

---

## **Gut microbiota may underlie the predisposition of healthy individuals to COVID-19**

Wanglong Gou<sup>1,4#</sup>, Yuanqing Fu<sup>1,4#</sup>, Liang Yue<sup>3,4#</sup>, Geng-dong Chen<sup>2#</sup>, Xue Cai<sup>3,4#</sup>, Menglei Shuai<sup>1,4#</sup>, Fengzhe Xu<sup>1,4#</sup>, Xiao Yi<sup>3,4</sup>, Hao Chen<sup>3,4</sup>, Yi Zhu<sup>3,4</sup>, Mian-li Xiao<sup>2</sup>, Zengliang Jiang<sup>1</sup>, Zelei Miao<sup>1</sup>, Congmei Xiao<sup>1</sup>, Bo Shen<sup>5</sup>, Xiaomai Wu<sup>5</sup>, Haihong Zhao<sup>5</sup>, Wenhua Ling<sup>2</sup>, Jun Wang<sup>6</sup>, Yu-ming Chen<sup>2\*</sup>, Tiannan Guo<sup>3,4\*</sup>, Ju-Sheng Zheng<sup>1,4,7\*</sup>

<sup>1</sup>Key Laboratory of Growth Regulation and Translation Research of Zhejiang Province, School of Life Sciences, Westlake University, Hangzhou 310024, China;

<sup>2</sup>Guangdong Provincial Key Laboratory of Food, Nutrition and Health; Department of Epidemiology, School of Public Health, Sun Yat-sen University, Guangzhou 510080, China;

<sup>3</sup>Key Laboratory of Structural Biology of Zhejiang Province, School of Life Sciences, Westlake University, Hangzhou 310024, China;

<sup>4</sup>Institute of Basic Medical Sciences, Westlake Institute for Advanced Study, Hangzhou 310024, China;

<sup>5</sup>Taizhou Hospital, Wenzhou Medical University, 150 Ximen Street, Linhai 317000, Zhejiang Province, China;

<sup>6</sup> CAS Key Laboratory for Pathogenic Microbiology and Immunology, Institute of Microbiology, Chinese Academy of Sciences, Beijing, 100101, China.

<sup>7</sup>Lead Contact

#These authors contributed equally

\*Correspondence: [chenyum@mail.sysu.edu.cn](mailto:chenyum@mail.sysu.edu.cn) (Y.M.C.);

[guotiannan@westlake.edu.cn](mailto:guotiannan@westlake.edu.cn) (T.G.); [zhengjusheng@westlake.edu.cn](mailto:zhengjusheng@westlake.edu.cn) (J.S.Z.)

---

1 **SUMMARY**

2 The COVID-19 pandemic is spreading globally with high disparity in the  
3 susceptibility of the disease severity. Identification of the key underlying factors for  
4 this disparity is highly warranted. Here we describe constructing a proteomic risk  
5 score based on 20 blood proteomic biomarkers which predict the progression to  
6 severe COVID-19. We demonstrate that in our own cohort of 990 individuals without  
7 infection, this proteomic risk score is positively associated with proinflammatory  
8 cytokines mainly among older, but not younger, individuals. We further discovered  
9 that a core set of gut microbiota could accurately predict the above proteomic  
10 biomarkers among 301 individuals using a machine learning model, and that these gut  
11 microbiota features are highly correlated with proinflammatory cytokines in another  
12 set of 366 individuals. Fecal metabolomic analysis suggested potential amino  
13 acid-related pathways linking gut microbiota to inflammation. This study suggests  
14 that gut microbiota may underlie the predisposition of normal individuals to severe  
15 COVID-19.  
16

---

## 17 **Introduction**

18 With the coronavirus disease 2019 (COVID-19) defined as ‘global pandemic’ and  
19 spreading worldwide at an unprecedented speed, more than two million individuals  
20 have been infected globally since its first detection in December 2019 to mid-April  
21 2020 (WHO, 2020). So far, many research papers have been published to characterize  
22 the clinical features of the COVID-19 patients, revealing that those individuals who  
23 are older, male or having other clinical comorbidities are more likely to develop into  
24 severe COVID-19 cases (Chen et al., 2020; Huang et al., 2020). Yet, little is known  
25 about the potential biological mechanisms or predictors for the susceptibility of the  
26 disease.

27

28 It is known that COVID-19 is caused by severe acute respiratory syndrome  
29 coronavirus 2 (SARS-CoV-2), which enters human cells by binding to angiotensin  
30 converting enzyme 2 (ACE2) as its receptor (Yan et al., 2020). Of note, ACE2 is an  
31 important regulator of intestinal inflammation, and that the expression of ACE2 is  
32 higher in the ileum and colon than in lung (Hashimoto et al., 2012; Zhang et al., 2020).  
33 ACE2 also has a major impact on the composition of gut microbiota, thus affecting  
34 cardiopulmonary diseases (Cole-Jeffrey et al., 2015). Moreover, over 60% of patients  
35 with COVID-19 report evidence of gastrointestinal symptoms, such as diarrhoea,  
36 nausea and vomiting, and that patients with gastrointestinal symptoms had overall  
37 more severe/critical diseases (Jin et al., 2020; Lin et al., 2020; Ng and Tilg, 2020).  
38 Taken together, the available evidence suggests a potential role of gut microbiota in  
39 the susceptibility of COVID-19 progression and severity.

40

41 Based on a recent investigation into the blood biomarkers of COVID-19 patients, we  
42 identified a set of proteomic biomarkers which could help predict the progression to  
43 severe COVID-19 among infected patients (Shen et al., 2020). The newly discovered  
44 proteomic biomarkers may help early prediction of severe COVID-19. However, the  
45 question remains as to whether this set of proteomic biomarkers could be used in  
46 healthy (non-infected) individuals to help explain the disease susceptibility. It is also

---

47 unclear whether gut microbiota could regulate these blood proteomic biomarkers  
48 among healthy individuals.

49

50 To address the above unresolved questions, we integrated blood proteomics data from  
51 31 COVID-19 patients and multi-omics data from a Chinese population without  
52 infection living in Guangzhou, involving 2413 participants (Figure 1; Figure S1;  
53 Table S1). Based on the COVID-19 patient data, we constructed a blood proteomic  
54 risk score (PRS) for the prediction of COVID-19 progression to clinically severe  
55 phase. Then, among 990 healthy individuals with the data of proteome and blood  
56 inflammatory biomarkers, we investigated the association of the COVID-19-related  
57 PRS with inflammatory biomarkers as a verification of the PRS with disease  
58 susceptibility in normal non-infected individuals. Next, we identified core gut  
59 microbiota features which predicted the blood proteomic biomarkers of COVID-19  
60 using a machine-learning model. We conducted further fecal metabolomics analysis to  
61 reveal potential biological mechanisms linking gut microbiota to the COVID-19  
62 susceptibility among non-infected individuals. Finally, we demonstrated the  
63 contribution of 40 host and environmental factors to the variance of the above  
64 identified core gut microbiota features.

65

## 66 **Results**

### 67 **Predictive proteomic profile for severe COVID-19 is correlated with** 68 **inflammatory factors among healthy individuals**

69 Based on a prior serum proteomic profiling of COVID-19 patients, 22 proteomic  
70 biomarkers contributed to the prediction of progression to severe COVID-19 status  
71 (Shen et al., 2020). Using this cohort, we constructed a blood PRS among the 31  
72 COVID-19 patients (18 non-severe cases and 13 severe cases) based on 20 proteomic  
73 biomarkers (Table S2). We only used 20 of the 22 proteins for our PRS construction  
74 because 2 proteins were unavailable in our large proteomics database among  
75 non-infected participants for the further analysis. Among the COVID-19 patients,  
76 Poisson regression analysis indicated that per 10% increment in the PRS there was

---

77 associated a 57% higher risk of progressing to clinically severe phase (RR, 1.57; 95%  
78 CI, 1.35-1.82; Figure 2A), in support of the PRS as being a valid proxy for the  
79 predictive biomarkers of severe COVID-19.

80

81 To explore the potential implication of the PRS among non-COVID-19 individuals,  
82 we constructed the PRS using the same set of 20 blood proteins among a cohort of  
83 non-infected participants with data of both proteomics and inflammatory markers  
84 (n=990). The blood proteomic data was based on the baseline serum samples of the  
85 cohort (Figure S1). We investigated the correlation between the PRS and blood  
86 inflammatory markers IL-1 $\beta$ , IL-6, TNF- $\alpha$  and hsCRP. The PRS had a significantly  
87 positive correlation with serum concentrations of hsCRP and TNF- $\alpha$  (p<0.001 and  
88 p<0.05, respectively), but not other markers (Figure 2B). As age and sex are very  
89 important factors related to the susceptibility to SARS-CoV-2 infection, we  
90 performed subgroup analysis stratified by age (<58 years vs.  $\geq$ 58 years, with 58 years  
91 as the median age of this cohort) and sex. Interestingly, we found that higher PRS was  
92 significantly correlated with higher serum concentrations of all the aforementioned  
93 inflammatory markers among older individuals (>58 years, n=493), but not among  
94 younger individuals ( $\leq$ 58 years, n=497) (Figure 2B and 2C). The PRS did not show  
95 any differential association with the inflammatory markers by sex (Figure S2).  
96 Whether the identified proteomic changes causally induce immune activation or  
97 consequences of the immune response are not clear at present, but the finding  
98 supports the hypothesis that the PRS may act as a biomarker of unbalanced host  
99 immune system, especially among older adults.

100

### 101 **Core microbiota features predict COVID-19 proteomic risk score and host** 102 **inflammation**

103 To investigate the potential role of gut microbiota in the susceptibility of healthy  
104 individuals to COVID-19, we next explored the relationship between the gut  
105 microbiota and the above COVID-19-related PRS in a sub-cohort of 301 participants  
106 with measurement of both gut microbiota (16s rRNA) and blood proteomics data

---

107 (Figure S1). Gut microbiota data were collected and measured during a follow-up  
108 visit of the cohort participants, with a cross-sectional subset of the individuals (n=132)  
109 having blood proteomic data at the same time point as the stool collection and another  
110 independent prospective subset of the individuals (n=169) having proteomic data at a  
111 next follow-up visit ~3 years later than the stool collection.

112

113 Among the cross-sectional subset, using a machine learning-based method:  
114 LightGBM and a very conservative and strict tenfold cross-validation strategy, we  
115 identified 20 top predictive operational taxonomic units (OTUs), and this subset of  
116 core OTUs explained an average 21.5% of the PRS variation (mean out-of-sample  
117  $R^2=0.215$  across ten cross-validations). The list of these core OTUs along with their  
118 taxonomic classification is provided in Table S3. These OTUs were mainly assigned  
119 to *Bacteroides* genus, *Streptococcus* genus, *Lactobacillus* genus, *Ruminococcaceae*  
120 family, *Lachnospiraceae* family and *Clostridiales* order.

121

122 To test the verification of the core OTUs, the Pearson correlation analysis showed the  
123 coefficient between the core OTUs-predicted PRS and actual PRS reached 0.59  
124 ( $p<0.001$ ), substantially outperforming the predictive capacity of other demographic  
125 characteristics and laboratory tests including age, BMI, sex, blood pressure and blood  
126 lipids (Pearson's  $r=0.154$ ,  $p=0.087$ ) (Figure 3A). Additionally, we used co-inertia  
127 analysis (CIA) to further test co-variance between the 20 identified core OTUs and 20  
128 predictive proteomic biomarkers of severe COVID-19, outputting a RV coefficient  
129 (ranged from 0 to 1) to quantify the closeness. The results indicated a close  
130 association of these OTUs with the proteomic biomarkers (RV=0.12,  $p<0.05$ ) (Figure  
131 S3A). When replicating this analysis stratified by age, significant association was  
132 observed only among older participants (age $\geq$ 58, n=66; RV=0.22,  $p<0.05$ ) (Figure  
133 S3B and S3C).

134

135 Importantly, the above results from cross-sectional analyses were successfully  
136 replicated in the independent prospective subset of 169 individuals, which showed a

---

137 Pearson's  $r$  of 0.18 between the core OTUs-predicted PRS versus actual PRS ( $p < 0.05$ ),  
138 also outperforming the predictive capacity of the above demographic characteristics  
139 and laboratory tests (Pearson's  $r = 0.08$ ,  $p = 0.31$ ) (Figure 3A). These findings support  
140 that change in the gut microbiota may precede the change in the blood proteomic  
141 biomarkers, inferring a potential causal relationship.

142

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

152

### 153 **Fecal metabolome may be the key to link the PRS-related core microbial features** 154 **and host inflammation**

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

---

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

178

### 179 **Host and environmental factors modulate the PRS-related core microbial OTUs**

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

192

### 193 **Discussion**

194 Our findings suggest that, among healthy non-infected individuals, gut microbial  
195 features are highly predictive of the blood proteomic biomarkers of severe COVID-19  
196 disease. The disruption of the corresponding gut microbiome features may potentially



---

197 predispose healthy individuals to abnormal inflammatory status, which may further  
198 account for the COVID-19 susceptibility and severity. The fecal metabolomics  
199 analysis reveals that amino acid-related pathway may provide the key link between  
200 the identified core gut microbiota, inflammation and COVID-19 susceptibility.  
201 Furthermore, modifications on host and environmental factors are likely to influence  
202 the above core gut microbiota compositions.

