Journal of Genetics and Genomics 48 (2021) 792–802

Contents lists available at ScienceDirect

Journal of Genetics and Genomics

Journal homepage: www.journals.elsevier.com/journal-of-genetics-and-genomics/

Original research

Gut microbiota, inflammation, and molecular signatures of host response to infection

Wanglong Gou [a, b, c, 1], Yuanqing Fu [b, c, d, 1], Liang Yue [d, e, 1], Geng-Dong Chen [f, 1], Xue Cai [d, e, 1], Menglei Shuai [b, c, 1], Fengzhe Xu [b, c, 1], Xiao Yi [d, e], Hao Chen [d, e], Yi Zhu [d, e], Mian-Li Xiao [f], Zengliang Jiang [b, c, d], Zelei Miao [b, c], Congmei Xiao [b, c], Bo Shen [g], Xiaomai Wu [g], Haihong Zhao [g], Wenhua Ling [f], Jun Wang [h, *], Yu-Ming Chen [f, *], Tiannan Guo [c, d, e, *], Ju-Sheng Zheng [a, b, c, d, *]

[a] College of Life Sciences, Zhejiang University, Hangzhou 310058, China
[b] Key Laboratory of Growth Regulation and Translational Research of Zhejiang Province, School of Life Sciences, Westlake University, Hangzhou 310024, China
[c] Center for Infectious Disease Research, Westlake Laboratory of Life Sciences and Biomedicine, Hangzhou 310024, China
[d] Institute of Basic Medical Sciences, Westlake Institute for Advanced Study, Hangzhou 310024, China
[e] Key Laboratory of Structural Biology of Zhejiang Province, School of Life Sciences, Westlake University, Hangzhou 310024, China
[f] Guangdong Provincial Key Laboratory of Food, Nutrition and Health, Department of Epidemiology, School of Public Health, Sun Yat-sen University, Guangzhou 510275, China
[g] Taizhou Hospital, Wenzhou Medical University, Linhai 325035, China
[h] CAS Key Laboratory for Pathogenic Microbiology and Immunology, Institute of Microbiology, Chinese Academy of Sciences, Beijing 100101, China

ARTICLE INFO

Article history:
Received 20 January 2021
Received in revised form
5 April 2021
Accepted 15 April 2021
Available online 3 May 2021

Keywords:
Gut microbiota
Proinflammatory cytokines
Proteomic biomarkers
COVID-19
Host infection response

ABSTRACT

Gut microbial dysbiosis has been linked to many noncommunicable diseases. However, little is known about specific gut microbiota composition and its correlated metabolites associated with molecular signatures underlying host response to infection. Here, we describe the construction of a proteomic risk score based on 20 blood proteomic biomarkers, which have recently been identified as molecular signatures predicting the progression of the COVID-19. We demonstrate that in our cohort of 990 healthy individuals without infection, this proteomic risk score is positively associated with proinflammatory cytokines mainly among older, but not younger, individuals. We further discover that a core set of gut microbiota can accurately predict the above proteomic biomarkers among 301 individuals using a machine learning model and that these gut microbiota features are highly correlated with proinflammatory cytokines in another independent set of 366 individuals. Fecal metabolomics analysis suggests potential amino acid-related pathways linking gut microbiota to host metabolism and inflammation. Overall, our multi-omics analyses suggest that gut microbiota composition and function are closely related to inflammation and molecular signatures of host response to infection among healthy individuals. These results may provide novel insights into the cross-talk between gut microbiota and host immune system.

Introduction

With the coronavirus disease 2019 (COVID-19) defined as a "global pandemic" and spreading worldwide at an unprecedented speed, more than 80 million individuals have been infected globally since its first detection in December 2019 to December 2020 (World Health Organization, 2020). Along with this pandemic and the high disparity in the disease severity after infection, it has

* Corresponding authors.
E-mail addresses: junwang@im.ac.cn (J. Wang), chenyum@mail.sysu.edu.cn (Y.-M. Chen), guotiannan@westlake.edu.cn (T. Guo), zhengjusheng@westlake.edu.cn (J.-S. Zheng).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.jgg.2021.04.002

become an increasingly interesting and important issue to investigate the factors underlying the risk of severity after infection among healthy individuals. Gut microbiome plays critical roles in the training of major components of the host's innate and adaptive immune system, with gut microbial dysbiosis linked to several diseases, such as autoimmune, allergic, and chronic inflammatory disorders (Duffy, 2020; Ruff et al., 2020; Zheng et al., 2020). Several studies have demonstrated that the patients with COVID-19 have alterations in gut microbiota composition compared with non-COVID-19 individuals (Gu et al., 2020; Zuo et al., 2020; Yeoh et al., 2021), and these alterations remained in the recovered patients (Yeoh et al., 2021).

Based on a recent investigation into the blood biomarkers of COVID-19 patients, researchers identified a set of molecular signatures (blood proteomic biomarkers), which could characterize and predict the interindividual variation in the disease severity after infection (Shen et al., 2020). The identification of these deep molecular signatures gives us an opportunity to examine the association of gut microbiota composition and function with the inflammation and host response to infection and to further understand the cross-talk between gut microbiome and host immune response.

We hypothesized that the gut microbiota composition and its correlated metabolites are associated with molecular signatures underlying host response to infection among healthy non-infected individuals. Here, we integrate blood proteomics data from 31 COVID-19 patients and multi-omics data (proteomic, gut microbiome, and fecal metabolome) and comprehensive phenotype data from a Chinese population living in Guangzhou, involving 2413 participants without infection (Figs. 1 and S1; Table S1). Based on the COVID-19 patient data, we constructed and validated a blood proteomic risk score (PRS) to represent the interindividual variation in the infection response. Then, among 990 individuals with the data of proteome and blood inflammatory biomarkers, we investigated the association of the PRS with inflammatory biomarkers. Next, we identified core gut microbiota features, which may be associated with the variability in infection response using a machine learning model. We conducted further fecal metabolomics analysis to reveal potential biological mechanisms linking gut microbiota,

