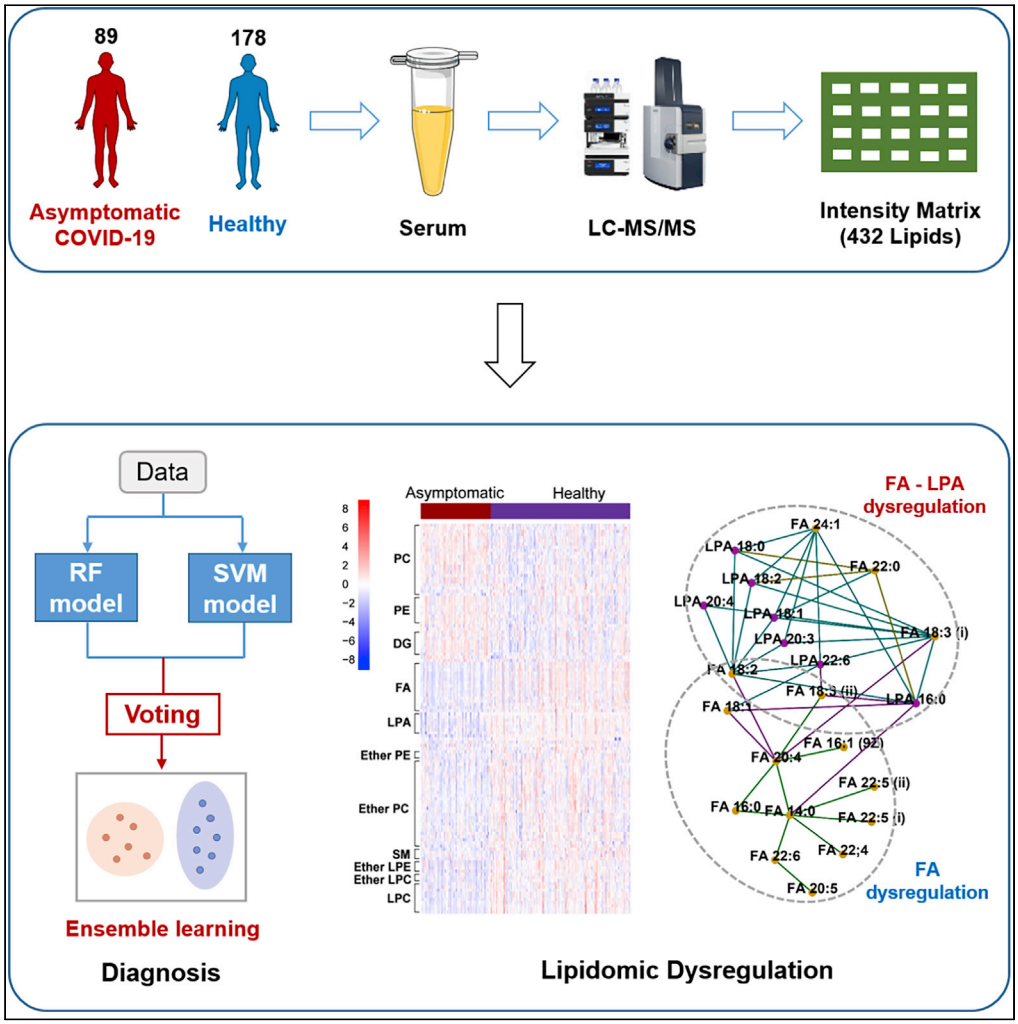
iScience

CellPress
OPEN ACCESS

Article

Distinct lipid metabolic dysregulation in asymptomatic COVID-19



Yanhong Hao, Zheng Zhang, Guifang Feng, ..., Liang Wu, Wenjing Nie, Suming Chen

sm.chen@whu.edu.cn

Highlights

Lipidomic profiling of asymptomatic COVID-19 serum was carried out

A panel of 15 serum lipids distinguished asymptomatic patients from healthy controls

A new ensemble learning model reduced the false negative rate for clinical diagnosis

Unique lipidomic dysregulation was identified for asymptomatic SARS-CoV-2 infection

Hao et al., iScience 24, 102974
September 24, 2021 © 2021
https://doi.org/10.1016/j.isci.2021.102974

iScience

CellPress
OPEN ACCESS

Article

Distinct lipid metabolic dysregulation in asymptomatic COVID-19

Yanhong Hao,[1,4] Zheng Zhang,[2,4] Guifang Feng,[1,4] Moran Chen,[1,4] Qiongqiong Wan,[1] Jie Lin,[3] Liang Wu,[1] Wenjing Nie,[1] and Suming Chen[1,5,*]

SUMMARY

Asymptomatic infection is a big challenge in curbing the spread of COVID-19. However, its identification and pathogenesis elucidation remain issues. Here, by performing comprehensive lipidomic characterization of serum samples from 89 asymptomatic COVID-19 patients and 178 healthy controls, we screened out a panel of 15 key lipids that could accurately identify asymptomatic patients using a new ensemble learning model based on stacking strategy with a voting algorithm. This strategy provided a high accuracy of 96.0% with only 3.6% false positive rate and 4.8% false negative rate. More importantly, the unique lipid metabolic dysregulation was revealed, especially the enhanced synthesis of membrane phospholipids, altered sphingolipids homeostasis, and differential fatty acids metabolic pattern, implicating the specific host immune, inflammatory, and antiviral responses in asymptomatic COVID-19. This study provides a potential prediagnostic method for asymptomatic COVID-19 and molecular clues for the pathogenesis and therapy of this disease.

INTRODUCTION

The coronavirus disease 2019 (COVID-19) is a newly emerged pandemic caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). The rapid spread of COVID-19 around the world presents an unprecedented threat to global public health. According to the data from the World Health Organization, by the time of 5th April, 2021, 131,020,967 cases have been confirmed worldwide, including 2,850,521 deaths. Although vaccination is ongoing, the shortage of vaccines and emergence of SARS-CoV-2 variants will make this disease threatening over a considerable period of time. A recent study estimated the B.1.1.7 SARS-CoV-2 variant is 43%–90% more transmissible than preexisting variants and predicted a higher number of coronavirus deaths across England in 2021 than those in 2020 (Davies et al., 2021).

Most patients confirmed with COVID-19 had different degrees of respiratory illness, ranging from mild to critical, with symptoms such as fever, cough, expectoration, and respiratory distress (Wang et al., 2020). However, there are a special group of patients who were diagnosed positive of SARS-CoV-2 nucleic acid by real-time polymerase chain reaction (RT-PCR) test but exhibited no typical clinical symptoms (Hu et al., 2020). The proportion of asymptomatic infections is hard to be precisely known, but appear to reach 51.7% for cases on board the Diamond Princess Cruise ship in Japan (Mizumoto et al., 2020). It has been known that these asymptomatic individuals are not inferior to spread the virus (Bai et al., 2020). Nevertheless, the lack of clinical oversight of these silent spreaders makes it more difficult for the prevention and control of the epidemic.

RT-PCR can detect the virus RNA in COVID-19 patients directly and is now widely used around the world, but its performance was usually compromised by the moderate sensitivity (~70%) and it needs repeat testing (Ai et al., 2020; Surkova et al., 2020; Watson et al., 2020). The false negative may lead to the omission of true SARS-CoV-2 carriers, especially the asymptomatic COVID-19 patients. Therefore, other techniques that could provide rapid and reliable diagnosis are still needed. Furthermore, the current treatment of COVID-19 remains empirical and the deep understanding of the pathogenesis of the disease is urgent.

Host cell lipids play important roles in the entire life cycle of the virus, from virus entry, virus replication, to virus release (Abu-Farha et al., 2020; Dadhich and Kapoor, 2020). Viruses are known to reprogram the lipid metabolism of host cells to support self-proliferation upon infection. Compounds targeting lipid

[1]Institute for Advanced Studies, Wuhan University, Wuhan, Hubei 430072, China

[2]School of Life Sciences, Central China Normal University, Wuhan, Hubei 430072, China

[3]Community Health Service Center of Shuiguohu Street, Wuhan, Hubei 430071, China

[4]These authors contributed equally

[5]Lead contact

*Correspondence: sm.chen@whu.edu.cn

https://doi.org/10.1016/j.isci.2021.102974

Table 1. Demographics and baseline characteristics of asymptomatic COVID-19

Variables	Asymptomatic (n = 89)	Healthy (n = 178)
Sex - no. (%)		
Male	47 (52.8%)	94 (52.8%)
Female	42 (47.2%)	84 (47.2%)
Age, years		
Range	19–88	19–91
Mean ± SD	45 ± 13	45 ± 13
Median (IQR)	43 (36–53)	43 (35–52)
Seropositive for antibodies against SARS-CoV-2 - no. (%)		
IgM	16 (18.0%)	0
IgG	83 (93.3%)	0

no. (%), number (percentage); SD, standard deviation; IQR, interquartile range.

See also Table S1.

metabolism have shown potential as antivirals for viruses including SARS-CoV-2 (Abu-Farha et al., 2020; Silvas et al., 2020). Several studies have been carried out to investigate the lipid alterations upon SARS-CoV-2 infection (Schwarz et al., 2020; Song et al., 2020; Thomas et al., 2020a; Wu et al., 2020). However, these studies are mostly focused on symptomatic patients. So far, to the best of our knowledge, the comparative study of lipidomic signature between the asymptomatic COVID-19 and healthy people has not been realized.

In this study, we hypothesized that SARS-CoV-2 infection would induce characteristic lipidomic alterations in the serum of asymptomatic patients. These molecular changes may contribute to the discrimination and deep understanding of asymptomatic COVID-19. To test this hypothesis, we performed mass-spectrometry (MS)-based lipidomic profiling of serum samples from 89 asymptomatic COVID-19 patients and 178 healthy controls. A total of 432 lipids including 19 subclasses were identified and relatively quantified. We used machine learning to process the MS data and discovered a panel of 15 distinct serum lipids that could effectively distinguish asymptomatic COVID-19 patients from healthy controls. A new ensemble learning model consisting of random forest (RF) and support vector machine (SVM) was developed, which provided a high mean accuracy of 96.0% with only 3.6% false positive rate and 4.8% false negative rate in 20 repeated nested fivefold cross-validation models. In addition, distinct molecular changes in several important lipid classes were identified in asymptomatic COVID-19, which indicated the unique and dysregulated lipid metabolism, especially the enhanced synthesis of membrane phospholipids, the altered sphingolipids homeostasis, and differential metabolic pattern of fatty acids (FAs). These molecular alterations provided mechanistic insights of asymptomatic COVID-19 and implicated the specific host immune, inflammatory, and antiviral responses during the virus invasion process in asymptomatic COVID-19.

