Supplementary Note 1. Additional analysis of four healthy patterns

Relations and differences with enterotypes. It should be emphasized that our definition of the healthy pattern was different from the previous concept and research on enterotypes, where the driver taxon analysis typically comes out with a total of three types [1, 2]: ET-B (*Bacteroides*-driven), ET-P (*Prevotella*-driven), and ET-F (*Ruminococcus*). It was argued that enterotypes could be variated in geographical regions and correlate with diet/lifestyle/health status, which in term could become an ambiguous indicator for describing the consistency of community taxa [3]. Different studies may also come out with partially inconsistent conclusions. For example, a recent inter-continental comparison study marked *Prevotella*, *Bacteroides*, and *Bifidobacterium* as the key enterotypes [4]. In fact, the hiPCA will boost the separation of healthy patterns by first eliminating those potentially unhealthy collections, which in terms could leverage a more distinct stratification of healthy gut community structures for useful clinical reference in practice. To explore the discriminative ability of four patterns, we further built the random forest classifier by using 66 species from six main genera as features and health patterns as labels. As shown in Table S4, the AUC can be over 99% which implies that there exist significant patterns in the population, and our division is reasonable in pattern recognition. Interestingly, we found *Bifidobacterium* contributed primarily to classification performance, and using the species features was found to perform over 99% of AUC. As a comparison with enterotype, we first divided enterotypes for 1909 health samples using an enterotype classification model trained with metaHIT and HMP [2]. A total of 1532 samples looked similar to the reference dataset and were labeled with ET-P, ET-B, or ET-F. Then the labeled enterotypes were reused as labels for random forest analysis. It was found that the RF could obtain the AUC over 98% with genus *Bacteroides* and *Prevotella*. After further investigating fine-grained species level, RF could

23 achieve AUC over 90% on enterotype classification using *Prevotella copri* and three *Bacteroides*

24 *species* (*B. uniformis*, *B. vulgatus*, and *B. stercoris*). We could conclude that the discrimination of our

25 health patterns was mostly driven by *Bifidobacterium* while the previous enterotypes were mainly

26 discriminated by *Bacteroides* and *Prevotella.*

27   Regional comparison. To investigate the interregional differences in healthy patterns, regional

28 factor was considered, including China, Denmark, Israel, Mongolian and Inner Mongolian,

29 Netherlands, and USA (Figure S1). It was revealed that for USA and China cohorts, both were

30 dominated by HP3 that exhibited the significant microbial richness of *Bacteriodes*, we speculate that

31 both cohorts should be characterized by the western dietary patterns. The Mongolian and Inner

32 Mongolian cohort was dominated by HP2, and the enriched species were *Faecalibacterium*

33 *prausnitzii*, and *Prevotella copri*, which was consistent with the previous related study [5]. For Israel

34 and Netherlands, the HP3 comprised nearly 50% in total, and the other patterns amounted to the rest

35 half. From the remaining pie regions, HP1 occupied more prevalence over HP2. Despite the lack of

36 background details and the varying degrees of adherence to the diet pattern among the population, the

37 Israel is a Mediterranean country, and the traditional Israeli diet is considered to closely fall within the

38 Mediterranean diet (MedDiet) with various adherence degrees. A recent study showed that the

39 MedDiet intervention in overweight and obese subjects was linked with the increased abundance of

40 fiber-degrading *Faecalibacterium prausnitzii* [6]. In another study [7], it was discovered a stronger

41 protective association of MedDiet to mitigate cardiometabolic disease risk with a lower abundance of

42 *Prevotella copri.* The Netherlands cohort showed putative signatures of adherence to the Israel, but

43 with more HP1 dominance, implying the potentially great similarity in the dietary pattern.

44   Relations between HPs and disease. To consider the associations between healthy patterns and

45 disease status, we used the RF classifier with two *Bifidobaterium* species to discriminate the HPs in

46 the disease population (See Table S5 for details). Undoubtedly, most of the disease patients were

47 classified into HP3, followed by HP1. Patients were hardly classified into HP2 and HP4. Among all

48 the disease patients classified into HP1, the morbidity of CA, IGT, and UC were the highest three

49 ones, while ACVD, SA, and CD were the lowest. The proportion of all patients classified into HP3 in

50 the total number of patients with their disease is higher than the proportion of healthy people

51 classified into HP3, and the proportion of ACVD, CD, and RA is even more than 90%. Among

52 patients classified as HP4, the proportion of SA also increased.

53 In human studies, it was found that the relative abundance of butyric acid producing bacteria

54 *Faecalibacterium* and *Clostridium* in the IGT patients' gut would significantly decrease [8]. In our

55 analysis, HP2 with the lowest probability of IGT development has the highest relative abundance of

56 *Faecalibacterium* among the four health patterns, while HP3 with the closest association with IGT has

57 the lowest relative abundance of *Clostridium*, which is consistent with the previous research.

58 Moreover, the occurrence of IGT is considered to be related to biotin metabolism enrichment.

59 Similarly, our metabolic subsystem statistics also showed the same result that the biotin degradation

60 ability of HP3 was stronger than that of HP1 and HP4 (Fig. 4) [8]. On the other way, IGT is

61 considered as the early stage of T2D. In the long term, supplementing the beneficial substrate to

62 produce butyrate by the gut microbiota of patients with pre-T2D may be a targeted therapy to prevent

63 or delay the progression of T2D. Previous research used first-degree relatives as comparators for RA

64 patients and showed an abundance of *Eggerthella lenta* in patients as compared with relatives [9]. In

65 our study, we found the similar result that the abundance of *Eggerthella lenta* in HP3 is the highest of

66 four health patterns, but almost absent in HP2 and HP4. The obesity can alter the composition and

67 structure of the gut microbiota was demonstrated by previous study. The diversity of microbes in the

68 gut of most obese patients is considered lower than the healthy people [10]. HP3 had the lowest

69 microbial diversity among the four health patterns, and the proportion of obese patients classified as

70 HP3 was also the highest.

71

Supplementary Note 2. The impact of different transformation methods on sparse data processing on model results

74 To evaluate the impact of different transformation methods on model results for sparse data, we

75 applied four types of transformations - logarithmic and square root transformation, logarithmic

76 transformation, centered log ratio transformation, and minimum abundance threshold transformation -

77 to our discovery and validation datasets. Specifically, we performed a total of 12 transformations,

78 including (1) only logarithmic and square root transformations; logarithmic and square root

79 transformation applied after two different minimum abundance threshold transformations, (2) only

80 logarithmic transformation; logarithmic transformation applied after two minimum abundance

81 threshold transformations, (3) only centered log ratio transformation; centered log ratio transformation

82 applied after two minimum abundance threshold transformations, and (4) only untransformed data;

83 untransformed data after two minimum abundance threshold transformations.

84 Based on our observations, we found that the balance accuracy of the models for the first three

85 transformation types differed slightly, but all performed better than the untransformed data. In terms

86 of individual data accuracy, the models using data transformed by the first transformation type

87 showed the most balanced and consistently high accuracy. Furthermore, we observed that there was

88 no significant difference in model results between data transformed by the minimum abundance

89　　threshold transformation and untransformed data, indicating that the minimum abundance threshold

90　　transformation may have little impact on our model results. All results are shown in **Table S6-17**.

91

92　　**Supplementary Note 3. Achieving a balance between predictive**

93　　**performance and model simplicity by selecting 92 features through the KS**

94　　**test**

95　　　　KS test was used to select features from the gut microbiota data. Then, the KS statistic for each

96　　feature between the healthy and disease groups, and ranked the features based on the p-values were

97　　calculated. Subsequently, we selected the top 92 features with balanced performance. To ensure

98　　consistency with the previous statement, we assessed the performance of hiPCA with 110 features and

99　　92 features, while considering different transformations such as CLR, and evaluated the results based

100　　on AUC values, as shown in Figure S2. Our findings indicate that the performance of our algorithm is

101　　comparable using either set of features, as there was only a slight difference in the AUC values.

102　　According to Occam's Razor "Entities should not be multiplied unnecessarily", we believe that

103　　selecting the top 92 features using the KS test strikes a good balance between dimensionality

104　　reduction and prediction performance, and provides a more parsimonious model with comparable

105　　accuracy to using 110 features.