203

204 Accumulating evidence suggests that “cytokine storm”, an excessive production of  
205 inflammatory cytokines, may be an important mechanism leading to the severity and  
206 death of COVID-19 patients (Huang et al., 2020; Yang et al., 2020). Therefore,  
207 anticytokine therapy for the suppression of the hyperinflammatory status of the  
208 patients is a recommended strategy to treat severe COVID-19 patients (Mehta et al.,  
209 2020; Monteleone et al., 2020). Among the 20 proteomic predictors of severe  
210 COVID-19, several most upregulated proteins are activated acute phase proteins,  
211 including serum amyloid A-1 (SAA1), SAA2, SAA4, alpha-1-antichymotrypsin  
212 (SERPINA3), complement 6 (C6) and complement factor B (CFB) (Shen et al., 2020).  
213 These proteins may be activated together with proinflammatory cytokines such as  
214 IL-6 and TNF- $\alpha$  following the invasion of the SARS-CoV-2. Therefore, this set of  
215 proteomic biomarkers may serve as an important biomarker or therapeutic target for  
216 treating SARS-CoV-2 infection. Beyond the previous data from the COVID-19  
217 patients, our current study based on data from healthy non-infected participants  
218 consistently supports that the proteomic biomarkers (integrated into a score) are  
219 positively associated with proinflammatory cytokines, especially among those with an  
220 older age. These results imply that the proteomic changes may precede the progression  
221 of COVID-19 to severe phase. Moreover, our finding of more significant associations  
222 between PRS and proinflammatory cytokines among older people agree with the  
223 observation during COVID-19 outbreak that older individuals are more susceptible to  
224 the virus, leading to severity of the disease, due to the induced hyperinflammation or  
225 “cytokine storm” (Chen et al., 2020; Zhou et al., 2020).

---

226 In the present study, the core gut microbial features (20 OTUs), with a satisfied  
227 performance, outperform demographic characteristics and laboratory tests in  
228 predicting the blood proteomic biomarkers, which highlights a potential role of gut  
229 microbiota in regulating the susceptibility of COVID-19 among normal individuals.  
230 In fact, maintaining gut homeostasis has been suggested as a treatment option in the  
231 “Diagnosis and Treatment Plan of Corona Virus Disease 2019 (Tentative Sixth  
232 Edition)” issued by National Health Commission of China, as to keep the equilibrium  
233 for intestinal microecology and prevent secondary bacterial infection (National Health  
234 Commission (NHC) of the PRC, 2020). Growing evidence has shown that microbiota  
235 plays a fundamental role on the induction, training and function of the host immune  
236 system, and the composition of the gut microbiota and its activity are involved in  
237 production of inflammatory cytokines (Belkaid and Hand, 2014; Cani and Jordan,  
238 2018). Prior studies reported that *Lactobacillus* genus was positively associated with  
239 IL-6 and IFN- $\gamma$ , while *Blautia* genus was positively associated with IL-10 (Jiang et al.,  
240 2012; Pohjavuori et al., 2004; Yoshida et al., 2001); these relationships were  
241 replicated in our study. Besides, we found the PRS-related OTUs belonging to  
242 *Bacteroides* genus and *Streptococcus* genus were negatively associated with most  
243 proinflammatory factors. These results further support the reliability of the selected  
244 core OTUs.

245  
246 Fecal metabolomics analyses for the identified core gut microbial OTUs suggest that  
247 these OTUs may be closely associated with amino acid metabolism, especially  
248 aminoacyl-tRNA biosynthesis pathway, arginine biosynthesis pathway, and valine,  
249 leucine and isoleucine biosynthesis pathway. As metabolic stress pathways and  
250 nutrient availability instruct immunity, amino acid levels in the tissue  
251 microenvironment are central to the maintenance of immune homeostasis (Murray,  
252 2016). Amino acid insufficiency will cause depletion of available aminoacylated  
253 tRNA, which is essential for the host to sense amino acid limitation and immune  
254 response (Brown et al., 2016, 2010; Harding et al., 2003). A recent study on several  
255 mammalian cell models reported that when aminoacyl-tRNA synthetase was inhibited,

---

256 the cytokine stimulated proinflammatory response would be substantially suppressed,  
257 and a single amino acid depletion, such as arginine or histidine, could also suppress  
258 the cytokine induced immune response (Kim et al., 2020). Thus the identified  
259 pathways regulating in aminoacyl-tRNA biosynthesis and arginine biosynthesis may  
260 be both involved in the inflammatory response. Additionally, arginine and BCAAs  
261 (i.e., valine, leucine and isoleucine), were also reported regulating innate and adaptive  
262 immune responses and enhancing intestinal development (Zhang et al., 2017).  
263 Collectively, these key roles that amino acids play in the immunoregulation may help  
264 explain how the PRS-related core OTUs modulate host inflammation via amino acid  
265 metabolism. Furthermore, given the high expression of ACE2 in the ileum and colon,  
266 and the role of ACE2 as a key regulator of dietary amino acid homeostasis and innate  
267 immunity (Hashimoto et al., 2012; Zhang et al., 2020), ACE2 may be another key  
268 mediator between gut microbiota and host inflammation. However, whether and how  
269 ACE2 may mediate the association between gut microbiota and COVID-19 severity  
270 warrants further mechanistic study.

271

272 We observed that several host demographic and clinical factors had a strong effect on  
273 the identified core OTU composition, among which drug use and metabolic  
274 phenotypes had been widely reported correlating with gut microbiome composition  
275 (Cabreiro et al., 2013; Gilbert et al., 2018; Vich Vila et al., 2020). Although these  
276 observations were quite crude, it gave us an overview of the potential influence of  
277 host and environmental factors on the PRS-related gut microbiota matrix. Those  
278 known factors contributed to the COVID-19 susceptibility also contributed to the  
279 variance of the gut microbiota, including age, sex, and indicators of clinical  
280 comorbidities (blood pressure, glucose triglycerides, high-density and low-density  
281 lipoprotein cholesterol, and diabetes medication).

282

283 In summary, our study provides novel insight that gut microbiota may underlie the  
284 susceptibility of the healthy individuals to the COVID-19. In the global crisis of  
285 COVID-19, a wide disparity in the susceptibility of the disease or disease progression

---

286 has been observed. Our results provide important evidence and suggestions about the  
287 potential biological mechanism behind the diverse susceptibility among different  
288 groups of people. The discovered core gut microbial features and related metabolites  
289 may serve as a potential preventive/treatment target for intervention especially among  
290 those who are susceptible to the SARS-CoV-2 infection. They could also serve as  
291 potential therapeutic targets for drug development.

292

### 293 **Acknowledgements**

294 This study was funded by the National Natural Science Foundation of China  
295 (81903316, 81773416, 81972492, 21904107, 81672086), Zhejiang Ten-thousand  
296 Talents Program (101396522001), Zhejiang Provincial Natural Science Foundation  
297 for Distinguished Young Scholars (LR19C050001), the 5010 Program for Clinical  
298 Researches (2007032) of the Sun Yat-sen University, Hangzhou Agriculture and  
299 Society Advancement Program (20190101A04), and Tencent foundation (2020). The  
300 funder had no role in study design, data collection and analysis, decision to publish, or  
301 writing of the manuscript. We thank Dr. C.R. Palmer for his invaluable comments to  
302 this study; and Westlake University Supercomputer Center for assistance in data  
303 storage and computation; and all the patients who consented to donate their clinical  
304 information and samples for analysis; and all the medical staff members who are on  
305 the front line fighting against COVID-19.

306

### 307 **Author Contributions**

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

### 314 **Declaration of Interests**

315 The authors declare no competing financial interests

---

## 316 **Figure Legends**

317 **Figure 1. Study design and analysis pipeline.** Study overview. 1) constructing a  
318 novel COVID-19 blood proteomic risk score (PRS) among 31 COVID-19 patients (18  
319 non-severe cases and 13 severe cases). 2) Applying the PRS in healthy participants,  
320 and further linking it to host inflammatory status (n=990). 3) Investigating the  
321 potential role of gut microbiota in predicting the PRS of COVID-19 based on a  
322 machine-learning method (n=301). 4) Assessing the relationships between the  
323 PRS-related gut microbiota and inflammatory factors (n=336). 5) Fecal metabolomics  
324 analysis reveals function of gut microbiota on host metabolism (n=987). 6)  
325 Investigating the impact of host and environmental factors on PRS-related core  
326 microbial OTUs (n=1729).

327

328 **Figure 2. Predictive proteomic profile for severe COVID-19 is correlated with**  
329 **pro-inflammatory factors among healthy individuals. (A)** The associations of  
330 COVID-19-related blood proteomic biomarkers and proteomic risk score (PRS) with  
331 host inflammatory markers. 990 participants were involved in this analysis. # protein  
332 down-regulated in severe patients, else, up-regulated.

333 **(B)** The correlation of the above blood proteomic biomarkers and PRS with host  
334 inflammatory markers stratified by the median age of participants (<58 years or ≥58  
335 years). The color of the heatmap indicates the Spearman correlation coefficients  
336 (blue-negative, red-positive). **(C)** The correlation of the PRS with individual host  
337 inflammatory markers stratified by the median age of participants (<58 years or ≥58  
338 years).

339

340 **Figure 3. Core microbiota features predict COVID-19 proteomic risk score (PRS)**  
341 **and host inflammation. (A)** Plots of out-of-sample predicted PRS versus actual PRS  
342 based on top 20 ranked OTUs or demographic/clinical factors (age, sex BMI, fasting  
343 glucose, HDL, LDL, TC, TG, DBP, and SBP) using LightGBM with 10-fold  
344 cross-validation. The plots in the first row indicate the model performance among  
345 cross-sectional subset of individuals (n=132); the plots in the second row indicate the

---

346 model performance among prospective subset of individuals (n=169). The mean  $R^2$   
347 across the 10 cross validations, Pearson r of predicted values versus actual values, and  
348 corresponding P-value are shown in the figures. **(B)** The correlation of the core  
349 microbial OTUs and host inflammatory cytokines (n=336). The color of the heatmap  
350 indicates the Spearman correlation coefficients (blue-negative, red-positive).

351

352 **Figure 4. Fecal metabolome may be the key to link the proteomic risk**

353 **score-related core microbial features and host inflammation**

354 **(A)** Associations of the core microbial OTUs with fecal metabolites (n=987). The  
355 relationships between microbial OTUs and fecal metabolites was assessed by a linear  
356 regression model adjusting for age, sex, BMI. Multiple testing was adjusted using  
357 Benjamini and Hochberg method, with a false discovery rate (FDR) of <0.05 being  
358 considered statistically significant. We only presented metabolites showing significant  
359 associations with more than half of the core microbial OTUs (n=20) in the figure.

360 Sizes of the nodes represent the number of OTUs related with fecal metabolites. Red  
361 edge,  $\beta$ -coefficient >0; blue edge,  $\beta$ -coefficient <0. **(B)** Pathway analysis for the core

362 fecal metabolites (shown in part A) using MetaboAnalyst 4.0 (Chong et al., 2019). **(C)**  
363 Metabolites enriched in the significant pathways (shown in part B).

364

365 **Figure 5. Host and environmental factors modulate the blood proteomic risk**

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

371

---

372 **STAR MEHTODS**

373 **RESOURCE AVAILABILITY**

374 *Lead Contact*

375 Further information and requests for resources and reagents should be directed to and  
376 will be fulfilled by the Lead Contact, Ju-Sheng Zheng  
377 ([zhengjusheng@westlake.edu.cn](mailto:zhengjusheng@westlake.edu.cn)).

378

379 *Materials Availability*

380 This study did not generate new unique reagents.

381

382 *Data and Code Availability*

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

385

386 **SUBJECT DETAILS**

387 *COVID-19 proteomics data set*

388 Detailed information about the COVID-19 patients and proteomics data set is  
389 described in our recent publication (Shen et al., 2020). Briefly, the proteome of sera  
390 from 46 COVID-19 patients and 53 control samples from Taizhou Public Health  
391 Medical Center were analyzed by TMTpro 16 plex-based quantitative proteomics  
392 technology. All the patients were diagnosed between January 23 and February 4, 2020.  
393 According to the Chinese Government Diagnosis and Treatment Guideline for  
394 COVID-19, the COVID-19 patients were classified into four groups, (1) mild (mild  
395 symptoms without pneumonia); (2) typical (fever or respiratory tract symptoms with  
396 pneumonia); (3) severe (fulfill any of the three criteria: respiratory distress,  
397 respiratory rate  $\geq 30$  times/min; mean oxygen saturation  $\leq 93\%$  in resting state;  
398 arterial blood oxygen partial pressure/oxygen concentration  $\leq 300$ mmHg); and (4)  
399 critical (fulfill any of the three criteria: respiratory failure and require mechanical  
400 ventilation; shock incidence; admission to ICU with other organ failure). We treated

---

401 mild and typical patients as a non-severe COVID-19 group, and the others a severe  
402 COVID-19 group.

403

404 *Healthy subjects, sample collection, and clinical metadata*

405 In the present study, the healthy (non-infected) subjects are from the  
406 community-based Guangzhou Nutrition and Health Study (GNHS), and the detailed  
407 study designs of GNHS have been reported previously (Zhang et al., 2014). Briefly,  
408 participants were enrolled between 2008 and 2013, and followed up to May 2018.  
409 Blood samples were collected at enrollment and follow-up visits, and stool samples  
410 were collected only during follow-up visits. All the blood samples were collected as  
411 venous whole blood in the early morning before diet using serum separation tubes.  
412 The blood samples were centrifuged at 3,500 rpm for 10 min for serum collection.  
413 The serum samples were frozen at -80°C. The stool samples were collected at a local  
414 study site within the School of Public Health at Sun Yat-sen University, and were  
415 transferred to a -80°C facility within 4 hours after collection.