RESULTS

Clinical characteristics and lipidomic profiling of asymptomatic COVID-19 serum

We procured a cohort of 267 subjects including 89 asymptomatic COVID-19 patients and age- and sex-matched healthy controls (n = 178). The demographic characteristics and laboratory findings of these participants are shown in Table 1 and Table S1. In the asymptomatic group, 18.0% (16/89) tested seropositive for IgM, while 93.3% (83/89) appeared to be seropositive for IgG. IgM is produced in the early stage of SARS-CoV-2 infection, while IgG is likely to exist for a longer period (Xu et al., 2020). The high positive rate for IgG and relatively low positive rate for IgM may indicate that most patients were in the middle or late stages of infection.

We used MS-based untargeted lipidomics approach to analyze the serum samples. Peaks with relative standard deviations (RSDs) < 30% across quality control samples after subtraction of background noise were further processed for structure confirmation. A total of 432 lipids including 19 subclasses were finally identified and relatively quantified (Tables S2 and S3). Specifically, the median RSD value for the identified lipids in quality control (QC) samples was 6.7% (Figure S1A), indicating good stability and reproducibility of

Starting pool: 432 identified lipids

1. Calculate the average importance of each lipid in 100 repeated random forest models.

Top 60 important lipids for classification

LPS	FA	LPA	LPC	Ether LPC	Ether LPE
LPS 18:1 LPS 18:0	FA 20:1 FA 18:1 FA 24:1 FA 18:2 FA 20:0	LPA 18:1 LPA 16:0 LPA 18:2 LPA 20:4 LPA 18:0	LPC 20:0 (ii) LPC 20:1 (i) LPC 22:1 LPC 20:1 (ii) LPC 18:1 (ii)	LPC O-18:1 (i)	LPE O-18:2 LPE O-16:1 LPE O-18:1

DG	PC	Ether PC	PE	SM	LPI
DG 30:0 (14:0_16:0) DG 36:5 (18:2_18:3) DG 32:2 (14:0_18:2) DG 32:1 (16:0_16:1, 14:0_18:1) DG 36:1 (18:0_18:1)	PC 32:3 (14:0_18:3) PC 34:3 (16:0_18:3) PC 36:5 (18:2_18:3) PC 31:0 (15:0_16:0) PC 31:0	PC O-35:4 (ii) PC O-36:5 (ii) PC O-40:8 (i) PC O-38:7 (iii) PC O-34:2 (16:1/18:1)	PE 40:4 (18:0_22:4) PE 36:2 (18:0_18:2) PE 38:6 (18:2_20:4)	SM 34:1;O2 SM 43:2;O2 (ii) SM 41:2;O2 (17:1;O2/24:1)	LPI 18:1

2. Randomly select one lipid from the top 3 important lipids in every subclass to form a series of panels.
3. Choose the panel with the highest mean accuracy of 20 repeated five-fold nested cross-validations in random forest models

Initial panel consisting of 12 lipids

4. Add each of the remaining lipids to the panel respectively and retain the lipid which lead to the highest increase in accuracy
5. Repeat step 4 until the resulting accuracy is no longer improved by adding new lipids.

Final panel consisting of 15 lipids

Final panel

LPS 18:1
FA 18:1; FA 20:0
LPA 18:1; LPA 18:0
LPC 22:1
LPC O-18:1 (i)
LPE O-18:2
DG 30:0 (14:0_16:0); DG 36:5 (18:2_18:3)
PC 36:5 (18:2_18:3)
PC O-35:4 (ii)
PE 36:2 (18:0_18:2)
SM 34:1;O2
LPI 18:1

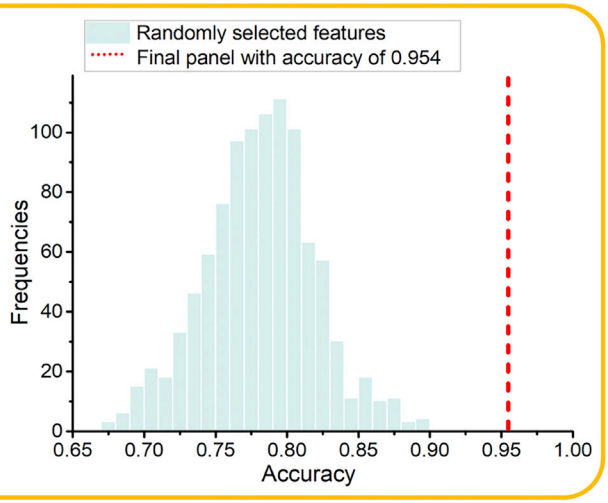


Figure 1. Scheme for construction of lipid panel to distinguish asymptomatic COVID-19 patients from healthy controls

The top 60 important lipids were picked out according to their average importance in 100 repeated random forest models. These lipids were divided into 12 sets according to their belonging subclasses. An initial panel consisting of 12 lipids was generated by first randomly selecting one lipid from the top 3 important lipids in each set to form a series of panels and then picking out the panel with the highest classification accuracy of 20 repeated fivefold nested cross-validation in random forest models. Add each of the remaining lipids to the initial panel respectively, and retain the lipid that lead to the highest increase in accuracy to form a new panel. Repeat this step until the resulting accuracy was no longer improved by adding a new lipid. A final panel consisting of 15 lipids was generated at last with an accuracy of 0.954 $\pm$ 0.029, much higher than the randomly selected 15 lipids from the starting pool (0.678–0.898 over 1,000 times).

See also Table S4.

the lipidomic analysis procedure. Besides, plots of QC samples were found to be well gathered in principal component analysis (PCA) plots based on the omics data, while no obvious separation was observed between asymptomatic COVID-19 and healthy controls (Figure S1B).

To investigate whether the sample size is enough for meaningful discrimination based on our MS data, we performed power analysis (a universal method to derive the optimal sample size by estimating statistical power in a hypothesis test) on a data set of randomly selected 90 samples (30 of asymptomatic patients and 60 of healthy controls, MS ID: 1–90 in Table S1) using 51 differentially expressed lipids (| log2 (asymptomatic/healthy) | > 0.25, adjusted p value <0.01) in MetaboAnalyst (http://www.metaboanalyst.ca/). The minimum number of samples resulted to be 264 (88/176, asymptomatic patients/healthy controls) with predicted power ~0.8 at a false discovery rate (FDR) of 0.1, which can be a sufficient confidence level to conclude the statistical meaningful results (Xia et al., 2015) (Figure S2).

A discovered panel of 15 serum lipids effectively distinguished asymptomatic COVID-19 patients from healthy controls

Viruses are known to induce profound molecular changes of host lipidomes (Ketter and Randall, 2019; Strating and van Kuppeveld, 2017). We hypothesized that the characteristic changes in serum lipids can be used as a signature of viral infection. However, the presence of 432 lipid peaks would be challenging for clinical application. Considering the feasibility of detection, we first screened the lipid pool to find out the lipids that were crucial for classification (Figure 1). We build a RF machine learning model (Liaw and Wiener, 2002) using the lipidomics data from all participants (Table S3). The importance (mean decrease in accuracy) for each lipid was calculated by averaging its importance values in 100 repeated RF models (Table S4).

The top 60 important lipids were then picked out and classified into 12 sets according to their belonging subclasses. Based on the hypothesis that each subclass of lipids has special biological functions, we randomly selected one lipid from the top 3 important lipids in every subclass to form a panel consisting of 12 individual lipids (if the number of lipids in a subclass is less than 3, all the lipids were used for selection). To avoid over-fitting by a particular composition of training and test data set and reduce bias of the resulting error rate, the performance of the resulting panel was evaluated by the average accuracy of 20 rounds of fivefold nested cross-validation in RF models (Krstajic et al., 2014). After all possible combinations were investigated, the panel with the highest accuracy was selected. In order to investigate whether there are any substances that are important for classification in the remaining lipids, we further added each of the remaining lipids to the panel, and the lipid that led to the highest increase in accuracy was retained to form a new panel. The step was repeated until the resulting accuracy was no longer improved by adding a new lipid. At last, a final panel consisting of 15 lipids was generated with an accuracy of 0.954 $\pm$ 0.029. For comparison, we also generated a series of panels consisting of 15 lipids randomly selected from the starting pool for 1,000 times. The resulting classifiers showed much lower rates of classification correctness with accuracies ranging from 0.678 to 0.898 (Figure 1). In order to further confirm the effectiveness of this feature selection strategy, we performed OPLS-DA analysis to discriminate the asymptomatic COVID-19 from the healthy group. The results (Figure S3) showed that the Orthogonal Projections to Latent Structures Discriminant Analysis (OPLS-DA) model could efficiently separate the asymptomatic group from the healthy one by using all 432 identified lipids. It presented R2Y and Q2 values equal to 0.846 ($p < 0.01$) and 0.718 ($p < 0.01$), respectively, after validation by permutation test. The asymptomatic COVID-19 could also be discriminated from the healthy group with OPLS-DA analysis by using the 15 selected lipids. The R2Y and Q2 values could reach 0.672 ($p < 0.01$) and 0.642 ($p < 0.01$), respectively. These results confirmed the effectiveness of the identified lipid biomarker panel in our study.