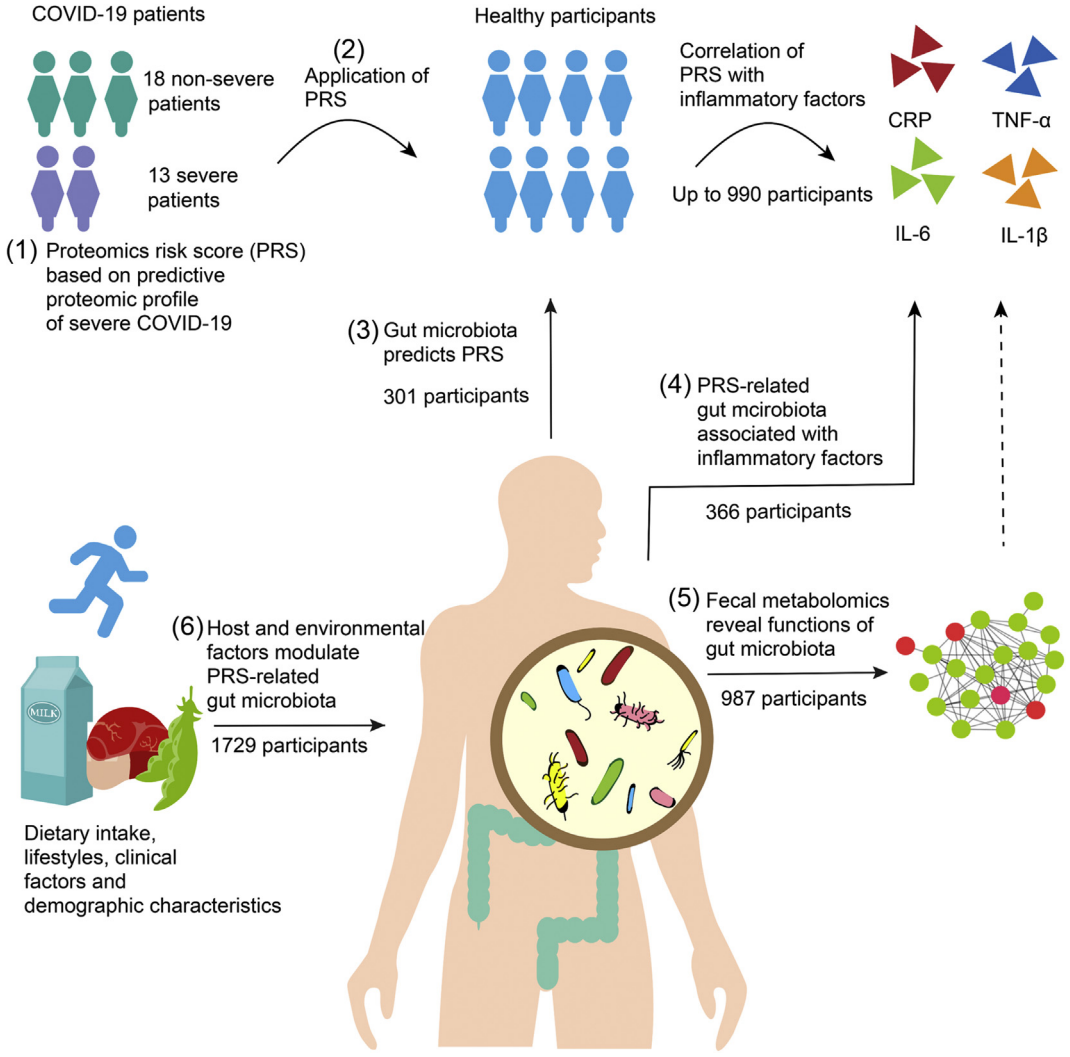

Fig. 1. Study design and analysis pipeline. Study overview: (1) Constructing a novel COVID-19 blood PSR among 31 COVID-19 patients (18 non-severe cases and 13 severe cases). (2) Applicating the PRS in healthy participants and further linking it to host inflammatory status ($n = 990$). (3) Investigating the potential role of gut microbiota in predicting the PRS 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 the function of gut microbiota on host metabolism ($n = 987$). (6) Investigating the impact of host and environmental factors on PRS-related core microbial OTU ($n = 1729$). PSR, proteomic risk score.

inflammation, and host infection response. Finally, we demonstrated the contribution of 40 host and environmental factors to the variance of the above-identified core gut microbiota features.

Results

PRS related to COVID-19 severity is correlated with inflammatory factors among healthy individuals

First, we derived a blood PRS that could characterize the inter-individual variation in the COVID-19 infection response using the set of 20 blood proteins (Shen et al., 2020). Among the COVID-19 patients (18 non-severe cases and 13 severe cases; Shen et al., 2020), Poisson regression analysis indicated that per 10% increment in the PRS, there was associated a 57% higher risk of progressing to clinically severe phase (risk ratio [RR], 1.57; 95% confidence interval [CI], 1.35–1.82; Fig. 2A), in support of the PRS as being a valid proxy for the inter-individual variation in the COVID-19 infection response.

Among a cohort of noninfection participants with data of both proteomics and inflammatory markers ($n = 990$), we investigated the correlation between the PRS and blood inflammatory markers interleukin (IL)-1β, IL-6, tumor necrosis factor (TNF)-α, and high-sensitivity C-reactive protein (hsCRP). The blood proteomic data were based on the baseline serum samples of the cohort (Fig. S1). Spearman correlation analysis indicated that PRS had a significant positive correlation with serum concentrations of hsCRP and TNF-α ($P < 0.001$ and $P < 0.05$, respectively), but not other markers (Fig. 2B). We also 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 associated 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 (Fig. S2).

Core microbiota features predict the molecular signatures of host response to infection

To investigate the potential role of gut microbiota in the previously mentioned proteomic biomarkers, we next explored the relation between the gut microbiota and the PRS in a subcohort of 301 participants with newly measurement of both gut microbiota (based on 16S rRNA gene amplicon sequencing) and blood proteomics data (Fig. 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 along with a highly conservative and strict 10-fold cross-validation (CV) 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 10-fold CVs). The predictive capacity for PRS based on the core OTUs substantially outperformed that of the demographic characteristics and laboratory tests (Pearson's $r = 0.154$, $P = 0.087$) (Fig. 3A) and that of the combination of core OTUs, demographic characteristics, and laboratory tests (Pearson's $r = 0.45$, $P < 0.001$). The list of these core OTUs along with their taxonomic classification is provided in Table S3. Demographic characteristics and laboratory tests include age, body mass index (BMI), sex, blood pressure, and blood lipids. These OTUs were mainly assigned to *Bacteroides* genus, *Streptococcus* genus, *Lactobacillus* genus, Ruminococcaceae family, Lachnospiraceae family, and Clostridiales order.

