Gut Microbiota May Underlie the Predisposition of Healthy Individuals to COVID-19-Sensitive Proteomic Biomarkers

Wanglong Gou
Westlake University
Yuanqing Fu
Westlake University
Liang Yue
Westlake University
Geng-dong Chen
Sun Yat-Sen University
Xue Cai
Westlake University
Menglei Shuai
Westlake University
Fengzhe Xu
Westlake University
Xiao Yi
Westlake University
Hao Chen
Westlake University
Yi Zhu
Westlake University
Mian-li Xiao
Sun Yat-Sen University
Zengliang Jiang
Westlake University
Zelei Miao
Westlake University
Congmei Xiao
Westlake University
Bo Shen
Wenzhou Medical University
Xiaomai Wu

Wenzhou Medical University
Haihong Zhao
Wenzhou Medical University
Wenhua Ling
Sun Yat-Sen University
Jun Wang
Chinese Academy of Sciences
Yu-ming Chen
Sun Yat-Sen University
Tiannan Guo
Westlake University
Ju-Sheng Zheng (zhengjusheng@westlake.edu.cn)
Westlake University https://orcid.org/0000-0001-6560-4890

Research

Keywords: COVID-19, Gut microbiota, proinflammatory cytokines, proteomic

Posted Date: July 22nd, 2020

DOI: https://doi.org/10.21203/rs.3.rs-45991/v1

Abstract

Background: The COVID-19 pandemic is spreading globally with high disparity in the susceptibility of the disease severity. Identification of the key underlying factors for this disparity is highly warranted.

Results: Here we describe constructing a proteomic risk score (PRS) based on 20 blood proteomic biomarkers which related to the progression to severe COVID-19. Among COVID-19 patients, per 10% increment in the PRS was associated with a 57% higher risk of progressing to clinically severe phase (RR=1.57; 95% CI, 1.35-1.82). We demonstrate that in our own cohort of 990 individuals without infection, this proteomic risk score is positively associated with proinflammatory cytokines mainly among older, but not younger, individuals. We further discovered that a core set of gut microbiota could accurately predict the blood proteomic biomarkers of COVID-19 using a machine learning model. The core OTU-predicted PRS had a significant correlation with actual PRS both cross-sectionally (n=132, $p<0.001$) and prospectively (n=169, $p<0.05$). Most of the core OTUs were highly correlated with proinflammatory cytokines. Fecal metabolomics analysis suggested potential amino acid-related pathways linking the above core gut microbiota to inflammation.

Conclusions: Our study suggests that gut microbiota may underlie the predisposition of healthy individuals to COVID-19-sensitive proteomic biomarkers.

Background

With the coronavirus disease 2019 (COVID-19) defined as ‘global pandemic’ and spreading worldwide at an unprecedented speed, more than eleven million individuals have been infected globally since its first detection in December 2019 to early-July 2020 [1]. So far, many research papers have been published to characterize the clinical features of the COVID-19 patients, revealing that those individuals who are older, male or having other clinical comorbidities are more likely to develop into severe COVID-19 cases [2, 3]. Yet, little is known about the potential biological mechanisms or predictors for the susceptibility of the disease.

It is known that COVID-19 is caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), which enters human cells by binding to angiotensin converting enzyme 2 (ACE2) as its receptor [4]. Of note, ACE2 is an important regulator of intestinal inflammation, and that the expression of ACE2 is higher in the ileum and colon than in lung [5, 6]. ACE2 also has a major impact on the composition of gut microbiota, thus affecting cardiopulmonary diseases [7]. Moreover, over 60% of patients with COVID-19 report evidence of gastrointestinal symptoms, such as diarrhoea, nausea and vomiting, and that patients with gastrointestinal symptoms had overall more severe/critical diseases [8–10]. Taken together, the available evidence suggests a potential role of gut microbiota in the susceptibility of COVID-19 progression and severity.

Based on a recent investigation into the blood biomarkers of COVID-19 patients, we identified a set of [illegible]lict the progression to severe COVID-19 among infected

patients [11]. The newly discovered proteomic biomarkers may help early prediction of severe COVID-19. However, the question remains as to whether this set of proteomic biomarkers could be used in healthy (non-infected) individuals to help explain the disease susceptibility. It is also unclear whether gut microbiota could regulate these blood proteomic biomarkers among healthy individuals.

To address the above unresolved questions, we integrated blood proteomics data from 31 COVID-19 patients and multi-omics data from a Chinese population without infection living in Guangzhou, involving 2413 participants (Fig. 1; Figure S1; Table S1). Based on the COVID-19 patient data, we constructed a blood proteomic risk score (PRS) for the prediction of COVID-19 progression to clinically severe phase. Then, among 990 healthy individuals with the data of proteome and blood inflammatory biomarkers, we investigated the association of the COVID-19-related PRS with inflammatory biomarkers as a verification of the PRS with disease susceptibility in normal non-infected individuals. Next, we identified core gut microbiota features which predicted the blood proteomic biomarkers of COVID-19 using a machine-learning model. We conducted further fecal metabolomics analysis to reveal potential biological mechanisms linking gut microbiota to the COVID-19 susceptibility among non-infected individuals. Finally, we demonstrated the contribution of 40 host and environmental factors to the variance of the above identified core gut microbiota features.

Methods

Study design

COVID-19 patients

Detailed information about the COVID-19 patients and proteomics data set is described in our recent publication [11]. In the present analysis, we included data from 31 COVID-19 patients, among which 13 were severe patients and 18 were non-severe patients from Taizhou Public Health Medical Center. Serum samples of these patients were analyzed by TMTpro 16 plex-based quantitative proteomics technology. All the patients were diagnosed between January 23 and February 4, 2020. According to the Chinese Government Diagnosis and Treatment Guideline for COVID-19, the COVID-19 patients were classified into four groups, (1) mild (mild symptoms without pneumonia); (2) typical (fever or respiratory tract symptoms with pneumonia); (3) severe (fulfill any of the three criteria: respiratory distress, respiratory rate $\geq$ 30 times/min; mean oxygen saturation $\leq$ 93% in resting state; arterial blood oxygen partial pressure/oxygen concentration $\leq$ 300 mmHg); and (4) critical (fulfill any of the three criteria: respiratory failure and require mechanical ventilation; shock incidence; admission to ICU with other organ failure). We treated mild and typical patients as a non-severe COVID-19 group, and the others a severe COVID-19 group.

Healthy participants, sample collection, and clinical metadata

A total of 2413 healthy non-infected individuals from the community-based Guangzhou Nutrition and Health Study (GNHS) are involved in the present study, which mainly consists of a subset of subjects with proteomic data at baseline (n = 990) and a subset of subjects with gut microbiome and metabolome data at a follow-up visit (n = 2172, within which 301 individuals also had proteomic data). The detailed study designs of GNHS have been reported previously [12]. Briefly, participants were enrolled between 2008 and 2013, and followed up to May 2018. Blood samples were collected at enrollment and follow-up visits, and stool samples were collected only during follow-up visits. All the blood samples were collected as venous whole blood in the early morning before diet using serum separation tubes. The blood samples were centrifuged at 3,500 rpm for 10 min for serum collection. The serum samples were frozen at -80 °C. The stool samples were collected at a local study site within the School of Public Health at Sun Yat-sen University, and were transferred to a -80 °C facility within 4 hours after collection.

Details method for the metadata measurements, proteomic analysis, measurement of inflammatory biomarkers, microbiome and metabolomic analysis for GNHS was provided in Supplemental text.

