# A multi-omic cohort as a reference point for promoting a healthy human gut microbiome

**Authors:** Zhuye Jie[1,2,3*,†], Suisha Liang[1,†], Qiuxia Ding[1,†], Fei Li[1], Shanmei Tang[1], Dan Wang[1], Yuxiang Lin[1], Peishan Chen[1], Kaiye Cai[1], Xuemei Qiu[1], Qiang Li[1], Yunli Liao[1], Dongsheng Zhou[1], Heng Lian[1], Yong Zuo[1], Xiaomin Chen[1], Weiqiao Rao[1], Yan Ren[1], Yuan Wang[1], Jin Zi[1], Rong Wang[1], Hongcheng Zhou[4], Haorong Lu[4], Xiaohan Wang[1], Wei Zhang[1], Tao Zhang[1,3], Liang Xiao[1,5,6], Yang Zong[1], Weibin Liu[1], Huanming Yang[1,7], Jian Wang[1,7], Yong Hou[1], Xiao Liu[1], Karsten Kristiansen[1,3], Huanzi Zhong[1,3], Huijue Jia[1,2,3,8,*], Xun Xu[1,*]

**Affiliations:**

[1]BGI-Shenzhen, Shenzhen, China.

[2]Shenzhen Key Laboratory of Human Commensal Microorganisms and Health Research, BGI-Shenzhen, Shenzhen, China.

[3]Department of Biology, Ole MaalØes Vej 5, University of Copenhagen, Copenhagen, Denmark.

[4]China National Genebank, BGI-Shenzhen, Shenzhen 518120, China.

[5]Shenzhen Engineering Laboratory of Detection and Intervention of human intestinal microbiome, BGI-Shenzhen, Shenzhen, China.

[6]BGI-Qingdao, BGI-Shenzhen, Qingdao, 266555, China.

[7]James D. Watson Institute of Genome Sciences, Hangzhou, China.

[8]Macau University of Science and Technology, Avenida Wai long, Taipa, Macau, China.

* To whom correspondence should be addressed. X.X. xuxun@genomics.cn ; Z.J. jiezhuye@genomics.cn; H.J. jiahuijue@genomics.cn

† These authors contributed equally to this work.

**Abstract:**

**More than a decade of gut microbiome studies have a common goal for human health. As most of the disease studies sample the elderly or the middle-aged, a reference cohort for young individuals has been lacking. It is also not clear what other omics data need to be measured to better understand the gut microbiome. Here we present high-depth metagenomic shotgun sequencing data for the fecal microbiome together with other omics data in a cohort of 2,183 adults, and observe a number of vitamins, hormones, amino acids and trace elements to correlate with the gut microbiome and cluster with T cell receptors. Associations with physical fitness, sleeping habits and dairy consumption are identified in this large multi-omic cohort. Many of the associations are validated in an additional cohort of 1,404 individuals. Our comprehensive data are poised to advise future study designs to better understand and manage our gut microbiome both in population and in mechanistic investigations.**

The gut microbiome has been implicated in a growing list of complex diseases, showing great potential for the diagnosis and treatment of metabolic, autoimmune and neurological diseases as well as cancer. While case-control studies have been

illuminating [1], recently published studies have emphasized difficulty in extrapolating to natural cohorts due to heterogeneity in location and ethnicity [2,3]. So far only a few cohorts made use of metagenomic shotgun sequencing instead of 16S rRNA gene amplicon sequencing, the largest being the LifeLines Deep cohort (n=1,135, 32 million reads per sample) from the Netherlands [4–7]. Fecal or plasma metabolites are more or less included in gut microbiome studies, but the conclusions usually did not go beyond short-chain fatty acids (SCFA), amino acids, vitamin B complex or bile acids. Levels of trace elements such as arsenic have been a health concern (https://www.usgs.gov/mission-areas/water-resources/science/arsenic-and-drinking-water?qt-science_center_objects=0#qt-science_center_objects, https://www.fda.gov/food/metals/arsenic-food-and-dietary-supplements), but are unexplored in the microbiome field. Biological sex is a strong determent for the gut microbiome in mice and livestock [8–10]. The impact of hormones on the human gut microbiome, or vice versa, remains unclear.