In addition, we used co-inertia analysis (CIA) to further test covariance between the 20 identified core OTUs and 20 proteomic biomarkers, outputting an RV coefficient (ranged from 0 to 1) to quantify the closeness. CIA analysis indicated a close association of these OTUs with the proteomic biomarkers ($RV = 0.12$, $P < 0.05$) (Fig. S3A). When replicating this analysis stratified by age, significant association was observed only among older participants (age ≥ 58 years, $n = 66$; $RV = 0.22$, $P < 0.05$) (Fig. S3B and S3C).

Importantly, the previously mentioned 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 *vs.* actual PRS ($P < 0.05$). It outperforms the predictive capacity of the above demographic characteristics and laboratory tests (Pearson's $r = 0.08$, $P = 0.31$) (Fig. 3A) and that of the combination of core OTUs, demographic characteristics, and laboratory tests (Pearson's $r = 0.13$, $P = 0.087$). These findings support that the 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 subcohort of 366 participants (Fig. 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 interferon (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 infection response

We hypothesized that the influences of the core microbial features on the PRS were driven by some specific microbial metabolites. We assessed the relationship between the core gut microbiota and fecal metabolome among 987 participants, whose fecal metabolome and 16S rRNA gene were measured at the same time point (Fig. S1). After correction for the multiple testing (false discovery rate [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 the infection-sensitive proteomic biomarkers.

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 was responsible for adding amino acid to nascent peptide chains and was a target for inhibiting cytokine-stimulated inflammation (Fig. 4C). In addition, four metabolites were associated with arginine biosynthesis pathway, and three metabolites were enriched in valine, leucine, and isoleucine (known as branch-chain amino acids [BCAAs]) biosynthesis pathway (Fig. 4C).

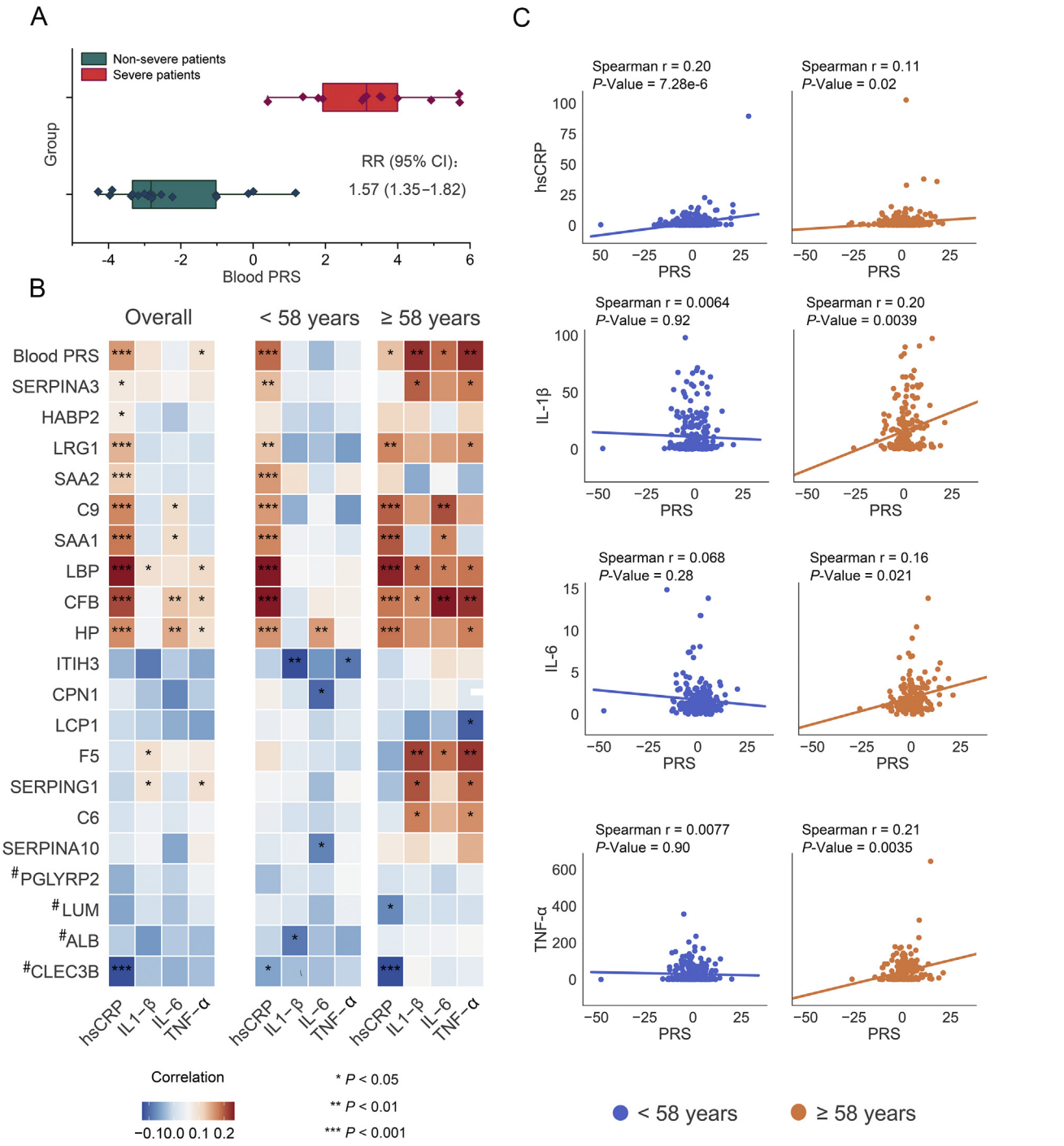

Fig. 2. PRS related to COVID-19 severity is correlated with inflammatory factors among healthy individuals. **A**: PRS related to the progression to severe COVID-19. Among the 31 COVID-19 patients, Poisson regression analysis indicated that per 10% increment in the PRS there was associated a 57% higher risk of progressing to clinically severe phase (RR, 1.57; 95% CI, 1.35−1.82). **B**: The Spearman correlation of proteomic biomarkers (i.e., molecular signatures of host infection response) with host inflammatory markers among the healthy participants (990 participants). We examined the Spearman correlation of the previously mentioned 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 downregulated in severe patients, else, upregulated. *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$. **C**: The Spearman correlation of the PRS with individual host inflammatory markers stratified by the median age of participants (<58 years or ≥58 years).