106

107　　**Supplementary Note 4. Supplementary analysis of individualized inference**

108　　**revealing disease-specific microbial responses**

109　　　　As previously mentioned, the following both PR-50 and KS-92 contribution plots will be used

110　　for a comprehensive analysis of the remaining eight diseases in the discovery cohort. The overlay bar

111 graph over microbial features regarding health and eight diseases are shown in Supplementary Figure

112 S3-10.

113   Atherosclerotic Cardiovascular Disease (ACVD) and Symptomatic Atherosclerosis (SA).

114 Notably, the top species with large contributions driven by PR-50 to the hiPCA index in ACVD were

115 *Atopobium parvulum*, *Gemella morbillorum*, *Lactobacillus salivarius*, *Fusobacterium nucleatum,*

116 *Peptostreptococcus stomatis, Solobacterium moorei.* The KS-92 driven model further added

117 *Oribacterium sinus*, *Fusobacterium varium*, *Scardovia wiggsiae*, *Streptococcus cristatus,* and

118 *Streptococcus tigurinus*. Our unsupervised diagnosis results were mostly consistent with [11], where

119 they used classifiers and found that most of the microorganisms presented in the oral cavity or

120 oropharyngeal area but were reported higher in patients with ACVD. Another species *candidate*

121 *division TM7 single-cell isolate TM7c* also showed a large contribution, which belonged to

122 *Saccharibacteria* (TM7) bacteria and also highly related to the oral cavity [12, 13]. There was

123 substantial evidence that the oral cavity bacteria could affect the development of ACVD [14]. Apart

124 from the oral cavity, the gut microbiota could also influence the development of ACVD, including

125 *Subdoligranulum sp.*, *Blautia producta*, *Granulicatella adiacens*, *Lachnospiraceae bacterium,*

126 *Clostridium spp*., and *Streptococcus spp*. Among them, *Subdoligranulum sp.* was thought to be

127 associated with *Akkermansia muciniphila* and showed positive correlations with healthy status

128 indicators like HDLC and microbial richness [15]. *G. adiacens* was reported to cause bacteremia and

129 endocarditis [16]. *O. sinus* was positively correlated with blood pressure in ACVD study [11]. *F.*

130 *varium* repressed the Reg3 antimicrobial family, which were important players in maintaining barrier

131 integrity [17]. *Lachnospiraceae bacterium* was related to the ACVD-depleted butyrate-producing

132 bacteria [11]. Lastly, for the variety of *Streptococcus spp*., our result was in line with [18], where they

133 also reported a high level abundance in ACVD cases. In general, our findings indicated hiPCA was a

134 promising inferential engine for ACVD diagnosis. For the SA cohort, the hiPCA cannot easily detect

135 case samples. However, our diagnostic plot still revealed several key biomarkers, some top ranked

136 species were shared with ACVD, such as *L. salivarius*, *S. tigurinus,* and *S. wiggsiae*. To this view, we

137 could speculate that both oral and gut microbiota contribute heavily to the health index of

138 atherosclerosis, which acted as a strong quantitative evidence to the previous hypothesis that both the

139 oral and gut microbiota contribute to atherosclerosis in humans [19].

140 Rheumatoid arthritis (RA). RA is a complex, polygenic, and autoimmune disorder [20]. From the

141 combined analysis of contribution plots PR-50 and KS-92, we could find that oral and gut microbial

142 like *Atopobium parvulum, Eubacterium infirmum, Granulicatella adiacens, Lactobacillus salivarius,*

143 *Scardovia wiggsiae, Streptococcus cristatus* showed top contributions to the PCA healthy indexes in

144 RA populations, and the other microbial like *Bifidobacterium dentium, Clostridium spp.* and

145 *Lachnospiraceae bacterium* added as high contributors, manifesting the oral-gut microbiome axis was

146 highly associated with the etiology of RA [21-24]. Early review summarized patients with RA were

147 enriched with *Atopobium, and Lactobacillus salivarius,* and decreased with *Streptococcus* [25].

148 Intriguingly, the *Subdoligranulum* sp and *Subdoligranulum variabile* invariably show the difference

149 between RA and healthy controls, which conforms to another comparative study of immune-mediated

150 inflammatory diseases [26]. However, our plot has not included *Porphyromonas gingivalis*, which is

151 regarded as a keystone pathogen and has a relatively clear mechanism in RA development [21]. In

152 general, our hiPCA findings were largely in line with the previous study, and more importantly, we

153 further revealed a comprehensive set of pathogenic microbiomes among the RA population compared

154 to healthy ones. Our results further demonstrated from the health inference that oral and gut

155 microbiomes were perturbed and contributed to the progression in RA [27].

156 Type 2 diabetes (T2D) and impaired glucose intolerance (IGT). T2D is a metabolic disorder, and

157 IGT cases are usually clinically regarded as prediabetes. Substantial studies have associated

158 microbiomes with the pathophysiology of T2D. According to PR-50 and KS-92, almost all species

159 were found in significantly different abundance levels in patients with T2D compared with healthy

160 subjects, such result is consistent with a previous study that T2D is not linked to any one specific

161 species or genus [28]. To name a few, those leading contributors were *Fusobacterium varium*,

162 *Megasphaera micronuciformis*, *Blautia producta*, *Lactobacillus salivarius*, *Solobacterium moorei*,

163 *Clostridium ramosum*, *Gemella morbillorum*, *Clostridium ramosum*, *Peptostreptococcus stomatis*,

164 *Lachnospiraceae bacterium*, *Clostridium clostridioforme*, *Subdoligranulum sp*, *Streptococcus*

165 *gordonii*, etc. Among them, we could judge that the *Blautia* and *Lactobacillus* species were widely

166 reported to be associated with T2D [29]. The study by [30] claimed that *S. moorei* was prone to be

167 found in the compromised immunity across all genders and age groups. T2D subjects upon metformin

168 were found to have higher levels of SCFA-producing bacteria *Megasphaera* than nondiabetic

169 individuals [31]. In a recent study [32], patients with T2D were reported to have a varied abundance

170 of *Lachnospiraceae* and a decreased abundance of *Subdoligranulum*. *Clostridium* species, like *C.*

171 *ramosum*, *C. hathewayi,* and *C. clostridioforme,* may all cause bacteraemia and intra-abdominal

172 infections [33, 34]. However, the key mechanistic resolution still remains unclear whether these

173 bacteria synergistically or asynchronously play a causal role in T2D [35]. Lastly, by inspecting the

174 IGT contribution plots, we can find that IGT shared most of the high contributors with T2D, like *S.*

175 *moorei*, *B. producta*, several *Streptococcus* species, *Clostridium* species, and *Lachnospiraceae* species,

176 implying the consistent etiology progress and mechanisms. Concomitantly, the KS-92 chart further

177   pointed out that the TOP 2 species *F. varium* and *M. micronuciformis* in T2D made relatively smaller

178   contributions in IGT. We could also find here that the *K. pneumoniae* was not an active contributor in

179   IGT as in T2D, manifesting that the increased dysregulated immune responses and insulin resistance

180   from IGT to T2D could also result in the increased level of susceptibility to infections [36].

