Supplementary method

Enrollment process

The isolation ward is divided into two areas in hospital, one for suspected patients ward and the other for confirmed patients ward. Patients with fever, cough, or other related symptoms, or clinical confirmed COVID-19 were admitted to suspected patients ward. After hospitalization, the throat swab specimens were collected for nucleic acid test by PT-PCR the next day. Patients were transferred to confirmed patients ward if the nucleic acid test was positive. Then our researcher evaluated the patient whether meeting our criteria and enrolled eligible patients after signing written informed consent. Patients with exposure to confirmed patients or Wuhan, but without symptoms, were isolated outside the hospital and collected throat swab specimens. Patients were admitted to hospital and transferred to confirmed patients ward if the result of nucleic acid test was positive. Then researchers screened and enrolled eligible patients.

Convalescent patients were transferred to another independent isolated region under observation for 15 days. If the nucleic acid test was still negative during this period, convalescent patients could departure from the independent isolated region. Samples from recovery patients were collected two days before discharge. The samples collection methods were consistent with our previous description.

Diagnostic, inclusion and exclusion criteria

Diagnostic criteria are based on the “Diagnosis and treatment program trial version 5 guidelines”[1] issued by the National Health Commission of the People’s Republic of China.

1. Suspected patients.

Combined with the following comprehensive analysis of epidemiological history and clinical manifestations: any one of the epidemiological history and any two of the clinical manifestations, or no clear epidemiological history and three of the clinical manifestations can be diagnosed as a suspected patient.

(1) History of Epidemiology

1) Travel history or residence history of Wuhan city and surrounding areas, or other communities with case reports within 14 days before onset.

2) A history of contact with a infected person (positive nucleic acid test) within 14 days before onset.

3) Have contacted patients with fever or respiratory symptoms from Wuhan city and surrounding areas, or from communities with case reports within 14 days before onset.

4) Aggregated onset (2 or more cases of fever and / or respiratory symptoms occurred within 2 weeks in a small area such as home, office, school class, etc.,).

(2) Clinical manifestations

1) Fever and / or respiratory symptoms.

2) With the characteristic image of COVID-19.

3) The total number of white blood cells in the early stage of the disease is normal or decreased, and the lymphocyte count is normal or decreased.

2. Confirmed patients.

Suspected cases with one of the following etiological or serological evidence:

(1) Fluorescence RT-PCR detects the positive of the COVID-19 virus nucleic acid.

(2) Viral gene sequencing finds highly homologous to COVID-19 virus.

3. Cured patients (Convalescent patients).

Need to meet the following conditions at the same time:

(1) Body temperature returned to normal for more than 3 days.

(2) Respiratory symptoms improved significantly.

(3) Pulmonary imaging showed a significant improvement in acute exudative lesions.

(4) Two consecutive sputum, nasopharyngeal swabs and other respiratory tract specimens with negative nucleic acid test (sampling time needs to be at least 24 hours apart).

The inclusion criteria for healthy volunteers referred to our previous study. In all

healthy controls, exclusion criteria included hypertension, diabetes, obesity, metabolic syndrome, irritable bowel syndrome (IBD), nonalcoholic fatty liver disease, coeliac disease and liver cirrhosis.

All enrolled individuals had healthy oral tissues and gingiva (determined by a professional dentist). Individuals who received antibiotics and/or probiotics within 8 weeks before enrolment were also excluded.

DNA extraction

Microbial DNA was extracted using a Qiagen Mini Kit (Qiagen, Hilden, Germany) following the manufacturer's instructions as described previously. The DNAs were quantified using a Qubit 2.0 Fluorometer (Invitrogen, Carlsbad, CA, USA), and molecular size was estimated using agarose gel electrophoresis. All microbial DNAs were diluted to 10 ng/μL for microbial analysis.

PCR amplification

Primers targeting the hypervariable V3-V5 region of the 16S rRNA gene were used to amplify the extracted DNA samples. The forward primer was 5'-ACTCCTACGGGAGGCAGCA-3' and the reverse primer was 5'-GGACTACHVGGGTWTCTAAT-3'. The PCR reaction system contained 2 μL of 2.5 mM dNTP (TransGen Biotech, Beijing, China), 4 μL of 5×Fast Pfu buffer, 0.4 μL of Fast Pfu polymerase, 0.8 μL of each primer (5 μM) and 10 ng of template DNA. Four reactions on each sample were performed in a PCR machine (ABI GeneAmp 9700): 95 °C for 2 minutes, 95 °C for 30 s, 55 °C for 30 s, 72 °C for 30 s, repeating 30 cycles and finally at 72 °C for 5 minutes. Agarose gel (Axygen Biosciences, Union City, CA) was used to separate, extract and purify the PCR products, and the products were quantified using a fluorescence assay kit (Quant-iT PicoGreen, Invitrogen)