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 to the identified core OTU composition. A total of 40 items belonging to two categories (i.e., demographic or clinical factors and dietary or 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 or clinical factors, which explained 2.4% of the variation, we observed

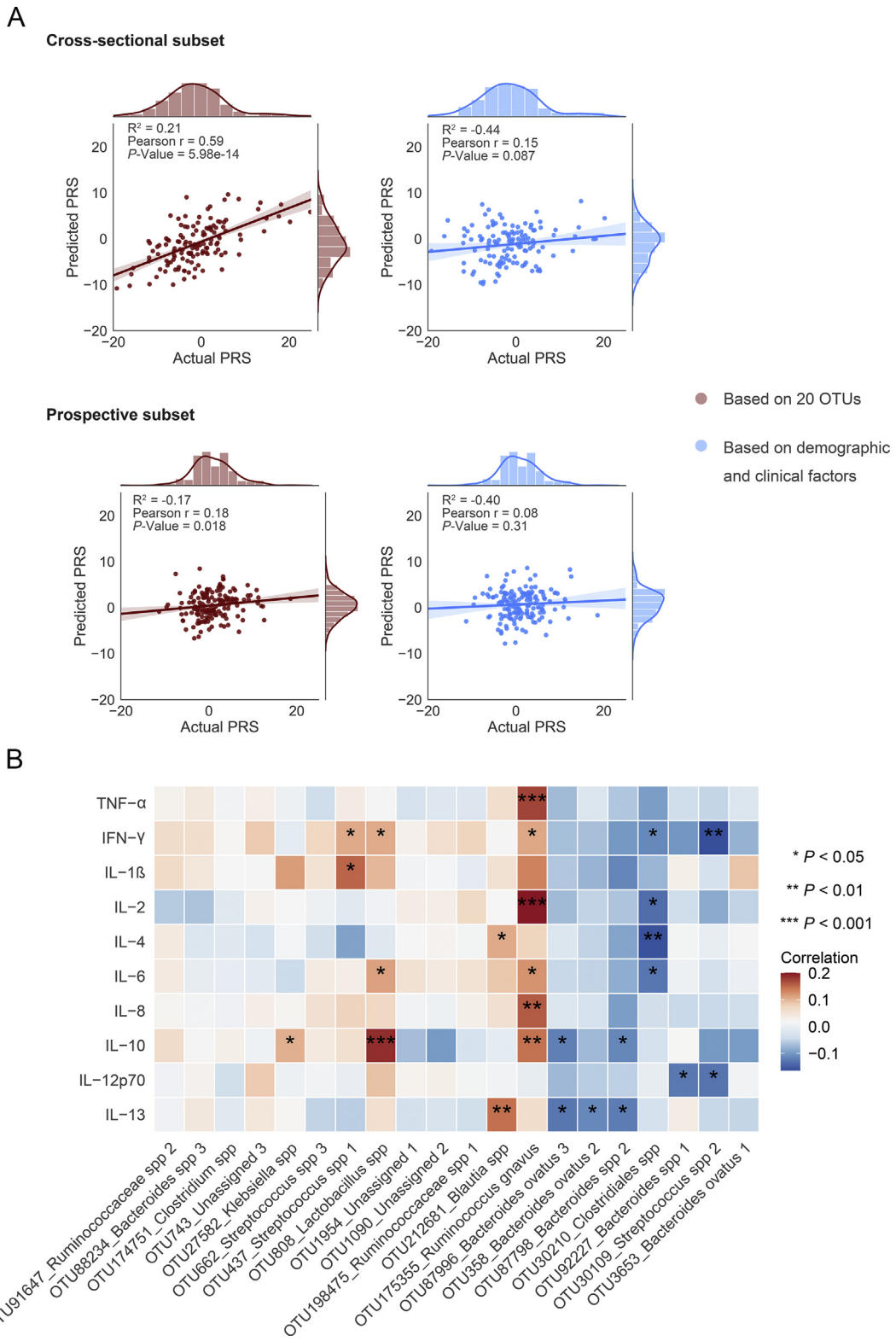

Fig. 3. Core microbiota features predict PRS and host inflammation. **A**: Plots of out-of-sample predicted PRS *vs.* actual PRS based on top 20 ranked OTUs or demographic or clinical factors (age, sex, BMI, fasting glucose, HDL, LDL, TC, TG, DBP, and SBP) using LightGBM with 10-fold CV. In this approach, the original sample is randomly partitioned into 10 equal size subsamples. LightGBM model is trained on 90% of the subsamples and PRS is predicted for the 10% of the subsamples which were not included in the model training. This process is repeated 10-fold, resulting in a predicted PRS set for each sample. 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's r of predicted values vs. actual values and the corresponding *P* value are shown in the figures. **B**: The Spearman 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). *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$. BMI, body mass index; HDL, high-density lipoprotein; LDL, low-density lipoprotein; TC, total cholesterol; TG, triglyceride; DBP, diastolic blood pressure; SBP, systolic blood pressure.