181   Obesity (OB), Overweight (OW), and underweight (UW). Overweight is defined by BMI greater

182   than or equal to 25, while obesity is greater than 30. As a chronic disease, obesity is associated with

183   various diseases, including cardiovascular diseases, and type 2 diabetes mellitus (T2DM). The human

184   gut microbiota is always regarded to play a critical role right through the adiposity and glucose

185   metabolism. With the contribution analysis, one can infer that there were almost identical contribution

186   spectrums among obesity and overweight populations. The top health contribution panel was

187   implicated by multiple *Lachnospiraceae bacterium*, *Gemella morbillorum*, *Peptostreptococcus*

188   *stomatis*, *Granulicatella adiacens*, *Oribacterium sinus*, *Megasphaera micronuciformis*, *Streptococcus*

189   *cristatus*, *Streptococcus tigurinus*, *Solobacterium moorei,* and *Atopobium parvulum.* Despite as the

190   main producers of SCFAs, the *Lachnospiraceae* family was highly associated with dietary pattern [37]

191   and obesity [38-40], also closely related to the intergenerational transmission of obesity or overweight

192   (OBOW) [41]. The *Clostridium* genus like *C. ramosum* and *C. clostridioforme* also leveraged a large

193   contribution to the health index. Previous investigations showed that *C. ramosum* could promote

194   high-fat diet-induced obesity in gnotobiotic mice models [42]. In a Danish cohort study, *C.*

195   *clostridioforme* was associated with the low community diversity that is more frequently obese [43],

196   and then found in a randomized controlled trial in a Canadian children study with OBOW subjects to

197   decrease significantly after the prebiotic intervention [44]. Apart from that, analysis of the hiPCA

198   contribution chart further indicated that the oral bacteria contribute a large proportion to the

199 development of obesity. For example, *Megasphaera* was among the key oral bacterial taxa associated

200 with OBOW and *Oribacterium* was another key oral microbe that may be co-occurred with

201 tryptophan biosynthesis [45]. In a MedDiet Intervention study in OWOB subjects, *S. cristatus* was

202 observed with increased levels and was regarded to inhibit virulence gene expression in

203 periodontopathogenic bacteria *P. gingivalis* [46]. Generally, to explain the possible mechanism of oral

204 bacteria to obesity, early researchers made three hypotheses, namely through increasing metabolic

205 efficiency, increasing appetite, and redirecting energy metabolism [47]. Both oral and gut bacteria

206 contributed to obesity, but the cross talks between oral microbiota with gut microbiota in obesity

207 development still need further endeavors [48, 49]. Lastly, we consider underweight, which is defined

208 by BMI less than the cut-off point of 18.5. It is usually related to malnutrition caused by food shortage

209 or disease. Compared with obesity, the relationship between underweight and gut microbiota is

210 underexplored [50]. Despite the appeared dissimilarity in definition and etiology, this study observed

211 great overlap in the considered microbiota contribution spectrum among the underweight and

212 overweight/obesity groups. Interestingly, our result may conform to a recent study in which they

213 found no statistically significant differences among the underweight, normal weight, and overweight

214 individuals at the genus level [50].

215

216 **Supplementary Note 5. hiPCA is largely reproducible on the additional**

217 **validation data**

218   To evaluate the generalization capability of hiPCA, the independent validation data consisting of

219 782 samples (118 healthy individuals and 664 unhealthy individuals) will be tested, which include

220 ankylosing spondylitis (AS), colorectal adenoma (CA), colorectal cancer (CRC), CD, liver cirrhosis

221　(LC), non-alcoholic fatty liver disease (NAFLD), RA, UW, OW, and OB.

222　　　As shown in Figure S11. Our hiPCA with KS-92 signal panel has a balanced accuracy of 70.26%,

223　and the detection accuracy for the healthy and unhealthy group were 70.34% and 70.18%, respectively.

224　Notably, the results achieved almost identical performance as on the discovery data (71.40% in

225　healthy group and 71.24% in unhealthy group). To make a close investigation, we separately

226　considered all sub-cohorts according to the published studies. The accuracies by hiPCA (KS-92) on

227　three healthy cohorts were 46.43% (H1), 84.48% (H2), and 65.63% (H3), which implied large

228　variations among healthy group background in three studies. As the matter of fact, through inspecting

229　the original studies, we found that only the H2 cohort declared that the control individuals were not

230　suffering from any diseases. H1 and H3 only ensured that the controls were not Parkinson's disease

231　patients (H1) or not colorectal cancer patients (H3), and they could not exclude other health problems

232　in the control group. In the meantime, we also found that the average ages in these healthy sub-groups

233　were 65.6 (H1), 31.9 (H2), and 57.3 (H3) respectively, which further demonstrated potential

234　accelerated age-related health loss in the elderly. In this regard, the hiPCA could also serve as an

235　additional indicator for the healthy people's initial assessment during recruitment. For those unhealthy

236　phenotype sub-cohorts, we had the following: 85.98% (141 of 164) for LC, 81.4% (70 of 86) for

237　NAFLD, 66.67% (10 of 15) for CD, 66.67% (18 of 27) for CA, 63.27% (31 of 49) for RA, 56.7% (55

238　of 97) for AS, 50% (7 of 14) for OB, 47.54% (29 of 61) for OW, 14.29% (4 of 28) for UW, and 92.5%

239　(37 of 40), 78.69% (48 of 61), and 72.73% (16 of 22) for three different cohorts of CRC. As can be

240　seen, our hiPCA provided well adverse health predictions (>75%) for LC, CRC, and NAFLD, where

241　GMHI could only perform well on LC and NAFLD. Comparing with GMHI, we could conclude

242　hiPCA is a more accurate and robust index for nonhealthy prediction (70.18% vs. 62.2%), and that

243   KS-92 is a more robust, generalized microbiome signal panel that can perform well even across

244   different diseases.

245   Like the discovery set, inferential analysis was made by bacteria-to-health-index contribution

246   analysis, and the plots were shown in Figure S12-21. CRC and CA: Our results showed that the

247   validation set highly reproduced the results of the discovery set, with only a slight difference in the

248   ranking of contributions of some species. By exploring the hiPCA performance for CRC diagnosis,

249   those marker species contribute most on average to the health index in PR-50, which could be listed

250   as *Peptostreptococcus stomatis*, *Gemella morbillorum*, *Granulicatella adiacens*, *Fusobacterium*

251   *nucleatum, Atopobium parvulum, Gemella sanguinis,* and *Clostridium clostridioforme*. KS-92 shared

252   most contributors, but also further supplements with *Streptococcus cristatus*, *Parvimonas micra*,

253   *Streptococcus tigurinus,* and *Eubacterium infirmum*. Among them, *Peptostreptococcus stomatis*,

254   *Fusobacterium nucleatum,* and *Parvimonas micra* have been well described before. Apart from that,

255   other species, such as *B. dentium*, *S. bovis*, *S. gallolyticus*, *S. Tigurinus,* and *Gemella morbillorum*

256   also need to be analyzed further. Previous studies showed that *Bifidobacterium* are enriched in people

257   without CA, but our results suggests that *B. dentium* has a great contribution to CA, which is contrary

258   to the current research. *B. dentium* is capable of enhancing the intestinal mucus layer and goblet cell

259   function via upregulation of gene expression and autophagy signaling pathways, with a net increase in

260   mucin production, and then provide a protective effect in the colitis model [51, 52]. However, a study

261   reported that the abundance of *Bifidobacterium* was increased in the benign group [53]. The effect and

262   mechanism of *B. dentium* on CA require further discussion. A variety of microbiota in *Streptococcus*

263   are thought to be associated with CRC and CA, including *S. bovis* and *S. gallolyticus* [54]. However,

264   this association is still under controversy. Some data do not support the idea that *S. bovis* colonizes

265 CRC and CA [55]. *S. gallolyticus* (previously known as S. bovis biotype I) is thought to be more

266 closely related to CRC and CA [56]. *S. Tigurinus* is suspected of causing bacteremia due to its

267 translocation. This species has recently been reported to be extremely pathogenic [57]. In another

268 study for predicting colon cancer, *Gemella morbillorum* was used as a biomarker of fecal bacteria,

269 and the predictive model with good sensitivity and specificity was established as a new diagnostic tool

270 for CRC [58]. *Gemella morbillorum* was shown to be enriched in the gut of patients with CRC by

271 metagenomics [59], and positively correlated with the enrichment of amino acids, cadaverine, and

272 creatine in CRC [60].