As part of the 4D-SZ (trans-omic, with more time points in future studies) cohort, here we present metagenomic shotgun sequencing data sufficient for high-resolution taxonomic and functional profiling (86.1 ± 23.3 million reads per sample) of the fecal microbiome in a cohort of 2,183 adults, along with questionnaire data, physical fitness tests, facial skin features, plasma metabolome and immune repertoire. Trans-omics analyses in this Han Chinese cohort put into context fecal microbiome disease markers, and uncover previously overlooked measurements such as aldosterone, testosterone, trace elements and vitamin A that influence the gut microbiome, which were validated in an additional cohort of 1,400 individuals. Trends for cardiometabolic diseases and colorectal

cancer can be seen, despite the average age of 29.6. This is also to our knowledge the largest cohort with facial skin data and immune repertoire data, which would also be of interest for general health management and disease studies.

A recent study casted doubt over the health benefits of probiotic consumption, concluding that colonization of the strains was highly variable between individuals [11]. Our large cohort unequivocally showed commercial yogurt strains, especially *Streptococcus thermophilus* and *Bifidobacterium animalis* in feces, and suggested beneficial effects in cardiometabolic health.

## Results

### Comprehensive gut microbiome data together with other omics

Fecal samples were collected during a physical examination, and 2,183 samples (Age, 29.6 ± 5.5, average ± stdev) were subjected to metagenomic shotgun sequencing, yielding 82.95 ± 24.26 million high-quality non-human reads per sample (Supplementary Table 1a), ensuring accurate taxonomic and functional profiling. The reads were mapped to a comprehensive human gut microbiome reference gene catalog containing 9.9 million genes (with a saturating mapping rate of 80.1 ± 4.9 %) and then assigned to 1,507 Metagenomic Species (MGSs) [12–14] and 2,981 metagenomic linkage groups (MLGs, Kendall's tau instead of Pearson's or Spearman's correlation between genes) (Qin et al, 2012, Jie et al, 2017), to include both known and unknown microbes.

Other omics data, including 104 plasma metabolites (3,980 samples), 634 immune indices (PBMC (Peripheral blood mononuclear cells) V(D)J usage and its shannon

diversity, 4,120 samples) from buffy coat, 72 basic medical data (body measurements and routine blood test, 2,715 samples), 49 facial skin imaging indices (2,049 samples), 24 physical fitness data (3,833 samples), 18 entries from psychological questionnaire (2,039 samples), and 56 entries from lifestyle questionnaire (3,820 samples) were collected from the same individuals (Fig. 1, Supplementary Tables 1b-d).

**The gut microbiome as a relatively independent dimension for health**

To get an overall idea of the relationship between omics, an inter-omics prediction value between omics data was calculated using a 5-fold cross-validated random forest model (RFCV, Fig. 2a). Basic medical data showed the highest global systematic association with other omics data. The accuracy of prediction from basic medical data to physical fitness data and from metabolites to basic medical data reaching 75% quantile showed RFCV R = 0.461 and 0.399, respectively (Fig. 2a, b, Supplementary Fig. 1). Basic medical data showed high prediction accuracy to metabolites (Fig. 2a, b**)**; on the other hand, serum creatinine, BMI, waist to hip ratio, hematocrit and triglyceride in basic medical data can be predicted by metabolites (Fig. 2c). Metabolites constituted the highest prediction accuracy to immune indices (R = 0.292) (Fig. 2b). Immune indices showed the second highest prediction accuracy to metabolites (Fig. 2a, b). Facial skin features can be predicted by basic medical data, metabolites, physical fitness data and lifestyle questionnaire (Fig. 2b, c). Among the lifestyle questionnaire, smoking, drinking (especially low concentration alcohol), sports habits (especially resistance training), high-sugar and high-fat dietary habit, and staying up until midnight can be predicted by other omics data (Fig. 2c).