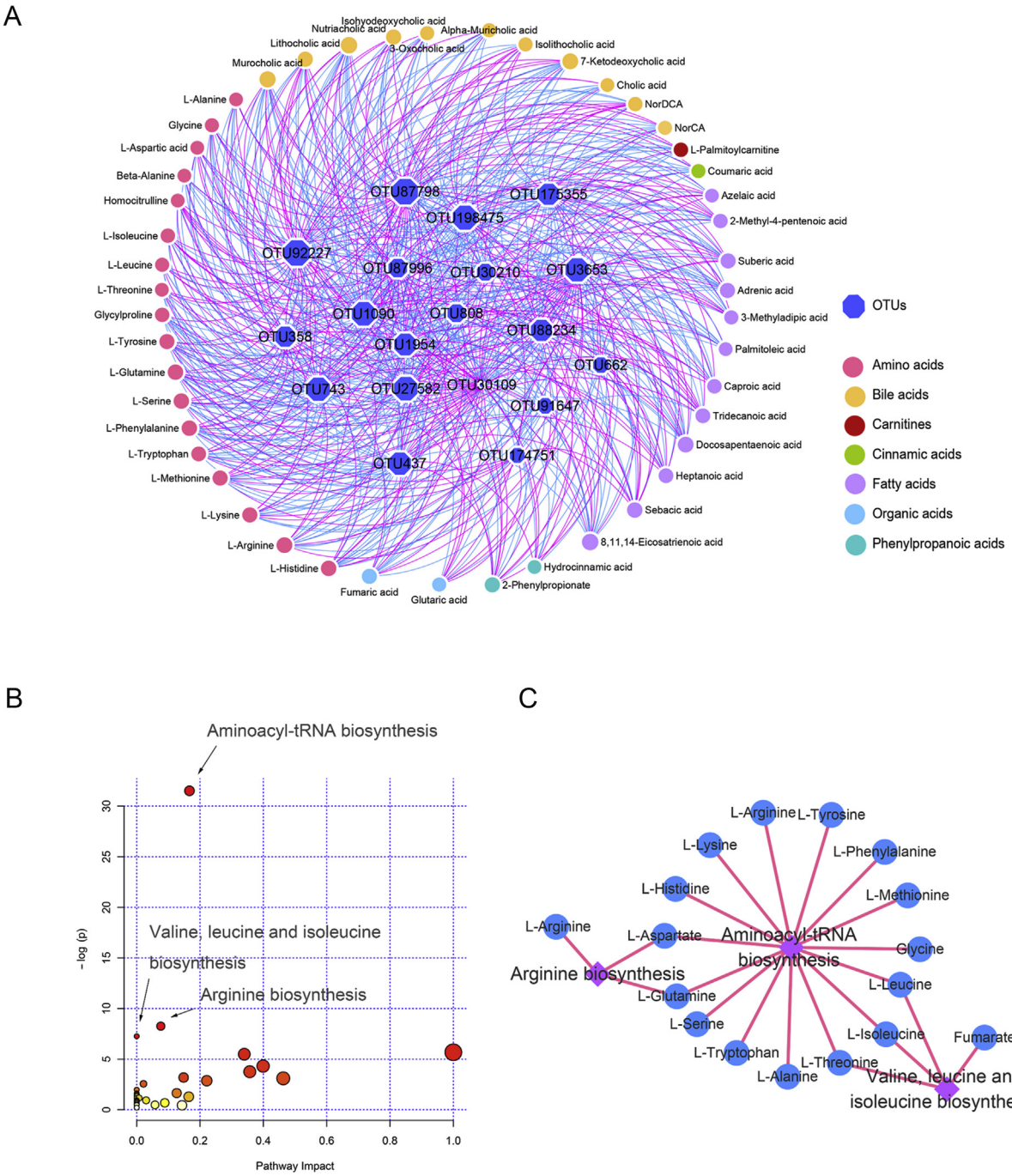

Fig. 4. Fecal metabolome may be the key to link the proteomic risk score-related core microbial OTUs and host inflammation. **A**: Associations of the core microbial OTUs with fecal metabolites ($n = 987$). The relationships between microbial OTUs and fecal metabolites were assessed by a linear regression model adjusting for age, sex, and 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 to fecal metabolites. Red edge, β-coefficient > 0; blue edge, β-coefficient < 0. **B**: Pathway analysis for the core fecal metabolites (shown in **A)** using MetaboAnalyst 4.0. (Chong et al., 2018) **C**: Metabolites enriched in the aminoacyl-tRNA biosynthesis pathway, arginine biosynthesis pathway, and valine, leucine, and isoleucine biosynthesis pathway, respectively.

associations of nine items (i.e., sex, education, physical activity, diastolic blood pressure, blood glucose, blood lipids, and medicine used for type 2 diabetes) with interindividual distances in the core OTU composition (PERMANOVA, $P < 0.05$) (Fig. 5). In the dietary or 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 individuals, gut microbial features are highly predictive of the blood molecular signatures of the host response to infection. These identified gut microbial features are also closely related to inflammatory markers. The fecal metabolomics analysis reveals that amino acid-related pathway may