Bioinformatic and statistical analysis

Data imputation and presentation

Missing values in proteomic features were imputed with 50% of the minimal value. Data are presented as mean ± SD or percentage as indicated. Statistical tests used to compare conditions are indicated in figure legends. Unless otherwise stated, statistical analysis was performed using Python 3.7, R software (version 3.6.1, R foundation for Statistical Computing, Austria), and Stata 15 (StataCorp, College Station, TX, USA).

Construction of proteomic risk score (PRS)

We used 20 out of 22 previously identified proteomic biomarkers to construct a proteomic risk score (PRS) for severe COVID-19 in COVID-19 patients and healthy participants.

$$PRS_i = \sum_{j=1}^{20} \beta_j x_{ij}$$

Where, PRS_i is a proteomic risk score for individual i, 20 is the number of proteins involved the score construction, x_{ij} is the Z score of the relative abundance of the protein j for individual i. β is 1 or -1 depending on the association between the protein j and risk of progressing to clinically severe phase (1, up-regulated in severe patients, -1, down-regulated in severe patients).

Association of PRS with the risk of progressing to clinically severe phase

Poisson regression model was used to examine the association of PRS with the risk of progressing to clinically severe phase among 31 COVID-19 patients (18 non-severe patients; 13 severe patients), adjusting for age, sex and BMI.

Correlation between PRS and pro-inflammatory biomarkers

Spearman correlation analysis was used to examine the correlation between PRS and pro-inflammatory biomarkers (i.e., hsCRP, IL-1β, IL-6 and TNF-α). $p < 0.05$ was considered as statistically significant.

Machine learning algorithms for identifying microbial features to predict PRS

A 10-fold cross-validation (CV) implementation of gradient boosting framework —LightGBM [13]and SHAP (Shapley Additive exPlanations) [14] was used to link input gut microbial features with PRS. Gradient boosting has been widely used in prediction for tabular data in biomedical fields [15, 16] and SHAP has been theoretically verified as the only consistent and locally accurate method to interpret machine learning results [17, 18]. A 10-fold CV predict implementation was used to generate a OTU-predicted PRS value for each participant. In this approach, each LightGBM model is trained on 90% of the cohort with 10-fold CV, and PRS is predicted for the 10% of the participants who were not used for model optimization. This process is repeated ten-fold resulting in a test PRS set for each participant and ten different average absolute SHAP value for each OTUs. The top 20 ranked OTUs based the sum of the average absolute SHAP value across ten-fold were included in further analysis. Pearson r was calculated using actual PRS and predicted PRS for the entire cohort. We also compared the predictive performance for the top 20 ranked OTUs, demographic characteristics and laboratory tests (age, BMI, sex, blood pressure and blood lipids). Our predictor is based on code adapted from the sklearn 0.15.2 lightgbm regression [19].

The relationship between the identified core OTUs and host inflammatory cytokines

Spearman correlation analysis was used to examine the correlation between PRS and cytokines (i.e., IL-1β, IL-2, IL-4, IL-6, IL-8, IL-10, IL-12p70, IL-13, TNF-α and IFN-γ). $p < 0.05$ was considered as statistically significant.

Relationship between OTUs and fecal metabolites

Prior to the analysis, we excluded the participants with T2D medication use, and all fecal metabolites were natural logarithmic transformed to reduce skewness of traits distributions. Similarly, to reduce skewness of the distribution of microbial taxa counts, we first added 1 to all OTUs and then performed natural log transformation. The relationship between fecal metabolites and microbial OTUs was assessed by linear regression analysis while adjusting for age, sex, BMI. Multiple testing was adjusted using Benjamini and Hochberg method, with a false discovery rate (FDR) of < 0.05 being considered

statistically significant. Metabolites showed significant associations with more than half of the selected microbial OTUs, were used for subsequent pathway analysis using MetaboAnalyst 4.0 [20].

Associations of host and environmental factors with gut microbial features

We assessed how many variations in the identified core OTUs composition (Bray-Curtis distance) can be explained by host and environmental factors (40 factors) using the function *adonis* from the R package vegan. The p value was determined by 1000x permutations. The total variation explained was also calculated per category (demographic/clinical category and dietary/nutritional factors) and for all factors together. Spearman correlation analysis was used to assess the potential effect of each factor on each of the core OTU. Multiple testing was adjusted using Benjamini and Hochberg method, with a false discovery rate (FDR) of < 0.05 being considered statistically significant.

Detailed information about the dataset at each step of analyses among the healthy individuals from GNHS was provided in Supplemental text.

Results

Predictive proteomic profile for severe COVID-19 is correlated with inflammatory factors among healthy individuals

First, we derived and validated a protein risk score which related to the progression to severe COVID-19. Based on a prior serum proteomic profiling of COVID-19 patients, 22 proteomic biomarkers contributed to the prediction of progression to severe COVID-19 status (18 non-severe cases and 13 severe cases) [11].Using this cohort, we constructed a blood PRS among the 31 COVID-19 patients based on 20 proteomic biomarkers (Table S2). We only used 20 of the 22 proteins for our PRS construction because 2 proteins were unavailable in our large proteomics database among non-infected participants for the further analysis. Among the COVID-19 patients, Poisson regression analysis indicated that per 10% increment in the PRS there was associated a 57% higher risk of progressing to clinically severe phase (RR, 1.57; 95% CI, 1.35–1.82; Fig. 2A), in support of the PRS as being a valid proxy for the predictive biomarkers of severe COVID-19.

To explore the potential implication of the PRS among non-COVID-19 individuals, we constructed the PRS using the same set of 20 blood proteins among a cohort of non-infected participants with data of both proteomics and inflammatory markers (n = 990). The blood proteomic data was based on the baseline serum samples of the cohort (Figure S1). We investigated the correlation between the PRS and blood inflammatory markers IL-1β, IL-6, TNF-α and hsCRP. The PRS had a significantly positive correlation with serum concentrations of hsCRP and TNF-α ($p < 0.001$ and $p < 0.05$, respectively), but not other markers (Fig. 2B). As age and sex are very important factors related to the susceptibility to SARS-CoV-2 infection,

we performed subgroup analysis stratified by age (< 58 years vs. ≥58 years, with 58 years as the median age of this cohort) and sex. Interestingly, we found that higher PRS was significantly correlated with higher serum concentrations of all the aforementioned inflammatory markers among older individuals (> 58 years, n = 493), but not among younger individuals (≤ 58 years, n = 497) (Fig. 2B and 2C). The PRS did not show any differential association with the inflammatory markers by sex (Figure S2). Whether the identified proteomic changes causally induce immune activation or consequences of the immune response are not clear at present, but the finding supports the hypothesis that the PRS may act as a biomarker of unbalanced host immune system, especially among older adults.

Core microbiota features predict COVID-19 proteomic risk score and host inflammation

To investigate the potential role of gut microbiota in the susceptibility of healthy individuals to COVID-19, we next explored the relationship between the gut microbiota and the above COVID-19-related PRS in a sub-cohort of 301 participants with measurement of both gut microbiota (16 s rRNA) and blood proteomics data (Figure S1). Gut microbiota data were collected and measured during a follow-up visit of the cohort participants, with a cross-sectional subset of the individuals (n = 132) having blood proteomic data at the same time point as the stool collection and another independent prospective subset of the individuals (n = 169) having proteomic data at a next follow-up visit ~ 3 years later than the stool collection.