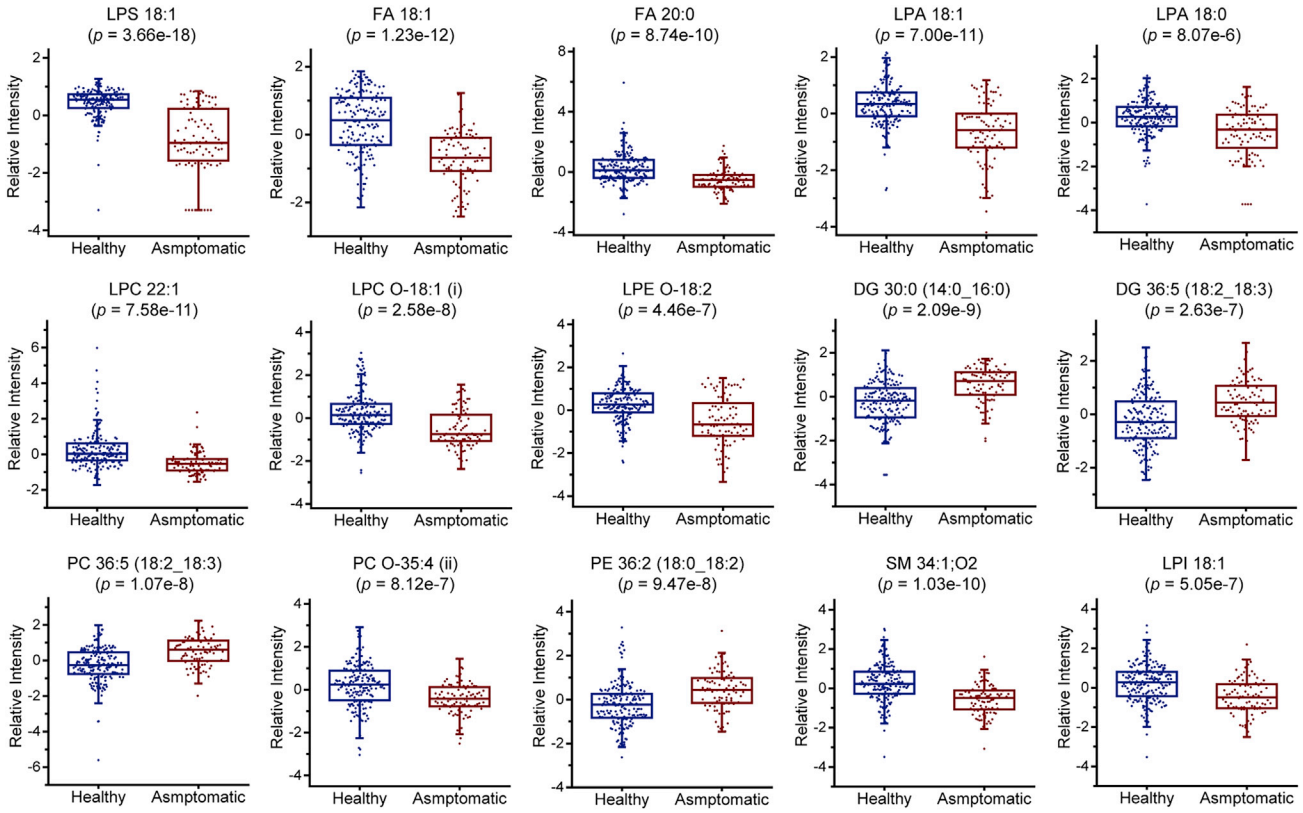


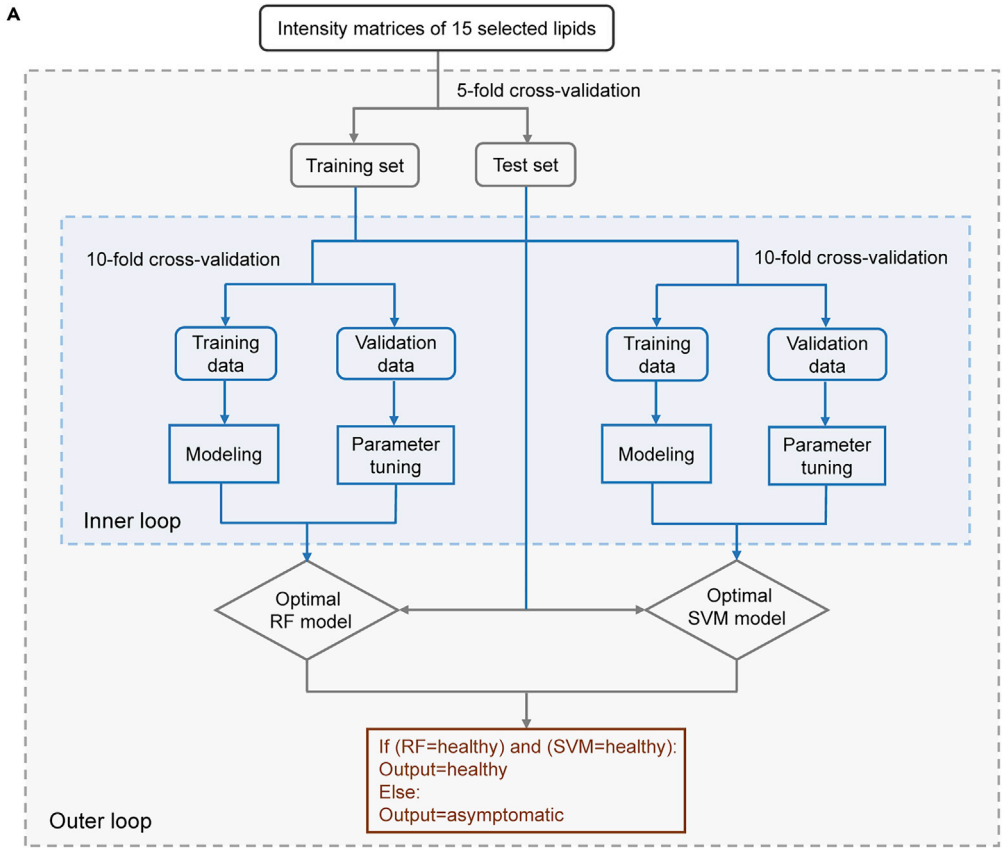
Figure 2. Changes (Z-scored log 2-scaled peak area value) of the 15 selected lipids in the serum of asymptomatic COVID-19 patients

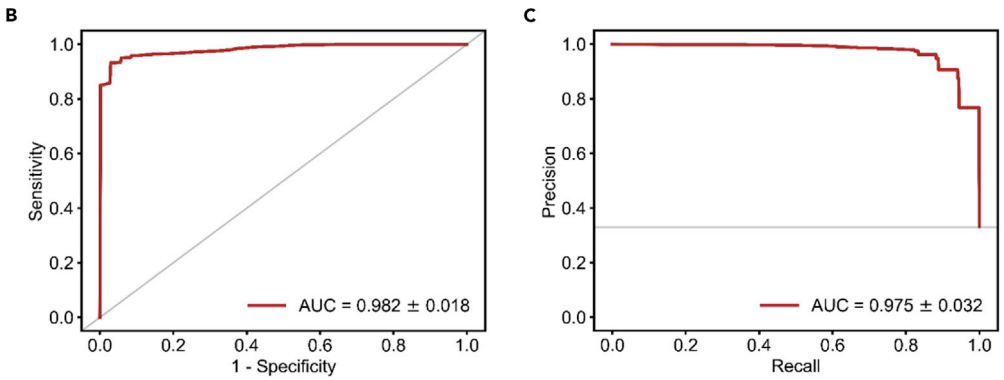
Two-sided Wilcoxon rank-sum test was performed for comparing groups, and adjusted p values were calculated by Benjamini and Hochberg correction. See also Table S6.

Among these lipids (Figure 2), lysophospholipids including lysophosphatidylserine (LPS) 18:1, lysophosphatidic acid (LPA) 18:1 and LPA 18:0, lysophosphatidylcholine (LPC) 22:1, and lysophosphatidylinositol (LPI) 18:1 were generally decreased in asymptomatic COVID-19 patients. Increases were observed in diacylglycerol (DG) 30:0 (14:0_16:0), DG 36:5 (18:2_18:3), phosphatidylcholine (PC) 36:5 (18:2_18:3), and phosphatidylethanolamine (PE) 36:2 (18:0_18:2). Ether lipids [including PC O-35:4 (i), LPC O-18:1 (i) and LPE O-18:2], sphingomyelin (SM34:1; O2), and FAs (including FA 18:1 and FA 20:0) were observed to be decreased.

A new ensemble learning model further improved the sensitivity for diagnosis of asymptomatic COVID-19

Based on the 15 lipids selected in the last step, we first investigated four machine learning algorithms including RF, SVM (Cortes and Vapnik, 1995), multilayer perceptron (MLP) (Gardner and Dorling, 1998), and logistic regression (LR) (Nick and Campbell, 2007) for diagnosis of asymptomatic infections. Nested cross-validation (10 repeated ten-fold cross-validation for the inner loop and 20 repeated fivefold cross-validation for the outer loop, 100 models in total) (Krstajic et al., 2014) was used to evaluate the model performance. For each algorithm, the hyperparameters were optimized in the inner loop by grid search, and then the models with the highest average accuracy were reported to the outer loop for classification of the test samples. Among the four models, RF exhibited the best performance with an average accuracy of 0.954 ± 0.029 and average specificity of 0.988 ± 0.016. However, the sensitivities of all models were found to be relatively low (<90%) (Figure 3D, Table S5). It has been reported that asymptomatic individuals had a weaker immune response to SARS-CoV-2 infection than the symptomatic patients (Long et al., 2020), and we infer that molecular changes of their serum lipids may be too mild for a separate machine learning model to classify these samples. Considering the high infectivity of asymptomatic patients, the consequences would be serious if they are not diagnosed in time.


A
Intensity matrices of 15 selected lipids
5-fold cross-validation
Training set
Test set
10-fold cross-validation
10-fold cross-validation
Training data
Validation data
Training data
Validation data
Modeling
Parameter tuning
Modeling
Parameter tuning
Inner loop
Optimal RF model
Optimal SVM model
If (RF=healthy) and (SVM=healthy):
Output=healthy
Else:
Output=asymptomatic
Outer loop


B
Sensitivity
1 - Specificity
AUC = 0.982 ± 0.018
C
Precision
Recall
AUC = 0.975 ± 0.032

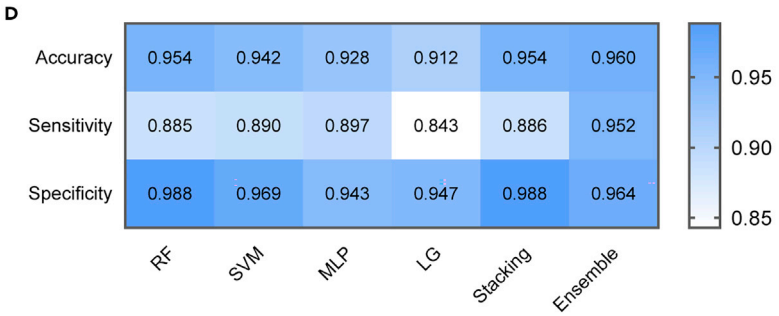

D
Accuracy 0.954 0.942 0.928 0.912 0.954 0.960
Sensitivity 0.885 0.890 0.897 0.843 0.886 0.952
Specificity 0.988 0.969 0.943 0.947 0.988 0.964
RF
SVM
MLP
LG
Stacking
Ensemble
0.95
0.90
0.85

Figure 3. Classification of asymptomatic patients and healthy controls by ensemble learning

(A) Workflow of the construction of ensemble models. The RF and SVM models were separately trained in the inner loop of nested cross-validation first, and then the output in the outer loop of the two models was ensembled by a new voting algorithm.