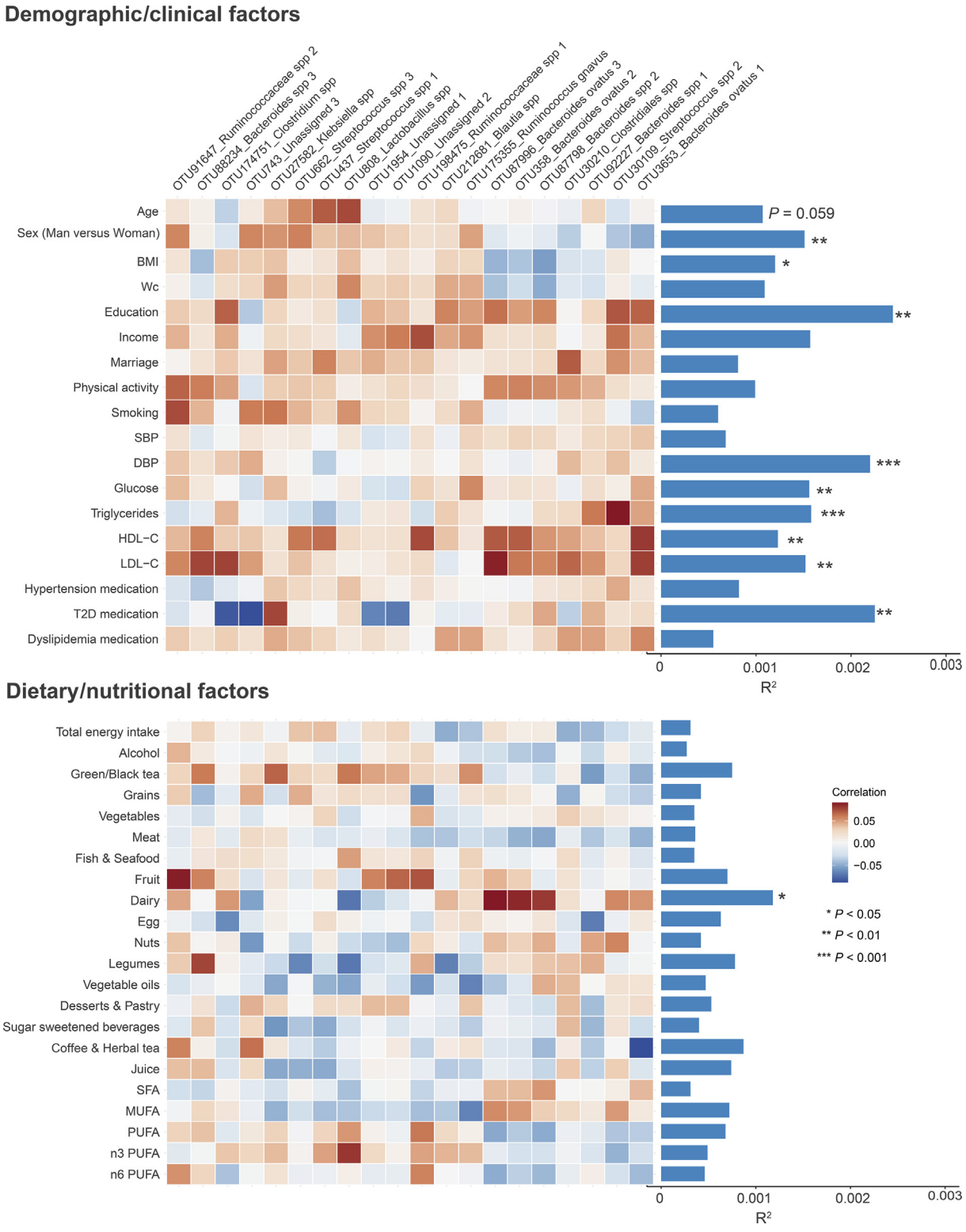

Fig. 5. Host and environmental factors are correlated with the proteomic risk score-related core microbial OTUs. PERMANOVA analysis was used to explore the variance contribution of the host and environmental factors for the core microbial OTUs. Spearman correlation analysis was used to assess the relationship between the host and environmental factors and core microbial OTUs. Multiple testing was adjusted using Benjamini and Hochberg method, with a false discovery rate (FDR) of <0.05 being considered statistically significant. Host and environmental factors, including 18 demographic or clinical items and 22 dietary or 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 coefficients of each item with the core OTUs.

provide the key link among the identified core gut microbiota, inflammation, and host infection response. Furthermore, modifications on host and environmental factors are likely to influence the previously mentioned core gut microbiota composition.

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 (Huang et al., 2020; Yang et al., 2020). 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) (Shen et al., 2020). These proteins may be activated together with proinflammatory cytokines, such as IL-6 and TNF-α, following the invasion of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). Previous study suggested that age may be an important confounding factor affecting the relationship between host metabolism and COVID-19 progress (Shen et al., 2020). The present study based on data from healthy noninfected participants supports that the proteomic biomarkers (integrated into a score) are positively associated with proinflammatory cytokines, especially among those with older age, which, to some extent, agreed with the observations during the COVID-19 outbreak that older individuals are more susceptible to the virus, leading to the severity of the disease, because of the induced hyperinflammation or "cytokine storm" (Chen et al., 2020; Zhou et al., 2020).

In the present study, the core gut microbial features (20 OTUs), with a satisfied performance, outperform demographic characteristics and laboratory tests in predicting the blood proteomic biomarkers, which highlight a potential role of gut microbiota in regulating the interindividual variation in the infection response. Growing evidence has shown that microbiota plays a fundamental role in the induction, training, and function of the host immune system, and the composition of the gut microbiota and its activity are involved in the production of inflammatory cytokines (Belkaid and Hand, 2014; Cani and Jordan, 2018). Prior studies reported that *Lactobacillus* genus was positively associated with IL-6 and IFN-γ, whereas *Blautia* genus was positively associated with IL-10 (Yoshida et al., 2001; Pohjavuori et al., 2004; Jiang et al., 2012); these relationships were replicated in our study. In addition, 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 (Murray, 2016). Amino acid insufficiency will cause depletion of available aminoacylated tRNA, which is essential for the host to sense amino acid limitation and immune response (Harding et al., 2003; Brown et al., 2010, 2016). 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 (Kim et al., 2020). Thus, the identified pathways regulating aminoacyl-tRNA biosynthesis and arginine biosynthesis may be both involved in the inflammatory response. In addition, arginine and BCAAs (i.e., valine, leucine, and isoleucine) were also reported, regulating innate and adaptive immune responses and enhancing intestinal development (Zhang et al., 2017). Collectively, these key roles that amino acids play in the immunoregulation may help explain how the PRS-related core OTUs impact host inflammation and infection response via amino acid metabolism.

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 (Cabreiro et al., 2013; Gilbert et al., 2018; Vich Vila et al., 2020). Although these observations were quite crude, they 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 host immune response also contributed to the variance of the gut microbiota, including age, sex, and indicators of clinical comorbidities (blood pressure, glucose triglycerides, high-density and low-density lipoprotein cholesterol, and diabetes medication).

Previous studies have demonstrated that the patients with COVID-19 have alterations in gut microbiota composition compared with non-COVID-19 individuals (Gu et al., 2020; Zuo et al., 2020; Yeoh et al., 2021). In our present study, we composed a multi-omics data set (blood proteomics, gut microbiota, and fecal metabolome) from a large healthy human cohort study and reported that the composition of core gut microbiota and its correlated metabolites may be closely associated with the host inflammation and risk of disease severity after the infection. The discovered core gut microbial features and metabolites may serve as valuable tools for prognosis as well as potential preventive or treatment targets for COVID-19 in the future.

The strength of our study lies in our analysis of multi-omics data set (blood proteomics, gut microbiota, and fecal metabolome) from a large healthy human cohort study. Our results lead to the hypothesis that gut microbiota may play an important role in the interindividual variation in inflammation and host infection response via affecting the fecal amino acid-related pathway, although the detailed mechanism is yet to be discovered. However, our study is descriptive and limited by its inability to determine causation. Further mechanistic studies are required to better understand the role of the gut microbiome in the interindividual variation in host infection response. Another limitation is that we did not directly investigate the association of PRS with the risk of disease severity after infection of the virus among healthy individuals. Further prospective studies are required to explore their relationship. Finally, for the prospective gut microbiome-PRS analysis, gut microbes were measured only once and may not represent long-term status. Changes in gut microbes over time are likely.

In summary, analyses of our multi-omics data set lead to the hypothesis that gut microbiota composition and function may be closely related to inflammation and molecular signatures of the host infection response. The discovered core gut microbial features and related metabolites may serve as a potential preventive or treatment target for intervention. These results may also provide novel insights about the cross-talk between the gut microbiota and the host immune system.

Materials and methods

Ethics

This study has been approved by the Ethics Committee of the School of Public Health at Sun Yat-sen University (2018048) and the Ethics Committee of Westlake University (20190114ZJS0003), and all participants provided written informed consent.

Study design and participants

COVID-19 patients

Detailed information about the COVID-19 patients and proteomics data set have been reported previously (Shen et al., 2020). 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 per minute; mean oxygen saturation $\leq$ 93% in resting state; arterial blood oxygen partial pressure or oxygen concentration $\leq$ 300 mm Hg); and (4) critical (fulfill any of the three criteria: respiratory failure and require mechanical ventilation; shock incidence; admission to intensive care unit with other organ failure). We treated mild and typical patients as a nonsevere COVID-19 group, and the others as a severe COVID-19 group.