A number of factors have been reported to explain gut microbial composition, while the total percentage of variance explained remained in single digits [4,15]. According to a RFCV predict model, we observe in this metagenomic cohort influence from lifestyle questionnaire factors such as defecation, yogurt, age, gender, smoking, milk, soymilk, drinking alcohol, fruit and vegetables on gut microbiome composition (Fig. 3), and the cumulative effect size was also in single digits (Supplementary Tables 2b,2c). The BMI distribution is narrow in this cohort (21.729±3.787, Supplementary Table 1b), so its effect size was 0.0015 (q-value=0.014, Supplementary Table 2b). ABO blood group could also predict fecal microbiome composition (RFCV R=0.2, Fig. 3), and specific differences include Lachospiraceae bacterium 3_1_46FAA in blood type A (q = 6.12E-5), *Ruminococcus torques* in blood type B (q = 1.59E-2), unnamed MGS209 in blood type AB (q = 1.59E-2) and *Megaspaera micronuciformis* in blood type O (q = 1.69E-2).

As our 'other genome', the gut microbiome could predict other omics in this cohort. Gut microbiome showed the greatest prediction power for metabolites, such as plasma vitamins (vitamin A, folic acid, vitamin B5, vitamin D), plasma hormones (testosterone, aldosterone), trace elements (mercury, selenium, arsenic) and plasma amino acids (branched chain amino acids (BCAA), glutamic acid, tryptophan, tyrosine, histidine, alanine) (Fig. 3). Interestingly, hand grip strength, vital capacity, speckles and pores on cheeks and staying up until midnight can also be predicted by the gut microbiome (Fig. 3).

We next included a validation cohort of 1404 individuals (mean age 29.515±5.248, 480 males and 570 females, 82.95 ± 24.26 million high-quality non-human reads per fecal sample), which differed by hometown location compared to the initial cohort (Supplementary Table 1a, Supplementary Table 1b, Fig. 1). The gut microbiome could

also predict these plasma metabolites, with greater effects from mercury, cysteine, selenium, iron and cobalt (Fig. 3, Supplementary Table 3), while other data such as physical fitness tests and facial skin features are not available.

**Defecation, hormone and gender**

We see that gender (female 1,016, male 1,007) was one of the most significant factors to diverge gut microbiome composition (Supplementary Fig.2a). *Eubacterium dolichum,* and *Blautia wexlerae* were significantly more abundant in males (Supplementary Fig. 2a), after adjusting for age, BMI, medication and dietary supplements (Supplementary Table 3b). *Fusobacterium mortiferum*, which positively associated with testosterone, was sensitive to the statistical adjustments (Supplementary Tables 3a, 3b). Compared to males, females showed a greater α-diversity (Supplementary Table 2a, Supplementary Fig. 2c). *Bifidobacterium longum*, *B. bifidum*, and *B. catenulatum*, *B. pseudocatenulatum* were all significantly enriched in females, as well as potentially oral or vaginal bacteria such as *Streptococcus parasanguinis*, *Prevotella bivia* (Supplementary Fig. 2a). Gut microbial functional potential for secondary bile acids strongly associated with self-reported defecation frequency, which were better validated than associations with sex hormones (Supplementary Fig. 2b), suggesting that these are stable patterns.

Aldosterone, one of the major adrenal gland mineralcorticoid, positively correlated with bacteria implicated in cardiometabolic health, such as *Bacteroides intestinalis, B. cellulosilyticus*, *B. stercorirosoris* and *Eubacterium eligens* (Supplementary Fig. 3) [16]. *E. eligens* and *Ruminococcus lactaris* scaled negatively with self-reported preference for a

salty diet, in contrast to *Blautia obeum* (Supplementary Fig. 2a), and mice on a high salt diet showed decrease in a number of commensal bacteria [17].

**The metabolome-immune-gut axis**

Among the strongest associations between different omics is that between immune repertoire and plasma metabolites (Fig. 2). More strikingly, when we plotted the associations in detail, the clusters of metabolites corresponded either to the same *TRBV* (T-cell receptor beta variable gene) or to the same *TRBJ* (T-cell receptor beta joining gene) (Supplementary Fig. 4a). Vitamin A, 5-methyl four hydrogen folic acid, selenium, mercury and serum aldosterone showed positive associations with a few TRBJ1-4 and TRBJ2-1, and negative association with TRBJ2-4. Vitamin B5, Vitamin E, phosphoserine, arginosuccinic acid and arsenic showed positive associations with TRBJ1-4, as well as negative associations with TRBJ2-4, TRBV20-1 and TRBV3-1. Glutamic acid and serine showed a pattern that were largely opposite to that of the vitamin A cluster, except for negative associations with TRBV20-1.