273   AS and RA: Through the analysis of contribution plots PR-50 and KS-92, we can clearly find

274 that AS and RA share many high-contribution species, although these high-contribution species have

275 differences in disease contribution. For AS, the contribution plot PR-50 reveals highly contributing

276 species, including *Clostridium clostridioforme*, *Clostridium symbiosum*, *Blautia producta*,

277 *Flavonifractor plautii*, *Clostridium asparagiforme*, *Anaerotruncus colihominis*, *Clostridium citroniae*,

278 *Clostridiales bacterium 1_7_47FAA*, *Clostridium hathewayi*, *Streptococcus mitis 11/5*, *Clostridium*

279 *ramosum*, *Lactobacillus salivarius*, *Lachnospiraceae bacterium 1_4_56FAA*, *Bifidobacterium dentium*,

280 *Clostridium bolteae*. The contribution plot KS-92 further reveals high-contribution species such as

281 *Megasphaera micronuciformis*, *Lactobacillus mucosae*, *Streptococcus cristatus*, *Rothia mucilaginosa*,

282 *and Fusobacterium ulcerans*. Yin found that *Clostridium 1_7_47FAA*, *Clostridium bolteae, and*

283 *Clostridium hatheway* were enriched in patients with AS through statistical analysis [61]. In previous

284 studies, *Streptococcus* and *Lactobacillus* were confirmed to be enriched in patients with AS [62].

285 *Bifidobacterium* is usually used in probiotics, and accumulated in patients with AS [63]. For the RA

286 cohort, the performance of the diagnostic plot is affected due to the impact of the batch effect,

287   resulting in a slight difference in high-contribution species between the RA of the validation set and

288   the RA of the discovery set. However, through the contribution plot, we can still find many common

289   high-contribution species, including *Clostridium asparagiforme*, *Clostridium ramosum*, *Clostridium*

290   *clostridioform*, *Clostridium citroniae*, *Clostridium sympiosum*, *Clostridiales bacterium 1_7_47FAA,*

291   and *Lachnospiraceae bacterium. Clostridium asparagiforme* and *Lachnospiraceae bacterium* were

292   enriched in patients with RA compared to healthy control groups [64]. This is consistent with the

293   above analysis results.

294   CD: In the analysis of validation set, we only discuss CD, nor UC. Although the exact cause is

295   unknown, CD appears to be due to a combination of environmental factors and genetic predisposition.

296   In this paper, the most genus in the discovery set and the validation set are similar. However,

297   compared with the other diseases we analyzed, the contribution degree of bacteria in the discovery

298   and validation sets to CD was different. For KS-92, eight of the 12 species with the highest

299   contribution were consistent. The difference in contribution between the discovery set and validation

300   set is mainly reflected in *Fusobacterium ulcerans*, *Lachnospiraceae bacterium 6_1_63FAA*,

301   *Fusobacterium mortiferum,* and *Clostridium scindens* of KS-92, *Granulicatella adiacens* of PR-50.

302   This difference may be due to batch effects and difficulty in diagnosing CD. The diagnosis of CD can

303   sometimes be challenging and often requires extensive testing to help doctors make the diagnosis. CD

304   may not be diagnosed with complete certainty, even after a battery of tests. In the validation set, the

305   degree of contribution of most microbiota was similar in KS-92 and PR-50. High contribution

306   candidates in both PR-50 and KS-92 include *Peptostreptococcus stomatis*, *Fusobacterium nucleatum,*

307   and *Clostridium clostridioforme*, while the degree of contribution of *Flavonifractor plautii*, *Gemella*

308   *morbillorum* is different between PR-50 and KS-92. There are some species have contributed greatly

309 to the diagnosis of diseases, but have not been well described in the previous content. Fusobacterium

310 species have been implicated recently in a broad spectrum of human pathologies, including Crohn's

311 disease, ulcerative colitis, and colorectal cancer [65, 66]. *F. nucleatum* targeted caspase activation and

312 recruitment domain 3 (CARD3) to activate the endoplasmic reticulum stress pathway and promote *F.*

313 *nucleatum*-mediated mucosal barrier damage *in vivo* and *in vitro* during CD development [67].

314 Moreover, previous studies demonstrated that highly invasive *F. nucleatum* isolates derived from the

315 inflamed guts of patients with CD evoked significantly greater MUC2 and tumor necrosis factor alpha

316 (TNF-α) gene expression than minimally invasive strains isolated from the noninflamed gut in human

317 colonic epithelial cells and in a rat ligated colonic loop model of infection [68]. In addition, *F.*

318 *nucleatum* accelerates the progression of colitis-associated colorectal cancer by promoting

319 epithelial-mesenchymal transition [69]. For *Clostridium*, *C. clostridioforme*, *C. scindens*, *C. ramosum*,

320 *C. symbiosum*, *C. citroniae,* and *C. asparagiforme* high contribution candidates in CD. Most of them

321 are commonly not dominant in the fecal microflora of healthy subjects [70]. *C. scindens* has been

322 previously associated with a case of an adolescent female with CD [71]. AS and CD are likely caused

323 by subclinical bowel infections with *Klebsiella* [72, 73]. This trend is also predicted in the results of

324 our validation set.

325   LC and NAFLD: NAFLD is a class of metabolic disorders that can progress from fatty liver to

326 steatohepatitis, further leading to cirrhosis [74]. According to the contribution plot PR-50, the main

327 contributing species of LC is *Fusobacterium nucleatum*, *Lactobacillus salivarius*, *Streptococcus*

328 *gordonii*, *Granulicatella adiacens*, *Atopobium parvulum*, *Peptostreptococcus stomatis*,

329 *Lachnospiraceae bacterium 1_4_56FAA*, *Streptococcus sanguinis*, *Clostridium symbiosum*,

330 *Clostridiales bacterium 1_7_47FAA*, *Streptococcus mitis 11/5*, *Lachnospiraceae bacterium*

331　*9_1_43BFAA*, *Clostridium ramosum*. The contribution plot KS-92 further reveals the

332　high-contributing species of LC, *Megasphaera micronuciformis*, *Streptococcus cristatus*,

333　*Streptococcus tigurinus*, *Fusobacterium varium* were added to high-contribution species. Previous

334　studies have identified increased abundances of the oral microbe *Fusobacterium* in the setting of LC

335　[75]. Our diagnosis is consistent with previous studies, where they used the Wilcoxon rank-sum test in

336　the healthy control group and the cirrhosis group and found that *Fusobacteria*, *Streptococcus*, and

337　*Clostridium* were enriched in the cirrhosis group [76]. *Streptococcus* are taxa that are part of the oral

338　bacterial ecosystem, while *Streptococcus* appeared in previous studies to distinguish between patients

339　with LC and healthy individuals, which may mean that the development of LC may have occurred

340　from the oral cavity, consistent with previous views [76-78]. For the NAFLD cohort,

341　high-contribution species mainly include *Atopobium parvulum*, *Streptococcus tigurinus*, *Scardovia*

342　*wiggsiae*, *Lachnospiraceae bacterium 4_1_37FAA*, *Lachnospiraceae bacterium 6_1_63FAA*, *Blautia*

343　*producta*, *Megasphaera micronuciformis*, *Clostridium ramosum, and Streptococcus cristatus*. We can

344　clearly find that the high-contribution species of NAFLD and LC are mostly the same, although their

345　variation in NAFLD is not as significant as in LC, which implies the close relationship between

346　NAFLD and LC [79]. Our results are consistent with previous studies that *Clostridium bolteae*,

347　*Clostridium nexills*, *Ruminococcus gnavusb, and Atopobium parvulum* are associated with blood

348　biomarkers of liver disease [80]. *Streptococcus* is a relatively important contributor genus in both

349　NAFLD and LC, and the variation of *M. micronuciformis* in LC is much higher than in NAFLD. This

350　is consistent with previous studies, which found that *Megasphaera* was only enriched in people with

351　NAFLD-cirrhosis, and *Streptococcus* was enriched in NAFLD-cirrhosis and NAFLD, which reflected

352　the differences between LC and NAFLD [81].

353   OB/OW/UW: The contribution of different strains to OB, OW, and UW was highly similar in the

354   discovery set, but this rule seemed to be weakened in the validation set. Most of the bacteria with a

355   high contribution in OB and OW are still the same, including *Gemella haemolysans*, *Solobacterium*

356   *moorei*, *Atopobium parvulum*, *Clostridium asparagus*, *Gamella Sanguinis*, *Granulicatella Adiacens*.