416

417 Demographic and lifestyle factors were all collected by questionnaire during on-site  
418 face-to-face interviews. Habitual dietary intakes over the past 12 months were  
419 assessed by a food frequency questionnaire, as previously described (Zhang CX,  
420 2009). Physical activity was assessed as a total metabolic equivalent for task (MET)  
421 hours per day on the basis of a validated questionnaire for physical activity (Liu et al.,  
422 2001).

423

424 Anthropometric factors were measured by trained nurses on site during the baseline  
425 interview. Fasting venous blood samples were taken at each recruitment or follow-up  
426 visit. Serum low-density lipoprotein cholesterol and glucose were measured by  
427 colorimetric methods using a Roche Cobas 8000 c702 automated analyzer (Roche  
428 Diagnostics GmbH, Shanghai, China). Intra-assay coefficients of variation (CV) was  
429 2.5% for glucose. Insulin was measured by electrochemiluminescence immunoassay  
430 (ECLIA) methods using a Roche cobas 8000 e602 automated analyzer (Roche



---

431 Diagnostics GmbH, Shanghai, China). High-performance liquid chromatography was  
432 used to measure glycated hemoglobin (HbA1c) using the Bole D-10 Hemoglobin A1c  
433 Program on a Bole D-10 Hemoglobin Testing System, and the intraassay CV was  
434 0.75%.

435

#### 436 *Ethics*

437 This study has been approved by the Ethical/Institutional Review Board of Taizhou  
438 Public Health Medical Center, the Ethics Committee of the School of Public Health at  
439 Sun Yat-sen University and Ethics Committee of Westlake University.

440

#### 441 **METHOD DETAILS**

##### 442 *Proteomic analysis*

443 1  $\mu$ L of serum sample from each patient was analyzed using proteomics technology.  
444 The serum was firstly denatured with 20  $\mu$ L of buffer containing 8 M urea (Sigma,  
445 #U1230) in 100 mM ammonium bicarbonate at 30°C for 30 min. The lysates were  
446 reduced with 10 mM tris (2-carboxyethyl) phosphine (TCEP, Sigma #T4708) at room  
447 temperature for 30 min, and were then alkylated with 40 mM iodoacetamide (IAA,  
448 Sigma, #SLCD4031) in darkness for 45 min. The solution was then diluted with 70  
449  $\mu$ L 100 mM ammonium bicarbonate to make sure urea concentration is less than 1.6M,  
450 and was subjected to two times of tryptic digestion (Hualishi Tech. Ltd, Beijing,  
451 China), each step with 2.5  $\mu$ L trypsin (0.4  $\mu$ g/ $\mu$ L), at 32°C for 4 hr and 12 hr,  
452 respectively. Thereafter, the solution was acidified with 1% trifluoroacetic (TFA)  
453 (Thermo Fisher Scientific, #T/3258/PB05) to pH 2–3 to stop the reaction. Peptides  
454 were cleaned using C18 (Thermo, #60209-001).

455

456 Peptide samples were then injected for LC-MS/MS analysis using an Eksigent  
457 NanoLC 400 System (Eksigent, Dublin, CA, USA) coupled to a TripleTOF 5600  
458 system (SCIEX, CA, USA). Briefly, peptides were loaded onto a trap column (5  $\mu$ m,  
459 120 Å, 10  $\times$  0.3 mm), and were separated along a 20 min LC gradient ( 5–32%

---

460 buffer B, 98% ACN, 0.1% formic acid in HPLC water; buffer A: 2% ACN, 0.1%  
461 formic acid in HPLC water) on an analytical column (3  $\mu\text{m}$ , 120  $\text{\AA}$ , 150  $\times$  0.3 mm) at  
462 a flow rate of 5  $\mu\text{L}/\text{min}$ . The SWATH-MS method is composed of a 100 ms of full  
463 TOF MS scan with the acquisition range of 350-1250  $m/z$ , followed by MS/MS scans  
464 performed on all precursors (from 100 to 1500 Da) in a cyclic manner (Gillet et al.,  
465 2012). A 55-variable-Q1 isolation window scheme was used in this study. The  
466 accumulation time was set at 30 ms per isolation window, resulting in a total cycle  
467 time of 1.9 s.

468  
469 After SWATH acquisition, the Wiff files were converted into mzXML format using  
470 msconvert (ProteoWizard 3.0) (Kessner et al., 2008) and analyzed using OpenSWATH  
471 (2.1) (Kessner et al., 2008) against a pan human spectral library (Kessner et al., 2008)  
472 that contains 43899 peptide precursors and 1667 unique Swiss-Prot proteins protein  
473 groups. The retention time extraction window was set at 120 seconds, and the  $m/z$   
474 extraction was performed with 30 ppm tolerance. Retention time was then calibrated  
475 using Common internal Retention Time standards (CiRT) peptides (Kessner et al.,  
476 2008). Peptide precursors were identified by OpenSWATH (version 2.0) and  
477 pyprophet (version 0.24) with  $\text{FDR} < 0.01$  to quantify the proteins in each sample.

478  
479 *Measurement of inflammatory biomarkers*

480 For samples collected at baseline, Human FlowCytomix (Simplex BMS8213FF and  
481 BMS8288FF, eBioscience, San Diego, CA, USA) and the Human Basic Kit  
482 FlowCytomix (BMS8420FF, eBioscience, San Diego, CA, USA) on a BD  
483 FACSCalibur instrument (BD Biosciences, Franklin Lakes, NJ, USA) were used for  
484 the measurements of serum tumor necrosis factor (TNF- $\alpha$ ), Interleukin-6 (IL-6), and  
485 Interleukin-1 $\beta$  (IL-1 $\beta$ ). High-sensitivity CRP was measured using a [Cardiac  
486 C-Reactive Protein (Latex) High Sensitive (CRPHS) kit], and detected on a Cobas  
487 c701 automatic analyzer. The between-plate CVs were 14.1% for MCP1, 6.6% for  
488 TNF- $\alpha$ , 2.5% for IL-6, and 10.2% for IL-1 $\beta$ .

489

---

490 For samples collected during follow-up visits, serum cytokine levels were assessed by  
491 electrochemiluminescence based immunoassays using the MSD V-Plex  
492 Proinflammatory Panel 1 (human) kit. 50  $\mu$ L of serum derived from whole blood by  
493 centrifugation (10 min, 3,500 rpm) was processed according to the manufacturer's  
494 instructions. Briefly, the serum was diluted at a minimum of 2-fold dilution at first,  
495 while the detection antibodies were combined and added to 2400  $\mu$ L of diluent.  
496 Thereafter, wash buffer and read buffer T were prepared as instructed. After finishing  
497 washing and adding samples, washing and adding detection antibody solution and  
498 washing plates again, then the plate could be analyzed on an MSD instrument.  
499 Accuracy and precision are evaluated by measuring calibrators across multiple runs  
500 and multiple lots. Intra-run coefficient of variations (CVs) are typically below 7% and  
501 inter-ran CVs are typically below 15%. In the present study, the inters-run CVs of  
502 calibrators were 2.2% for IL-1 $\beta$ , 2.5% for IL-2, 1.6% for IL-4, 1.7 for IL-6, 3.6 for  
503 IL-8, 2.5 for 1.3% for IL-10, 1.1% for IL-12p70, 0.87% for IL-13, TNF- $\alpha$ , 2.0% for  
504 IFN- $\gamma$ .

505

#### 506 *Microbiome analysis ---- DNA extraction*

507 Total bacterial DNA was extracted using the QIAamp® DNA Stool Mini Kit (Qiagen,  
508 Hilden, Germany) following the manufacturer's instructions. DNA concentrations  
509 were measured using the Qubit quantification system (Thermo Scientific, Wilmington,  
510 DE, US). The extracted DNA was then stored at -20 °C.

511

#### 512 *Microbiome analysis ---- 16S rRNA gene amplicon sequencing*

513 The 16S rRNA gene amplification procedure was divided into two PCR steps, in the  
514 first PCR reaction, the V3-V4 hypervariable region of the 16S rRNA gene was  
515 amplified from genomic DNA using primers 341F(CCTACGGGNGGCWGCAG) and  
516 805R(GACTACHVGGGTATCTAATCC). Amplification was performed in 96-well  
517 microtiter plates with a reaction mixture consisting of 1X KAPA HiFi Hot start Ready  
518 Mix, 0.1  $\mu$ M primer 341 F, 0.1  $\mu$ M primer 805 R, and 12.5 ng template DNA giving a  
519 total volume of 50  $\mu$ L per sample. Reactions were run in a T100 PCR thermocycle

---

520 (BIO-RAD) according to the following cycling program: 3 min of denaturation at  
521 94 °C, followed by 18 cycles of 30 s at 94 °C (denaturing), 30 s at 55 °C (annealing),  
522 and 30 s at 72 °C (elongation), with a final extension at 72 °C for 5 min. Subsequently,  
523 the amplified products were checked by 2% agarose gel electrophoresis and ethidium  
524 bromide staining. Amplicons were quantified using the Qubit quantification system  
525 (Thermo Scientific, Wilmington, DE, US) following the manufacturers' instructions.  
526 Sequencing primers and adaptors were added to the amplicon products in the second  
527 PCR step as follows 2 µL of the diluted amplicons were mixed with a reaction  
528 solution consisting of 1×KAPA HiFi Hotstart ReadyMix, 0.5µM fusion forward and  
529 0.5µM fusion reverse primer, 30 ng Meta-gDNA(total volume 50 µL). The PCR was  
530 run according to the cycling program above except with cycling number of 12. The  
531 amplification products were purified with Agencourt AMPure XP Beads (Beckman  
532 Coulter Genomics, MA, USA) according to the manufacturer's instructions and  
533 quantified as described above. Equimolar amounts of the amplification products were  
534 pooled together in a single tube. The concentration of the pooled libraries was  
535 determined by the Qubit quantification system. Amplicon sequencing was performed  
536 on the Illumina MiSeq System (Illumina Inc., CA, USA). The MiSeq Reagent Kits v2  
537 (Illumina Inc.) was used. Automated cluster generation and 2 × 250 bp paired-end  
538 sequencing with dual-index reads were performed.

539

#### 540 *Microbiome analysis ---- 16S rRNA gene sequence data processing*

541 Fastq-files were demultiplexed by the MiSeq Controller Software (Illumina Inc.). The  
542 sequence was trimmed for amplification primers, diversity spacers, and sequencing  
543 adapters, merge-paired and quality filtered by USEARCH. UPARSE was used for  
544 OTU clustering equaling or above 97%. Taxonomy of the OTUs was assigned and  
545 sequences were aligned with RDP classifier. The OTUs were analyzed by  
546 phylogenetic and operational taxonomic unit (OTU) methods in the Quantitative  
547 Insights into Microbial Ecology (QIIME) software version 1.9.0 (Caporaso et al.,  
548 2010).

549

---

550 *Metabolomic analysis ---- sample preparation and instrumentation*

551 Targeted metabolomics approach was used to analyze fecal samples, with a total of  
552 198 metabolites quantified. Feces samples were thawed on ice-bath to diminish  
553 degradation. About 10mg of each sample was weighed and transferred to a new  
554 1.5mL tube. Then 25 $\mu$ L of water was added and the sample was homogenated with  
555 zirconium oxide beads for 3 minutes. 185 $\mu$ L of ACN/Methanol (8/2) was added to  
556 extract the metabolites. The sample was centrifuged at 18000g for 20 minutes. Then  
557 the supernatant was transferred to a 96-well plate. The following procedures were  
558 performed on a Biomek 4000 workstation (Biomek 4000, Beckman Coulter, Inc.,  
559 Brea, California, USA). 20 $\mu$ L of freshly prepared derivative reagents was added to  
560 each well. The plate was sealed and the derivatization was carried out at 30°C for 60  
561 min. After derivatization, 350 $\mu$ L of ice-cold 50% methanol solution was added to  
562 dilute the sample. Then the plate was stored at -20°C for 20 minutes and followed by  
563 4000g centrifugation at 4 °C for 30 minutes. 135 $\mu$ L of supernatant was transferred to  
564 a new 96-well plate with 15 $\mu$ L internal standards in each well. Serial dilutions of  
565 derivatized stock standards were added to the left wells. Finally the plate was sealed  
566 for LC-MS analysis.

567

568 An ultra-performance liquid chromatography coupled to tandem mass spectrometry  
569 (UPLC-MS/MS) system (ACQUITY UPLC-Xevo TQ-S, Waters Corp., Milford, MA,  
570 USA) was used to quantitate the microbial metabolite in the present study. The  
571 optimized instrument settings are briefly described below. ACQUITY UPLC BEH  
572 C18 1.7  $\mu$ M VanGuard pre-column (2.1 $\times$ 5 mm) and ACQUITY UPLC BEH C18 1.7  
573  $\mu$ M analytical column (2.1  $\times$  100 mm) were used. The column temperature was 40°C  
574 and sample manager temperature was 10°C. Mobile phase A was water with 0.1%  
575 formic acid, and B was acetonitrile / IPA (90:10). The gradient conditions were as  
576 follows: 0-1 min (5% B), 1-12 min (5-80% B), 12-15 min (80-95% B), 15-16 min  
577 (95-100%B), 16-18 min (100%B), 18-18.1 min (100-5% B), 18.1-20 min (5% B), at a  
578 flow rate of 0.40 mL/min. The capillary of mass spectrometer were 1.5 (ESI+) and 2.0  
579 (ESI-), while the source temperature and desolvation temperature was 150°C and

---

580 550°C, respectively. The desolvation gas flow was 1000 L/hour.