Among the cross-sectional subset, using a machine learning-based method: LightGBM and a very conservative and strict tenfold cross-validation strategy, we identified 20 top predictive operational taxonomic units (OTUs), and this subset of core OTUs was strongly predictive of PRS (Cross-validated Pearson's $r = 0.59$, $p < 0.001$ across ten cross-validations). The predictive capacity for PRS based on the core OTUs substantially outperformed that of demographic characteristics and laboratory tests including age, BMI, sex, blood pressure and blood lipids (Pearson's $r = 0.154$, $p = 0.087$) (Fig. 3A). The list of these core OTUs along with their taxonomic classification is provided in Table S3. These OTUs were mainly assigned to *Bacteroides* genus, *Streptococcus* genus, *Lactobacillus* genus, *Ruminococcaceae* family, *Lachnospiraceae* family and *Clostridiales* order.

Additionally, we used co-inertia analysis (CIA) to further test co-variance between the 20 identified core OTUs and 20 predictive proteomic biomarkers of severe COVID-19, outputting a RV coefficient (ranged from 0 to 1) to quantify the closeness. The results indicated a close association of these OTUs with the proteomic biomarkers (RV = 0.12, $p < 0.05$) (Figure S3A). When replicating this analysis stratified by age, significant association was observed only among older participants (age ≥ 58, n = 66; RV = 0.22, $p < 0.05$) (Figure S3B and S3C).

Importantly, the above results from cross-sectional analyses were successfully replicated in the independent prospective subset of 169 individuals, which showed a Pearson's r of 0.18 between the core

OTUs-predicted PRS versus actual PRS ($p < 0.05$), also outperforming the predictive capacity of the above demographic characteristics and laboratory tests (Pearson's $r = 0.08$, $p = 0.31$) (Fig. 3A). These findings support that change in the gut microbiota may precede the change in the blood proteomic biomarkers, inferring a potential causal relationship.

To further verify the reliability of these core OTUs, in another larger independent sub-cohort of 366 participants (Figure S1), we examined the cross-sectional relationship between the core OTUs and 10 host inflammatory cytokines including IL-1β, IL-2, IL-4, IL-6, IL-8, IL-10, IL-12p70, IL-13, TNF-α and IFN-γ, and found 11 microbial OTUs were significantly associated with the inflammatory cytokines (Fig. 3B). Specifically, *Bacteroides* genus, *Streptococcus* genus and *Clostridiales* order were negatively correlated with most of the tested inflammatory cytokines, whereas *Ruminococcus* genus, *Blautia* genus and *Lactobacillus* genus showed positive associations.

Fecal metabolome may be the key to link the PRS-related core microbial features and host inflammation

We hypothesized that the influences of the core microbial features on the PRS and host inflammation were driven by some specific microbial metabolites. So we assessed the relationship between the core gut microbiota and fecal metabolome among 987 participants, whose fecal metabolomics and 16 s rRNA microbiome data were collected and measured at the same time point during the follow-up visit of the participants (Figure S1). After correction for the multiple testing (FDR < 0.05), a total of 183 fecal metabolites had significant correlations with at least one selected microbial OTU. Notably, 45 fecal metabolites, mainly within the categories of amino acids, fatty acids and bile acids, showed significant associations with more than half of the selected microbial OTUs (Fig. 4A), these metabolites might play a key role in mediating the effect of the core gut microbiota on host metabolism and inflammation.

Based on these key metabolites, we performed metabolic pathway analysis to elucidate possible biological mechanisms. The results showed that these 45 fecal metabolites were mainly enriched in three pathways, namely aminoacyl-tRNA biosynthesis pathway, arginine biosynthesis pathway, and valine, leucine and isoleucine biosynthesis pathway (Fig. 4B). There were 15 fecal metabolites involved in the aminoacyl-tRNA biosynthesis pathway, which is responsible for adding amino acid to nascent peptide chains and is a target for inhibiting cytokine stimulated inflammation (Fig. 4C). Additionally, 4 metabolites were associated with arginine biosynthesis pathway and 3 metabolites were enriched in valine, leucine and isoleucine (known as branch-chain amino acids, BCAAs) biosynthesis pathway (Fig. 4C).

Host and environmental factors are correlated with the PRS-related core microbial OTUs

As demographic, socioeconomic, dietary and lifestyle factors may all be closely related to the gut microbiota, we explored the variance contribution of these host and environmental factors for the identified core OTU composition. A total of 40 items belonging to two categories (i.e., demographic/clinical factors and dietary/nutritional factors) were tested (Fig. 5), which together explained 3.6% of the variation in interindividual distance of the core OTU composition (Bray-Curtis distance). In the demographic/clinical factors which explained 2.4% of the variation, we observed associations of 9 items (i.e., sex, education, physical activity, diastolic blood pressure, blood glucose, blood lipids and medicine use for type 2 diabetes) with inter-individual distances in the core OTU composition (PERMANOVA, $p < 0.05$; Fig. 5). While in the dietary/nutritional category (1.1% variance was explained), only dairy consumption significantly contributed to the variance of the core OTU composition.

Discussion

Our findings suggest that, among healthy non-infected individuals, gut microbial features are highly predictive of the blood proteomic biomarkers of severe COVID-19 disease. The disruption of the corresponding gut microbiome features may potentially predispose healthy individuals to abnormal inflammatory status, which may further account for the COVID-19 susceptibility and severity. The fecal metabolomics analysis reveals that amino acid-related pathway may provide the key link between the identified core gut microbiota, inflammation and COVID-19 susceptibility. Furthermore, modifications on host and environmental factors are likely to influence the above core gut microbiota compositions.

Accumulating evidence suggests that “cytokine storm”, an excessive production of inflammatory cytokines, may be an important mechanism leading to the severity and death of COVID-19 patients [2, 21]. Therefore, anticytokine therapy for the suppression of the hyperinflammatory status of the patients is a recommended strategy to treat severe COVID-19 patients [22, 23]. Among the 20 proteomic predictors of severe COVID-19, several most upregulated proteins are activated acute phase proteins, including serum amyloid A-1 (SAA1), SAA2, SAA4, alpha-1-antichymotrypsin (SERPINA3), complement 6 (C6) and complement factor B (CFB) [11]. These proteins may be activated together with proinflammatory cytokines such as IL-6 and TNF-α following the invasion of the SARS-CoV-2. Therefore, this set of proteomic biomarkers may serve as an important biomarker or therapeutic target for treating SARS-CoV-2 infection. Beyond the previous data from the COVID-19 patients, our current study based on data from healthy non-infected participants consistently supports that the proteomic biomarkers (integrated into a score) are positively associated with proinflammatory cytokines, especially among those with an older age. These results imply that the proteomic changes my precede the progression of COVID-19 to severe phase. Moreover, our finding of more significant associations between PRS and proinflammatory cytokines among older people agree with the observation during COVID-19 outbreak that older individuals are more susceptible to the virus, leading to severity of the disease, due to the induced hyperinflammation or “cytokine storm” [3, 24].