357   However, only a few bacteria of OW and UW had high contribution and were similar, namely

358   *Ruminococcus torques* and *Megasphaera micronuciformis*. *C. ramosum*, *C. clostridioforme*,

359   *Megasphaera*, *Oribacterium*, *S. cristatus*, and *P. gingivalis* have been described in detail above, so we

360   will not repeat them in the validation set. *R. torques* appeared in three disease states simultaneously

361   with a high contribution. It is the most abundant intestinal bacteria in obese people, and has been

362   investigated to be linked to obesity [82]. A meta-transcriptomic analysis revealed *R. torques* as a

363   bacterium known to reduce intestinal barrier integrity [83]. In three studies of intervention

364   improvement in experimental animals fed a high-fat diet, the results showed that the relative

365   abundance of *R. torques* in the intestines of the animals was reduced, and the inflammation of the

366   animals was improved [84-86]. However, the relationship between *R. torques* and UW has not been

367   proved well. *Veillonella* are a group of bacteria that can't ferment in the gut, and they cannot use

368   carbohydrates or amino acids. In our validation set, it has a high contribution to UW, but there are few

369   studies on *Veillonella* and UW at present.

370

Supplementary Note 6. The results further demonstrate strong
371

372   **reproducibility of hiPCA on the test data.**

373   In order to further evaluate the generalization ability of hiPCA, a separate test dataset consisting

374   of 605 samples (292 healthy individuals and 313 unhealthy individuals) was used, which included

375 End-stage renal disease (ESRD) and Schizophrenia (SCZ). As shown in Figure S22, our KS-92 signal

376 panel hiPCA achieved a balanced accuracy of 71.60%, with detection accuracies of 67.81% and

377 75.40% for the healthy and unhealthy groups, respectively. Further analysis of the unhealthy

378 phenotype subgroups revealed that ESRD had a good adverse health prediction (85.20% in 190 out of

379 223 individuals), while SCZ had a detection rate of 51.11% (46 out of 90 individuals). Overall, it can

380 be concluded that hiPCA demonstrates strong reproducibility on the test data and provides good

381 predictions for adverse health outcomes, particularly for ESRD.

382 Like the discovery and validation set, inferential analysis was made by bacteria-to-health-index

383 contribution analysis, and the plots were shown in Figure S23-24.

384 ESRD: In test cohort, *Gemella morbillorum*, *Granulicatella adiacens*, *Lachnospiraceae*

385 *bacterium 9_1_43BFAA*, *Clostridium clostridioforme*, *Gemella sanguinis*, *Clostridium ramosum*, and

386 *Lactobacillus salivarius* were the highest contribution biomarkers in ESRD with PR-50 features.

387 KS-92 shared most contributors but was further supplemented with *Streptococcus cristatus* and

388 *Scardovia wiggsiae*. Gut microbiota can be a potential therapeutic target for kidney diseases, such as

389 chronic kidney disease (CKD), ESRD, peritoneal dialysis, hemodialysis and renal transplantation.

390 Among them, the diversity of *Scardovia wiggsiae* has been reported to be associated with indoxyl

391 sulfate, which is an inflammatory biomarker of ESRD [87]. A previous study found that the absolute

392 quantity of total bacteria and *Clostridium* were signifcantly reduced in ESRD patients [88]. In our

393 study, similarly, *Clostridium ramosum* and *Clostridium clostridioforme* were the important

394 contributors in ESRD. However, there are still many strains that have not been reported, and it seems

395 that we can refer to the degree of contribution to explore more potential diagnostic or therapeutic

396 targets.

397   SCZ: In the test cohort, *Lactobacillus salivarius*, *Anaerotruncus colihominis*, *Clostridium*

398   *clostridioforme*, *Atopobium parvulum*, *Ruminococcus gnavus*, *Lachnospiraceae bacterium*

399   *1_4_56FAA*, and *Clostridium symbiosum* are the highest contribution biomarkers in SCZ with PR-50

400   features. KS-92 indicated that *Megasphaera micronuciformis*, *Lactobacillus mucosae*, *Lactobacillus*

401   *salivaryus*, *Fusobacterium varium*, *Atopobium parvulum*, *Lachnospiraceae bacterium 9_1_43BFAA*

402   and *Anaerotruncus colihominis* were the top seven contributing species to SCZ. Among them,

403   *Lactobacillus salivarius*, *Anaerotruncus colihominis*, and *Atopobium parvulum* were the main

404   contributors shared by PR-50 and KS-92. Previous studies have reported that *Lachnospiraceae* and

405   *Ruminococcaceae* were associated with SCZ [89], which is similar to our results. Gut microbiota

406   disturbances in patients with depression, bipolar disorder, schizophrenia, and anxiety all tend to one

407   pattern [90]. *Lactobacillus salivarius* as a probiotic seemed to have positive effects on the mild

408   chronic inflammation processes in major depressive disorder (MDD) [91]. In future work, we can

409   mine more information on other potential target strains and their intervention effects.

410

References

411

412 1. Arumugam M, Raes J, Pelletier E, Le Paslier D, Yamada T, Mende DR, Fernandes GR, Tap J,
413   Bruls T, Batto J-M: **Enterotypes of the human gut microbiome**. *nature* 2011,
414   **473**(7346):174-180.

415 2. Costea PI, Hildebrand F, Arumugam M, Bäckhed F, Blaser MJ, Bushman FD, De Vos WM,
416   Ehrlich SD, Fraser CM, Hattori M: **Enterotypes in the landscape of gut microbial**
417   **community composition**. *Nature microbiology* 2018, **3**(1):8-16.

418 3. Gorvitovskaia A, Holmes SP, Huse SM: **Interpreting Prevotella and Bacteroides as**
419   **biomarkers of diet and lifestyle**. *Microbiome* 2016, **4**(1):1-12.

420 4. Mobeen F, Sharma V, Tulika P: **Enterotype variations of the healthy human gut**
421   **microbiome in different geographical regions**. *Bioinformation* 2018, **14**(9):560.

422 5. Liu W, Zhang J, Wu C, Cai S, Huang W, Chen J, Xi X, Liang Z, Hou Q, Zhou B: **Unique**
423   **features of ethnic Mongolian gut microbiome revealed by metagenomic analysis**.
424   *Scientific reports* 2016, **6**(1):1-13.

425 6. Meslier V, Laiola M, Roager HM, De Filippis F, Roume H, Quinquis B, Giacco R, Mennella I,
426   Ferracane R, Pons N: **Mediterranean diet intervention in overweight and obese subjects**
427   **lowers plasma cholesterol and causes changes in the gut microbiome and metabolome**
428   **independently of energy intake**. *Gut* 2020, **69**(7):1258-1268.

429 7. Wang DD, Nguyen LH, Li Y, Yan Y, Ma W, Rinott E, Ivey KL, Shai I, Willett WC, Hu FB:
430   **The gut microbiome modulates the protective association between a Mediterranean diet**
431   **and cardiometabolic disease risk**. *Nature medicine* 2021, **27**(2):333-343.

432 8. Wu H, Tremaroli V, Schmidt C, Lundqvist A, Olsson LM, Kramer M, Gummesson A, Perkins
433   R, Bergstrom G, Backhed F: **The Gut Microbiota in Prediabetes and Diabetes: A**
434   **Population-Based Cross-Sectional Study**. *Cell Metabolism* 2020, **32**(3):379-390.e373.

435 9. Bodkhe R, Balakrishnan B, Taneja V: **The role of microbiome in rheumatoid arthritis**
436   **treatment**. *Therapeutic advances in musculoskeletal disease* 2019, **11**:1759720x19844632.

437 10. Thingholm LB, Rühlemann MC, Koch M, Fuqua B, Laucke G, Boehm R, Bang C, Franzosa
438   EA, Hübenthal M, Rahnavard A *et al*: **Obese Individuals with and without Type 2 Diabetes**
439   **Show Different Gut Microbial Functional Capacity and Composition**. *Cell host &*
440   *microbe* 2019, **26**(2):252-264.e210.

441 11. Jie Z, Xia H, Zhong S-L, Feng Q, Li S, Liang S, Zhong H, Liu Z, Gao Y, Zhao H: **The gut**
442   **microbiome in atherosclerotic cardiovascular disease**. *Nature communications* 2017,
443   **8**(1):1-12.