581

582 *Metabolome analysis ---- Analytical quality control procedures*

583 The rapid turnover of many intracellular metabolites makes immediate metabolism  
584 quenching necessary. The extraction solvents are stored in -20°C freezer overnight  
585 and added to the samples immediately after the samples were thawed. We use ice-salt  
586 bath to keep the samples at a low temperature and minimize sample degradation  
587 during sample preparation. All the prepared samples should be analyzed within 48  
588 hours after sample extraction and derivatization.

589

590 A comprehensive set of rigorous quality control/assurance procedures is employed to  
591 ensure a consistently high quality of analytical results, throughout controlling every  
592 single step from sample receipt at laboratory to final deliverables.

593 The ultimate goal of QA/QC is to provide the reliable data for biomarker discovery  
594 study and/ or to aid molecular biology research. To achieve this, three types of quality  
595 control samples i.e., test mixtures, internal standards, and pooled biological samples  
596 are routinely used in the metabolomics platform. In addition to the quality controls,  
597 conditioning samples, and solvent blank samples are also required for obtaining  
598 optimal instrument performance.

599

600 Test mixtures comprise a group of commercially available standards with a mass  
601 range across the system mass range used for the study samples. These samples were  
602 analyzed at the beginning and end of each batch run to ensure that the instruments  
603 were performing within laboratory specifications (retention time stability,  
604 chromatographic peak shape, and peak signal intensity). The retention time shift  
605 should be within 4 sec. and the difference of peak intensity should be within 15% for  
606 LC-MS.

607

608 Internal standards were added to the test samples in order to monitor analytical  
609 variations during the entire sample preparation and analysis processes. The Pooled

---

610 QC samples were prepared by mixing aliquots of the study samples such that the  
611 pooled samples broadly represent the biological average of the whole sample set. The  
612 QC samples for this project were prepared with the test samples and injected at  
613 regular intervals (after every 14 test samples for LC-MS) throughout the analytical  
614 run.

615

616 Reagent blank samples are a mixture of solvents used for sample preparation and  
617 are commonly processed using the same procedures as the samples to be analyzed.  
618 The reagent blanks serve as a useful alert to systematic contamination. As the reagent  
619 blanks consist of high purity solvents and are analyzed using the same methods as the  
620 study samples, they are also used to wash the column and remove cumulative matrix  
621 effects throughout the study.

622

623 The calibrators consist of a blank sample (matrix sample processed without  
624 internal standard), a zero sample (matrix sample processed with internal standard),  
625 and a series of seven concentrations covering the expected range for the metabolites  
626 present in the specific biological samples. LLOQ and ULOQ are the lowest and  
627 highest concentration of the standard curve that can be measured with acceptable  
628 accuracy and precision.

629

630 To diminish analytical bias within the entire analytical process, the samples were  
631 analyzed in group pairs but the groups were analyzed randomly. The QC samples,  
632 calibrators, and blank samples were analyzed across the entire sample set.

633

634 *Metabolome analysis ----Software and quantitation*

635 The raw data files generated by UPLC-MS/MS were processed using the  
636 QuanMET software (v2.0, Metabo-Profile, Shanghai, China) to perform peak  
637 integration, calibration, and quantification for each metabolite. The current QuanMET  
638 is hosted on Dell PowerEdge R730 Servers operated with Linux Ubuntu 16.10 OS.

---

639 The secured Java UI (User Interface) permits the user have access to use a great  
640 variety of statistical tools for viewing and exploring project data.

641

642 Mass spectrometry-based quantitative metabolomics refers to the determination  
643 of the concentration of a substance in an unknown sample by comparing the unknown  
644 to a set of standard samples of known concentration (i.e., calibration curve). The  
645 calibration curve is a plot of how the analytical signal changes with the concentration  
646 of the analyte (the substance to be measured). For most analyses a plot of instrument  
647 response vs. concentration will show a linear relationship and the concentration of the  
648 measured samples were calculated.

649

## 650 **STATISTICAL ANALYSIS**

651 *Dataset at each step of analyses among the healthy individuals from Guangzhou*  
652 *Nutrition and Health Study*

653 **Dataset 1 (n=990):** data from the baseline of GNHS. A total of 990 subjects with  
654 measurement of serum proteomics at baseline of the GNHS cohort were included in  
655 the initial discovery cohort. Among this initial discovery cohort, 455 subjects had data  
656 of serum IL-1 $\beta$  and IL-6, 456 subjects had data of serum TNF- $\alpha$ , and 953 subjects had  
657 serum hsCRP data. These data were used to investigate the relationship between the  
658 PRS and host inflammatory status.

659

660 **Dataset 2 (n=301):** data from a follow-up visit of GNHS. A sub-cohort of 301  
661 participants with measurement of both gut microbiota (16s rRNA) and blood  
662 proteomics data. Gut microbiota data were collected and measured during a follow-up  
663 visit of the cohort participants, with a cross-sectional subset of the individuals (n=132)  
664 having blood proteomic data at the same time point as the stool collection and another  
665 independent prospective subset of the individuals (n=169) having proteomic data at a  
666 next follow-up visit ~3 years later than the stool collection. Data from these subjects  
667 were used to explore the predictive capacity of the gut microbiota for PRS.

668



---

669 **Dataset 3 (n=366):** data from a follow-up visit of GNHS. Independent from the  
670 above 301 subjects in dataset 2, there were additionally 336 subjects with both fecal  
671 16s rRNA sequencing and serum inflammatory cytokines data, at the same time point  
672 during follow-up. The data from these 336 subjects were used to examine the  
673 relationships between the core OTUs and 10 host inflammatory cytokines.

674

675 **Dataset 4 (n=987):** data from a follow-up visit of GNHS. A total of 987 individuals  
676 had received fecal metabolomics and fecal microbiome examination at the same time  
677 point during the follow-up visit. These subjects were included in the analysis to assess  
678 the relationships between the core gut microbiota and fecal metabolomics.

679

680 **Dataset 5 (n=1729):** data from a follow-up visit of GNHS. In total, 1729 participants  
681 finished food frequency questionnaire, demographic questionnaire and medical  
682 examination, and provided stool samples during follow-up. Thus, this subset of 1729  
683 subjects were included to test how the dietary habits, lifestyle and health status  
684 influence the gut microbiota composition.

685

686 In summary, a sum of 2413 healthy non-infected individuals are involved in the  
687 present study, which mainly consists of a subset of subjects with proteomic data at  
688 baseline (n=990) and a subset of subjects with gut microbiome and metabolome data  
689 at a follow-up visit (n=2172, within which 301 individuals also had proteomic data).

690

691 *Data imputation and presentation*

692 Missing values in proteomic features were imputed with 50% of the minimal value.

693 Data are presented as mean  $\pm$  SD or percentage as indicated. Statistical tests used to

694 compare conditions are indicated in figure legends. Unless otherwise stated, statistical

695 analysis was performed using Python 3.7, R software (version 3.6.1, R foundation for

696 Statistical Computing, Austria), and Stata 15 (StataCorp, College Station, TX, USA).

697

698

---

699 *Construction of proteomic risk score (PRS)*

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

$$703 \quad PRS_i = \sum_{j=1}^{20} \beta_j x_{ij}$$

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

709

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

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

714

715 *Correlation between PRS and pro-inflammatory biomarkers*

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

719

720 *Machine learning algorithms for identifying microbial features to predict PRS*

721 A 10-fold cross-validation (CV) implementation of gradient boosting framework  
722 —LightGBM and SHAP (Shapley Additive exPlanations) was used to link input gut  
723 microbial features with PRS (Ke et al., 2017; Lee, 2017). A 10-fold CV predict  
724 implementation was used to generate a OTU-predicted PRS value for each participant.  
725 In this approach, each LightGBM model is trained on 90% of the cohort with 10-fold  
726 CV, and PRS is predicted for the 10% of the participants who were not used for model  
727 optimization. This process is repeated ten-fold resulting in a test PRS set for each

---

728 participant and ten different average absolute SHAP value for each OTUs. The top 20  
729 ranked OTUs based the sum of the average absolute SHAP value across ten-fold were  
730 included in further analysis. The  $R^2$  score was computed by taking the mean of all the  
731  $R^2$  scores across the 10 out-of-sample predictions. Pearson r was calculated using  
732 actual PRS and predicted PRS for the entire cohort. We also compared the predictive  
733 performance for the top 20 ranked OTUs, demographic characteristics and laboratory  
734 tests (age, BMI, sex, blood pressure and blood lipids). Our predictor is based on code  
735 adapted from the sklearn 0.15.2 lightgbm regression (Pedregosa et al., 2011).

736

737 *The relationship between the identified core OTUs and host inflammatory cytokines*

738 Spearman correlation analysis was used to examine the correlation between PRS and  
739 cytokines (i.e., IL-1 $\beta$ , IL-2, IL-4, IL-6, IL-8, IL-10, IL-12p70, IL-13, TNF- $\alpha$  and  
740 IFN- $\gamma$ ).  $p < 0.05$  was considered as statistically significant.

741

742 *Relationship between OTUs and fecal metabolites*

743 Prior to the analysis, we excluded the participants with T2D medication use, and all  
744 fecal metabolites were natural logarithmic transformed to reduce skewness of traits  
745 distributions. Similarly, to reduce skewness of the distribution of microbial taxa  
746 counts, we first added 1 to all OTUs and then performed natural log transformation.  
747 The relationship between fecal metabolites and microbial OTUs was assessed by  
748 linear regression analysis while adjusting for age, sex, BMI. Multiple testing was  
749 adjusted using Benjamini and Hochberg method, with a false discovery rate (FDR) of  
750  $< 0.05$  being considered statistically significant. Metabolites showed significant  
751 associations with more than half of the selected microbial OTUs, were used for  
752 subsequent pathway analysis using MetaboAnalyst 4.0 (Chong et al., 2019).

753

754 *Associations of host and environmental factors with gut microbial features*

755 We assessed how many variations in the identified core OTUs composition  
756 (Bray-Curtis distance) can be explained by host and environmental factors (40 factors)

---

757 using the function *adonis* from the R package *vegan*. The p value was determined by  
758 1000x permutations. The total variation explained was also calculated per category  
759 (demographic/clinical category and dietary/nutritional factors) and for all factors  
760 together. Spearman correlation analysis was used to assess the potential effect of each  
761 factor on each of the core OTU. Multiple testing was adjusted using Benjamini and  
762 Hochberg method, with a false discovery rate (FDR) of <0.05 being considered  
763 statistically significant.  
764

---

765 **Supplementary Figure Legends**

766 **Figure S1. Timeline of the participants enrollment, follow-up visit and sample**  
767 **collection in the Guangzhou Health and Nutrition Study. (A)** At baseline, 4048  
768 subjects provided completed metadata (required for analysis in the present study), and  
769 1114 subjects provided blood samples (1114 for proteomic analysis and 990 for  
770 measurement of inflammatory factors). At follow-up visit, 2172 subjects provided  
771 stool samples (n=1729 for 16s rRNA sequencing; n=987 for metabolomic analysis) ,  
772 among which 667 subjects provided blood samples (n=301 for proteomic analysis;  
773 n=366 for measurement of inflammatory factors). (B) Detail information about the  
774 number of participants in the dataset1-dataset4 used in the present study. (C) Detail  
775 information about the number of participants in the dataset5 used in the present study.

776

777 **Figure S2. The correlation of the blood proteomic biomarkers and PRS with host**  
778 **inflammatory markers stratified by sex (n=990).** The color of the heatmap  
779 indicates the Spearman correlation coefficients (blue-negative, red-positive). # protein  
780 down-regulated in severe patients, else, up-regulated.

781

782 **Figure S3. Co-variation of the core OTUs abundance and predictive proteomic**  
783 **biomarkers of COVID-19 (n=132).** Co-inertia analysis of the relationship between  
784 20 identified OTUs and 20 predictive proteomic biomarkers among a cross-sectional  
785 subset of 132 individuals. Each sample is represented with an arrow. The sample  
786 projection in the OTUs and the proteomic biomarkers space are represented by the  
787 starting point and the end of the arrow, respectively. Length of the arrow represents  
788 distance between the projections. (A, overall; B, age  $\geq$  58; C, age <58).