In the present study, the core gut microbial features (20 OTUs), with a satisfied performance, outperform tests in predicting the blood proteomic biomarkers, which

highlights a potential role of gut microbiota in regulating the susceptibility of COVID-19 among normal individuals. Some of our identified gut microbes was consistent with those from recent study of COVID-19 patients [25]. In fact, maintaining gut homeostasis has been suggested as a treatment option in the "Diagnosis and Treatment Plan of Corona Virus Disease 2019 (Tentative Sixth Edition)" issued by National Health Commission of China, as to keep the equilibrium for intestinal microecology and prevent secondary bacterial infection [26]. Growing evidence has shown that microbiota plays a fundamental role on the induction, training and function of the host immune system, and the composition of the gut microbiota and its activity are involved in production of inflammatory cytokines [27, 28]. Prior studies reported that *Lactobacillus* genus was positively associated with IL-6 and IFN-γ, while *Blautia* genus was positively associated with IL-10 [29–31]; these relationships were replicated in our study. Besides, we found the PRS-related OTUs belonging to *Bacteroides* genus and *Streptococcus* genus were negatively associated with most proinflammatory factors. These results further support the reliability of the selected core OTUs.

Fecal metabolomics analyses for the identified core gut microbial OTUs suggest that these OTUs may be closely associated with amino acid metabolism, especially aminoacyl-tRNA biosynthesis pathway, arginine biosynthesis pathway, and valine, leucine and isoleucine biosynthesis pathway. As metabolic stress pathways and nutrient availability instruct immunity, amino acid levels in the tissue microenvironment are central to the maintenance of immune homeostasis [32]. Amino acid insufficiency will cause depletion of available aminoacylated tRNA, which is essential for the host to sense amino acid limitation and immune response [33–35]. A recent study on several mammalian cell models reported that when aminoacyl-tRNA synthetase was inhibited, the cytokine stimulated proinflammatory response would be substantially suppressed, and a single amino acid depletion, such as arginine or histidine, could also suppress the cytokine induced immune response [36].Thus the identified pathways regulating in aminoacyl-tRNA biosynthesis and arginine biosynthesis may be both involved in the inflammatory response. Additionally, arginine and BCAAs (i.e., valine, leucine and isoleucine), were also reported regulating innate and adaptive immune responses and enhancing intestinal development [37]. Collectively, these key roles that amino acids play in the immunoregulation may help explain how the PRS-related core OTUs modulate host inflammation via amino acid metabolism. Furthermore, given the high expression of ACE2 in the ileum and colon, and the role of ACE2 as a key regulator of dietary amino acid homeostasis and innate immunity [5, 6], ACE2 may be another key mediator between gut microbiota and host inflammation. However, whether and how ACE2 may mediate the association between gut microbiota and COVID-19 severity warrants further mechanistic study.

We observed that several host demographic and clinical factors had a strong effect on the identified core OTU composition, among which drug use and metabolic phenotypes had been widely reported correlating with gut microbiome composition [38–40]. Although these observations were quite crude, it gave us an overview of the potential influence of host and environmental factors on the PRS-related gut microbiota matrix. Those known factors contributed to the COVID-19 susceptibility also contributed to the variance of the gut microbiota, including age, sex, and indicators of clinical comorbidities (blood pressure, glucose [illegible] oprotein cholesterol, and diabetes medication).

The strength of our study lies in our analysis of multi-omics dataset (blood proteomics, gut microbiota, fecal metabolome) from a large healthy human cohort study, and cross-referenced with COVID-19 specific proteomic risk score identified in recent study[11]. Our results lead to the hypothesis that gut microbiota may play an important role in regulating the predisposition of healthy individual to the COVID-19 via affecting the fecal amino acid-related pathway and host inflammation, although the detailed mechanism is yet to be discovered. A major limitation of the present study is that we don't directly investigated the association of gut microbiota with the COVID-19 susceptibility among COVID-19 patients. Nevertheless, some of our identified gut microbes are consistent with those from recent study of COVID-19 patients[25], and our study among non-infected healthy participants may be more generalizable in terms of the prevention of the disease and disease severity.

Conclusions

In the global crisis of COVID-19, a wide disparity in the susceptibility of the disease or disease progression has been observed. Our results provide important evidence and suggestions about the potential biological mechanism behind the diverse susceptibility among different groups of people. The discovered core gut microbial features and related metabolites may serve as a potential preventive/treatment target for intervention especially among those who are susceptible to the SARS-CoV-2 infection. They could also serve as potential therapeutic targets for drug development.

Abbreviations

COVID-19: coronavirus disease 2019; PRS: proteomic risk score; OTUs: Operational taxonomic units; SARS-CoV-2: Severe acute respiratory syndrome coronavirus 2; ACE2: Angiotensin converting enzyme 2; GNHS: Guangzhou Nutrition and Health Study; CV: Cross-validation; SHAP: Shapley additive explanations; FDR: False discovery rate; BCAA: Branch-chain amino acids;

Declarations

Ethics approval and consent to participate

This study has been approved by the Ethical/Institutional Review Board of Taizhou Public Health Medical Center (K20200209), the Ethics Committee of the School of Public Health at Sun Yat-sen University (2018048) and Ethics Committee of Westlake University (20190114ZJS0003), and all participants provided written informed consent.

Consent for publication

Not applicable.

Availability of data and materials

The raw data of 16 S rRNA gene sequences are available at CNSA (https://db.cngb.org/cnsa/) of CNGBdb at accession number CNP0000829. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) via the iProX partner repository with the dataset identifier PXD019675.

Competing interests

The authors declare no competing financial interests.

Funding

This study was supported by the National Natural Science Foundation of China (81903316, 81773416, 81972492, 21904107, 81672086), Zhejiang Ten-thousand Talents Program (101396522001), Zhejiang Provincial Natural Science Foundation for Distinguished Young Scholars (LR19C050001), the 5010 Program for Clinical Researches (2007032) of the Sun Yat-sen University, Hangzhou Agriculture and Society Advancement Program (20190101A04), and Tencent foundation (2020). The funder had no role in study design, data collection and analysis, decision to publish, or writing of the manuscript.

Author Contributions

Conceptualization, J.S.Z.; Methodology, W.G. and Y.F.; Formal Analysis, W.G., Y.F., L.Y., G.D.C, X.C., M.S., and F.X.; Investigation, M.L.X., B.S, X.W., H.Z., and W.H.L.; Data curation, X.Y., H.C., Y.Z., Z.J., Z.M. and C.X.; Resources, Y.M.C., T.G., J.S.Z; Writing, Y.F. and J.S.Z.; Writing-Review & Editing, J.S.Z, T.G., J.W, Y.F. and W.L.G; Visualization, M.S. and F.X.; Supervision, J.S.Z., Y.M.C., and T.G.; Funding Acquisition, J.S.Z., Y.M.C. and T.G.

Acknowledgements

We thank all the participants for their collaboration as well as Dr. C.R. Palmer for his invaluable comments to this study; and Westlake University Supercomputer Center for assistance in data storage and computation; and all the medical staff members who are on the front line fighting against COVID-19.

References

1. World Health Organization (2020). Coronavirus disease (COVID-19) Pandemic. https://www.who.int/emergencies/diseases/novel-coronavirus-2019. 2020;Accessed A:2020.