444 12. Baker JL, Morton JT, Dinis M, Alvarez R, Tran NC, Knight R, Edlund A: **Deep**
445   **metagenomics examines the oral microbiome during dental caries, revealing novel taxa**
446   **and co-occurrences with host molecules**. *Genome research* 2021, **31**(1):64-74.

447 13. Murugkar PP, Collins AJ, Chen T, Dewhirst FE: **Isolation and cultivation of candidate**
448   **phyla radiation Saccharibacteria (TM7) bacteria in coculture with bacterial hosts**.
449   *Journal of oral microbiology* 2020, **12**(1):1814666.

450 14. Jonsson AL, Bäckhed F: **Role of gut microbiota in atherosclerosis**. *Nature Reviews*
451   *Cardiology* 2017, **14**(2):79-87.

452 15. Van Hul M, Le Roy T, Prifti E, Dao MC, Paquot A, Zucker J-D, Delzenne NM, Muccioli GG,
453   Clément K, Cani PD: **From correlation to causality: the case of Subdoligranulum**. *Gut*
454   *microbes* 2020, **12**(1):1849998.

455 16. Christensen JJ, Facklam RR: **Granulicatella and Abiotrophia species from human clinical**
456   **specimens**. *Journal of Clinical Microbiology* 2001, **39**(10):3520-3523.

457 17. Geva-Zatorsky N, Sefik E, Kua L, Pasman L, Tan TG, Ortiz-Lopez A, Yanortsang TB, Yang L,
458   Jupp R, Mathis D: **Mining the human gut microbiota for immunomodulatory organisms**.
459   *Cell* 2017, **168**(5):928-943. e911.

460 18. Wang J, Li W, Wang C, Wang L, He T, Hu H, Song J, Cui C, Qiao J, Qing L *et al*: **Enterotype**
461   **Bacteroides Is Associated with a High Risk in Patients with Diabetes: A Pilot Study**.
462   *Journal of diabetes research* 2020, **2020**:6047145.

463 19. Koren O, Spor A, Felin J, Fåk F, Stombaugh J, Tremaroli V, Behre CJ, Knight R, Fagerberg B,
464   Ley RE: **Human oral, gut, and plaque microbiota in patients with atherosclerosis**.
465   *Proceedings of the National Academy of Sciences* 2011, **108**(Supplement 1):4592-4598.

466 20. Scher JU, Abramson SB: **The microbiome and rheumatoid arthritis**. *Nature reviews*
467   *rheumatology* 2011, **7**(10):569-578.

468 21. du Teil Espina M, Gabarrini G, Harmsen HJ, Westra J, van Winkelhoff AJ, van Dijl JM: **Talk**
469   **to your gut: the oral-gut microbiome axis and its immunomodulatory role in the etiology**
470   **of rheumatoid arthritis**. *FEMS microbiology reviews* 2019, **43**(1):1-18.

471 22. Wang J, Jia H: **Metagenome-wide association studies: fine-mining the microbiome**. *Nature*
472   *Reviews Microbiology* 2016, **14**(8):508-522.

473 23. Blod C, Schlichting N, Schülin S, Suttkus A, Peukert N, Stingu CS, Hirsch C, Elger W,
474   Lacher M, Bühligen U: **The oral microbiome—the relevant reservoir for acute pediatric**
475   **appendicitis?** *International journal of colorectal disease* 2018, **33**(2):209-218.

476 24. Liu X, Tong X, Zhu J, Tian L, Jie Z, Zou Y, Lin X, Liang H, Li W, Ju Y:
477   **Metagenome-genome-wide association studies reveal human genetic impact on the oral**
478   **microbiome**. *Cell discovery* 2021, **7**(1):1-16.

479 25. Lu M, Xuan S, Wang Z: **Oral microbiota: A new view of body health**. *Food Science and*
480   *Human Wellness* 2019, **8**(1):8-15.

481 26. Forbes JD, Chen C-y, Knox NC, Marrie R-A, El-Gabalawy H, de Kievit T, Alfa M, Bernstein
482   CN, Van Domselaar G: **A comparative study of the gut microbiota in immune-mediated**
483   **inflammatory diseases—does a common dysbiosis exist?** *Microbiome* 2018, **6**(1):1-15.

484 27. Zhang X, Zhang D, Jia H, Feng Q, Wang D, Liang D, Wu X, Li J, Tang L, Li Y: **The oral and**
485   **gut microbiomes are perturbed in rheumatoid arthritis and partly normalized after**
486   **treatment**. *Nature medicine* 2015, **21**(8):895-905.

487 28. Tierney BT, Tan Y, Kostic AD, Patel CJ: **Gene-level metagenomic architectures across**
488   **diseases yield high-resolution microbiome diagnostic indicators**. *Nature communications*
489   2021, **12**(1):1-12.

490 29. Gurung M, Li Z, You H, Rodrigues R, Jump DB, Morgun A, Shulzhenko N: **Role of gut**
491   **microbiota in type 2 diabetes pathophysiology**. *EBioMedicine* 2020, **51**:102590.

492 30. Barrak I, Stájer A, Gajdács M, Urbán E: **Small, but smelly: the importance of**
493   **Solobacterium moorei in halitosis and other human infections**. *Heliyon* 2020,
494   **6**(10):e05371.

495 31. Aron-Wisnewsky J, Warmbrunn MV, Nieuwdorp M, Clément K: **Metabolism and metabolic**
496   **disorders and the microbiome: The intestinal microbiota associated with obesity, lipid**
497   **metabolism, and metabolic health—Pathophysiology and therapeutic strategies**.
498   *Gastroenterology* 2021, **160**(2):573-599.

499 32. Alvarez-Silva C, Kashani A, Hansen TH, Pinna NK, Anjana RM, Dutta A, Saxena S, Støy J,
500   Kampmann U, Nielsen T: **Trans-ethnic gut microbiota signatures of type 2 diabetes in**
501   **Denmark and India**. *Genome medicine* 2021, **13**(1):1-13.

502 33. Qin J, Li Y, Cai Z, Li S, Zhu J, Zhang F, Liang S, Zhang W, Guan Y, Shen D: **A**
503   **metagenome-wide association study of gut microbiota in type 2 diabetes**. *Nature* 2012,
504   **490**(7418):55-60.

505 34. Karlsson FH, Tremaroli V, Nookaew I, Bergström G, Behre CJ, Fagerberg B, Nielsen J,
506   Bäckhed F: **Gut metagenome in European women with normal, impaired and diabetic**
507   **glucose control**. *Nature* 2013, **498**(7452):99-103.

508 35. Upadhyaya S, Banerjee G: **Type 2 diabetes and gut microbiome: at the intersection of**
509   **known and unknown**. *Gut microbes* 2015, **6**(2):85-92.

510 36. Jin L, Liu Y, Jing C, Wang R, Wang Q, Wang H: **Neutrophil extracellular traps**
511   **(NETs)-mediated killing of carbapenem-resistant hypervirulent Klebsiella pneumoniae**
512   **(CR-hvKP) are impaired in patients with diabetes mellitus**. *Virulence* 2020,
513   **11**(1):1122-1130.

514 37. Salonen A, Lahti L, Salojärvi J, Holtrop G, Korpela K, Duncan SH, Farquharson F, Johnstone
515   AM, Lobley GE, Louis P: **Impact of diet and individual variation on intestinal microbiota**
516   **composition and fermentation products in obese men**. *The ISME journal* 2014,
517   **8**(11):2218-2230.

518 38. Perry RJ, Peng L, Barry NA, Cline GW, Zhang D, Cardone RL, Petersen KF, Kibbey RG,
519   Goodman AL, Shulman GI: **Acetate mediates a microbiome–brain–β-cell axis to promote**
520   **metabolic syndrome**. *Nature* 2016, **534**(7606):213-217.

521 39. Vacca M, Celano G, Calabrese FM, Portincasa P, Gobbetti M, De Angelis M: **The**

522   **controversial role of human gut lachnospiraceae**. *Microorganisms* 2020, **8**(4):573.