789

790 **Supplementary Tables**

791 **Table S1. Characteristics of the data analysis sets**

792 **Table S2. List of the 20 proteomic biomarkers integrated in the proteomic risk**  
793 **score**

794 **Table S3. List and taxonomic classifications of the PRS-related core microbial**  
795 **OTUs**

796

---

797 **References**

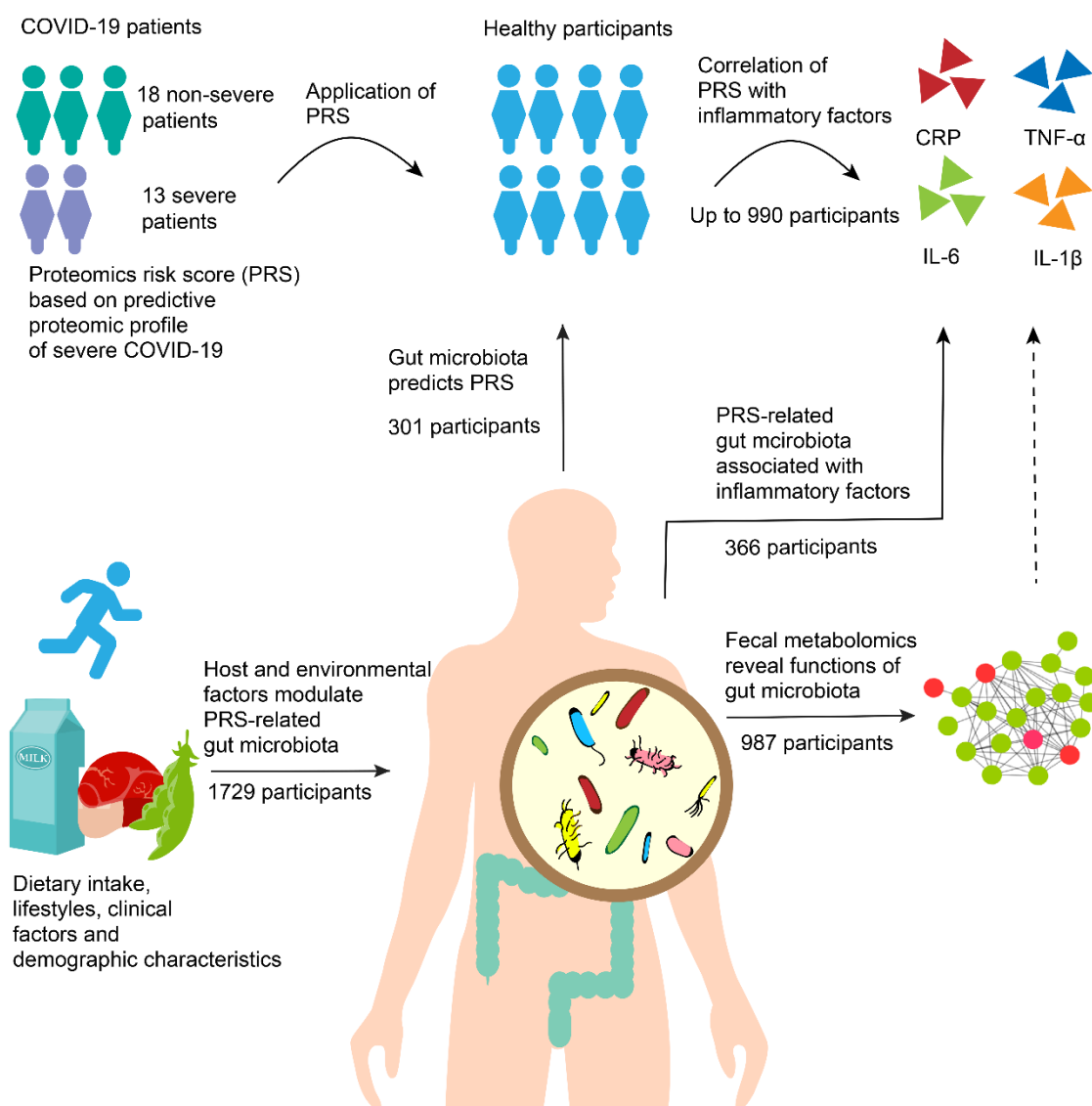
- 798 Belkaid, Y., and Hand, T.W. (2014). Role of the microbiota in immunity and  
799 inflammation. *Cell* 157, 121–141.
- 800 Brown, A., Fernández, I.S., Gordiyenko, Y., and Ramakrishnan, V. (2016).  
801 Ribosome-dependent activation of stringent control. *Nature* 534, 277–280.
- 802 Brown, M. V, Reader, J.S., and Tzima, E. (2010). Mammalian aminoacyl-tRNA  
803 synthetases: cell signaling functions of the protein translation machinery. *Vascular*  
804 *Pharmacology* 52, 21–26.
- 805 Cabreiro, F., Au, C., Leung, K.-Y., Vergara-Irigaray, N., Cochemé, H.M., Noori, T.,  
806 Weinkove, D., Schuster, E., Greene, N.D.E., and Gems, D. (2013). Metformin retards  
807 aging in *C. elegans* by altering microbial folate and methionine metabolism. *Cell* 153,  
808 228–239.
- 809 Cani, P.D., and Jordan, B.F. (2018). Gut microbiota-mediated inflammation in obesity:  
810 a link with gastrointestinal cancer. *Nature Reviews. Gastroenterology & Hepatology*  
811 15, 671–682.
- 812 Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello,  
813 E.K., Fierer, N., Peña, A.G., Goodrich, K., Gordon, J.I., et al. (2010). QIIME allows  
814 analysis of high-throughput community sequencing data. *Nat Methods* 7, 335–336.
- 815 Chen, N., Zhou, M., Dong, X., Qu, J., Gong, F., Han, Y., Qiu, Y., Wang, J., Liu, Y.,  
816 Wei, Y., et al. (2020). Epidemiological and clinical characteristics of 99 cases of 2019  
817 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *Lancet (London,*  
818 *England)* 395, 507–513.
- 819 Chong, J., Wishart, D.S., and Xia, J. (2019). Using MetaboAnalyst 4.0 for  
820 Comprehensive and Integrative Metabolomics Data Analysis. *Current Protocols in*  
821 *Bioinformatics* 68, e86.
- 822 Cole-Jeffrey, C.T., Liu, M., Katovich, M.J., Raizada, M.K., and Shenoy, V. (2015).  
823 ACE2 and Microbiota: Emerging Targets for Cardiopulmonary Disease Therapy.  
824 *Journal of Cardiovascular Pharmacology* 66, 540–550.
- 825 Gilbert, J.A., Blaser, M.J., Caporaso, J.G., Jansson, J.K., Lynch, S. V, and Knight, R.  
826 (2018). Current understanding of the human microbiome. *Nature Medicine* 24, 392–  
827 400.
- 828 Gillet, L.C., Navarro, P., Tate, S., Röst, H., Selevsek, N., Reiter, L., Bonner, R., and  
829 Aebersold, R. (2012). Targeted data extraction of the MS/MS spectra generated by  
830 data-independent acquisition: a new concept for consistent and accurate proteome  
831 analysis. *Mol. Cell. Proteomics*. 11, O111.016717.
- 832 Harding, H.P., Zhang, Y., Zeng, H., Novoa, I., Lu, P.D., Calfon, M., Sadri, N., Yun,  
833 C., Popko, B., Paules, R., et al. (2003). An integrated stress response regulates amino  
834 acid metabolism and resistance to oxidative stress. *Molecular Cell* 11, 619–633.
- 835 Hashimoto, T., Perlot, T., Rehman, A., Trichereau, J., Ishiguro, H., Paolino, M., Sigl,  
836 V., Hanada, T., Hanada, R., Lipinski, S., et al. (2012). ACE2 links amino acid  
837 malnutrition to microbial ecology and intestinal inflammation. *Nature* 487, 477–481.
- 838 Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., Zhang, L., Fan, G., Xu, J., Gu,  
839 X., et al. (2020). Clinical features of patients infected with 2019 novel coronavirus in  
840 Wuhan, China. *Lancet (London, England)* 395, 497–506.

- 841 Jiang, Y., Lü, X., Man, C., Han, L., Shan, Y., Qu, X., Liu, Y., Yang, S., Xue, Y., and  
842 Zhang, Y. (2012). Lactobacillus acidophilus induces cytokine and chemokine  
843 production via NF- $\kappa$ B and p38 mitogen-activated protein kinase signaling pathways in  
844 intestinal epithelial cells. *Clinical and Vaccine Immunology : CVI* 19, 603–608.
- 845 Jin, X., Lian, J.-S., Hu, J.-H., Gao, J., Zheng, L., Zhang, Y.-M., Hao, S.-R., Jia, H.-Y.,  
846 Cai, H., Zhang, X.-L., et al. (2020). Epidemiological, clinical and virological  
847 characteristics of 74 cases of coronavirus-infected disease 2019 (COVID-19) with  
848 gastrointestinal symptoms. *Gut*.
- 849 Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., and Liu, T.-Y.  
850 (2017). LightGBM: A Highly Efficient Gradient Boosting Decision Tree. *Nips '17* 9.
- 851 Kessner, D., Chambers, M., Burke, R., Agus, D., and Mallick, P. (2008).  
852 ProteoWizard: open source software for rapid proteomics tools development.  
853 *Bioinformatics (Oxford, England)* 24, 2534–2536.
- 854 Kim, Y., Sundrud, M.S., Zhou, C., Edenius, M., Zocco, D., Powers, K., Zhang, M.,  
855 Mazitschek, R., Rao, A., Yeo, C.-Y., et al. (2020). Aminoacyl-tRNA synthetase  
856 inhibition activates a pathway that branches from the canonical amino acid response  
857 in mammalian cells. *Proceedings of the National Academy of Sciences of the United*  
858 *States of America*.
- 859 Lee, S.M.L.S.-I. (2017). A Unified Approach to Interpreting Model Predictions.  
860 *NIPS*.
- 861 Lin, L., Jiang, X., Zhang, Z., Huang, S., Zhang, Z., Fang, Z., Gu, Z., Gao, L., Shi, H.,  
862 Mai, L., et al. (2020). Gastrointestinal symptoms of 95 cases with SARS-CoV-2  
863 infection. *Gut*.
- 864 Liu, B., Woo, J., Tang, N., Ng, K., Ip, R., and Yu, A. (2001). Assessment of total  
865 energy expenditure in a Chinese population by a physical activity questionnaire:  
866 examination of validity. *International Journal of Food Sciences and Nutrition* 52,  
867 269–282.
- 868 Mehta, P., McAuley, D.F., Brown, M., Sanchez, E., Tattersall, R.S., and Manson, J.J.  
869 (2020). COVID-19: consider cytokine storm syndromes and immunosuppression.  
870 *Lancet (London, England)* 395, 1033–1034.
- 871 Monteleone, G., Sarzi-puttini, P.C., and Ardizzone, S. (2020). Preventing  
872 COVID-19-induced pneumonia with anticytokine therapy. *The Lancet Rheumatology*  
873 9913, 30092–30098.
- 874 Murray, P.J. (2016). Amino acid auxotrophy as a system of immunological control  
875 nodes. *Nature Immunology* 17, 132–139.
- 876 National Health Commission (NHC) of the PRC, N.A. of T.C.M. of the P. (2020).  
877 Diagnosis and Treatment Plan of Corona Virus Disease 2019 (Tentative Sixth  
878 Edition). Guidance for Corona Virus Disease 2019: Prevention, Control and  
879 Diagnosis and Management. Beijing, China: People's Medical Publishing House,.
- 880 Ng, S.C., and Tilg, H. (2020). COVID-19 and the gastrointestinal tract: more than  
881 meets the eye. *Gut*.
- 882 Pedregosa, F., Weiss, R., and Brucher, M. (2011). Scikit-learn : Machine Learning in  
883 Python. 12, 2825–2830.
- 884 Pohjavuori, E., Viljanen, M., Korpela, R., Kuitunen, M., Tiittanen, M., Vaarala, O.,