2. 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 (London, England). 2020;395:497–506.
3. 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 (London, England). 2020;395:507–13.
4. Yan R, Zhang Y, Li Y, Xia L, Guo Y, Zhou Q. Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. Science (New York, NY). United States; 2020;367:1444–8.
5. Hashimoto T, Perlot T, Rehman A, Trichereau J, Ishiguro H, Paolino M, et al. ACE2 links amino acid malnutrition to microbial ecology and intestinal inflammation. Nature. 2012;487:477–81.
6. Zhang H, Kang Z, Gong H, Xu D, Wang J, Li Z, et al. Digestive system is a potential route of COVID-19 : an analysis of single- cell coexpression pattern of key proteins in viral entry process. 2020;1–9.
7. Cole-Jeffrey CT, Liu M, Katovich MJ, Raizada MK, Shenoy V. ACE2 and Microbiota: Emerging Targets for Cardiopulmonary Disease Therapy. Journal of cardiovascular pharmacology. 2015;66:540–50.
8. Lin L, Jiang X, Zhang Z, Huang S, Zhang Z, Fang Z, et al. Gastrointestinal symptoms of 95 cases with SARS-CoV-2 infection. Gut. England; 2020;
9. Jin X, Lian J-S, Hu J-H, Gao J, Zheng L, Zhang Y-M, et al. Epidemiological, clinical and virological characteristics of 74 cases of coronavirus-infected disease 2019 (COVID-19) with gastrointestinal symptoms. Gut. 2020;
10. Ng SC, Tilg H. COVID-19 and the gastrointestinal tract: more than meets the eye. Gut. England; 2020;
11. Shen B, Yi X, Sun Y, Bi X, Du J, Zhang C, et al. Proteomic and Metabolomic Characterization of COVID-19 Patient Sera. Cell. 2020;182:1–14.
12. Zhang Z-Q, He L-P, Liu Y-H, Liu J, Su Y-X, Chen Y-M. Association between dietary intake of flavonoid and bone mineral density in middle aged and elderly Chinese women and men. Osteoporosis international : a journal established as result of cooperation between the European Foundation for Osteoporosis and the National Osteoporosis Foundation of the USA. England; 2014;25:2417–25.
13. Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, et al. LightGBM: A Highly Efficient Gradient Boosting Decision Tree. NIPS. 2017;9.
14. Lee SMLS-I. A Unified Apporach to Interpreting Model Predictions. NIPS. 2017;
15. Lundberg SM, Nair B, Vavilala MS, Horibe M, Eisses MJ, Adams T, et al. Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. Nature Biomedical Engineering. Springer US; 2018;2:749–60.
16. Artzi NS, Shilo S, Hadar E, Rossman H, Barbash-Hazan S, Ben-Haroush A, et al. Prediction of gestational diabetes based on nationwide electronic health records. Nature medicine. United States; 2020;26:71–6.
17. Lundberg SM, Erion GG, Lee S-I. Consistent Individualized Feature Attribution for Tree Ensembles. NIPS. 2018;

18. Lundberg SM, Erion G, Chen H, DeGrave A, Prutkin JM, Nair B, et al. From Local Explanations to Global Understanding with Explainable AI for Trees. Nature machine intelligence. 2020;2:56–67.
19. Pedregosa F, Weiss R, Brucher M. Scikit-learn : Machine Learning in Python. 2011;12:2825–30.
20. Chong J, Soufan O, Li C, Caraus I, Li S, Bourque G, et al. MetaboAnalyst 4.0: towards more transparent and integrative metabolomics analysis. Nucleic acids research. 2018;46:W486–94.
21. Yang Y, Shen C, Li J, Yuan J, Yang M, Wang F, et al. Exuberant elevation of IP-10, MCP-3 and IL-1ra during SARS-CoV-2 infection is associated with disease severity and fatal outcome. medRxiv [Internet]. 2020;https://doi.org/10.1101/2020.03.02.20029975. Available from: http://medrxiv.org/content/early/2020/03/06/2020.03.02.20029975.abstract
22. Monteleone G, Sarzi-puttini PC, Ardizzone S. Preventing COVID-19-induced pneumonia with anticytokine therapy. The Lancet Rheumatology. Elsevier Ltd; 2020;9913:30092–8.
23. Mehta P, McAuley DF, Brown M, Sanchez E, Tattersall RS, Manson JJ. COVID-19: consider cytokine storm syndromes and immunosuppression. Lancet (London, England). England; 2020. p. 1033–4.
24. Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. Lancet (London, England). England; 2020;395:1054–62.
25. Zuo T, Zhang F, Lui GCY, Yeoh YK, Li AYL, Zhan H, et al. Alterations in Gut Microbiota of Patients With COVID-19 During Time of Hospitalization. Gastroenterology. 2020;
26. National Health Commission (NHC) of the PRC NA of TCM of the P. Diagnosis and Treatment Plan of Corona Virus Disease 2019 (Tentative Sixth Edition). Guidance for corona virus disease 2019: prevention, control and diagnosis and management. 2020;Beijing, China: People's Medical Publishing House,.
27. Belkaid Y, Hand TW. Role of the microbiota in immunity and inflammation. Cell. 2014;157:121–41.
28. Cani PD, Jordan BF. Gut microbiota-mediated inflammation in obesity: a link with gastrointestinal cancer. Nature reviews Gastroenterology & hepatology. England; 2018;15:671–82.
29. Yoshida K, Matsumoto T, Tateda K, Uchida K, Tsujimoto S, Yamaguchi K. Induction of interleukin-10 and down-regulation of cytokine production by Klebsiella pneumoniae capsule in mice with pulmonary infection. Journal of medical microbiology. England; 2001;50:456–61.
30. Pohjavuori E, Viljanen M, Korpela R, Kuitunen M, Tiittanen M, Vaarala O, et al. Lactobacillus GG effect in increasing IFN-gamma production in infants with cow's milk allergy. The Journal of allergy and clinical immunology. United States; 2004;114:131–6.
31. Jiang Y, Lü X, Man C, Han L, Shan Y, Qu X, et al. Lactobacillus acidophilus induces cytokine and chemokine production via NF-κB and p38 mitogen-activated protein kinase signaling pathways in intestinal epithelial cells. Clinical and vaccine immunology : CVI. 2012;19:603–8.
32. Murray PJ. Amino acid auxotrophy as a system of immunological control nodes. Nature immunology. 2016;17:132–9.

33. Brown A, Fernández IS, Gordiyenko Y, Ramakrishnan V. Ribosome-dependent activation of stringent control. Nature. 2016;534:277–80.
34. Harding HP, Zhang Y, Zeng H, Novoa I, Lu PD, Calfon M, et al. An integrated stress response regulates amino acid metabolism and resistance to oxidative stress. Molecular cell. United States; 2003;11:619–33.
35. Brown M V, Reader JS, Tzima E. Mammalian aminoacyl-tRNA synthetases: cell signaling functions of the protein translation machinery. Vascular pharmacology. United States; 2010;52:21–6.
36. Kim Y, Sundrud MS, Zhou C, Edenius M, Zocco D, Powers K, et al. Aminoacyl-tRNA synthetase inhibition activates a pathway that branches from the canonical amino acid response in mammalian cells. Proceedings of the National Academy of Sciences of the United States of America. United States; 2020;
37. Zhang S, Zeng X, Ren M, Mao X, Qiao S. Novel metabolic and physiological functions of branched chain amino acids: a review. Journal of animal science and biotechnology. 2017;8:10.
38. Vich Vila A, Collij V, Sanna S, Sinha T, Imhann F, Bourgonje AR, et al. Impact of commonly used drugs on the composition and metabolic function of the gut microbiota. Nature communications. 2020;11:362.
39. Gilbert JA, Blaser MJ, Caporaso JG, Jansson JK, Lynch S V, Knight R. Current understanding of the human microbiome. Nature medicine. 2018;24:392–400.
40. Cabreiro F, Au C, Leung K-Y, Vergara-Irigaray N, Cochemé HM, Noori T, et al. Metformin retards aging in C. elegans by altering microbial folate and methionine metabolism. Cell. 2013;153:228–39.