Sequence data process

Based on the sample-specific barcode, the filtered readings were distributed to different samples, and then the barcode and primers were trimmed. The amplified reads were

processed through the following steps: (a) use FLASH v1.2.10 software to overlap pair end sequenced reads of each library. (b) use the customization of each program to perform more specific quality control on overlapping reads generated by FLASH 1) Ambiguous bases (N); 2) the mismatch rate in the overlap region was no more than 0.05; 3) mismatches in barcode and primer regions were not allowed; (c) reads were de-multiplexed and assigned into different samples according to barcodes; (d) detect and remove the chimeric sequences by UCHIME version 4.2.40 (version microbiome util-r20110519, http://drive5.com/uchime/gold.fa) to match Operational Taxonomy Units (OTUs).

Microbial diversity and taxonomic analysis

Microbial diversity, including alpha diversity and beta diversity were determined based on sampling OTU analysis. We used R program package 'vegan' to calculate Shannon index, observed_otus, ACE index, Chao index and Simpson index for alpha diversity. We used R program package (http://www.R-project.org/) to conduct Principal coordinate analysis (PCoA) for beta diversity. A heatmap for the key variables was constructed by the Heatmap Builder.

Wilcoxon rank sum test was used to analyze and compare the bacterial taxonomic. Through linear discriminant analysis (LDA) effect size (LEfSe) method (http://huttenhower.sph.harvard.edu/lefse/)e/), the microbial characteristics between different groups were analyzed[2]. We used the Kruskal-Wallis rank sum test ($p < 0.05$) to identify the significantly different taxa by LEfSe. Then, we used LDA to evaluate the effect size of each feature, and the cut-off value was set as LDA score (log10) =2 or 2.5 [3].

Construction of POD

We used the abundance profile of the optimal OTUs markers in the discovery set to perform fivefold cross-validation on a random forest model (R 3.4.1, randomForest 4.6-12 package). Then we acquired the cross-validation error curve through five trials of the fivefold cross-validation. We defined the point with the minimum cross-validation error as the cut-off point through the minimum error plus the standard deviation (SD). The sets of OTU markers with the error less than the cut-off value were listed and the set with the smallest number of OTUs were defined as the optimal set. Then we used the optimal set of OTUs to calculate the probability of disease (POD) index in discovery cohort and validation cohort. The constructed models was evaluated using the receiver operating characteristic (ROC) curve (R 3.3.0, pROC package), and the ROC effect was showed using AUC.

Detection of immunoglobulin G (IgG) against SARS-CoV-2

Levels of antibodies against SARS-CoV-2 in serum were tested in recovered subjects based on the chemiluminescence immunoassay using kits (Shenzhen Mairui Biomedical Electronics Co., Ltd., Guangdong). IgG tests were conducted according to the manufacturer's instructions and under stringent biosafety conditions. The positive judgment value of the kit was 10 U/ml (a value > 10 U/ml was defined as positive, and a value < 10 U/ml was defined as negative). The antibody levels were calculated as $\log_2$(value).

Lipidomics detection

100 μL serum were placed in a 96-well plate with 300 μL isopropanol (precooled at-20 °C) and 10 μL SPLASH internal standard storage solution (330707, SPLASHTM Lipidomix Mass Spec Standard, Avanti Polar Lipids, USA), vortexed for 1 min, and stored overnight at -20 °C. Then the extraction mixture was centrifugated at 4,000 g for 20 min and the The supernatant was placed in the upper sample bottle. The quality control (QC) sample was consist of 10 μL supernatant of each sample to evaluate the repeatability and stability of the Ultra-performance liquid chromatography-mass spectrometry (UPLC-MS) analysis process.

UPLC-MS analysis for lipidomics

The separation and detection of metabolites were employed with a Waters 2D UPLC (waters, USA) connected with to a Q Exactive high resolution mass spectrometer (Thermo Fisher Scientific, USA). Chromatographic separation was employed with an CSH C18 column (1.7 μm 2.1x100 mm, Waters, USA). The positive ion mode mobile phase is an aqueous solution containing 10 mM formic acid ammonia, 0.1% formic acid and 60% acetonitrile (liquid A) and a solution containing 10 mM formic acid ammonia, 0.1% formic acid, 90% isopropanol and 10% acetonitrile (liquid B). The negative ion mode mobile phase is an aqueous solution containing 10 mM fformic acid ammonia and 60% acetonitrile (liquid A) and a solution containing 10 mM formic acid ammonia, 90% isopropanol and 10% acetonitrile (liquid B). The following gradients were used for elution: 0-2 min, 40%-43% B solution; 2-2.1 min, 43%-50% B solution; 2.1-7 min, 50%-54% B solution; 7-7.1 min, 54%-70% B solution; 7.1-13 min, 70%-99% B solution; 13-13.1 min, 99%-40% B solution; 13.1-15 min, 40% B solution. The flow rate was 0.35 ml/min, the column temperature was 55°C, and the injection volume was 5 μL. Q Exactive mass spectrometer (Thermo Fisher Scientific, USA) was used to collect primary and secondary mass spectrometry data. The range of mass-nucleus ratio of mass spectrometry scanning is 200-2000, the first-order resolution is 70,000, the