We next explored how the gut microbiome might help put the metabolome-immune associations into context. Vitamin A is central to a healthy immune system but is typically studied for its role in early development [18]. A recent mice study reported modulation of retinol dehydrogenase 7 expression and dampened antimicrobial response in the gut by Clostridiale [19]. Consistently, we observed associations between Clostridia species (Clostridia MGS0123, MGS0560, MGS0558, Lachnospiraceae bacterium 1_4_56FAA, Lachnospiraceae bacterium 6_1_63FAA, Lachnospiraceae bacterium 9_1_43BFAA, *C, bolteae, Clostridium* sp. AT4, *Clostridium* sp. M62.1) and vitamin A in adult humans both with Spearman's corelation and with Masaslin associated

153 (Supplementary Fig. 3, Supplementary Table 3a). 5-methyl four hydrogen folic acid
154 exhibited a positive correlation with *Eubacterium eligens* (Supplementary Fig. 4a,
155 Supplementary Fig. 3), a butyrate-producing bacterium that was relatively depleted in
156 atherosclerotic cardiovascular disease [16]. 5-methyl four hydrogen folic acid also
157 negatively associated with *Dorea* and *Blautia* species (Supplementary Fig. 4a), which
158 have been implicated in obesity and could metabolize formate or hydrogen [20–22]
159 (Supplementary Fig. 4a**)**. Associations between the gut microbiome and trace elements
160 including mercury, selenium and arsenic might be surprising (Supplementary Fig. 4a).
161 Selenium-containing rice is commercially promoted as anti-cancer, and we found that the
162 association pattern largely followed arsenic, consistent with these two trace elements'
163 similar function in anaerobic respiration [23]. Selenium and mercury also correlated with
164 disease-associated species such as *Clostridium bolteae* and *Ruminococcus gnavus* in the
165 gut microbiome.

166 The metabolome-immune cluster represented by phosphoserine, and argininosuccinic
167 acidnegatively associated with *Bacteroides coprophilus* (Supplementary Table 3i), a
168 prevalent but not very abundant species from the *Bacteroides* genus. MGSs from
169 *Faecalibacterium prausnitzii* (Supplementary Fig. 4a, Supplementary Fig. 3), a bacterium
170 reported to produce butyrate and metabolize arsenic [24], positively associated with
171 L−homocitrulline, phosphoserine, negatively associated with vitamin A, mercury, as well
172 as with specific TCR V(D)J including positive correlation with TRBV27_TRBJ2.3 and
173 TRBV27_TRBJ2.5 and negative correlation with TRBV20−1:TRBJ2−4 (Supplementary
174 Fig. 4a, Supplementary Table 3). The third cluster represented by glutamic acid showed
175 negative associations with previously reported bacteria implicated with lower BMI such

as *Alistipes shahii*, *Bacteroides cellulosilyticus*, *Ruminococcus lactis* and *Eubacterium eligens* [16] in this large cohort (Supplementary Fig. 4a), consistent with higher glutamic acid in individuals with obesity or insulin resistance [21,25], and here we tentatively identified their associated TCRs (Supplementary Fig. 4a).

Moreover, gut microbiome functional potential showed specific associations with TCR immune repertoire. The gut microbial module (GMM) [26] for homoacetogenesis (production of acetate from hydrogen and carbon dioxide) displayed widespread negative associations, most notably with TRBV7-8:TRBJ2-2 (Supplementary Fig. 4b). TRBV7-8 frequency had been reported to be higher in Pima Indian individuals with Type 2 diabetes [27] (Supplementary Table 3i). Modules for degradation of arginine and lysine, degradation of lactose and galactose, also associated with a number of VJs (Supplementary Fig. 4b). In the validation cohort, associations with fecal microbiome modules such as lysine degradation, mucin degradation, lactose and galactose degradation, sulfate reduction were validated (Supplementary Fig. 4b, Supplementary Table 3f), which was impressive given the differences in trace metals and other metabolites between the two cohorts (Fig. 3, Supplementary Table 1). So, from both taxonomic and functional points-of-view, the gut microbiome is involved in the metabolome-immune interplay in circulation, with important new leads for experimental investigations.