Figures

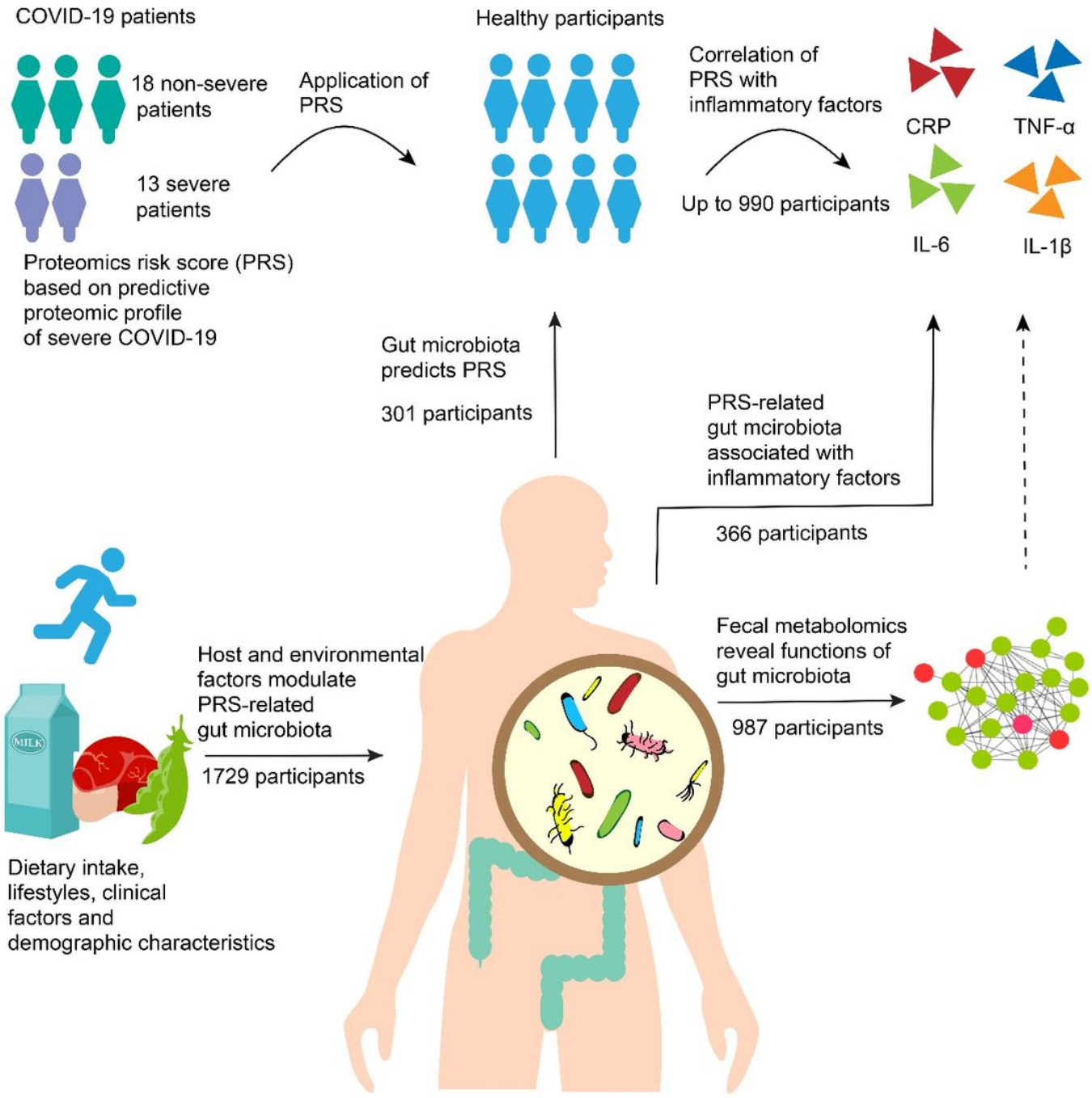


Figure 1

Study design and analysis pipeline. Study overview. 1) constructing a novel COVID-19 blood proteomic risk score (PRS) among 31 COVID-19 patients (18 non-severe cases and 13 severe cases). 2) Applicating the PRS in healthy participants, and further linking it to host inflammatory status (n=990). 3) Investigating the potential role of gut microbiota in predicting the PRS of COVID-19 based on a machine-learning method (n=301). 4) Assessing the relationships between the PRS-related gut microbiota and inflammatory factors (n=336). 5) Fecal metabolomics analysis reveals function of gut microbiota on host

metabolism (n=987). 6) Investigating the impact of host and environmental factors on PRS-related core microbial OTUs (n=1729).