AGC is 3e6, and the maximum injection time (IT, injection time) is 100ms. According to the strength of the parent ion, Top3 is selected for fragmentation, and the secondary information is collected. The secondary resolution is 17,500, the AGC is 1e5, the maximum injection time (IT, injection time) is 50ms, and the fragmentation energy (stepped nce) is set as follows: 15, 30, 45eV. The (ESI) parameters of ion source are as follows: Sheath gas flow rate is 40, auxiliary gas flow rate is 10, spray voltage (|KV|) positive ion mode is 3.80, negative ion mode is 3.20, ion transport tube temperature (Capillary temp) is 320C, auxiliary gas heater temperature is 350°C. Moreover, during testing, every 10 samples were interspersed with a QC sample to provide more reliable experimental results.

Data analysis

The original data (raw file) generated by UPLC-MS were imported into LipidSearch v.4.1 (Thermo Fisher Scientific, USA) for mass spectrometry data analysis, and the data matrix including lipid molecular identification results and quantitative results was obtained. LipidSearch v.4.1 first identifies and raises the peak of a single sample and then aligns the peaks of all samples. A single sample is first identified, and the peak of a single sample is raised. Then, peak alignment is performed on all samples.

The following parameters are used for lipid identification and peak lifting: the identification type is selected as Product, the search mass deviation of parent ion and daughter ion is 5ppm, the response threshold is set to 5.0% of the relative response deviation of the sub-ion; the quantitative parameter is set to calculate the peak area of all identified lipids, and the mass deviation of peak lifting is set to 5ppm. The filter is set to top rank, all isomer peak, FA priority, M-score is set to 5.0, and the c-score is set to 2.0, and the identification level is selected as "A", "B", "C", "D"; select all lipid categories for identification; the adduct form of the positive ion mode is selected as [M+H]+，[M+NH4]+，[M+Na]+, and the adduct form of the negative ion mode is selected as [M-H]-，[M-2H]-，[M-HCOO]-. Peak alignment was performed on all lipids with identified results, and the results that were not marked as "reject" were considered for inclusion in the next analysis. The method of peak alignment is set to mean,

retention time deviation is set to 0.1 min, peak filtering is set to New Filter,top rank,all isomer peak, identification level is selected as "A", "B", "C", "D".

The original data exported by LipidSearch is imported into metaX[4] for data preprocessing and subsequent analysis. The contents of data preprocessing include: 1. delete the QC samples that are more than 50% lipid molecules missing and experimental samples that are more than 80% lipid molecules missing; 2. use KNN3 algorithm (k-Nearest Neighbor) to fill the missing values; 3. the probability quotient normalization method (Probabilistic Quotient Normalization, PQN[5]) is used to normalize the data, and the relative peak area is obtained.; use QCRLSC[6] (Quality control-based robust LOESS signal correction, local polynomial regression fitting signal correction) to correct batch effect; delete the lipid molecules with a relative peak area of CV (Coefficient of Variation, coefficient of variation) greater than 30% in all QC samples.

1 http://www.nhc.gov.cn/yzygj/s7653p/202002/d4b895337e19445f8d728fcaf1e3e13a/files/ab6bec7f93e64e7f998d802991203cd6.pdf. Diagnosis and treatment of pneumonia infected by novel coronavirus (revised version of the fifth edition on trial).

2 Brown CT, Davis-Richardson AG, Giongo A, Gano KA, Crabb DB, Mukherjee N, *et al.* Gut microbiome metagenomics analysis suggests a functional model for the development of autoimmunity for type 1 diabetes. PloS one 2011;**6**:e25792.

3 Ling Z, Liu X, Jia X, Cheng Y, Luo Y, Yuan L, *et al.* Impacts of infection with different toxigenic Clostridium difficile strains on faecal microbiota in children. Scientific reports 2014;**4**:7485.

4 Wen B, Mei Z, Zeng C, Liu S. metaX: a flexible and comprehensive software for processing metabolomics data. BMC bioinformatics 2017;**18**:183.

5 Di Guida R, Engel J, Allwood JW, Weber RJ, Jones MR, Sommer U, *et al.* Non-targeted UHPLC-MS metabolomic data processing methods: a comparative investigation of normalisation, missing value imputation, transformation and scaling. Metabolomics : Official journal of the Metabolomic Society 2016;**12**:93.

6 Dunn WB, Broadhurst D, Begley P, Zelena E, Francis-McIntyre S, Anderson N, *et al.* Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. Nature protocols 2011;**6**:1060-83.