Healthy participants, sample collection, and clinical metadata

A total of 2413 healthy individuals from the community-based Guangzhou Nutrition and Health Study (GNHS) are involved in the present study, which mainly consists of a subset of individuals with proteomic data at baseline ($n = 990$) and a subset of individuals 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 (Zhang et al., 2014). Briefly, participants were enrolled between 2008 and 2013 and followed up to May 2018. Blood samples were collected at enrollment between 2008 and 2013 and follow-up visits between 2014 and 2018, and stool samples were collected only during follow-up visits between 2014 and 2018. 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 3500 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 h after collection.

Details of the method for the metadata measurements, proteomic analysis, measurement of inflammatory biomarkers, microbiome, and metabolomic analysis for GNHS were 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 ± standard deviation (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 PRS

We used 20 of 22 previously identified proteomic biomarkers to construct a PRS in healthy participants as molecular signatures of host infection response (Shen et al., 2020). We only used 20 of the 22 proteins for our PRS construction because 2 proteins were unavailable in our large proteomics database among healthy participants for further analysis.

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

where PRS_i is a PRS 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*, and β is 1 or −1 depending on the association between the protein *j* and risk of progressing to clinically severe phase (1, upregulated in severe patients; −1, downregulated 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 and 13 severe patients), adjusting for age, sex, and BMI.

Correlation between PRS and proinflammatory biomarkers

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

Machine learning algorithms for identifying microbial features to predict PRS

A 10-fold CV implementation of gradient boosting framework—LightGBM (Ke et al., 2017)—and SHAP (Shapley Additive exPlanations) (Lee, 2017) was used to link input gut microbial features with PRS. Gradient boosting has been widely used in prediction for tabular data in biomedical fields (Lundberg et al., 2018a; Artzi et al., 2020), and SHAP has been theoretically verified as the only consistent and locally accurate method to interpret machine learning results (Lundberg et al., 2018b; Lundberg et al., 2020). A 10-fold CV prediction implementation was used to generate 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 10-fold, resulting in a test PRS set for each participant and 10 different average absolute SHAP values for each OTUs. The top 20 ranked OTUs based on the sum of the average absolute SHAP value across 10-fold were included in further analysis. Pearson's 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), and the combination of both. Our predictor is based on code adapted from the sklearn 0.15.2 LightGBM regression (Pedregosa et al., 2011).

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 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 the skewness of trait distributions. Similarly, to reduce the 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, and BMI. Multiple testing was adjusted using Benjamini and Hochberg method, with an 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 (Chong et al., 2018).

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 1000 permutations. The total variation explained was also calculated per category (demographic or clinical category and dietary or 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 an FDR of <0.05 being considered statistically significant.

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

Data availability

The raw data of 16S rRNA gene sequences are available at CNSA (https://db.cngb.org/cnsa/) of CNGBdb at accession number CNP0000829.

CRediT authorship contribution statement

Wanglong Gou: Methodology, Data analysis, Data visualization, Writing - Original draft. **Yuanqing Fu**: Methodology, Writing - Original draft. **Liang Yue**: Data curation, Experimentation. **Geng-Dong Chen**: Sample collection, Data curation. **Xue Cai**: Data curation, Experimentation. **Menglei Shuai**: Experimentation, Data visualization. **Fengzhe Xu**: Data visualization. **Xiao Yi**: Experimentation. **Hao Chen**: Data analysis. **Yi Zhu**: Experimentation, Data curation. **Mian-Li Xiao**: Sample collection. **Zengliang Jiang**: Data analysis. **Zelei Miao**: Data curation. **Congmei Xiao**: Experimentation. **Bo Shen**: Sample collection. **Xiaomai Wu**: Sample collection. **Haihong Zhao**: Sample collection. **Wenhua Ling**: Sample collection. **Jun Wang**: Writing - Review & Editing, Data interpretation. **Yu-Ming Chen**: Writing - Review & Editing, Resources. **Tiannan Guo**: Writing - Review & Editing, Data curation. **Ju-Sheng Zheng**: Conceptualization, Project administration, Resources, Supervision, Writing - Review & Editing.

Conflicts of interest

The authors declare no competing financial interests.

Acknowledgments

This study was supported by the National Natural Science Foundation of China (82073529, 81903316, 81773416, 81972492, 21904107, 81672086), Zhejiang Ten-thousand Talents Program (2019R52039), 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). 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.

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jgg.2021.04.002.

