.06.28.20141556.
- 554 14. Zheng Q, Hou J, Zhou Y, Yang Y, Xie B, and Cao X. Siglec1 suppresses antiviral
555 innate immune response by inducing TBK1 degradation via the ubiquitin ligase
556 TRIM27. *Cell Res*. 2015;25(10):1121-36.
- 557 15. Yan L, Zhang H-T, Goncalves J, Xiao Y, Wang M, Guo Y, et al. An interpretable
558 mortality prediction model for COVID-19 patients. *Nature Machine*
559 *Intelligence*. 2020;2(5):283-8.
- 560 16. Messner CB, Demichev V, Wendisch D, Michalick L, White M, Freiwald A, et
561 al. Ultra-High-Throughput Clinical Proteomics Reveals Classifiers of COVID-
562 19 Infection. *Cell Syst*. 2020.
- 563 17. De Vlaminck I. Blood-borne biomarkers may help predict COVID-19 mortality.

-
- 564 *Science Translational Medicine*. 2020;12(545):eabb7102.
- 565 18. Li Y, Liu H, Xu Q, Wu R, Zhang Y, Li N, et al. OASL as a Diagnostic Marker
566 for Influenza Infection Revealed by Integrative Bioinformatics Analysis With
567 XGBoost. *Frontiers in Bioengineering and Biotechnology*. 2020;8.
- 568 19. Love MI, Wolfgang H, and Simon A. Moderated estimation of fold change and
569 dispersion for RNA-seq data with DESeq2. *Genome Biology*. 2014.
- 570 20. Langfelder P, and Horvath S. WGCNA package FAQ.
571 [https://horvath.genetics.ucla.edu/html/CoexpressionNetwork/Rpackages/WGC](https://horvath.genetics.ucla.edu/html/CoexpressionNetwork/Rpackages/WGCNA/faq.html)
572 [NA/faq.html](https://horvath.genetics.ucla.edu/html/CoexpressionNetwork/Rpackages/WGCNA/faq.html). Accessed 7.7, 2020.
- 573 21. Langfelder P, and Horvath S. WGCNA: an R package for weighted correlation
574 network analysis. *BMC bioinformatics*. 2008;9(1):559.
- 575 22. Schriml LM, Mitra E, Munro J, Tauber B, Schor M, Nickle L, et al. Human
576 Disease Ontology 2018 update: classification, content and workflow expansion.
577 *Nucleic Acids Res*. 2019;47(D1):D955-d62.
- 578 23. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al.
579 Cytoscape: a software environment for integrated models of biomolecular
580 interaction networks. *Genome research*. 2003;13(11):2498-504.
- 581 24. Chen T, and Guestrin C. *Proceedings of the 22nd acm sigkdd international*
582 *conference on knowledge discovery and data mining*. 2016:785-94.
- 583 25. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al.
584 Scikit-learn: Machine learning in Python. *Journal of machine learning research*.
585 2011;12(Oct):2825-30.
- 586 26. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez J-C, et al. pROC:
587 an open-source package for R and S+ to analyze and compare ROC curves.
588 *BMC bioinformatics*. 2011;12(1):77.
- 589 27. Liao M, Liu Y, Yuan J, Wen Y, Xu G, Zhao J, et al. Single-cell landscape of
590 bronchoalveolar immune cells in patients with COVID-19. *Nat Med*.
591 2020;26(6):842-4.
- 592 28. Morse C, Tabib T, Sembrat J, Buschur KL, Bittar HT, Valenzi E, et al.
593 Proliferating SPP1/MERTK-expressing macrophages in idiopathic pulmonary
594 fibrosis. *Eur Respir J*. 2019;54(2).
- 595 29. Kono N, and Arai H. Platelet-activating factor acetylhydrolases: An overview
596 and update. *Biochim Biophys Acta Mol Cell Biol Lipids*. 2019;1864(6):922-31.