(B) Receiver operating characteristic (ROC) curve obtained by averaging 20 rounds of fivefold nested cross-validations (100 models in total) with a mean AUC of 0.982.

(C) Precision recall (PR) curve obtained by averaging 20 rounds of fivefold nested cross-validations (100 models in total) with a mean AUC of 0.975.

(D) Performance indicators of four separate machine learning models, stacking ensemble model with meta-learner, and the new ensemble model with a voting algorithm.

See also Table S5.

To improve the sensitivity of the classifiers, an ensemble learning scheme named stacking (Li et al., 2019) was attempted. Stacking is a two-level framework that consisted of multiple machine learning models as the first level whose output was input to a meta-learner in the second level to generate the final output. Owing to the poor performance of MLP and LG, RF and SVM were selected to form the first layer of the ensemble model, RF was used as the meta-learner, but the sensitivity was not improved as expected (0.886 ± 0.087) (Figure 3D, Table S5). Given that a reliable diagnosis method for COVID-19 should minimize the false negative rate, we further replaced the meta-learner in the second level by a new voting algorithm as described in the following: a sample will be predicted as healthy only when the outputs of all classifiers in first level are healthy. If the output of any one classifier is asymptomatic, the sample will be predicted as asymptomatic (Figure 3A). The resulting new ensemble model reached a high average sensitivity of 0.952 ± 0.049 while maintaining a high average specificity of 0.964 ± 0.029 (Figure 3D, Table S5). The mean accuracy, mean area under the curve (AUC) of the receiver operating characteristic (ROC) curve and the mean AUC of the precision recall (PR) curve of the new ensemble model was 0.960 ± 0.025 (Figure 3D, Table S5), 0.982 ± 0.018 (Figure 3B), and 0.975 ± 0.032 (Figure 3C), respectively, which demonstrated the high performance of the ensemble learning model with a new voting algorithm.

Enhanced biosynthetic pathway of membrane phospholipids in asymptomatic SARS-CoV-2 infection

To further investigate the molecular basis for this high-performance diagnosis, lipids with $|\log2(\text{asymptomatic/healthy})| > 0.25$ and adjusted p value <0.05 (Table S6) were shortlisted and visualized in a heatmap (Figure 4) to reveal the differential lipid metabolism in asymptomatic COVID-19. Among these lipids, 83 lipids were found to be downregulated in the serum of asymptomatic COVID-19 patients compared with healthy controls, while 41 lipids were upregulated. We found that the glycerophospholipids which constitute most eukaryotic membranes (van Meer et al., 2008) are increased in the serum of asymptomatic COVID-19 patients, including PC and PE. The increased circulating levels may indicate the elevated intracellular biosynthesis and extracellular transport. Research studies have revealed that enveloped viruses including SARS-CoV-2 utilize host lipid membranes to assemble viral replication complexes (VRCs) for viral genome amplification and hijack the lipid metabolism of hosts to produce lipids for their envelopes (Strating and van Kuppeveld, 2017). Viruses can promote PC synthesis and/or accumulation at the viral replication sites (Zhang et al., 2016). The inclusion of PE in PC bilayers may contribute to the induction of negative membrane curvature, which is important for membrane fusion, budding, and fission during virus propagation (van Meer et al., 2008). The replication of tomato bushy stunt virus (TBSV) was reported to depend on local enrichment of PE at replication sites (Xu and Nagy, 2015). Besides, an overall decrease of lysophospholipids were observed in asymptomatic COVID-19 patients, including LPS, LPA, LPC, and LPI, which may indicate the compromised activity of phospholipase upon virus infection. Specifically, LPA is essential for synthesis of other phospholipids. It is the precursor of phosphatidic acid (PA), which could be converted to DG or cytidine diphosphate-DG and then form most subclasses of phospholipids in the endoplasmic reticulum (Han, 2016). We also identified the increase of DG. These integrated data may indicate that the membrane glycerophospholipids biosynthesis pathway was enhanced upon SARS-CoV-2 infection to support virus entry and replication.

Altered sphingolipids homeostasis in asymptomatic COVID-19

Sphingomyelin (SM), the most abundant sphingolipid component of the mammalian plasma membrane, was observed to be decreased in asymptomatic patients. It has known that SMs preferentially associate with cholesterol to form lipid rafts which is important for virus entry on the cellular surface (Otsuki et al., 2018; Radenkovic et al., 2020). The role of lipid rafts has been reported in promoting SARS-CoV entry by

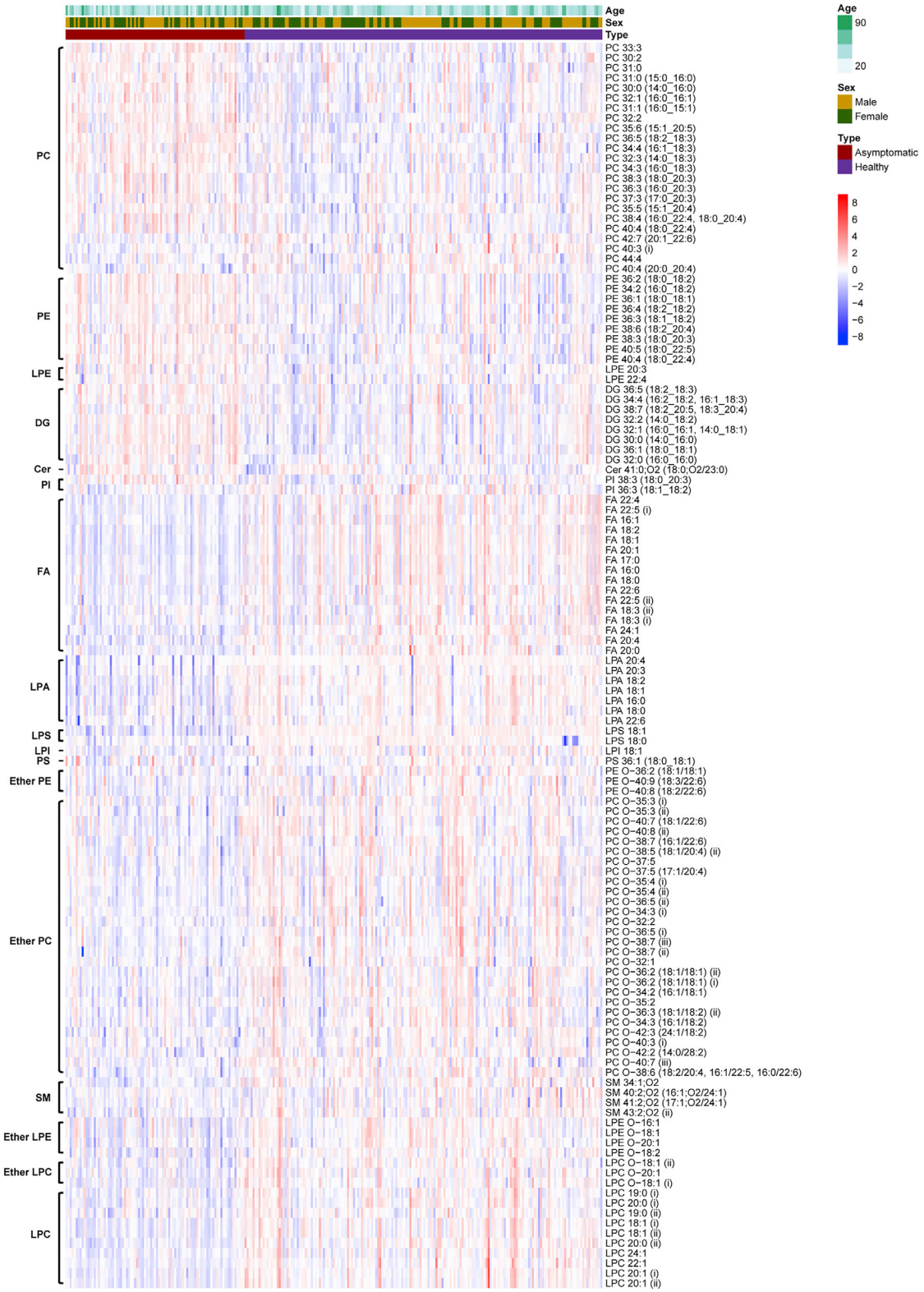


Figure 4. Heatmap of 124 dysregulated lipids belonging to 17 subclasses as indicated

See also Table S6.

providing a platform facilitating the interaction of the virus S-protein with the cellular receptor ACE2 (Lu et al., 2008). Therefore, the decrease in SMs may resist the SARS-CoV-2 infection by blocking the formation of lipid rafts. In contrast to SMs, the level of ceramide (Cer) was observed to be increased in the serum of asymptomatic COVID-19 patients. SM is biosynthesized from Cer by donation of phosphorylcholine from phosphatidylcholine catalyzed by sphingomyelin synthase (SMS), whereas it could also be hydrolyzed to Cer in the catalysis of sphingomyelinase (SMase). The altered sphingolipid homeostasis may correspond to either the suppression of SMS or elevated activation of SMase during the virus invasion. This change is opposite to that observed in symptomatic COVID-19 patients (Song et al., 2020), which implies certain distinct metabolic regulations may happen in asymptomatic SARS-CoV-2 infection.

Suppressed biosynthesis of ether phospholipids in asymptomatic COVID-19

In addition to diacyl phospholipids, the dysregulation of ether phospholipids was also observed in asymptomatic COVID-19 patients. Ether lipids are synthesized in peroxisomes through acyl-dihydroxyacetone (DHAP) pathway by DHAP and FAs. They play important roles in facilitating membrane fusion and stabilizing lipid raft microdomains (Dean and Lodhi, 2018). We found the overall decrease of both ether lipids including ether PC, ether PE, and ether lysophospholipids such as ether LPC and ether LPE in the serum of asymptomatic patients. This change indicated the suppression of the synthetic pathway of ether phospholipids, which might be caused by the competition of the enhanced pathway of regular phospholipids biosynthesis. The decreased ether phospholipids in membrane may further affect the replication of SARS-CoV-2 virus.