**Biomarkers for hyperuricemia and cardiometabolic diseases**

Hyperuricemia is common in the East Asian population, and urate is excreted in urine or through the gastrointestinal tract. In our cohort, serum uric acid showed negative correlations with gut bacteria such as *Faecalibacterium prausnitzii*, *Alistipes shahii*, *Oscillospiraceae* and *Bacteroides intestinalis* (Fig. 4), adjusted for medication and

dietary supplements. Moreover, serum uric acid positively correlated with vitamins (vitamin A, B5, D3 and E), amino acids (glutamic acid and alanine), trace elements (arsenic and mercury), while negatively associated with testosterone (Fig. 4). The negative associations between fecal *Butyricimonas virosa*, *Odoribacter splanchnicus* and plasma alanine were consistent with butyrate production from amino acids (Supplementary Table 3i)[28,29], which together with methylhistidines hinted at a meat-excess diet [30]. Self-reported dietary structure indeed showed association with serum uric acid (Supplementary Table 3j). This is the first set of large-scale evidence for gut microbiome dysbiosis in hyperuricemia, together with hormonal, metabolic and potentially immunological differences.

We next defined a score according to 8 routine blood parameters and 80 fecal microbiome features for cardiometabolic disease risk (see Methods) in this young cohort and tested it in previously published case-control samples. With the fecal markers alone, metagenomic samples from Chinese patients with atherosclerotic cardiovascular disease (ACVD), liver cirrhosis, obesity and Crohn's disease all scored higher compared to control samples without the disease ($P < 0.05$) (Supplementary Fig. 5a**)**, while those from diseases such as colorectal cancer, rheumatoid arthritis and medication-unstratified T2D did not (Supplementary Fig. 5a) [16,21,31–35]. The clinical parameters help clarified T2D and Crohn's disease (Supplementary Fig. 5b). Thus, although regional differences and misidentifications remain a concern, we illustrate the potential for population-wide screens of cardiometabolic diseases using the fecal microbiome.

**Biomarkers for colorectal cancer**

This young multi-omic cohort also provide more insight into the relationship between gut microbiome, plasma metabolome and colorectal cancer (CRC). Both the microbiome and the plasma metabolome are being actively studied for CRC biomarkers, but to our knowledge they have not been investigated in the same cohort. We see here that previously reported CRC-enriched bacteria [1,33,36,37] showed associations with plasma metabolites regardless of statistical adjustment for covariates (Supplementary Table 3a). *Peptostreptococcus stomatis* positively associated with plasma leucine, phenylalanine, alanine, tyrosine, as well as sarcosine, a metabolite studied for prostate cancer and a degradation intermediate of glycine betaine [38,39] (Supplementary Fig. 6). Enterobactericeae including *Escherichia coli, Klebsiella pneumoniae, Enterobacter cloacae* and *Citrobacter freundii* positively associated with sarcosine, hydroxylysine, branched chain amino acids, tyrosine, tryptophan, 1−methylhistidine, hydroxyproline, and argininosuccinic acid (Supplementary Fig. 6). 1-methylhistidine is a marker for habitual meat intake, especially red meat [30]. Bacteria such as *Bacteroides thetaiotaomicron*, *Butyricimonas virosa* were more associated with 3-methylhistidine (Supplementary Table 3a). Besides, the butyrate-producing *E. eligens* positively associated with fruit and vegetable intake, while negatively associated with plasma alanine (Fig. 4a, Supplementary Fig. 6). A number of these associations were also observed in the validation cohort, e.g. *Enterobacter cloacae* and hydroxylysine, *E. eligens* and alanine (Supplementary Table 3a). These results corroborate fecal markers of CRC with plasma metabolites, and suggest further studies on the long-term interplay between dietary metabolites and bacteria for CRC etiology and threshold for intervention.