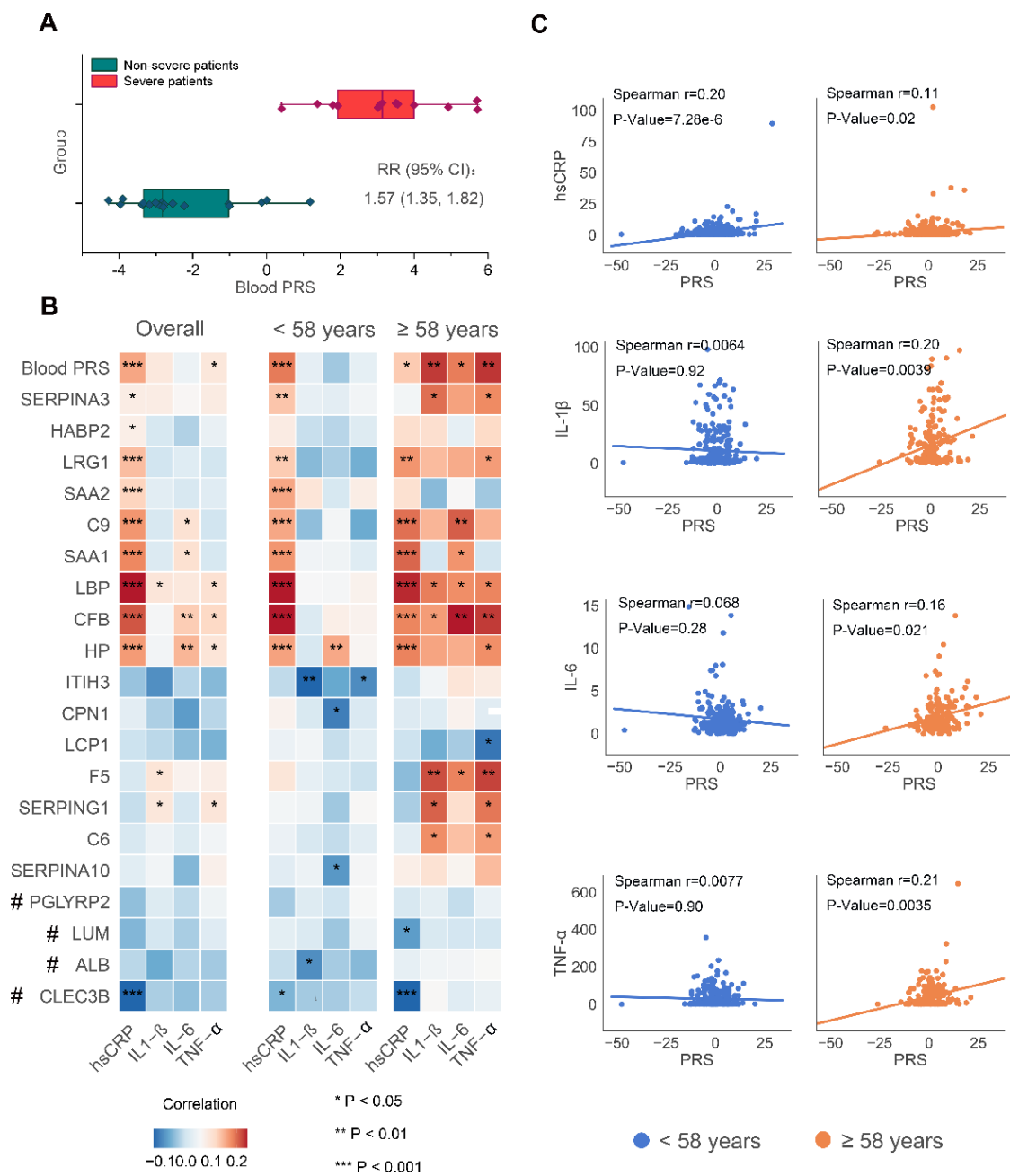
- 885 and Savilahti, E. (2004). Lactobacillus GG effect in increasing IFN-gamma  
886 production in infants with cow's milk allergy. *The Journal of Allergy and Clinical*  
887 *Immunology* *114*, 131–136.
- 888 Shen, B., Yi, X., Sun, Y., Bi, X., Du, J., Zhang, C., Quan, S., Zhang, F., Sun, R., Qian,  
889 L., et al. (2020). Proteomic and Metabolomic Characterization of COVID-19 Patient  
890 Sera. *MedRxiv*, <https://doi.org/10.1101/2020.04.07.20054585>.
- 891 Vich Vila, A., Collij, V., Sanna, S., Sinha, T., Imhann, F., Bourgonje, A.R., Mujagic,  
892 Z., Jonkers, D.M.A.E., Masclee, A.A.M., Fu, J., et al. (2020). Impact of commonly  
893 used drugs on the composition and metabolic function of the gut microbiota. *Nature*  
894 *Communications* *11*, 362.
- 895 World Health Organization (2020). Coronavirus disease (COVID-19) Pandemic.  
896 <https://www.who.int/emergencies/diseases/novel-coronavirus-2019>; Accessed Apr. 18,  
897 2020
- 898 Yan, R., Zhang, Y., Li, Y., Xia, L., Guo, Y., and Zhou, Q. (2020). Structural basis for  
899 the recognition of SARS-CoV-2 by full-length human ACE2. *Science* (New York,  
900 N.Y.) *367*, 1444–1448.
- 901 Yang, Y., Shen, C., Li, J., Yuan, J., Yang, M., Wang, F., Li, G., Li, Y., Xing, L., Peng,  
902 L., et al. (2020). Exuberant elevation of IP-10, MCP-3 and IL-1ra during  
903 SARS-CoV-2 infection is associated with disease severity and fatal outcome.  
904 *MedRxiv*, <https://doi.org/10.1101/2020.03.02.20029975>.
- 905 Yoshida, K., Matsumoto, T., Tateda, K., Uchida, K., Tsujimoto, S., and Yamaguchi,  
906 K. (2001). Induction of interleukin-10 and down-regulation of cytokine production by  
907 *Klebsiella pneumoniae* capsule in mice with pulmonary infection. *Journal of Medical*  
908 *Microbiology* *50*, 456–461.
- 909 Zhang, C.-X., and Ho, S.C. (2009). Validity and reproducibility of a food frequency  
910 Questionnaire among Chinese women in Guangdong province. *Asia Pacific Journal of*  
911 *Clinical Nutrition* *18*, 240–250.
- 912 Zhang, H., Kang, Z., Gong, H., Xu, D., Wang, J., Li, Z., Li, Z., Cui, X., Xiao, J., Zhan,  
913 J., et al. (2020). Digestive system is a potential route of COVID-19: an analysis of  
914 single- - cell coexpression pattern of key proteins in viral entry process. 1–9.
- 915 Zhang, S., Zeng, X., Ren, M., Mao, X., and Qiao, S. (2017). Novel metabolic and  
916 physiological functions of branched chain amino acids: a review. *Journal of Animal*  
917 *Science and Biotechnology* *8*, 10.
- 918 Zhang, Z.-Q., He, L.-P., Liu, Y.-H., Liu, J., Su, Y.-X., and Chen, Y.-M. (2014).  
919 Association between dietary intake of flavonoid and bone mineral density in middle  
920 aged and elderly Chinese women and men. *Osteoporosis International: A Journal*  
921 *Established as Result of Cooperation between the European Foundation for*  
922 *Osteoporosis and the National Osteoporosis Foundation of the USA* *25*, 2417–2425.
- 923 Zhang CX, H.S. (2009). Validity and reproducibility of a food frequency  
924 Questionnaire among Chinese women in Guangdong province. *Asia Pac J Clin Nutr*  
925 *18*, 240–250.
- 926 Zhou, F., Yu, T., Du, R., Fan, G., Liu, Y., Liu, Z., Xiang, J., Wang, Y., Song, B., Gu,  
927 X., et al. (2020). Clinical course and risk factors for mortality of adult inpatients with  
928 COVID-19 in Wuhan, China: a retrospective cohort study. *Lancet* *395*, 1054–1062.
- 929



**Figure 1**



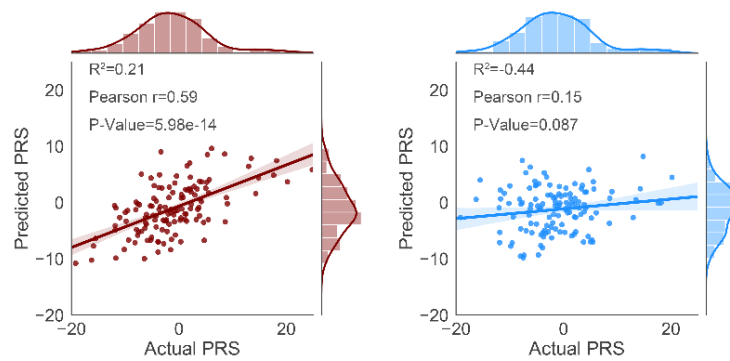
**Figure 2**



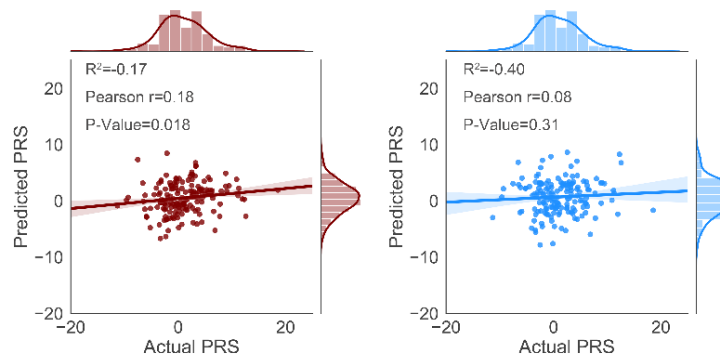
**Figure 3**

**A**

**Cross-sectional subset**

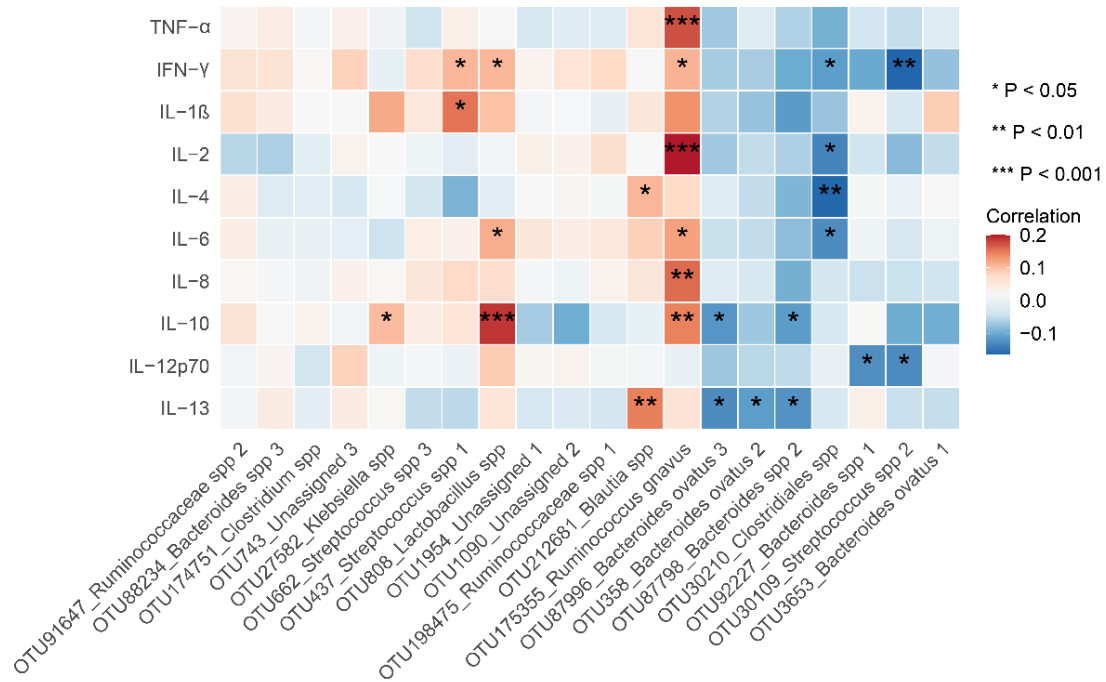


**Prospective subset**



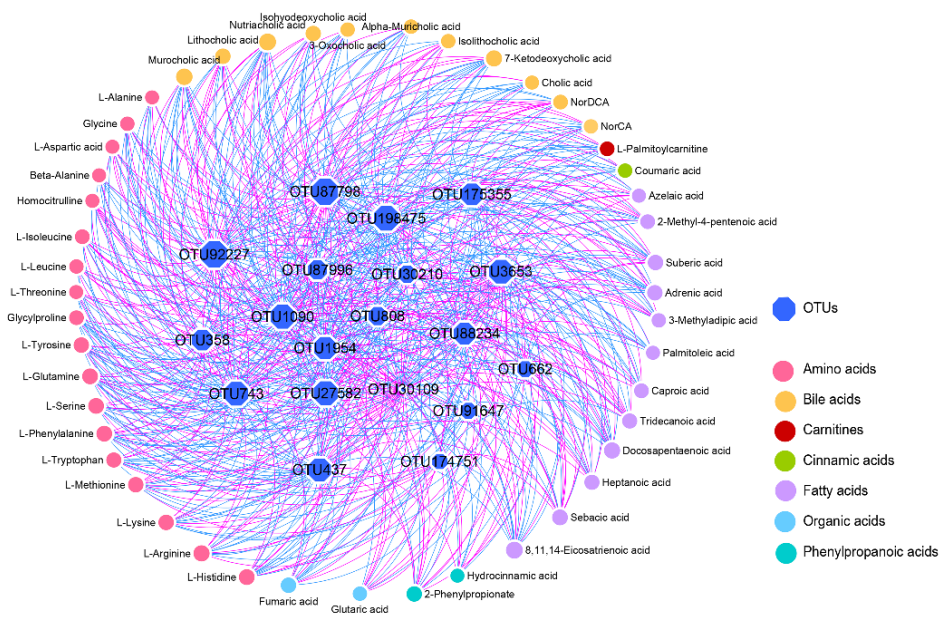
● Based on 20 OTUs  
● Based on demographic and clinical factors