Differential metabolic pattern of FAs in asymptomatic COVID-19

Fatty acids are essential for viral infection because they provide building blocks for various membranes lipids during virus proliferating, and poly-unsaturated fatty acids (PUFAs) can be converted to hundreds of lipid mediators such as eicosanoids which play important roles in immune and inflammatory response (Bennett and Gilroy, 2016). It is known that the SARS-CoV-2 infection could cause systematic hyperinflammatory response and potentially life-threatening cytokine storms (Mehta et al., 2020). Elevated levels of PUFAs and derived lipid mediators have been observed in COVID-19 patients in several studies (Hammock et al., 2020; Schwarz et al., 2020; Thomas et al., 2020b), which may be associated with the activated biosynthetic pathways due to the hyperinflammatory response in severe SARS-CoV-2 infection. Interestingly, all FAs including a variety of PUFAs were observed to be significantly decreased in asymptomatic COVID-19 patients in this study. The overall decrease of FAs may be related to the enhanced synthetic pathways of membrane phospholipids, which was consistent with the enrichment of PC, PE, and DG consisting of both saturated and unsaturated fatty acyl chains.

To further investigate the metabolism of FAs, we performed multiscale embedded correlation analysis to investigate the perturbations of FAs coregulation upon asymptomatic SARS-CoV-2 infection (Table S7). Coregulated lipids may display similar patterns in lipid metabolism, which may result in strong correlations between their lipid levels. Changes in lipid-lipid correlation patterns between disease and healthy groups may reveal pathologically related metabolic disorders (Lu et al., 2019; Song et al., 2020). Here, the lipid pairs with significant differential correlations (Benjamini & Hochberg adjusted p values <0.05) in asymptomatic patients compared with healthy controls were displayed in a network (Figure 5). Three notable modules related to FAs were identified from the global network.

Module I comprises myristic acid (FA 14:0) as the hub connected to palmitic acid and several PUFAs including FA 22:5, FA 22:4, FA 20:4, FA 22:6, and FA 22:5, represented by green (++/+) and purple lines (+/0). The weakened positive correlation of myristic acid with other PUFAs was observed, which was confirmed by the larger decrease of PUFAs (1.47–1.89-fold) in asymptomatic COVID-19 than that of FA 14:0 (1.38-fold) (Table S7), which may indicate the enhanced conversion of PUFAs to downstream lipid mediator products. This result further indicated the induced inflammatory response during the asymptomatic SARS-CoV-2 infection and the involvement of PUFAs in this process. Arachidonic acid pathway is a central regulator of inflammatory response. We also identified the weakened positive correlation of FA 20:4 with other unsaturated FAs, such as FA 18:2, FA 18:3, FA 18:3, FA 16:1, and FA 18:1 (module II). Compared with these FAs (decrease of 1.61–2.64-fold), the decrease of FA 20:4 (1.47-fold) was compromised, which may be due to its relatively enhanced synthesis in the inflammatory process even in the context of overall decrease of FAs. In addition, differential correlation between multiple LPAs (LPA 16:0, LPA 18:0, LPA 18:1, LPA 18:2, LPA 22:6) and FAs, especially FA 18:2, FA 18:3, and FA 24:1, was also observed (module III). Most

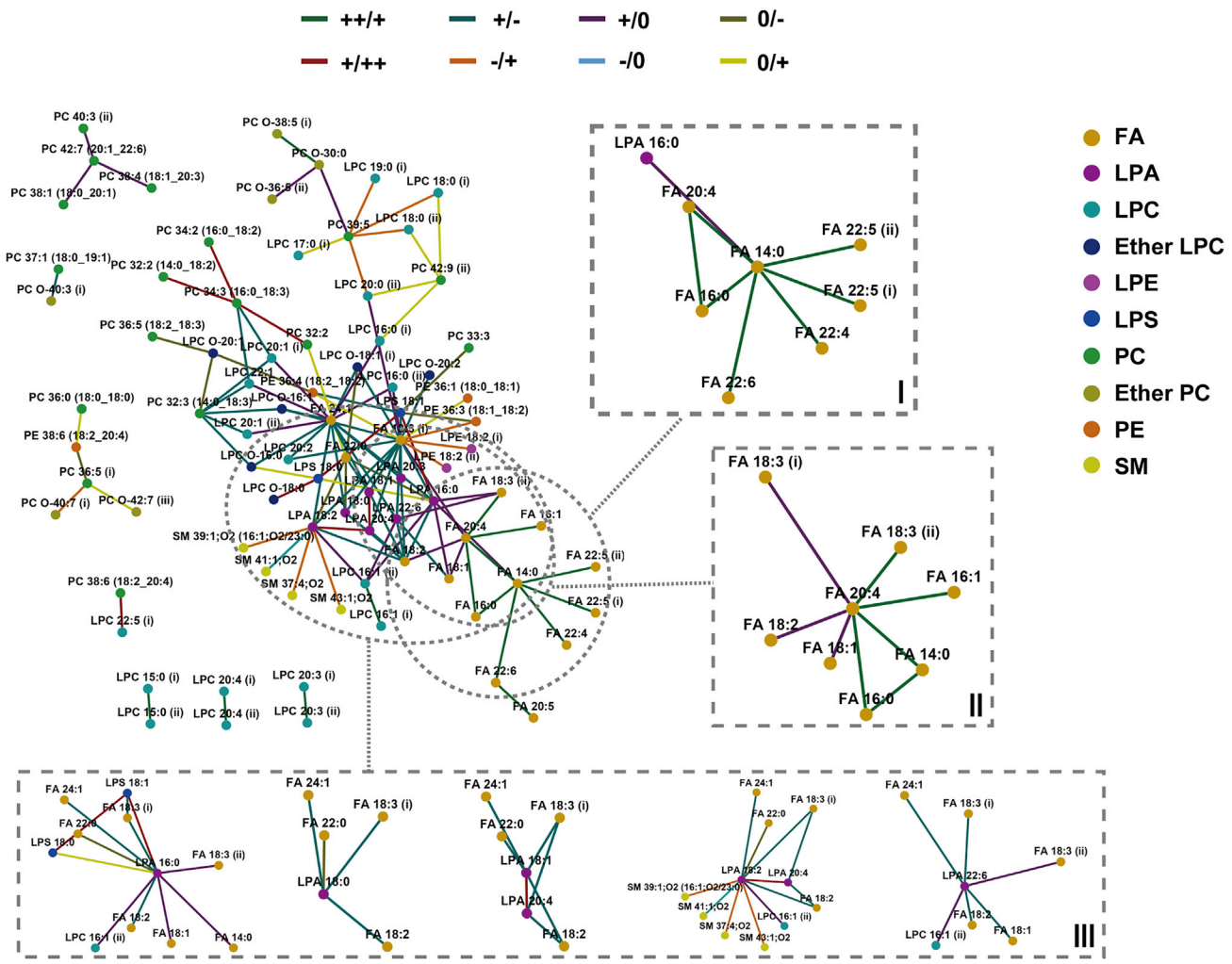


Figure 5. Differential correlation network analyses of serum lipids in asymptomatic patients relative to healthy controls

Lipid pairs with significant differential correlations (Benjamini and Hochberg adjusted p values <0.05) were connected by lines of different colors according to their changes in correlation patterns. Sign/sign indicates the direction and strength of the correlation in control/asymptomatic COVID-19. For instance, red line +/++ indicates that correlation between two connected lipids was positive (+) in control samples and became more strongly positive (++) in asymptomatic COVID-19 patients, as defined by statistically significant increase in correlation coefficients between the lipid pair in asymptomatic patients compared with healthy controls. Three modules (I–III) of interest were circled and expanded for better visualization. Module I comprises myristic acid as the hub connected to palmitic acid and several poly-unsaturated fatty acids (PUFAs) by green (++/+) and purple lines (+/0). Module II comprise AA as the hub connected to several 16-carbon and 18-carbon FAs by green (++/+) and purple lines (+/0). Module III showed the differential correlation network between multiple LPAs and FAs.

See also Table S7.

LPAs-FAs showed positive correlations in healthy controls but converted to negative correlations in asymptomatic patients. Given that both FAs and LPAs were observed to be decreased in patients, the altered correlation might indicate the enhanced utilization of FAs in other pathways, such as the β-oxidation to provide energy during virus invasion.

Association of lipid metabolism with serum antibody levels of IgM and IgG

To investigate whether lipid metabolism relates to IgM and IgG levels in this study, we performed spearman correlation analysis between levels of serum lipids and IgG/IgM, but found that no lipids showed correlation coefficient >0.4 and p value <0.05 with IgM or IgG (Figure S4), suggesting no significant correlation between them. We also tried to divide asymptomatic patients into different groups according to the levels of serum IgM (IgM <1 versus IgM >1) and IgG (1 < IgG <10 versus IgG >10) to see if there are

significant differences in serum lipids between these groups (Table S8). Unfortunately, we did not find significant differences between them using the criterion of | log2 FC | > 0.25 and adjusted p value <0.05 (Figure S5).

DISCUSSION

The rapid spread of COVID-19 around the world presents an unprecedented threat to global public health. The emergence of asymptomatic COVID-19 makes it more difficult for the prevention and control of the epidemic because these silent spreaders are hard to identify. Although SARS-CoV-2 can be detected by RT-PCR, the attempts to develop alternative methods that could diagnose asymptomatic COVID-19 with high sensitivity and convenience are still meaningful. The subtle changes in asymptomatic SARS-CoV-2 infection make its molecular diagnosis more challengeable. In this study, we show that asymptomatic patients can be distinguished from healthy controls based on a panel of only 15 of the key serum lipids, which provides a potential method for future clinical application. We noted that low sensitivity is a common problem in many early diagnostic methods (Delafiori et al., 2021; Kourou et al., 2015). Therefore, a new ensemble learning model was developed using stacking strategy with a new voting algorithm. This ensemble model showed much better performance than that of separate machine learning models or current stacking models for improving the sensitivity (Krstajic et al., 2014). The improved sensitivity with only 4.8% false negative rate is critical for identifying the asymptomatic COVID-19 patients.