**Physical fitness, exercising and sleeping reflected in the gut microbiome**

Vital capacity, a commonly used index to assess lung function, positively associated with bacteria such as *A. shahii, F. prausnitzii* and *Bifidobacterium adolescentis*, while negatively correlated with disease-related bacteria including *Clostridium clostridioforme*, *Ruminococcus gnavus* and *E. coli*, regardless of statistical adjustments (Fig. 2, Fig. 5, Supplementary Table 3e). Hand grip strength, a protective factor for cardiovascular casualty [40], negatively associated with *E. coli* (Fig. 5). Age and sex stratified vertical jump score (Supplementary Table 4) negatively associated with *E. coli*, while positively associated with *B. cellulosilyticus*, *B. intestinalis*, *Eubacterium rectale*, etc. *Bacteroides cellulosilyticus* and *B. stercorirosoris*, which associated with exercise intensity, even correlated with a faster reaction time (Fig. 5), reminding us with associations between *B. cellulosilyticus* and aldosterone, *B. stercorirosoris* and folic acid in both cohorts (Supplementary Table 3a). Moreover, gut microbiome diversity (Shannon index) associated with favorable scores in most of the fitness tests (Supplementary Table 2a).

Besides, individuals who stay up until after midnight also showed negative correlations with *Holdemania filiformis, Veillonella atypica* and 25-hydroxy vitamin D3/D, while positively correlated with *Clostridium hatheway*, *Clostridium phoceensis*, mercury, selenium, arsenic, vitamin A, hydroxyproline and phosphoserine (Supplementary Fig. 2a, Supplementary Fig. 5, Supplementary Table 3b). Thus, sleeping is also a factor to consider for a complete understanding of the gut microbiome.

**Species from yogurt in the healthy gut microbiome**

Besides defecation frequency and gender, yogurt consumption explained a notable portion of variances in the gut microbiome (Fig. 3, Supplementary Table 2b). A recent study casted doubt over the health benefits of probiotics, concluding that colonization of

the bacteria was highly variable between individuals [11]. In both our large cohorts, *Streptococcus thermophilus*, a species included in commercial yogurt mainly for its thermal stability and metabolic support for other strains, was consistently detected in yogurt eaters, and scaled with self-reported frequency of yogurt consumption (Fig. 6, Supplementary Fig. 7). *Bifidobacterium animalis*, likely representing the star strain from CHR HANSEN, *B. animalis* subsp. lactis BB-12, was also enriched in yogurt eaters, and fecal relative abundance of *B. animalis* associated with less stress, less bilirubin, lower diastolic blood pressure, as well as with TCR V(D)J combinations (Fig. 6d), suggesting immune modulation. The association between *B. animalis* and TRBV5.6:TRBJ2.5 was also observed in the validation cohort (Supplementary Table 3c), while the other parameters were unfortunately not available. In contrast to *S. thermophilus*, *B. animalis*, and *Veillonella*, there was no significant increase in any *Lactobacillus* strains (Fig. 6). Those who used to take yogurt also showed less *Clostridium bolteae*, a bacterium known to be elevated in a number of cardiometabolic diseases [1,16]. Intriguingly, fecal *C. bolteae* associated with plasma triglyceride, uric acid, phosphoserine, vitamin A, and mercury (Fig. 6e), offering an explanation for epidemiological evidence of yogurt consumption and reduced risk of gout [41]. In the validation cohort, *C. bolteae* also associated with mercury and to a lesser extent vitamin A (the vitamin A association was sensitive to covariates, Supplementary Table 3a). Besides, yogurt consumption was associated with a number of favorable measurements such as higher HDL (high-density lipoprotein) cholesterols, lower uric acid and triglycerides, less cysteine, mercury and hydroxyproline (Fig. 6a).

Regarding *Bifidobacterium* in the gut microbiome, however, individuals who consumed milk enriched for *B. longum, B. catenulatum* and *B. pseudocatenulatum* (Fig. 6a, b), implying that some of the yogurt-associated differences come from its exogenous strains such as *S. thermophilus* and *B. animalis*, as well as less *C. bolteae*. The higher *Bifidobacterium* spp., and lower *Blautia wexlerae* and *Ruminococcus* sp. 5_1_39BFAA associated with milk intake were validated in the additional 1404 individuals (Supplementary Table 3). Milk drinking also associated with vitamin B2, B5, B6, HDL, lymphocytes, etc. in the blood, vital capacity, and psychological scores (Fig. 6a, b, Supplementary Fig. 7).