**B**

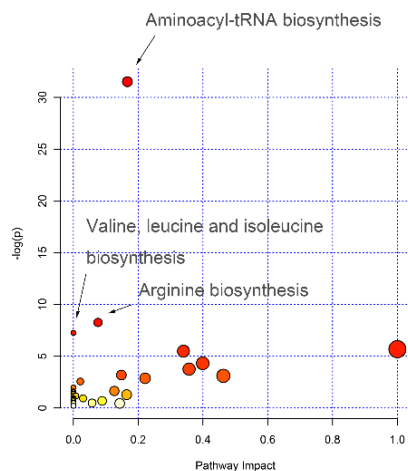


**Figure 4**

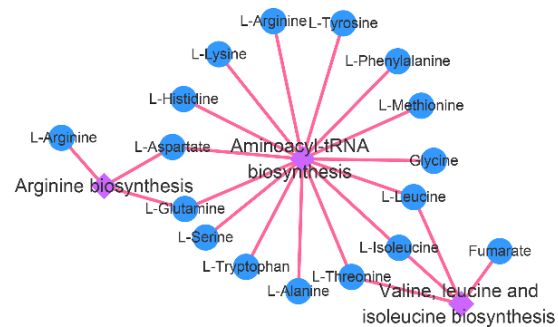
**A**



**B**

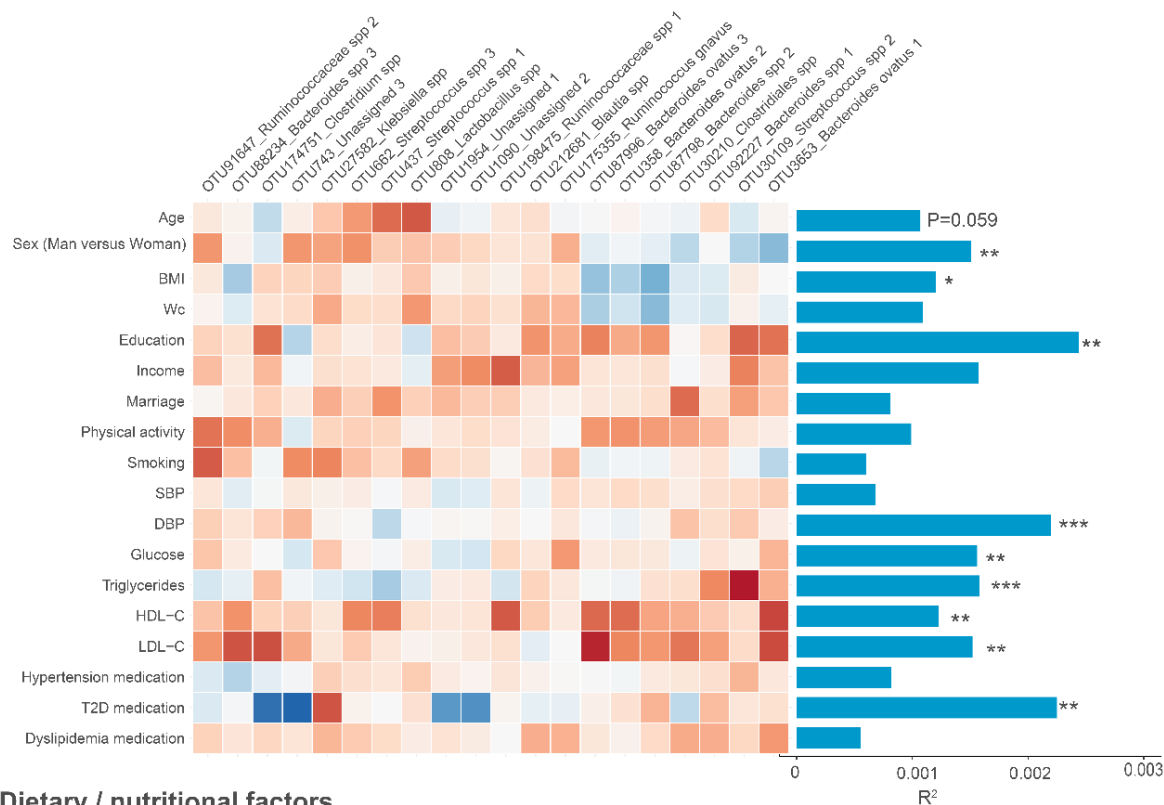


**C**

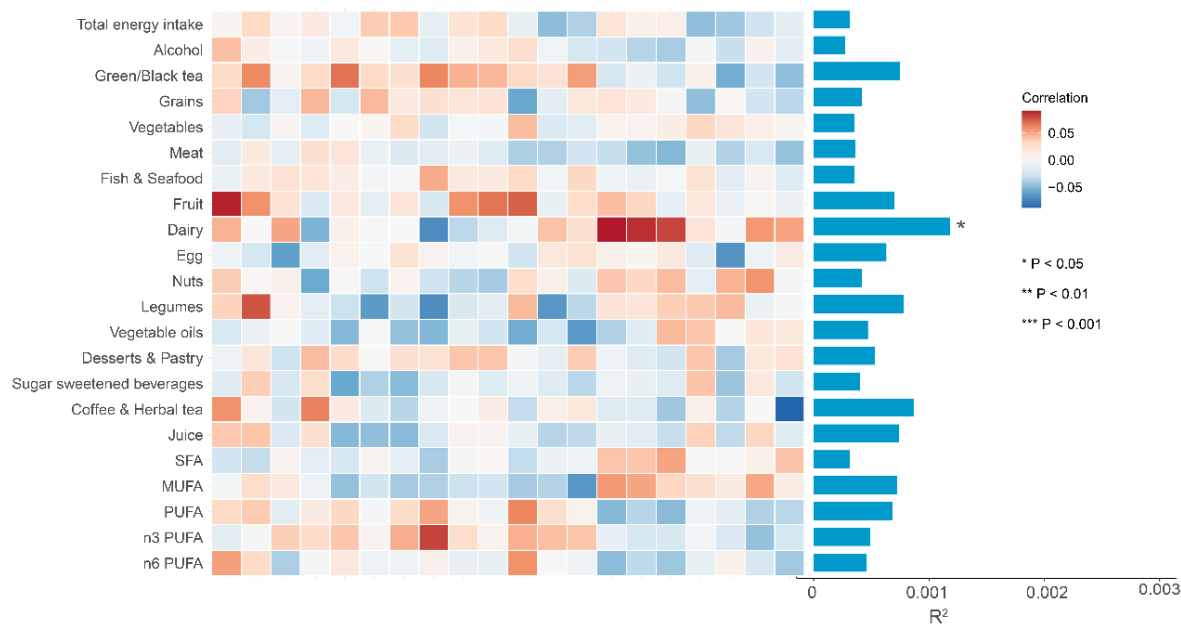


**Figure 5**

**Demographic / clinical factors**

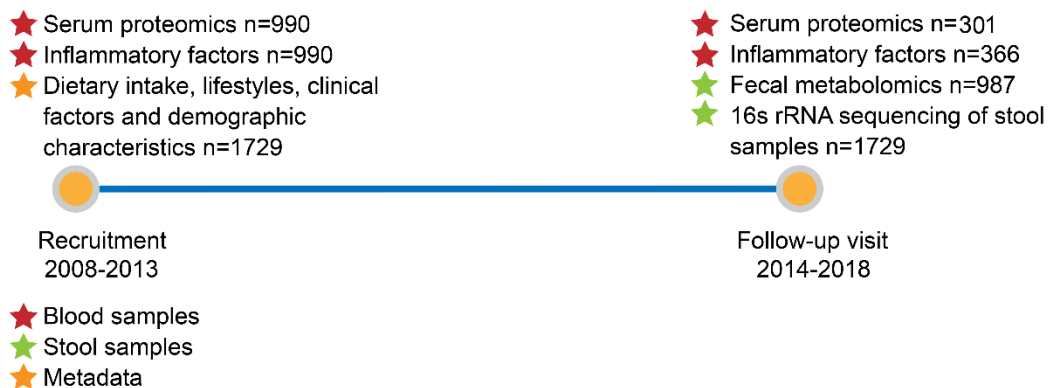


**Dietary / nutritional factors**

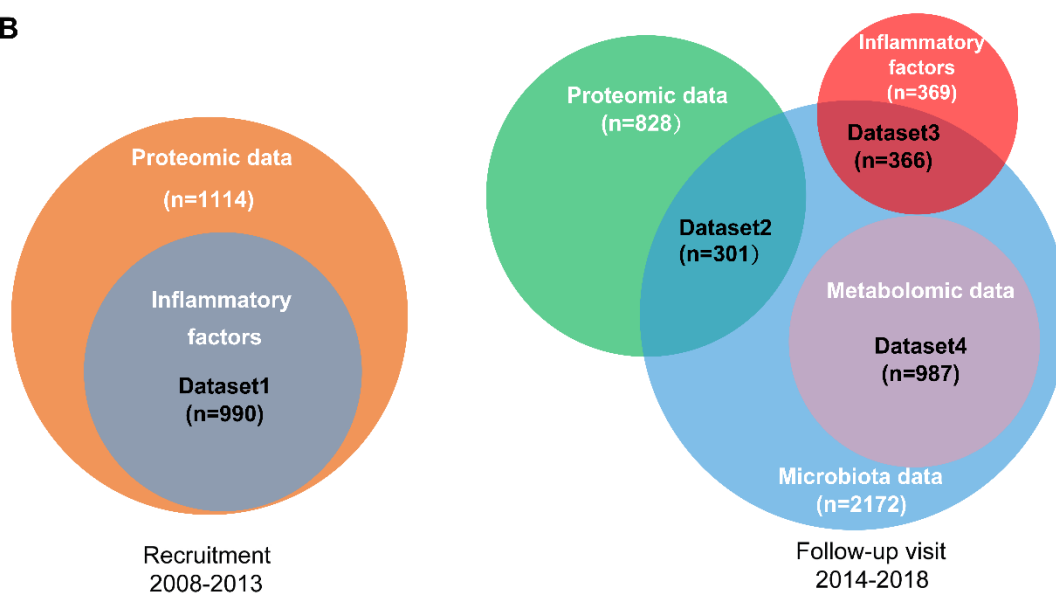


**Figure S1**

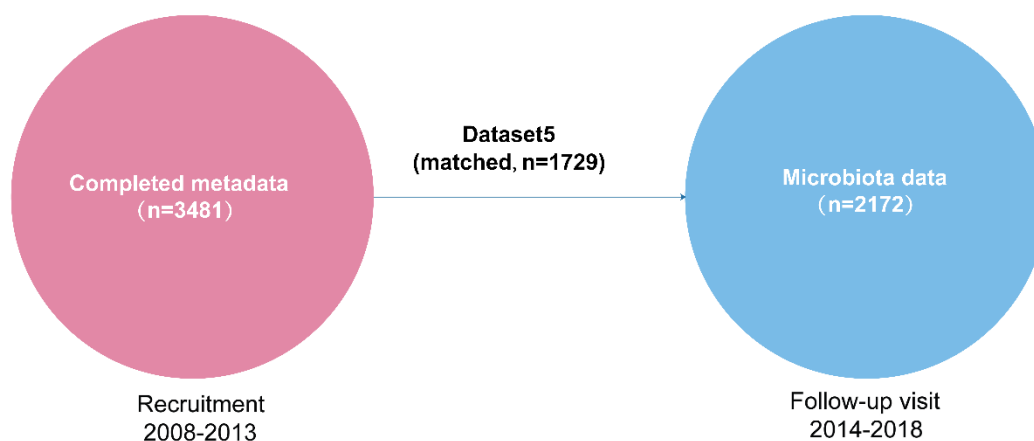
**A**



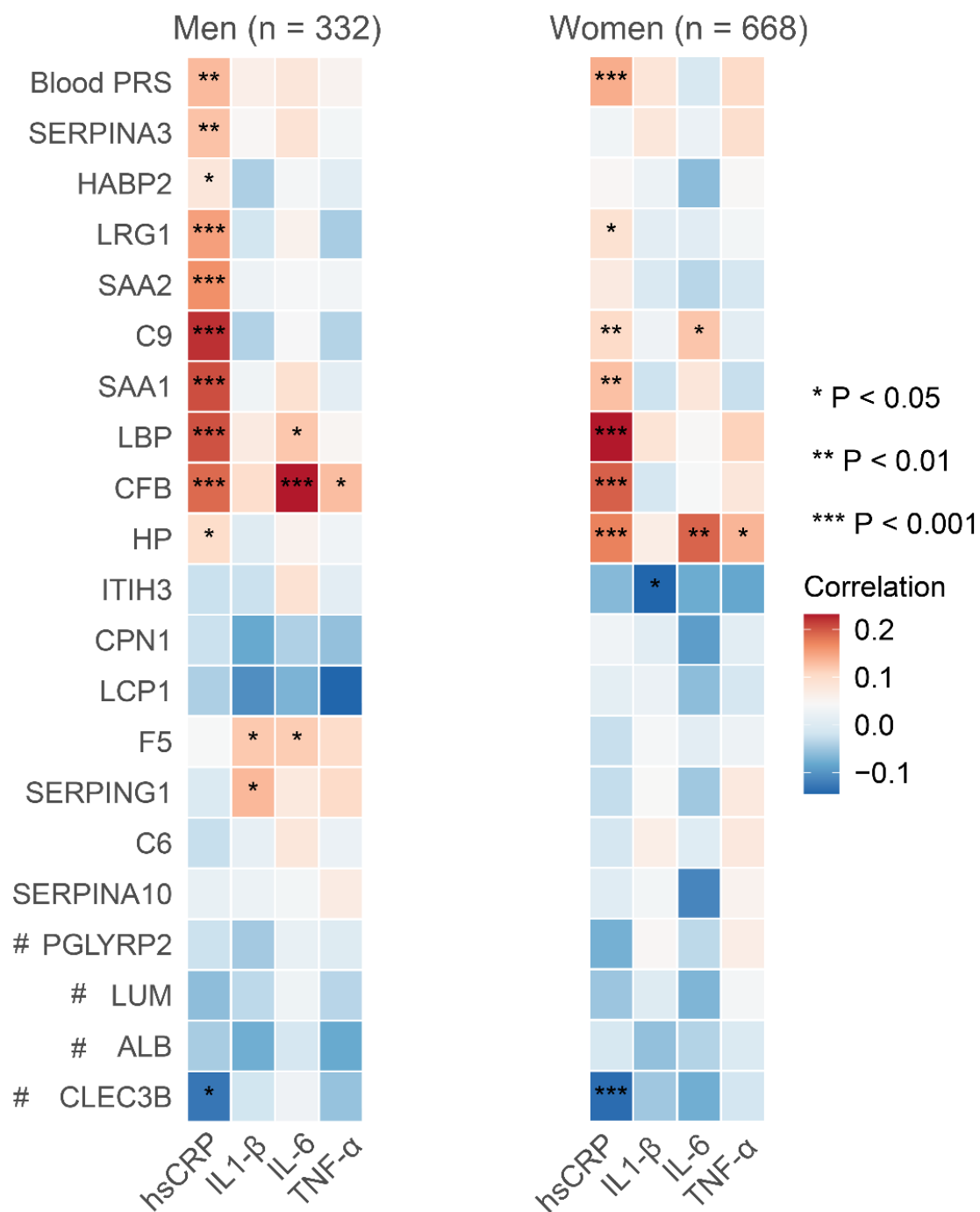
**B**



**C**

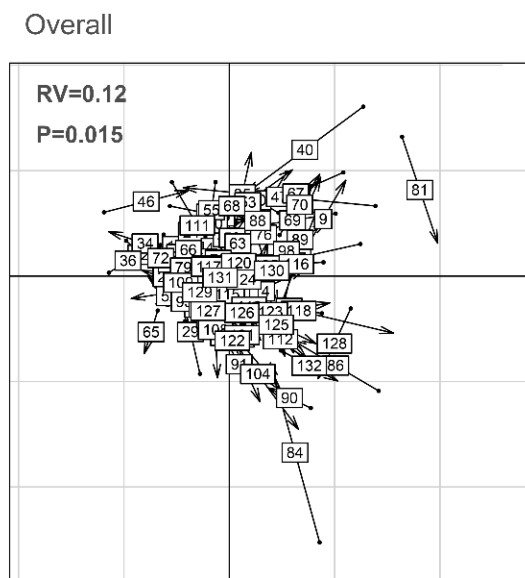


**Figure S2**

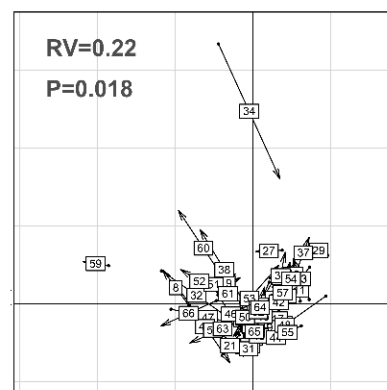


**Figure S3**

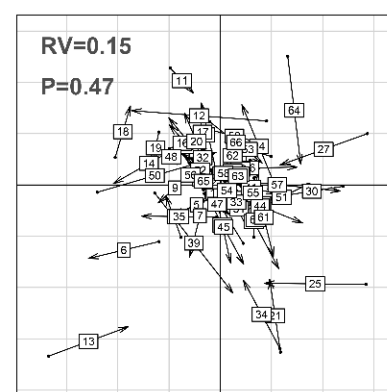
**A**



**B**  $\geq 58$  years



**C**  $< 58$  years





**Table S1. Characteristics of the data analysis set<sup>a</sup>**

Characteristics	dataset1	dataset2	dataset3	dataset4	dataset5
No. of participants	990	301	366	987	1729
Age (year)	58.79 (5.64)	66.21 (6.41)	66.09 (5.27)	64.71 (5.77)	58.54 (6.06)
Sex					
Women	668 (67.5%)	227 (75.4%)	241 (65.8%)	690 (69.9%)	1159 (67.0%)
Men	322 (32.5%)	74 (24.6%)	125 (34.2%)	297 (30.1%)	570 (33.0%)
BMI, kg/m <sup>2</sup>	23.55 (3.23)	23.57 (3.04)	23.90 (3.65)	23.63 (3.26)	23.27 (3.08)
Marital status,%					
Married	911 (92.0%)	261 (87.0%)	342 (93.4%)	884 (89.6%)	1572 (90.9%)
Others	79 (8.0%)	39 (13.0%)	24 (6.6%)	103 (10.4%)	157 (9.1%)
Education,%					
Middle school or lower	294 (29.7%)	82 (27.3%)	88 (24.0%)	254 (25.7%)	469 (27.1%)
High school or professional college	433 (43.7%)	116 (38.7%)	161 (44.0%)	434 (44.0%)	800 (46.3%)
University	263 (26.6%)	102 (34.0%)	117 (32.0%)	299 (30.3%)	460 (26.6%)
Income (Yuan/month/person), %					
≤500	21 (2.1%)	3 (1.0%)	7 (1.9%)	11 (1.1%)	25 (1.4%)
501-1500	258 (26.1%)	70 (23.3%)	112 (30.6%)	235 (23.8%)	385 (22.3%)
1501-3000	565 (57.1%)	136 (45.2%)	177 (48.4%)	567 (57.4%)	1091 (63.1%)
>3000	146 (14.7%)	92 (30.6%)	70 (19.1%)	174 (17.6%)	228 (13.2%)
Current smoking status					
No	823 (83.1%)	268 (89.0%)	310 (84.7%)	858 (86.9%)	1458 (84.3%)
Yes	167 (16.9%)	33 (11.0%)	56 (15.3%)	129 (13.1%)	271 (15.7%)
Current alcohol drinking					
No	924 (93.3%)	289 (96.0%)	341 (93.2%)	914 (92.6%)	1604 (92.8%)
Yes	66 (6.7%)	12 (4.0%)	25 (6.8%)	73 (7.4%)	125 (7.2%)

<sup>a</sup>Values are numbers (percentages) or mean (SD).

**Table S2: List of the 20 proteomic biomarkers integrated in the proteomic risk score**

Uniprot ID	Name	Annotation
Q9UK55	SERPINA10	Protein Z-dependent protease inhibitor; Inhibits activity of the coagulation protease factor Xa in the presence of PROZ, calcium and phospholipids. Also inhibits factor XIa in the absence of cofactors; Belongs to the serpin family
Q96PD5	PGLYRP2	N-acetylmuramoyl-L-alanine amidase; May play a scavenger role by digesting biologically active peptidoglycan (PGN) into biologically inactive fragments. Has no direct bacteriolytic activity; Belongs to the N-acetylmuramoyl-L-alanine amidase 2 family
Q14520	HABP2	Hyaluronan-binding protein 2; Cleaves the alpha-chain at multiple sites and the beta-chain between 'Lys-53' and 'Lys-54' but not the gamma-chain of fibrinogen and therefore does not initiate the formation of the fibrin clot and does not cause the fibrinolysis directly. It does not cleave (activate) prothrombin and plasminogen but converts the inactive single chain urinary plasminogen activator (pro- urokinase) to the active two chain form. Activates coagulation factor VII. May function as a tumor suppressor negatively regulating cell proliferation and cell migration
Q06033	ITIH3	Inter-alpha-trypsin inhibitor heavy chain H3; May act as a carrier of hyaluronan in serum or as a binding protein between hyaluronan and other matrix protein, including those on cell surfaces in tissues to regulate the localization, synthesis and degradation of hyaluronan which are essential to cells undergoing biological processes
P51884	LUM	Lumican; Small leucine rich repeat proteoglycans; Belongs to the small leucine-rich proteoglycan (SLRP) family. SLRP class II subfamily
P18428	LBP	Lipopolysaccharide-binding protein; Plays a role in the innate immune response. Binds to the lipid A moiety of bacterial lipopolysaccharides (LPS), a glycolipid present in the outer membrane of all Gram-negative bacteria. Acts as an affinity enhancer for CD14, facilitating its association with LPS. Promotes the release of cytokines in response to bacterial lipopolysaccharide; BPI fold containing
P15169	CPN1	Carboxypeptidase N catalytic chain; Protects the body from potent vasoactive and inflammatory peptides containing C-terminal Arg or Lys (such as kinins or anaphylatoxins) which are released into the circulation; M14 carboxypeptidases
P13796	LCPI	Plastin-2; Actin-binding protein. Plays a role in the activation of T-cells in response to costimulation through TCR/CD3 and CD2 or CD28. Modulates the cell surface expression of IL2RA/CD25 and CD69; EF-hand domain containing
P13671	C6	Complement component C6; Constituent of the membrane attack complex (MAC) that plays a key role in the innate and adaptive immune response by forming pores in the plasma membrane of target cells; Complement system
P12259	F5	Coagulation factor V; Central regulator of hemostasis. It serves as a critical cofactor for the prothrombinase activity of factor Xa that results in the activation of prothrombin to thrombin
P0DJ19	SAA2	Serum amyloid A-2 protein; Major acute phase reactant. Apolipoprotein of the HDL complex; Belongs to the SAA family