- 597 30. Vijay R, Hua X, Meyerholz DK, Miki Y, Yamamoto K, Gelb M, et al. Critical
598 role of phospholipase A2 group IID in age-related susceptibility to severe acute
599 respiratory syndrome-CoV infection. *J Exp Med*. 2015;212(11):1851-68.
- 600 31. Wu Q, Zhou L, Sun X, Yan Z, Hu C, Wu J, et al. Altered Lipid Metabolism in
601 Recovered SARS Patients Twelve Years after Infection. *Sci Rep*.
602 2017;7(1):9110.
- 603 32. Tjoelker LW, Wilder C, Eberhardt C, Stafforini DM, Dietsch G, Schimpf B, et
604 al. Anti-inflammatory properties of a platelet-activating factor acetylhydrolase.
605 *Nature*. 1995;374(6522):P.549-53.
- 606 33. Merad M, and Martin JC. Pathological inflammation in patients with COVID-
607 19: a key role for monocytes and macrophages. *Nature Reviews Immunology*.
608 2020;20(6):355-62.
- 609 34. Rossi G, Cavazza A, Spagnolo P, Bellafiore S, Kuhn E, Carassai P, et al. The
610 role of macrophages in interstitial lung diseases: Number 3 in the Series
611 "Pathology for the clinician" Edited by Peter Dorfmueller and Alberto Cavazza.
612 *Eur Respir Rev*. 2017;26(145).
- 613 35. Brack MC, Lienau J, Kuebler WM, and Witzernath M. Cardiovascular sequelae
614 of pneumonia. *Curr Opin Pulm Med*. 2019;25(3):257-62.
- 615 36. Restrepo MI, and Reyes LF. Pneumonia as a cardiovascular disease.
616 *Respirology*. 2018;23(3):250-9.
- 617 37. Corrales-Medina VF, Alvarez KN, Weissfeld LA, Angus DC, Chirinos JA,
618 Chang CC, et al. Association between hospitalization for pneumonia and
619 subsequent risk of cardiovascular disease. *Jama*. 2015;313(3):264-74.
- 620 38. Puntmann VO, Carerj ML, Wieters I, Fahim M, Arendt C, Hoffmann J, et al.
621 Outcomes of Cardiovascular Magnetic Resonance Imaging in Patients Recently
622 Recovered From Coronavirus Disease 2019 (COVID-19). *JAMA Cardiol*. 2020.
- 623 39. Zheng YY, Ma YT, Zhang JY, and Xie X. COVID-19 and the cardiovascular
624 system. *Nat Rev Cardiol*. 2020;17(5):259-60.
- 625 40. Inciardi RM, Lupi L, Zacccone G, Italia L, Raffo M, Tomasoni D, et al. Cardiac
626 Involvement in a Patient With Coronavirus Disease 2019 (COVID-19). *JAMA*
627 *Cardiol*. 2020.
- 628 41. Deng Q, Hu B, Zhang Y, Wang H, Zhou X, Hu W, et al. Suspected myocardial
629 injury in patients with COVID-19: Evidence from front-line clinical observation

630 in Wuhan, China. *Int J Cardiol.* 2020;311:116-21.

631 42. JT L, WE J, HS P, AE J, and JD S. The sva package for removing batch effects
632 and other unwanted variation in high-throughput experiments. *Bioinformatics*
633 (*Oxford, England*). 2012;28(6):882-3.

634 43. Yu G, Wang LG, Han Y, and He QY. clusterProfiler: an R package for comparing
635 biological themes among gene clusters. *Omics-a Journal of Integrative Biology.*
636 2012;16(5):284-7.

637 44. Butler A, Hoffman P, Smibert P, Papalexi E, and Satija R. Integrating single-cell
638 transcriptomic data across different conditions, technologies, and species. *Nat*
639 *Biotechnol.* 2018;36(5):411-20.

640

641

643 **Figure 1. In silico discovery of PLA2G7 as the biomarker of SARS-COV-2**
644 **infection.** (A) Heatmap of the correlation between module eigengenes and the clinical
645 traits recorded in GSE152075. (B) Distribution of average gene significance and errors
646 in the modules associated with positivity of SARS-COV-2. (C) Disease Ontology (DO)
647 enrichment results on genes in red module. (D) Visualization of the network

648 connections among the most connected genes in the red module. The size of circles was
649 equal to the log2 fold change. (E) Heatmap based on unsupervised clustering of the
650 selected 6 genes after scaling. Each row represented one gene; each column represents
651 one patient. Expression intensity is indicated by color. (F) Volcano plot of differentially
652 expressed genes (DEGs) for SARS-COV-2 infection contrasting positive cases with
653 negative ones in GSE152075.
654

659 **(B)** UMAP plot showing the expression of PLA2G7 across the immune cells. **(C)**
660 UMAP maps showing the expression of PLA2G7, along with the markers (SPP1,
661 FCN1 and FABP4) in re-integrated macrophages. **(D)** Expression of SPP1 and
662 PLA2G7 by four macrophages groups. **(E)** UMAP projection of four macrophage
663 groups among the health controls, moderate and severe COVID-19 patients. **(F)**
664 Average normalized expression of PLA2G7 was correlated with progression of
665 COVID-19. UMAP, Uniform Manifold Approximation and Projection.

666

668 **Figure 3. PLA2G7 was detected in a retrospective collection of nasal swabs from**
669 **both COVID-19 and pneumonia patients. (A)** Density plot of SARS-COV-2 Ct
670 values distribution in GSE152075 and present cohorts. **(B)** Correlation between
671 positive predictive value (PPV) of PLA2G7 and Ct values of SARS-COV-2. **(C)**
672 Positive rates of PLA2G7 in six patients' group. **(D)** Ct values distribution of PLA2G7
673 in patients' group; The reference was COVID-19 group with Ct of SARS-COV-2 ≤ 25.
674 ns: $P \geq 0.05$; *: $P < 0.05$; **: $P \leq 0.01$; ***: $P \leq 0.001$; ****: $P \leq 0.0001$.

Figure 4. Abnormal plasma levels of Lp-PLA₂ in COVID-19 patients. (A) Plasma levels of Lp-PLA₂ in health controls (n=36), hospitalized COVID-19 patients (n=47), discharged and re-positive COVID-19 patients (n=19) and fully recovered COVID-19 patients (n=15). **(B)** Scatter plot of plasma levels of Lp-PLA₂ in hospitalization group along with time from the initial symptoms onset. **(C)** Evaluation of classification performance of plasma levels of Lp-PLA₂ between COVID-19 patients and health controls.