The lack of clinical oversight of asymptomatic COVID-19 means we know very little about the proportions of people throughout the course of their infection. Current studies mainly focus on symptomatic COVID-19 patients partially owing to the difficulty to recruit cohort of asymptomatic COVID-19 patients. Acquisition of molecular information is critical for deep understanding of its mechanism and pathogenesis. Our data shed light on the lipidomic changes ref lected in the serum asymptomatic COVID-19. The results revealed not only common dysregulation of lipids observed in SARS-CoV-2 infection but also the distinct changes of lipid metabolism in asymptomatic COVID-19. For example, increase in PC, PE, and DG indicated enhanced biosynthesis of membrane glycerophospholipids upon SARS-CoV-2 infection to support entry and replication of virus, which is consistent with other symptomatic COVID-19 studies (Xu and Nagy, 2015; Zhang et al., 2016). More importantly, distinct molecular changes of several major lipid classes were identified in the serum of asymptomatic COVID-19 patients. We observed all FAs including a variety of PUFAs to be significantly decreased in asymptomatic COVID-19 patients, whereas the severe COVID-19 patients were known by systemic hyperinflammatory response and potentially life-threatening cytokine storms with elevated FAs levels (Schwarz et al., 2020; Thomas et al., 2020b). It has been reported that only a specific mild immune response is caused by the SARS-CoV-2 invasion in asymptomatic patients (Long et al., 2020), and immune suppression was observed in the early stage of COVID-19 disease (Tian et al., 2020). In the context of the overall decrease of FAs, interestingly, we observed the relatively enhanced synthesis of FA 20:4 (Figure 5). This finding may highlight the arachidonic acid pathway as a central regulator of inflammatory response (Calder, 2003). Compared with the activated immune response that emerges in severely infected patients, the milder immune response in asymptomatic and mild-type patients may help them escape from potentially life-threatening cytokine storms caused by the systemic hyperinflammatory response. In addition, we also observed the altered sphingolipids homeostasis in asymptomatic COVID-19 patients, which was opposite to that of previous report in mild to severe COVID-19 patients. SMs can associate with cholesterol to form lipid rafts which could promote virus entry on cellular surface (Lu et al., 2008; Otsuki et al., 2018; Radenkovic et al., 2020), whereas depletion of host and viral SMs was reported to impair influenza virus infection (Audi et al., 2020). Decrease in SMs and increase in Cer may implicate the resistance of the host to the entry and internalization of SARS-CoV-2 virus. Other characteristic changes also include the impaired synthesis of both ether and lyso-ether phospholipids in asymptomatic COVID-19 patients in our research. This distinct lipidomic landscape of asymptomatic SARS-CoV-19 infection provided underlying basis for its molecular identification, new insights into its underlying mechanism, and the clues for the treatment of COVID-19.

In conclusion, our study presents a systematic lipidomic investigation of serum samples from a cohort of asymptomatic COVID-19 patients and healthy controls. We demonstrated the potential for the accurate identification of asymptomatic COVID-19 using a discovered panel of 15 lipids with the developed ensemble learning method. In addition, the distinct lipidomic landscape revealed the comprehensive changes of bioactive lipids during the asymptomatic SARS-CoV-2 infection. These results implicated the specific host immune, inflammatory, and antiviral responses during the virus invasion in asymptomatic

COVID-19. The revealed molecular signature provided clues for precise diagnosis and therapy of this disease.

Limitations of the study

Gender and age were matched between asymptomatic patients and healthy controls in this cohort, whereas the information of body mass index and other clinical symptoms were not included owing to the constraints in collecting these medical records during the outbreak. These might be potential confounders in this study. In addition, restricted by the controlled healthcare resource during the pandemic, collection of controls with other infections was not available, which might influence the specificity of lipidomics profile observed in asymptomatic COVID-19. Although the symptomatic COVID-19 patients were not available in this cohort, the comparative strategy adopted in this study enable the comparison of our results with those of other reported studies. Finally, our results were based on a single Chinese cohort of asymptomatic COVID-19 patients, and future studies in different racial, ethnic, and geographical cohorts will be indispensable for extending our current understanding of lipid metabolic dysregulation in asymptomatic COVID-19 pathogenesis.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Cohort enrollment and data collection
- METHOD DETAILS
  - Serum collection and lipids extraction
  - Untargeted lipidomic analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Statistical analysis
  - Principal Component Analysis
  - Power analysis
  - Heatmap
  - Machine learning
  - Differential correlation analysis

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2021.102974.

ACKNOWLEDGMENTS

This work was financially supported by the start-up funds of Wuhan University and the "Youth Thousand Talents Plan" of China. We also thank the support of the National Natural Science Foundation of China (22074111, 22004093, 22004092), the Fellowship of Wuhan Municipal Health Commission (XG202005080009), and the Special Fund for COVID-19 Control and Prevention from Wuhan (EG20B04).

AUTHOR CONTRIBUTIONS

S.M.C. and Y.H.H. conceived and designed the experiments. Z.Z. and J.L. collected the samples and clinical data. G.F.F. and Y.H.H. conducted the lipidomic analysis. M.R.C. and Y.H.H. performed the statistical analysis and machine learning. Y.H.H. interpreted data. Y.H.H. and S.M.C. cowrote the manuscript. All authors discussed the results and commented on the manuscript. S.M.C. supervised the overall research.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: April 6, 2021
Revised: July 31, 2021
Accepted: August 8, 2021
Published: September 24, 2021

REFERENCES

Abu-Farha, M., Thanaraj, T.A., Qaddoumi, M.G., Hashem, A., Abubaker, J., and Al-Mulla, F. (2020). The role of lipid metabolism in COVID-19 virus infection and as a drug target. Int. J. Mol. Sci. *21*, 3544.

Ai, T., Yang, Z.L., Hou, H.Y., Zhan, C.N., Chen, C., Lv, W.Z., Tao, Q., Sun, Z.Y., and Xia, L.M. (2020). Correlation of chest CT and RT-PCR testing for coronavirus disease 2019 (COVID-19) in China: a report of 1014 cases. Radiology *296*, E32–E40.

Audi, A., Soudani, N., Dbaibo, G., and Zaraket, H. (2020). Depletion of host and viral sphingomyelin impairs influenza virus infection. Front. Microbiol. *11*, 612.

Bai, Y., Yao, L.S., Wei, T., Tian, F., Jin, D.Y., Chen, L.J., and Wang, M.Y. (2020). Presumed asymptomatic carrier transmission of COVID-19. JAMA *323*, 1406–1407.

Bennett, M., and Gilroy, D.W. (2016). Lipid mediators in inflammation. In Myeloid Cells in Health and Disease, S. Gordon, ed. (ASM Press), pp. 343–366.

Calder, P.C. (2003). n-3 polyunsaturated fatty acids and inflammation: from molecular biology to the clinic. Lipids *38*, 343–352.

Cortes, C., and Vapnik, V. (1995). Support-vector networks. Mach. Learn. *20*, 273–297.

Dadhich, R., and Kapoor, S. (2020). Various facets of pathogenic lipids in infectious diseases: exploring virulent lipid-host interactome and their druggability. J. Membr. Biol. *253*, 399–423.

Davies, N.G., Abbott, S., Barnard, R.C., Jarvis, C.I., Kucharski, A.J., Munday, J.D., Pearson, C.A.B., Russell, T.W., Tully, D.C., Washburne, A.D., et al. (2021). Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. Science *372*, eabg3055. https://doi.org/10.1126/science.abg3055.

Dean, J.M., and Lodhi, I.J. (2018). Structural and functional roles of ether lipids. Protein Cell *9*, 196–206.

Delafiori, J., Navarro, L.C., Siciliano, R.F., de Melo, G.C., Busanello, E.N.B., Nicolau, J.C., Sales, G.M., de Oliveira, A.N., Val, F.F.A., de Oliveira, D.N., et al. (2021). Covid-19 automated diagnosis and risk assessment through metabolomics and machine learning. Anal. Chem. *93*, 2471–2479.

Gardner, M.W., and Dorling, S.R. (1998). Artificial neural networks (the multilayer perceptron) - a review of applications in the atmospheric sciences. Atmos. Environ. *32*, 2627–2636.

Hammock, B.D., Wang, W.C., Gilligan, M.M., and Panigrahy, D. (2020). The overlooked storm in coronavirus disease 2019 (COVID-19)? Am. J. Pathol. *190*, 1782–1788.

Han, X.L. (2016). Lipidomics for studying metabolism. Nat. Rev. Endocrinol. *12*, 668–679.

Hu, Z.L., Song, C., Xu, C.J., Jin, G.F., Chen, Y.L., Xu, X., Ma, H.X., Chen, W., Lin, Y., Zheng, Y.S., et al. (2020). Clinical characteristics of 24 asymptomatic infections with COVID-19 screened among close contacts in Nanjing, China. Sci. China. Life Sci. *63*, 706–711.

Ketter, E., and Randall, G. (2019). Virus impact on lipids and membranes. Annu. Rev. Virol. *6*, 319–340.

Kourou, K., Exarchos, T.P., Exarchos, K.P., Karamouzis, M.V., and Fotiadis, D.I. (2015). Machine learning applications in cancer prognosis and prediction. Comput. Struct. Biotechnol. J. *13*, 8–17.

Krstajic, D., Buturovic, L.J., Leahy, D.E., and Thomas, S. (2014). Cross-validation pitfalls when selecting and assessing regression and classification models. J. Cheminform. *6*, 10.

Li, W.Z., Miao, W., Cui, J.X., Fang, C., Su, S.T., Li, H.Z., Hu, L.H., Lu, Y.H., and Chen, G.H. (2019). Efficient corrections for DFT noncovalent interactions based on ensemble learning models. J. Chem. Inf. Model. *59*, 1849–1857.

Liaw, A., and Wiener, M. (2002). Classification and regression by randomForest. R. News *23*, 18–22.

Long, Q.X., Tang, X.J., Shi, Q.L., Li, Q., Deng, H.J., Yuan, J., Hu, J.L., Xu, W., Zhang, Y., Lv, F.J., et al. (2020). Clinical and immunological assessment of asymptomatic SARS-CoV-2 infections. Nat. Med. *26*, 1200–1204.

Lu, J.L., Lam, S.M., Wan, Q., Shi, L.X., Huo, Y.N., Chen, L.L., Tang, X.L., Li, B.W., Wu, X.Y., Peng, K., et al. (2019). High-coverage targeted lipidomics reveals novel serum lipid predictors and lipid pathway dysregulation antecedent to type 2 diabetes onset in normoglycemic Chinese adults. Diabetes Care *42*, 2117–2126.