P0DJ18	SAA1	Serum amyloid A-1 protein; Major acute phase protein; Belongs to the SAA family

---

P05452	CLEC3B	Tetranectin; Tetranectin binds to plasminogen and to isolated kringle 4. May be involved in the packaging of molecules destined for exocytosis; C-type lectin domain containing
P05155	SERPING1	Plasma protease C1 inhibitor; Activation of the C1 complex is under control of the C1-inhibitor. It forms a proteolytically inactive stoichiometric complex with the C1r or C1s proteases. May play a potentially crucial role in regulating important physiological pathways including complement activation, blood coagulation, fibrinolysis and the generation of kinins. Very efficient inhibitor of FXIIa. Inhibits chymotrypsin and kallikrein; Serpin peptidase inhibitors
P02768	ALB	Serum albumin; Serum albumin, the main protein of plasma, has a good binding capacity for water, Ca(2+), Na(+), K(+), fatty acids, hormones, bilirubin and drugs. Its main function is the regulation of the colloidal osmotic pressure of blood. Major zinc transporter in plasma, typically binds about 80% of all plasma zinc; Belongs to the ALB/AFP/VDB family
P02750	LRG1	Leucine rich alpha-2-glycoprotein 1
P02748	C9	Complement component C9; Constituent of the membrane attack complex (MAC) that plays a key role in the innate and adaptive immune response by forming pores in the plasma membrane of target cells. C9 is the pore-forming subunit of the MAC; Belongs to the complement C6/C7/C8/C9 family
P01011	GIG25	Serpin peptidase inhibitor, clade A (alpha-1 antiprotease, antitrypsin), member 3; Although its physiological function is unclear, it can inhibit neutrophil cathepsin G and mast cell chymase, both of which can convert angiotensin-1 to the active angiotensin-2; Serpin peptidase inhibitors
P00751	CFB	Complement factor B; Factor B which is part of the alternate pathway of the complement system is cleaved by factor D into 2 fragments: Ba and Bb. Bb, a serine protease, then combines with complement factor 3b to generate the C3 or C5 convertase. It has also been implicated in proliferation and differentiation of preactivated B- lymphocytes, rapid spreading of peripheral blood monocytes, stimulation of lymphocyte blastogenesis and lysis of erythrocytes. Ba inhibits the proliferation of preactivated B-lymphocytes; Belongs to the peptidase S1 family
P00738	HP	Haptoglobin; As a result of hemolysis, hemoglobin is found to accumulate in the kidney and is secreted in the urine. Haptoglobin captures, and combines with free plasma hemoglobin to allow hepatic recycling of heme iron and to prevent kidney damage. Haptoglobin also acts as an Antimicrobial; Antioxidant, has antibacterial activity and plays a role in modulating many aspects of the acute phase response. Hemoglobin/haptoglobin complexes are rapidly cleared by the macrophage CD163 scavenger receptor expressed on the surface of liver Kupfer cells through an endocytic lysosomal degradation.

---

**Table S3: List and taxonomic classifications of the PRS-related core microbial OTUs**

OTUs	Taxa annotation
OTU30210_Clostridiales <i>spp</i>	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__; g__; s__
OTU1090_Unassigned 2	Unassigned
OTU174751_Clostridiu <i>m spp</i>	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Clostridiaceae; g__Clostridium; s__
OTU175355_Ruminococ <i>cus gnavus</i>	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Lachnospiraceae; g__[Ruminococcus]; s__gnavus
OTU1954_Unassigned 1	Unassigned
OTU198475_Ruminococ <i>caceae spp 1</i>	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Ruminococcaceae; g__; s__
OTU212681_Blautia <i>spp</i>	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Lachnospiraceae; g__Blautia; s__
OTU27582_Klebsiella <i>spp</i>	k__Bacteria; p__Proteobacteria; c__Gammaproteobacteria; o__Enterobacteriales; f__Enterobacteriaceae; g__Klebsiella; s__
OTU30109_Streptococcu <i>s spp 2</i>	k__Bacteria; p__Firmicutes; c__Bacilli; o__Lactobacillales; f__Streptococcaceae; g__Streptococcus; s__
OTU358_Bacteroides <i>ovatus 2</i>	k__Bacteria; p__Bacteroidetes; c__Bacteroidia; o__Bacteroidales; f__Bacteroidaceae; g__Bacteroides; s__ovatus
OTU3653_Bacteroides <i>ovatus 1</i>	k__Bacteria; p__Bacteroidetes; c__Bacteroidia; o__Bacteroidales; f__Bacteroidaceae; g__Bacteroides; s__ovatus
OTU437_Streptococcus <i>spp 1</i>	k__Bacteria; p__Firmicutes; c__Bacilli; o__Lactobacillales; f__Streptococcaceae; g__Streptococcus; s__
OTU662_Streptococcus <i>spp 3</i>	k__Bacteria; p__Firmicutes; c__Bacilli; o__Lactobacillales; f__Streptococcaceae; g__Streptococcus; s__
OTU743_Unassigned 3	Unassigned
OTU808_Lactobacillus <i>spp</i>	k__Bacteria; p__Firmicutes; c__Bacilli; o__Lactobacillales; f__Lactobacillaceae; g__Lactobacillus; s__
OTU87798_Bacteroides <i>spp 2</i>	k__Bacteria; p__Bacteroidetes; c__Bacteroidia; o__Bacteroidales; f__Bacteroidaceae; g__Bacteroides; s__
OTU87996_Bacteroides <i>ovatus 3</i>	k__Bacteria; p__Bacteroidetes; c__Bacteroidia; o__Bacteroidales; f__Bacteroidaceae; g__Bacteroides; s__ovatus
OTU88234_Bacteroides <i>spp 3</i>	k__Bacteria; p__Bacteroidetes; c__Bacteroidia; o__Bacteroidales; f__Bacteroidaceae; g__Bacteroides; s__
OTU91647_Ruminococ <i>aceae spp 2</i>	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Ruminococcaceae; g__; s__
OTU92227_Bacteroides <i>spp 1</i>	k__Bacteria; p__Bacteroidetes; c__Bacteroidia; o__Bacteroidales; f__Bacteroidaceae; g__Bacteroides; s__