References

Artzi, N.S., Shilo, S., Hadar, E., Rossman, H., Barbash-Hazan, S., Ben-Haroush, A., Balicer, R.D., Feldman, B., Wiznitzer, A., Segal, E., 2020. Prediction of gestational diabetes based on nationwide electronic health records. Nat. Med. 26, 71–76.
Belkaid, Y., Hand, T.W., 2014. Role of the microbiota in immunity and inflammation. Cell 157, 121–141.
Brown, A., Fernández, I.S., Gordiyenko, Y., Ramakrishnan, V., 2016. Ribosome-dependent activation of stringent control. Nature 534, 277–280.
Brown, M.V., Reader, J.S., Tzima, E., 2010. Mammalian aminoacyl-tRNA synthetases: cell signaling functions of the protein translation machinery. Vasc. Pharmacol. 52, 21–26.
Cabreiro, F., Au, C., Leung, K.-Y., Vergara-Irigaray, N., Cochemé, H.M., Noori, T., Weinkove, D., Schuster, E., Greene, N.D.E., Gems, D., 2013. Metformin retards aging in *C. elegans* by altering microbial folate and methionine metabolism. Cell 153, 228–239.
Cani, P.D., Jordan, B.F., 2018. Gut microbiota-mediated inflammation in obesity: a link with gastrointestinal cancer. Nat. Rev. Gastroenterol. Hepatol. 15, 671–682.
Chen, N., Zhou, M., Dong, X., Qu, J., Gong, F., Han, Y., Qiu, Y., Wang, J., Liu, Y., Wei, Y., et al., 2020. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. Lancet (London, England) 395, 507–513.
Chong, J., Soufan, O., Li, C., Caraus, I., Li, S., Bourque, G., Wishart, D.S., Xia, J., 2018. MetaboAnalyst 4.0: towards more transparent and integrative metabolomics analysis. Nucleic Acids Res. 46, W486–W494.
Duffy, D., 2020. Understanding immune variation for improved translational medicine. Curr. Opin. Immunol. 65, 83–88.
Gilbert, J.A., Blaser, M.J., Caporaso, J.G., Jansson, J.K., Lynch, S.V., Knight, R., 2018. Current understanding of the human microbiome. Nat. Med. 24, 392–400.
Gu, S., Chen, Yanfei, Wu, Z., Chen, Yunbo, Gao, H., Lv, L., Guo, F., Zhang, X., et al., 2020. Alterations of the gut microbiota in patients with coronavirus disease 2019 or H1N1 influenza. Clin. Infect. Dis. 71, 2669–2678.
Harding, H.P., Zhang, Y., Zeng, H., Novoa, I., Lu, P.D., Calfon, M., Sadri, N., Yun, C., Popko, B., Paules, R., et al., 2003. An integrated stress response regulates amino acid metabolism and resistance to oxidative stress. Mol. Cell 11, 619–633.
Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., Zhang, L., Fan, G., Xu, J., Gu, X., et al., 2020. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet (London, England) 395, 497–506.
Jiang, Y., Lü, X., Man, C., Han, L., Shan, Y., Qu, X., Liu, Y., Yang, S., Xue, Y., Zhang, Y., 2012. Lactobacillus acidophilus induces cytokine and chemokine production via NF-κB and p38 mitogen-activated protein kinase signaling pathways in intestinal epithelial cells. Clin. Vaccine Immunol. 19, 603–608.
Ke, G., Qi, M., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., Liu, T.-Y., 2017. LightGBM: a highly efficient gradient boosting decision tree. NIPS (News Physiol. Sci.) 30, 3149–3157.
Kim, Y., Sundrud, M.S., Zhou, C., Edenius, M., Zocco, D., Powers, K., Zhang, M., Mazitschek, R., Rao, A., Yeo, C.-Y., et al., 2020. Aminoacyl-tRNA synthetase inhibition activates a pathway that branches from the canonical amino acid response in mammalian cells. Proc. Natl. Acad. Sci. U. S. A. 117, 8900–8911.
Lee, S.M.L.S.-I., 2017. A Unified Apporach to Interpreting Model Predictions, 07874 arXiv: 1705.
Lundberg, S.M., Erion, G., Chen, H., DeGrave, A., Prutkin, J.M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., Lee, S.-I., 2020. From local explanations to global understanding with explainable ai for trees. Nat. Mach. Intell. 2, 56–67.
Lundberg, Scott M., Erion, G.G., Lee, S.-I., 2018a. Consistent individualized feature attribution for tree ensembles. arXiv, 1802.03888.
Lundberg, Scott M., Nair, B., Vavilala, M.S., Horibe, M., Eisses, M.J., Adams, T., Liston, D.E., Low, D.K.W., Newman, S.F., et al., 2018b. Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. Nat. Biomed. Eng. 2, 749–760.
Murray, P.J., 2016. Amino acid auxotrophy as a system of immunological control nodes. Nat. Immunol. 17, 132–139.
Pedregosa, F., Varoquaux, G., Weiss, R., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., et al., 2011. Scikit-learn : machine learning in Python. J. Mach. Learn. Res. 12, 2825–2830.
Pohjavuori, E., Viljanen, M., Korpela, R., Kuitunen, M., Tiittanen, M., Vaarala, O., Savilahti, E., 2004. Lactobacillus GG effect in increasing IFN-gamma production in infants with cow's milk allergy. J. Allergy Clin. Immunol. 114, 131–136.
Ruff, W.E., Greiling, T.M., Kriegel, M.A., 2020. Host-microbiota interactions in immune-mediated diseases. Nat. Rev. Microbiol. 18, 521–538.
Shen, B., Yi, X., Sun, Y., Bi, X., Du, J., Zhang, C., Quan, S., Zhang, F., Sun, R., Qian, L., et al., 2020. Proteomic and metabolomic characterization of COVID-19 patient sera. Cell 182, 1–14.
Vich Vila, A., Collij, V., Sanna, S., Sinha, T., Imhann, F., Bourgonje, A.R., Mujagic, Z., Jonkers, D.M.A.E., Masclee, A.A.M., Fu, J., et al., 2020. Impact of commonly used drugs on the composition and metabolic function of the gut microbiota. Nat. Commun. 11, 362.

World Health Organization, 2020. Coronavirus disease (COVID-19) pandemic. https://www.who.int/emergencies/diseases/novel-coronavirus-2019. (Accessed December 2020).

Yang, Y., Shen, C., Li, J., Yuan, J., Wei, J., Huang, F., Wang, F., Li, G., Li, Y., Xing, L., et al., 2020. Plasma IP-10 and MCP-3 levels are highly associated with disease severity and predict the progression of COVID-19. J. Allergy Clin. Immunol. 146, 119–127.

Yeoh, Y.K., Zuo, T., Lui, G.C.-Y., Zhang, F., Liu, Q., Li, A.Y., Chung, A.C., Cheung, C.P., Tso, E.Y., Fung, K.S., et al., 2021. Gut microbiota composition reflects disease severity and dysfunctional immune responses in patients with COVID-19. Gut 70, 698–706.

Yoshida, K., Matsumoto, T., Tateda, K., Uchida, K., Tsujimoto, S., Yamaguchi, K., 2001. Induction of interleukin-10 and down-regulation of cytokine production by Klebsiella pneumoniae capsule in mice with pulmonary infection. J. Med. Microbiol. 50, 456–461.

Zhang, S., Zeng, X., Ren, M., Mao, X., Qiao, S., 2017. Novel metabolic and physiological functions of branched chain amino acids: a review. J. Anim. Sci. Biotechnol. 23, 8–10.

Zhang, Z.-Q., He, L.-P., Liu, Y.-H., Liu, J., Su, Y.-X., Chen, Y.-M., 2014. Association between dietary intake of flavonoid and bone mineral density in middle aged and elderly Chinese women and men. Osteoporos. Int. 25, 2417–2425.

Zheng, D., Liwinski, T., Elinav, E., 2020. Interaction between microbiota and immunity in health and disease. Cell Res. 30, 492–506.

Zhou, F., Yu, T., Du, R., Fan, G., Liu, Y., Liu, Z., Xiang, J., Wang, Y., Song, B., Gu, X., et al., 2020. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. Lancet (London, England) 395, 1054–1062.

Zuo, T., Zhang, F., Lui, G.C.Y., Yeoh, Y.K., Li, A.Y.L., Zhan, H., Wan, Y., Chung, A., Cheung, C.P., Chen, N., et al., 2020. Alterations in gut microbiota of patients with COVID-19 during time of hospitalization. Gastroenterology 159, 944–955.