523 40.   Hildebrandt MA, Hoffmann C, Sherrill–Mix SA, Keilbaugh SA, Hamady M, Chen YY,
524   Knight R, Ahima RS, Bushman F, Wu GD: **High-fat diet determines the composition of the**
525   **murine gut microbiome independently of obesity**. *Gastroenterology* 2009,
526   **137**(5):1716-1724. e1712.

527 41.   Tun HM, Bridgman SL, Chari R, Field CJ, Guttman DS, Becker AB, Mandhane PJ, Turvey
528   SE, Subbarao P, Sears MR: **Roles of birth mode and infant gut microbiota in**
529   **intergenerational transmission of overweight and obesity from mother to offspring**.
530   *JAMA pediatrics* 2018, **172**(4):368-377.

531 42.   Woting A, Pfeiffer N, Loh G, Klaus S, Blaut M: **Clostridium ramosum promotes high-fat**
532   **diet-induced obesity in gnotobiotic mouse models**. *MBio* 2014, **5**(5):e01530-01514.

533 43.   Le Chatelier E, Nielsen T, Qin J, Prifti E, Hildebrand F, Falony G, Almeida M, Arumugam M,
534   Batto J-M, Kennedy S: **Richness of human gut microbiome correlates with metabolic**
535   **markers**. *Nature* 2013, **500**(7464):541-546.

536 44.   Nicolucci AC, Hume MP, Martínez I, Mayengbam S, Walter J, Reimer RA: **Prebiotics reduce**
537   **body fat and alter intestinal microbiota in children who are overweight or with obesity**.
538   *Gastroenterology* 2017, **153**(3):711-722.

539 45.   Aleti G, Kohn JN, Troyer EA, Weldon K, Huang S, Tripathi A, Dorrestein PC, Swafford AD,
540   Knight R, Hong S: **Salivary bacterial signatures in depression-obesity comorbidity are**
541   **associated with neurotransmitters and neuroactive dipeptides**. *BMC microbiology* 2022,
542   **22**(1):1-17.

543 46.   Laiola M, De Filippis F, Vitaglione P, Ercolini D: **A mediterranean diet intervention**
544   **reduces the levels of salivary periodontopathogenic bacteria in overweight and obese**
545   **subjects**. *Applied and environmental microbiology* 2020, **86**(12):e00777-00720.

546 47.   Goodson J, Groppo D, Halem S, Carpino E: **Is obesity an oral bacterial disease?** *Journal of*
547   *dental research* 2009, **88**(6):519-523.

548 48.   Benahmed AG, Gasmi A, Doşa A, Chirumbolo S, Mujawdiya PK, Aaseth J, Dadar M,
549   Bjørklund G: **Association between the gut and oral microbiome with obesity**. *Anaerobe*
550   2021, **70**:102248.

551 49.   Gupta A, Osadchiy V, Mayer EA: **Brain–gut–microbiome interactions in obesity and food**
552   **addiction**. *Nature Reviews Gastroenterology & Hepatology* 2020, **17**(11):655-672.

553 50.   Wan Y, Yuan J, Li J, Li H, Yin K, Wang F, Li D: **Overweight and underweight status are**
554   **linked to specific gut microbiota and intestinal tricarboxylic acid cycle intermediates**.
555   *Clinical Nutrition* 2020, **39**(10):3189-3198.

556 51.   Engevik MA, Luk B, Chang-Graham AL, Hall A, Herrmann B, Ruan W, Endres BT, Shi Z,
557   Garey KW, Hyser JM *et al*: **Bifidobacterium dentium Fortifies the Intestinal Mucus Layer**
558   **via Autophagy and Calcium Signaling Pathways**. *mBio* 2019, **10**(3):e01087-01019.

559 52.   Engevik M, Ruan W, Ihekweazu F, Versalovic J: **MODULATION OF THE MUCUS**

560 **LAYER BY BIFIDOBACTERIUM DENTIUM PROVIDES PROTECTION IN A**
561 **MODEL OF COLITIS**. *Inflammatory Bowel Diseases* 2020, **26**:S38-S38.

562 53. Yu X, Zhang Y, Yu E, Wang N, Shuai Q, Yan F, Jiang L, Wang H, Liu J, Chen Y *et al*:
563 **Changes in gut microbiota of patients with colorectal adenoma and polyps:a case control**
564 **study**. *Fudan University Journal of Medical Sciences* 2018, **45**(5):658-663.

565 54. Abdulamir AS, Hafidh RR, Abu Bakar F: **The association of Streptococcus**
566 **bovis/gallolyticus with colorectal tumors: the nature and the underlying mechanisms of**
567 **its etiological role**. *Journal of experimental & clinical cancer research : CR* 2011, **30**(1):11.

568 55. Norfleet RG, Mitchell PD: **Streptococcus bovis does not selectively colonize colorectal**
569 **cancer and polyps**. *Journal of clinical gastroenterology* 1993, **17**(1):25-28.

570 56. Boleij A, Tjalsma H: **The itinerary of Streptococcus gallolyticus infection in patients with**
571 **colonic malignant disease**. *Lancet Infectious Diseases* 2013, **13**(8):719-724.

572 57. Hirai J, Sakanashi D, Hagihara M, Haranaga S, Uechi K, Kato H, Hamada H, Nishiyama N,
573 Koizumi Y, Suematsu H *et al*: **Bacteremia due to Streptococcus tigurinus: A case report**
574 **and literature review**. *Journal of Infection and Chemotherapy* 2016, **22**(11):762-766.

575 58. Yao Y, Ni H, Wang X, Xu Q, Zhang J, Jiang L, Wang B, Song S, Zhu X: **A New Biomarker**
576 **of Fecal Bacteria for Non-Invasive Diagnosis of Colorectal Cancer**. *Frontiers in cellular*
577 *and infection microbiology* 2021, **11**:744049.

578 59. Chang HB, Mishra R, Cen CP, Tang YQ, Ma CC, Wasti S, Wang YY, Ou QY, Chen KN,
579 Zhang JC: **Metagenomic Analyses Expand Bacterial and Functional Profiling**
580 **Biomarkers for Colorectal Cancer in a Hainan Cohort, China**. *Current Microbiology*
581 2021, **78**(2):705-712.

582 60. Avuthu N, Guda C: **Meta-Analysis of Altered Gut Microbiota Reveals Microbial and**
583 **Metabolic Biomarkers for Colorectal Cancer**. *Microbiology spectrum* 2022,
584 **10**(4):e0001322.

585 61. Yin J, Sternes PR, Wang M, Song J, Morrison M, Li T, Zhou L, Wu X, He F, Zhu J *et al*:
586 **Shotgun metagenomics reveals an enrichment of potentially cross-reactive bacterial**
587 **epitopes in ankylosing spondylitis patients, as well as the effects of TNFi therapy upon**
588 **microbiome composition**. *Annals of the rheumatic diseases* 2020, **79**(1):132-140.

589 62. Wang Y, Wei J, Zhang W, Doherty M, Zhang Y, Xie H, Li W, Wang N, Lei G, Zeng C: **Gut**
590 **dysbiosis in rheumatic diseases: A systematic review and meta-analysis of 92**
591 **observational studies**. *EBioMedicine* 2022, **80**:104055.

592 63. Wen C, Zheng Z, Shao T, Liu L, Xie Z, Le Chatelier E, He Z, Zhong W, Fan Y, Zhang L *et al*:
593 **Quantitative metagenomics reveals unique gut microbiome biomarkers in ankylosing**
594 **spondylitis**. *Genome biology* 2017, **18**(1):142.

595 64. Zhang X, Zhang D, Jia H, Feng Q, Wang D, Liang D, Wu X, Li J, Tang L, Li Y *et al*: **The**
596 **oral and gut microbiomes are perturbed in rheumatoid arthritis and partly normalized**
597 **after treatment**. *Nature medicine* 2015, **21**(8):895-905.

598   65.   Sekizuka T, Ogasawara Y, Ohkusa T, Kuroda M: **Characterization of Fusobacterium**
599   **varium Fv113-g1 isolated from a patient with ulcerative colitis based on complete**
600   **genome sequence and transcriptome analysis**. *PloS one* 2017, **12**(12):e0189319.