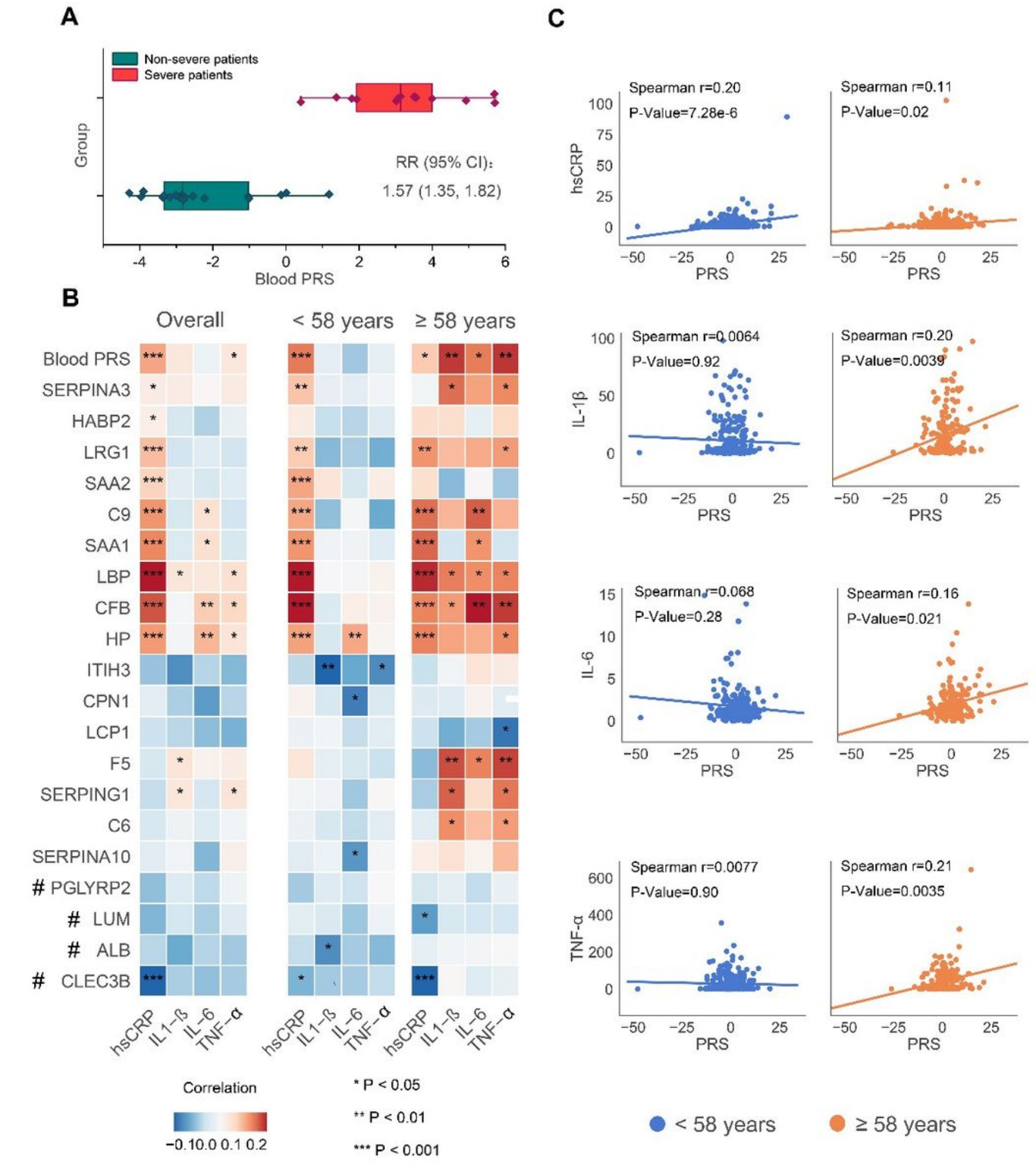


Figure 2

Predictive proteomic profile for severe COVID-19 is correlated with pro-inflammatory factors among healthy individuals. (A) Protein risk score (PRS) related to the progression to severe COVID-19. Among the [MathJax]/jax/output/CommonHTML/jax.js alysis indicated that per 10% increment in the PRS there was

associated a 57% higher risk of progressing to clinically severe phase. (B) The associations of COVID-19-related blood proteomic biomarkers and proteomic risk score (PRS) with host inflammatory markers among the non-infected participants (990 participants). We also examined the correlation of the above blood proteomic biomarkers and PRS with host inflammatory markers stratified by the median age of participants (<58 years or ≥58 years). The color of the heatmap indicates the spearman correlation coefficients (blue-negative, red-positive). # protein down-regulated in severe patients, else, up-regulated. (C) The correlation of the PRS with individual host inflammatory markers stratified by the median age of participants (<58 years or ≥58 years).

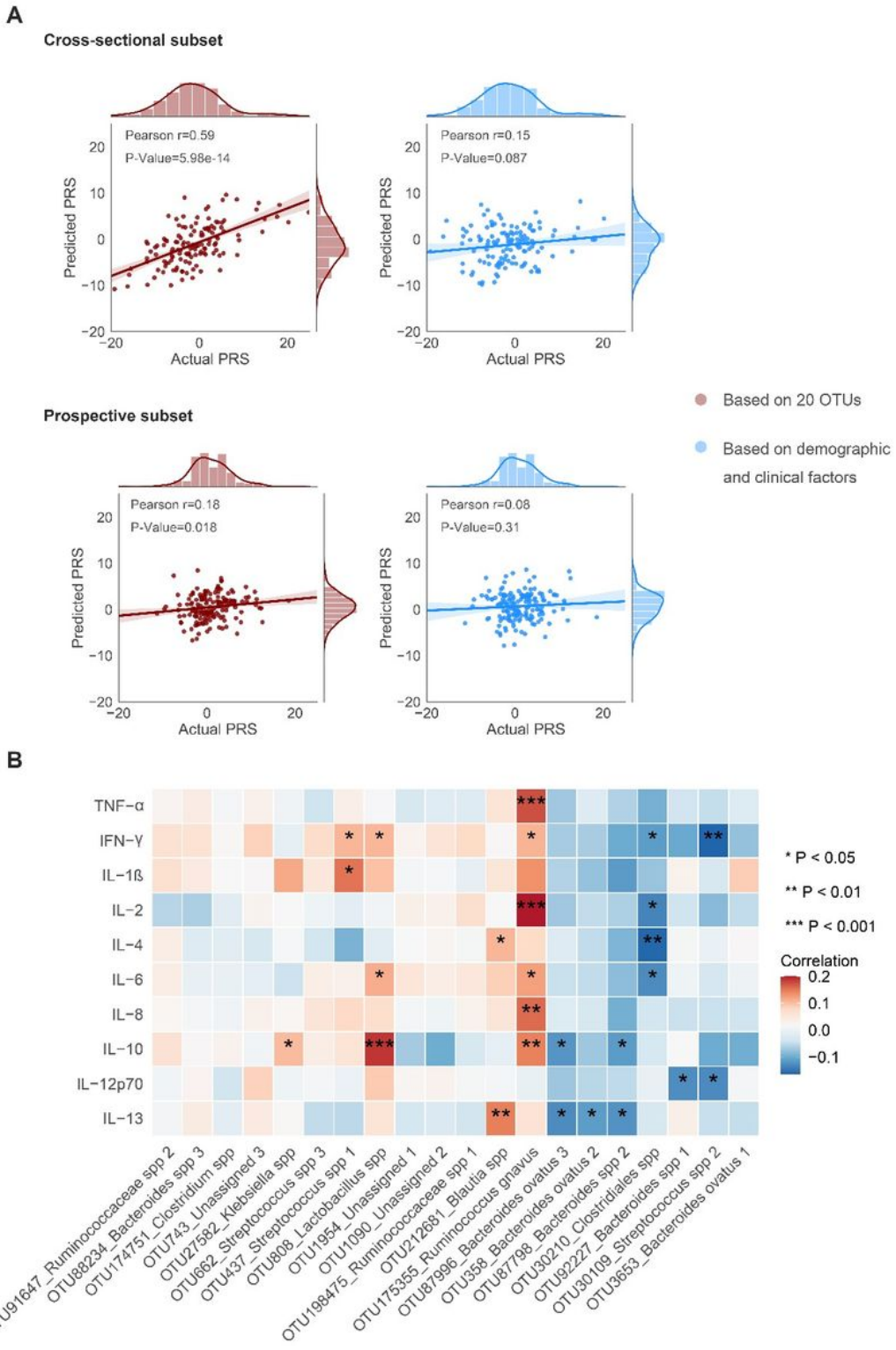
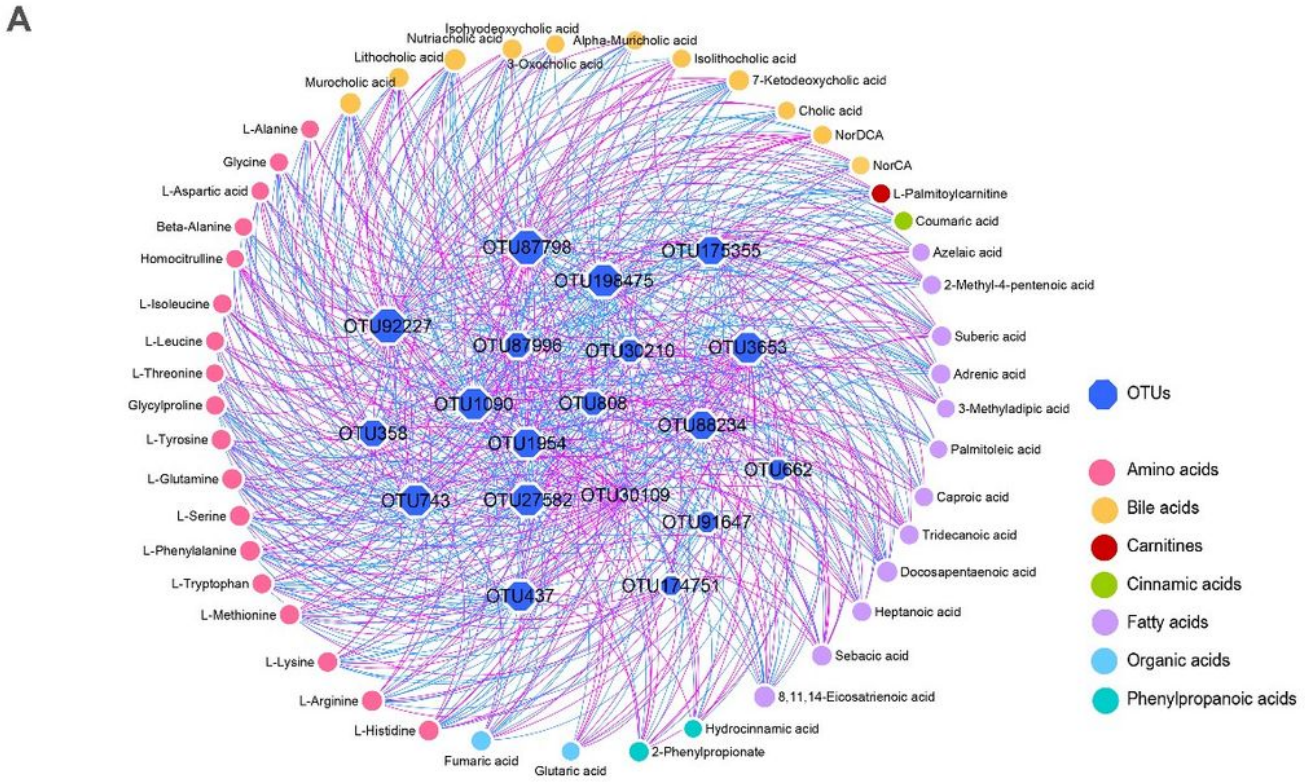