Lu, Y.N., Liu, D.X., and Tam, J.P. (2008). Lipid rafts are involved in SARS-CoV entry into Vero E6 cells. Biochem. Biophys. Res. Commun. *369*, 344–349.

Mehta, P., McAuley, D.F., Brown, M., Sanchez, E., Tattersall, R.S., Manson, J.J., and Speciality, H.A. (2020). COVID-19: consider cytokine storm syndromes and immunosuppression. Lancet *395*, 1033–1034.

Mizumoto, K., Kagaya, K., Zarebski, A., and Chowell, G. (2020). Estimating the asymptomatic proportion of coronavirus disease 2019 (COVID-19) cases on board the Diamond Princess cruise ship, Yokohama, Japan, 2020. Euro. Surveill. *25*, 2–6.

Nick, T.G., and Campbell, K.M. (2007). Logistic regression. In Topics in Biostatistics, Walter T. Ambrosius, ed. (Humana Press), pp. 237–301.

Otsuki, N., Sakata, M., Saito, K., Okamoto, K., Mori, Y., Hanada, K., and Takeda, M. (2018). Both sphingomyelin and cholesterol in the host cell membrane are essential for rubella virus entry. J. Virol. *92*, e01130–01117.

Radenkovic, D., Chawla, S., Pirro, M., Sahebkar, A., and Banach, M. (2020). Cholesterol in relation to COVID-19: should we care about it? J. Clin. Med. *9*, 1909.

Schwarz, B., Sharma, L., Roberts, L., Peng, X., Bermejo, S., Leighton, I., Massana, A.C., Farhadian, S., Ko, A., DelaCruz, C., et al. (2020). Severe SARS-CoV-2 infection in humans is defined by a shift in the serum lipidome resulting in dysregulation of eicosanoid immune mediators. medRxiv. https://doi.org/10.1101/2020.07.09.20149849.

Silvas, J.A., Jureka, A.S., Nicolini, A.M., Chvatal, S.A., and Basler, C.F. (2020). Inhibitors of VPS34 and lipid metabolism suppress SARS-CoV-2 replication. bioRxiv. https://doi.org/10.1101/2020.07.18.210211.

Song, J.W., Lam, S.M., Fan, X., Cao, W.J., Wang, S.Y., Tian, H., Chua, G.H., Zhang, C., Meng, F.P., Xu, Z., et al. (2020). Omics-driven systems interrogation of metabolic dysregulation in COVID-19 pathogenesis. Cell. Metab. *32*, 188–202.e5.

Strating, J.R.P.M., and van Kuppeveld, F.J.M. (2017). Viral rewiring of cellular lipid metabolism to create membranous replication compartments. Curr. Opin. Cell. Biol. *47*, 24–33.

Surkova, E., Nikolayevskyy, V., and Drobniewski, F. (2020). False-positive COVID-19 results: hidden problems and costs. Lancet Resp. Med. *8*, 1167–1168.

Thomas, T., Stefanoni, D., Dzieciatkowska, M., Issaian, A., Nemkov, T., Hill, R.C., Francis, R.O., Hudson, K.E., Buehler, P.W., Zimring, J.C., et al. (2020a). Evidence for structural protein damage and membrane lipid remodeling in red blood cells from COVID-19 patients. medRxiv. https://doi.org/10.1101/2020.06.29.20142703.

Thomas, T., Stefanoni, D., Reisz, J.A., Nemkov, T., Bertolone, L., Francis, R.O., Hudson, K.E., Zimring, J.C., Hansen, K.C., Hod, E.A., et al. (2020b). COVID-19 infection alters kynurenine and fatty acid metabolism, correlating with IL-6 levels and renal status. JCI Insight *5*, e140327.

Tian, W.M., Zhang, N., Jin, R.H., Feng, Y.M., Wang, S.Y., Gao, S.X., Gao, R.Q., Wu, G.Z., Tian, D., Tan, W.J., et al. (2020). Immune suppression in

the early stage of COVID-19 disease. Nat. Commun. *11*, 5859.

Tsugawa, H., Ikeda, K., Takahashi, M., Satoh, A., Mori, Y., Uchino, H., Okahashi, N., Yamada, Y., Tada, I., Bonini, P., et al. (2020). A lipidome atlas in MS-DIAL 4. Nat. Biotechnol. *38*, 1159–1163.

van Meer, G., Voelker, D.R., and Feigenson, G.W. (2008). Membrane lipids: where they are and how they behave. Nat. Rev. Mol. Cell Biol. *9*, 112–124.

Wang, D.W., Hu, B., Hu, C., Zhu, F.F., Liu, X., Zhang, J., Wang, B.B., Xiang, H., Cheng, Z.S., Xiong, Y., et al. (2020). Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus-infected pneumonia in Wuhan, China. JAMA *323*, 1061–1069.

Watson, J., Whiting, P.F., and Brush, J.E. (2020). Interpreting a covid-19 test result. BMJ *369*, m1808.

Wu, D., Shu, T., Yang, X.B., Song, J.X., Zhang, M.L., Yao, C.Y., Liu, W., Huang, M.H., Yu, Y., Yang, Q.Y., et al. (2020). Plasma metabolomic and lipidomic alterations associated with COVID-19. Natl. Sci. Rev. *7*, 1157–1168.

Xia, J.G., Sinelnikov, I.V., Han, B., and Wishart, D.S. (2015). MetaboAnalyst 3.0-making metabolomics more meaningful. Nucleic Acids Res. *43*, W251–W257.

Xu, K., and Nagy, P.D. (2015). RNA virus replication depends on enrichment of phosphatidylethanolamine at replication sites in subcellular membranes. Proc. Natl. Acad. Sci. U S A *112*, E1782–E1791.

Xu, X., Sun, J., Nie, S., Li, H.Y., Kong, Y.Z., Liang, M., Hou, J.L., Huang, X.Z., Li, D.F., Ma, T., et al. (2020). Seroprevalence of immunoglobulin M and G antibodies against SARS-CoV-2 in China. Nat. Med. *26*, 1193–1195.

Zhang, J.T., Zhang, Z.L., Chukkapalli, V., Nchoutmboube, J.A., Li, J.H., Randall, G., Belov, G.A., and Wang, X.F. (2016). Positive-strand RNA viruses stimulate host phosphatidylcholine synthesis at viral replication sites. Proc. Natl. Acad. Sci. U S A *113*, E1064–E1073.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
Serum samples from 89 asymptomatic COVID-19 patients and 178 healthy individuals	WuHan Prevention and Treatment Center for Occupational Diseases	This paper(Table S1-Patient ID)
Chemicals, peptides, and recombinant proteins		
Myristic acid FA 14:0	Sigma-Aldrich	Cat # 70082-50G
Palmitic acid FA 16:0	Sigma-Aldrich	Cat # 27734-1KG
Palmitoleic acid FA 16:1 (9Z)	Sigma-Aldrich	Cat # P9417-100MG
Margaric acid FA 17:0	Sigma-Aldrich	Cat # H3500-1G
Stearic acid FA 18:0	Sigma-Aldrich	Cat # S4751-1G
Oleic acid FA 18:1 (9Z)	Sigma-Aldrich	Cat # O1008-1G
Linoleic acid FA 18:2 (9Z, 12Z)	Sigma-Aldrich	Cat # L1376-10MG
γ-Linolenic acid FA 18:3 (6Z, 9Z, 12Z)	Sigma-Aldrich	Cat # L2378-10MG
α-Linolenic acid FA 18:3 (9Z, 12Z, 15Z)	Sigma-Aldrich	Cat # L2376-500MG
Arachidic acid FA 20:0	Sigma-Aldrich	Cat # A3631-1G
Eicosenoic acid (11Z) FA 20:1 (11Z)	Sigma-Aldrich	Cat # E3635-100MG
Arachidonic acid (AA) FA 20:4 (5Z, 8Z, 11Z, 14Z)	Sigma-Aldrich	Cat # 10931-250MG
Icosapentaenoic acid (EPA) FA 20:5 (5Z, 8Z, 11Z, 14Z, 17Z)	Sigma-Aldrich	Cat # E2011-10MG
Behenic acid FA 22:0	Sigma-Aldrich	Cat # 216941-5G
Adrenic acid FA 22:4 (7Z, 10Z, 13Z, 16Z)	Sigma-Aldrich	Cat # D3659-25MG
Omega-3 Docosapentaenoic acid (DPA-3) FA 22:5 (7Z, 10Z, 13Z, 16Z, 19Z)	Sigma-Aldrich	Cat # D1797-10MG
Omega-6 Docosapentaenoic acid (DPA-6) FA 22:5 (4Z, 7Z, 10Z, 13Z, 16Z)	Supelco	Cat # 18566-10MG
Docosahexaenoic acid (DHA) FA 22:6 (4Z, 7Z, 10Z, 13Z, 16Z, 19Z)	Sigma-Aldrich	Cat # D2534-25MG
Nervonic acid FA 24:1 (15Z)	Sigma-Aldrich	Cat # N1514-100MG
Water	Made in house by Millipore Direct-Q5	Cat # ZRQS VP500
Acetonitrile	Fisher chemical	Cat # A998-4
Isopropanol	Fisher chemical	Cat # A451-4
Ethonal	Fisher chemical	Cat # A995-4
Formic acid	Aladdin	Cat # F112034-100ml
Ammonium formate	Sinopharm Chemical Reagent	Cat # 30011661
Critical commercial assays		
SARS-CoV-2 nucleic acid extraction kit	Shanghai Zhijiang	NO. P20200201
SARS-CoV-2 nucleic acid detection kit	Shanghai Zhijiang	NO. P20200203
IgM antibody detection kit	Orienter	http://www.scwwt.com
IgG antibody detection kit	Orienter	http://www.scwwt.com

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Lipidomics raw datasets	This paper	https://www.iprox.org/ Project ID: IPX0002853000
Data analysis codes	This paper	https://github.com/Chen-micslab/Covid19_TIMS
Software and algorithms		
DataAnalysis	Bruker	Cat # 1867357
Reifycs file converter	Reifycs	http://www.reifycs.com/AbfConverter/index.html
CompassXtract (V 3.2.201)	Bruker	https://www.bruker.com/cn/service/support-upgrades/software-downloads/mass-spectrometry/compass-tools.html
MS-DIAL (version 4.24)	(Tsugawa et al., 2020)	http://prime.psc.riken.jp/compms/msdial/main.html
R (version 4.0.2)	R Foundation for Statistical Computing	https://www.r-project.org
Python (version 3.7.7)	Python Software Foundation	https://www.python.org/
Compass Hystar	Bruker	Cat # 1850838
Other		
ACQUITY BEH C18 column, 2.1 mm × 100 mm, 1.7 μm	Waters Corporation	Cat # 186008316

RESOURCE AVAILABILITY

Lead contact

Further information should be directed to and will be fulfilled by the Lead Contact Suming Chen (sm.chen@whu.edu.cn).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- Data

The lipidomics data are deposited in ProteomeXchange Consortium (https://www.iprox.org/) and are publicly available as of the date of publication. Project ID: IPX0002853000.