601   66.   Manson McGuire A, Cochrane K, Griggs AD, Haas BJ, Abeel T, Zeng Q, Nice JB,
602   MacDonald H, Birren BW, Berger BW *et al*: **Evolution of invasion in a diverse set of**
603   **Fusobacterium species**. *mBio* 2014, **5**(6):e01864.

604   67.   Cao P, Chen Y, Guo X, Chen Y, Su W, Zhan N, Dong W: **Fusobacterium nucleatum**
605   **Activates Endoplasmic Reticulum Stress to Promote Crohn's Disease Development via**
606   **the Upregulation of CARD3 Expression**. *Frontiers in pharmacology* 2020, **11**:106.

607   68.   Dharmani P, Strauss J, Ambrose C, Allen-Vercoe E, Chadee K: **Fusobacterium nucleatum**
608   **Infection of Colonic Cells Stimulates MUC2 Mucin and Tumor Necrosis Factor Alpha**.
609   *Infection and Immunity* 2011, **79**(7):2597-2607.

610   69.   Yu MR, Kim HJ, Park HR: **Fusobacterium nucleatum Accelerates the Progression of**
611   **Colitis-Associated Colorectal Cancer by Promoting EMT**. *Cancers* 2020, **12**(10):2728.

612   70.   Mangin I, Bonnet R, Seksik P, Rigottier-Gois L, Sutren M, Bouhnik Y, Neut C, Collins MD,
613   Colombel JF, Marteau P *et al*: **Molecular inventory of faecal microflora in patients with**
614   **Crohn's disease**. *Fems Microbiology Ecology* 2004, **50**(1):25-36.

615   71.   Elsayed S, Zhang KY: **Isolation and 16S ribosomal RNA gene sequence-based**
616   **identification of Clostridium scindens from an intra-abdominal abscess**. *Anaerobe* 2006,
617   **12**(1):13-16.

618   72.   Rashid T, Ebringer A: **Autoimmunity in Rheumatic Diseases Is Induced by Microbial**
619   **Infections via Crossreactivity or Molecular Mimicry**. *Autoimmune diseases* 2012,
620   **2012**:539282.

621   73.   Sung YK, Lee JK, Lee KH, Lee KT, Kang CI: **The Clinical Epidemiology and Outcomes of**
622   **Bacteremic Biliary Tract Infections Caused by Antimicrobial-Resistant Pathogens**.
623   *American Journal of Gastroenterology* 2012, **107**(3):473-483.

624   74.   Zhang X, Coker OO, Chu ES, Fu K, Lau HCH, Wang YX, Chan AWH, Wei H, Yang X, Sung
625   JJY *et al*: **Dietary cholesterol drives fatty liver-associated liver cancer by modulating gut**
626   **microbiota and metabolites**. *Gut* 2021, **70**(4):761-774.

627   75.   Engevik MA, Danhof HA, Ruan W, Engevik AC, Chang-Graham AL, Engevik KA, Shi Z,
628   Zhao Y, Brand CK, Krystofiak ES *et al*: **Fusobacterium nucleatum Secretes Outer**
629   **Membrane Vesicles and Promotes Intestinal Inflammation**. *mBio* 2021,
630   **12**(2):e02706-02720.

631   76.   Qin N, Yang F, Li A, Prifti E, Chen Y, Shao L, Guo J, Le Chatelier E, Yao J, Wu L *et al*:
632   **Alterations of the human gut microbiome in liver cirrhosis**. *Nature* 2014,
633   **513**(7516):59-64.

634   77.   Chen Y, Ji F, Guo J, Shi D, Fang D, Li L: **Dysbiosis of small intestinal microbiota in liver**
635   **cirrhosis and its association with etiology**. *Scientific reports* 2016, **6**:34055.

78. Aron-Wisnewsky J, Vigliotti C, Witjes J, Le P, Holleboom AG, Verheij J, Nieuwdorp M, Clément K: **Gut microbiota and human NAFLD: disentangling microbial signatures from metabolic disorders**. *Nature reviews Gastroenterology & hepatology* 2020, **17**(5):279-297.

79. Tilg H, Adolph TE, Dudek M, Knolle P: **Non-alcoholic fatty liver disease: the interplay between metabolism, microbes and immunity**. *Nature metabolism* 2021, **3**(12):1596-1607.

80. Li RJ, Jie ZY, Feng Q, Fang RL, Li F, Gao Y, Xia HH, Zhong HZ, Tong B, Madsen L *et al*: **Network of Interactions Between Gut Microbiome, Host Biomarkers, and Urine Metabolome in Carotid Atherosclerosis**. *Frontiers in cellular and infection microbiology* 2021, **11**:708088.

81. Caussy C, Tripathi A, Humphrey G, Bassirian S, Singh S, Faulkner C, Bettencourt R, Rizo E, Richards L, Xu ZZ *et al*: **A gut microbiome signature for cirrhosis due to nonalcoholic fatty liver disease**. *Nat Commun* 2019, **10**(1):1406.

82. Xie YH, Zhang CY, Wang ZH, Wei C, Liao NJ, Wen X, Niu CW, Yi L, Wang ZJ, Xi Z: **Fluorogenic Assay for Acetohydroxyacid Synthase: Design and Applications**. *Analytical Chemistry* 2019, **91**(21):13582-13590.

83. Deaver JA, Eum SY, Toborek M: **Circadian Disruption Changes Gut Microbiome Taxa and Functional Gene Composition**. *Frontiers in microbiology* 2018, **9**:737.

84. Zeng Z, Guo X, Zhang J, Yuan Q, Chen S: **Lactobacillus paracasei modulates the gut microbiota and improves inflammation in type 2 diabetic rats**. *Food & function* 2021, **12**(15):6809-6820.

85. Yue C, Li M, Li J, Han X, Zhu H, Yu G, Cheng J: **Medium-, long- and medium-chain-type structured lipids ameliorate high-fat diet-induced atherosclerosis by regulating inflammation, adipogenesis, and gut microbiota in ApoE(-/-) mice**. *Food & function* 2020, **11**(6):5142-5155.

86. Iversen KN, Dicksved J, Zoki C, Fristedt R, Pelve EA, Langton M, Landberg R: **The Effects of High Fiber Rye, Compared to Refined Wheat, on Gut Microbiota Composition, Plasma Short Chain Fatty Acids, and Implications for Weight Loss and Metabolic Risk Factors (the RyeWeight Study)**. *Nutrients* 2022, **14**(8):1669.

87. Asgharian M, Gholizadeh P, Samadi Kafil H, Ghojazadeh M, Samadi A, Soleymani J, Jouyban A, Tayebi Khosroshahi H: **Correlation of inflammatory biomarkers with the diversity of Bacteroidaceae, Bifidobacteriaceae, Prevotellaceae and Lactobacillaceae families in the intestinal microbiota of patients with end stage renal disease**. *Advances in medical sciences* 2022, **67**(2):304-310.

88. Jiang S, Xie S, Lv D, Wang P, He H, Zhang T, Zhou Y, Lin Q, Zhou H, Jiang J *et al*: **Alteration of the gut microbiota in Chinese population with chronic kidney disease**. *Scientific reports* 2017, **7**(1):2870.

89. Wang D, Russel WA, Sun Y, Belanger K, Ay A: **Machine learning and network analysis of the gut microbiome from patients with schizophrenia and non-psychiatric subject**

675   **controls reveal behavioral risk factors and bacterial interactions**. *Schizophrenia Research*
676   2022, **251**:49-58.

677 90.   Nikolova VL, Smith MRB, Hall LJ, Cleare AJ, Stone JM, Young AH: **Perturbations in Gut**
678   **Microbiota Composition in Psychiatric Disorders: A Review and Meta-analysis**. *JAMA*
679   *psychiatry* 2021, **78**(12):1343-1354.

680 91.   Reiter A, Bengesser SA, Hauschild AC, Birkl-Töglhofer AM, Fellendorf FT, Platzer M,
681   Färber T, Seidl M, Mendel LM, Unterweger R *et al*: **Interleukin-6 Gene Expression**
682   **Changes after a 4-Week Intake of a Multispecies Probiotic in Major Depressive**
683   **Disorder-Preliminary Results of the PROVIT Study**. *Nutrients* 2020, **12**(9):2575.

684