Figure 3

Core microbiota features predict COVID-19 proteomic risk score (PRS) and host inflammation. (A) Plots of out-of-sample predicted PRS versus actual PRS based on top 20 ranked OTUs or demographic/clinical factors (age, sex BMI, fasting glucose, HDL, LDL, TC, TG, DBP, and SBP) using LightGBM with 10-fold cross-validation. In this approach, each LightGBM model is trained on 90% of the cohort with 10-fold CV, icipants who were not used for model optimization. The plots

in the first row indicate the model performance among cross-sectional subset of individuals (n=132); the plots in the second row indicate the model performance among prospective subset of individuals (n=169). Pearson r of predicted values versus actual values, and corresponding P-value across the 10 cross validations are shown in the figures. (B) The correlation of the core microbial OTUs and host inflammatory cytokines (n=336). The color of the heatmap indicates the Spearman correlation coefficients (blue-negative, red-positive).



Fecal metabolome may be the key to link the proteomic risk score-related core microbial features and host inflammation (A) Associations of the core microbial OTUs with fecal metabolites (n=987). The relationships between microbial OTUs and fecal metabolites was assessed by a linear regression model adjusting for age, sex, BMI. Multiple testing was adjusted using Benjamini and Hochberg method, with a false discovery rate (FDR) of <0.05 being considered statistically significant. We only presented metabolites showing significant associations with more than half of the core microbial OTUs (n=20) in the figure. Sizes of the nodes represent the number of OTUs related with fecal metabolites. Red edge, β-coefficient >0; blue edge, β-coefficient <0. (B) Pathway analysis for the core fecal metabolites (shown in part A) using MetaboAnalyst 4.0. [20] (C) Metabolites enriched in the significant pathways (shown in part B).

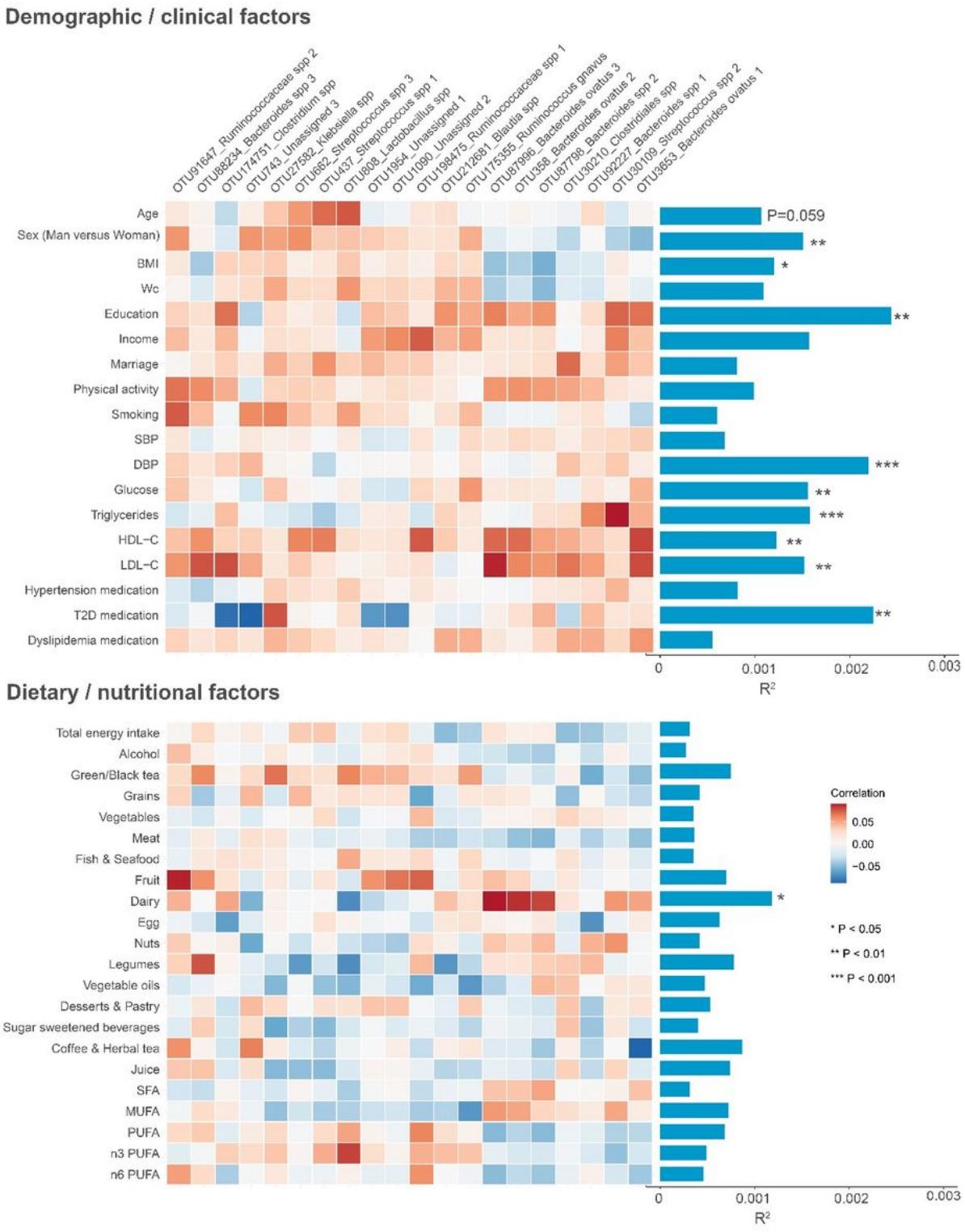


Figure 5

Host and environmental factors are correlated with the blood proteomic risk score-related core microbial OTUs. Host and environmental factors including 18 demographic/clinical items and 22 dietary/nutritional items were used in this analysis (n=1729). The bar plot indicates the explained variation of the core OTUs composition (Bray-Curtis distance) by each item. The heatmap next to the bar plot shows the correlation

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementarydata.docx