## Discussion

### Insights from multi-omics

In summary, our trans-omic investigation of thousands volunteers establish an unprecedented reference data set for the human gut microbiome. Judging from the associations, it appears as though a number of factors in circulation crosstalk with the gut microbiome, and then manifest on the face, in the head and in fitness tests. Levels of trace elements, such as mercury, arsenic and selenium, as important cofactors for bacteria respiration and other functions [23], should be measured even in uncontaminated regions, and in individuals showing normal levels of these elements. Although rice is often studied for such contaminants, exposure can be from other food, drink, air and soil sources[42](https://www.fda.gov/food/metals/arsenic-food-and-dietary-supplements). Our results suggest that commensal microbial metabolism of trace elements might help determine their levels in the blood, and influence immune functions.

The PBMC TCRβ CDR3 V(D)J usage in such a large cohort is a great resource for discovering microbial antigens other than those from traditional pathogens. While some *TRBV* and *TRBJ* segments are more frequent than others [43], we do not yet know how they correspond to T cell sub-populations. Existing studies on TCR have been focusing on pathogens, autoimmune diseases and cancer. For example, TCR profiles of tumor-resident $T_{reg}$ (regulatory T) cells have been shown to significantly overlap with those of circulating $T_{reg}$ cells [44]; immune phenotype of peripheral blood $T_{reg}$ II cells was not only similar to that of intratumoral $T_{reg}$ cells, but also predicted future relapse of breast cancer patients [45]. A high diversity in the T cell immune repertoire is believed to be preferable, but the T cell immune repertoire diversity has been reported to be unchanged after a 3-month switch from omnivorous to vegetarian and lower in long-term vegetarians [46]. In our analyses of this cohort, the overall diversity (Immunity index, Methods) was not the most important factor that predicted other omics, yet could be reflected by metabolites, physical fitness tests, lifestyle and skin features (RFCV R ~ 0.2, Fig. 2,3). We identify clustering patterns of specific TCRβ CDR3 VJ joining with plasma metabolites including vitamins, trace elements and amino acids (Supplementary Fig. 4a). The chains of causality remain to be fully elucidated; yet, it is likely to be a two-way interplay for metabolite-gut microbiome, metabolite-T cells, and gut microbiome-T cells. Our results imply long-term differences in these features in apparently healthy individuals. A similar speculation could be made for facial skin features, which we expect to be resilient against topical interventions judging from the strong associations with blood parameters.

We have tentatively identified gut bacteria associated with each ABO blood type. A larger proportion of blood type A in Europeans compared to East Asians might help

explain the greater abundance of Lachospiraceae bacterium [12,47]. Blood type B is more prevalent in northern Chinese, and the blood type B-enriched mucin-degrading bacterium *R. torques* has recently been reported to show an association with blood glucose [48] and was also associated with ulcerative colitis [49] and a Bristol stool score of 1 or 2 (Supplementary Fig. 2a). *Megaspaera micronuciformis*, seen in association with blood type O, can produce butyrate from acetate [50]. Genetic studies of the gut microbiota have not yet reported genome-wide significant associations with ABO blood type genes themselves [51–54], while multiple studies have reported impact of *FUT2* secretor/non-secretor status on gut microbiota composition [55–57]. Tentative associations here are yet to be matched with *in vitro* studies with the glycans [58,59].

Differences in gut microbiome composition between sexes and a greater microbial diversity in females have recently been reported in the LifeLines Deep cohort, yet the gut microbiome in females was influenced by oral contraceptives, ovariectomy as well as antibiotics for vaginal or pelvic infections [60]. Males of Hadza hunter-gatherers showed differences in gut microbiota compared to females [61], including higher *Eubacterium* and *Blautia* in men which were also recapitulated in our Chinese cohort (*E. dolichum, B. wexlerae*). Interestingly, *E. dolichum* associated with a dietary structure of more meat instead of fruit and vegetables, while *B. wexlerae* scaled negatively with milk consumption (Supplementary Fig. 2, Fig. 6). The evolutionary implications remain unclear.