- Code

The project data analysis codes are deposited in GitHub (https://github.com/Chen-micslab/Covid19_TIMS) and are publicly available now.

- Additional Information

Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Cohort enrollment and data collection

We retrospectively recruited a total of 89 asymptomatic COVID-19 patients who were under quarantine in Wuhan from January to March, 2020. They were diagnosed as asymptomatic COVID-19 according to the Chinese Government Diagnosis and Treatment Guideline (Trial 6th version) (NHCPRC, 2020). Briefly, asymptomatic COVID-19 referred to those people who meet the following two clinical criteria: 1) exhibited no typical clinical symptoms, 2) tested either positive for SARS-CoV-2 nucleic acid test in respiratory

specimens or seropositive for IgM antibody test. All enrolled patients were confirmed positive for SARS-CoV-2 nucleic acid except for three individuals whose RT-PCR test turned from positive to negative at the blood collection time (Table S1), they were also classified as asymptomatic due to the seropositive of IgM or IgG. According to the epidemiologic investigations, none of these patients developed symptoms until they were tested negative for SARS-CoV-2. For each confirmed patient, two healthy people of the same gender and approximate age was matched as controls. These healthy controls come from the physical examination population in Wuhan Prevention and Treatment Center for Occupational Diseases during the sample collection period. Baseline characteristics including age and gender and library findings including serum IgM and IgG of both asymptomatic patients and healthy controls are summarized in Tables 1 and S1.

For identification of SARS-CoV-2 infections, throat swabs were collected and tested by real-time polymerase chain reaction (RT–PCR) using virus nucleotide acid extraction kit (Shanghai Zhijiang, China, NO. P20200201) and detection kit (triple fluorescence PCR, Shanghai Zhijiang, China, NO. P20200203) according to manufacturer instructions. Briefly, target genes including RdRp, E and N were simultaneously amplified and tested during RT–PCR. Patients were diagnosed as positive if RdRp gene was positive (Ct < 43), and one of E or N was positive (Ct < 43). Patients were also diagnosed as positive if two sequential tests of RdRp were positive while E and N were negative.

IgG and IgM against SARS-CoV-2 were detected in serum samples using chemiluminescence immunoassay kits (Orienter Biotechnology Co., Ltd, Sichuan, China) and Access2 automatic microparticle chemiluminescence immunoassay system (BechmanCoulter, California, USA). Antibody levels were expressed as the ratio of the chemiluminescence signal over the cutoff value (S/CO). The result was defined as positive if the S/CO value is higher than 1.00.

This study was approved by the Ethics Review Commission of WuHan Prevention and Treatment Center for Occupational Diseases (reference no. 202002).

METHOD DETAILS

Serum collection and lipids extraction

Peripheral blood samples for all participants were collected using serum separation tubes after an overnight fast. Serum was separated by centrifugation at 1,500 g for 10 min and then stored at − 80°C after standard diagnostic tests. Before used, the frozen serum was slowly thawed at 4°C overnight. For virus inactivation and lipids extraction, pre-chilled ethanol was added to each sample to make a final solution of 75% (v/v) ethanol. The mixture was shaken vigorously for 5 min to ensure inactivation of virus. The supernatant was collected by centrifuged at 1,2000 g for 15 min. A pooled sample was generated by taking equal aliquot of each experimental sample to serve as a technical replicate which was run multiple times throughout the experiment. Extracted water samples served as blanks. All samples were stored at − 80°C until analysis.

Untargeted lipidomic analysis

The lipidomics analysis was performed on an UltiMate 3000 UHPLC System (DIONEX, Thermo Fisher Scientific, U.S.A.) coupled with a TIMS-TOF mass spectrometer (Bruker, Germany). In detail, the UHPLC separation was performed on a Waters ACQUITY UPLC BEH C18 Column (2.1 mm × 100 mm, 1.7 μm) at 35°C with a flow rate of 0.3 mL/min. ACN/H_2O (6:4, v/v) and IPA/ACN (9:1, v/v), both containing 10 mM NH_4COOH and 0.1% (v%) FA, were employed as mobile phase A and B, separately. Gradient elution was achieved with the following program: 0 min, 30% B, 2 min, 30% B, 3.5 min, 52% B, 5 min, 63% B, 6 min, 68% B, 12.5 min, 74.5% B, 13.5 min, 80% B, 14.5 min, 30% B, 19 min, 30% B. MS detection was performed on a TIMS-TOF mass spectrometer equipped with an electrospray ionization (ESI) source in both positive and negative ion modes, respectively. Data was acquired using an auto MS/MS (data dependent acquisition) method with scan range of 100-1000 m/z for both MS1 and MS2. The time was set as 0.5 second for a single cycle. Dynamic exclusion was activated by excluding the precursor ions for MS/MS acquisition after they had been acquired 3 times and releasing them after 0.2min. The precursor ion was reconsidered if its current intensity was 2 fold of the previous intensity. The collision energy for MS/MS was set at 30 eV. A solution of sodium formate was injected to the mass spectrometer at the beginning of each sample analysis process.

For quantitative and qualitative analysis of lipids, the spectra were first calibrated post-run by the sodium formate cluster ions using DataAnalysis software (Bruker) to improve the mass accuracy. The resulting data files (.d format) were then converted to ABF format using Reifycs file converter and CompassXtract (Bruker). Then the abf files were imported into MS-DIAL (version 4.24) (Tsugawa et al., 2020) for data processing including peak extraction, alignment and annotation. MS1 tolerance was set as 0.005 Da and retention time tolerance were set as 0.15min for peak alignment. Peaks presented in less than 30% of a sample group were excluded for subsequent analysis. Fatty acids identification was achieved by accurate mass and retention time (RT) match with standards. Identification of other lipids was achieved by searching the internal theoretical MS/MS spectra library of MS-DIAL, the accurate mass tolerance for MS1 and MS2 were set as 0.01 Da and 0.05 Da, respectively. Lipids identified in positive and negative ion mode were finally integrated to generate a matrix containing lipid name and peak area information of all samples. If a lipid was identified in both modes, peak information of either mode was retained for further analysis according to their peak intensity and reliability of MS/MS spectra match.

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analysis

Missing values were replaced with 1/10 of minimum peak height over all samples. Lipids with relative standard deviations (RSDs) of over 30% in the pooled quality control samples were removed for subsequent analysis. Log2 fold-change (log2 FC) was calculated by log2-scaling the ratio of mean peak area in asymptomatic and healthy group. Two-sided Wilcoxon rank-sum test was performed for comparing groups and adjusted p values were calculated by Benjamini & Hochberg correction. The significant differential expressed lipids were defined using the criteria of adjust p value less than 0.05 and absolute log2 FC larger than 0.25.

Principal Component Analysis

PCA was performed using Python package scikit-learn (version 0.22.1) and matplotlib (version 3.1.3).

Power analysis

Power analysis was performed on MetaboAnalyst (http://www.metaboanalyst.ca/) by uploading a dataset of randomly selected 90 samples (30/60: asymptomatic patients/healthy controls, MS ID: 1-90 in Table S1) using 51 differentially expressed lipids (| log2 (asymptomatic/healthy) | > 0.25, adjusted p value < 0.01). The false discovery rate (FDR) was set as 0.1.

Heatmap

Heatmap of the discriminative lipids was generated using R (version 4.0.2) with package pheatmap. The original peak area values were log 2-scaled and Z-scored stepwise for better visualization.

Machine learning

Separate machine learning algorithms were carried out using Random Forest (RF), Support vector machine (SVM), Multi-Layer Perceptron (MLP) and Logistic Regression (LR). Before imported to the model, the intensity matrix of the 15 selected lipids was preprocessed using different methods. Zero-mean normalization coupled with PCA (n_components=0.99) were used for SVM. Single zero-mean normalization was used for MLP and LR. RF used the original intensity matrix without preprocessing. For each model, nested cross-validation (10 repeated ten-fold cross-validation for the inner-loop and 20 repeated five-fold cross-validation for the outer loop, 100 models in total) was used for performance evaluation. The ratio of asymptomatic/healthy was kept as 1:2 in each split of both the inner loop and the outer loop. The hyperparameters (n_estimators and max_depth for RF, C and gamma for SVM, the number of nodes for each layer for MLP (three layers), C and penalty for LR) of each algorithm were optimized in the inner loop by grid search, then the models with the highest average accuracy were reported to the outer loop for classification of the test samples. The mean accuracy, mean sensitivity, mean specificity, mean AUC of ROC curve and mean AUC of PR curve of the test samples were calculated based on the 100 models.

For ensemble learning named stacking, SVM and RF were selected to form the first layer of the ensemble model, RF was used as the meta-learner in the second layer. Model training and test were performed using nested cross-validation as described above in both layers.

The new voting algorithm which replaced the meta-learner in the second level of stacking followed the rule below:

```
If (rf == healthy, svm == healthy):

output = healthy

else:

output = asymptomatic
```

In above algorithm, rf and svm represented the predictions of the same sample obtained by the two classifiers separately, and output was the prediction of the ensemble model.

All machine learning algorithms, zero-mean normalization and PCA were carried out in Python (version 3.7.7) with the package scikit-learn (version 0.22.1). ROC curve and PR curve were generated by the package matplotlib (version 3.1.3).

Differential correlation analysis

Differential correlation analysis was carried out in R (version 4.0.2) using package DGCA (version 1.0.2). Cytoscape (version 3.8.0) was used to build correlation networks from differentially correlated lipid pairs (Benjamini & Hochberg adjusted p values < 0.05) in asymptomatic COVID-19 patients relative to healthy controls.