**A baseline of the gut microbiome with deviations towards diseases**

Metagenome-wide association studies (MWAS) have documented gut microbial perturbations in a growing list of diseases by comparing cases versus controls. Here we

provide a high-depth metagenomic cohort, the mean age for which did not exceed 30 years old. Alarmingly enough, trends for cardiometabolic diseases and colorectal cancer can already been seen from the fecal microbiome and a few parameters in the blood. The set of healthy gut microbes for leanness are increasingly clear [16], such as *A. shahii*, *F. prausnitzii*, *E. eligens* and *B. cellulosilyticus*. And we have a better idea how to increase their relative abundances. Interestingly, we observed few association with *Akkermansia*, which may indeed be too diverse among individuals [5,62] or require mucosal sampling. The list of potentially harmful gut microbes are also increasingly clear; future studies are needed to confirm whether we can decrease *E. coli* and *R. gnavus* with exercising and diet, fend off *C. boltae* with yogurt, etc.

While an older cohort would be needed to look at type 2 diabetes [63,64], hyperuricemia is common in this cohort (Supplementary Table 1). *A. shahii* negatively associated with plasma tryptophan (Fig. 4, Supplementary Table 3i), and hyperuricemia has been reported to skew tryptophan metabolism towards kynurenine production in mice models [65], instead of indole reported for *A. shahii* [66], potentially modulating signaling through aryl hydrocarbon receptors (AhR) [67]. One of the bacteria negatively associated with serum uric acid, *F. prausnitzii*, has been reported to encode a methyltransferase for arsenic detoxification (Supplementary Table 3i) [24]. IL-1β, the major cytokine responsible for gout [68], has been associated with urinary level of arsenic [69]. Co-stimulation of patient-derived PBMCs with monosodiurm urate crystals and TLR2 or TLR4 (toll-like receptors) ligands have been shown to disrupt IL-1β/ IL-1Ra (IL1 receptor antagonist) balance [70], consistent with involvement of microbes in gout.

Genetic potential for histidine degradation instead of synthesis have been observed to increase in CRC relative to healthy controls according to metagenomic studies [37,71]. 1-methylhistidine, a marker for habitual meat intake [30], could be metabolized into histidine. Plasma level of the amino acid proline was reported to increase in a mouse model of CRC [72], but found in another study to decrease in human CRC [73]. In this young cohort from China, we did not see significant associations between proline and known gut microbiome markers of CRC. Hydroxyproline, on the other hand, is better predicted by the gut microbiome composition compared to proline (Fig. 3), and associated with meat consumption, staying up until after 0 am (Supplementary Fig. 7). Enterobacteriaceae such as *Escherichia coli* and *Klebsiella pneumoniae* positively associated with hydroxyproline in this cohort. A recent study analyzed fecal metabolites together with fecal microbiome and reported among others an increase in branched chain amino acids and aromatic amino acids in CRC [74]. Here we observe plasma levels of these amino acids to associate with CRC markers such as *P. stomatis*, and *E. coli*, while the fecal metagenomic potential for leucine biosynthesis was control-enriched [37,74], implying that leucine was normally not in excess.

We also find it intriguing that decarboxylases appear generally important for bacterial stress response in the microbiome, i.e. to maintain a balanced pH for themselves. The top one for gut microbes may be glutamate decarboxylase (produces GABA (γ-aminobutyric acid) from glutamate), while histidine decarboxylase in the female reproductive tract might contribute to menstrual pains [75]. Besides, recent studies identified tyrosine decarboxylases in gut microbes that could digest the medication levodopa used to treat Parkinson's disease [76,77].

**Behavioral changes to be trialed for a healthy gut microbiome?**

Although effects of sleep fragmentation on hemopoiesis have been seen despite antibiotic treatment [78], our results nonetheless suggest that the gut microbiome may have an additional role, together with trace elements, vitamins, and host genetics [79]. The less hypocretin in mice subjected to sleep fragmentation promoted atherosclerosis [78]. The increased adiposity and decreased lean mass with sleep loss also involved toll-like receptors (TLRs) [80,81], and we identify cardiometabolic disease-associated species including *Clostridium hatheway* here.

Potential influence of physical activity on the gut microbiota has been analyzed in small cohorts of rugby athletes [82] and colorectal cancer [37]. Although more detailed information for physical activity is preferable, compliance to recordings such as Fitbit is notoriously bad in healthy individuals [83]. Results from this large cohort at least suggest that exercising might help improve cardio-pulmonary function (grip strength, vital capacity) to decrease incidence of cardiometabolic diseases. Intense exercise, explored for application to individuals with diseases such as prediabetics and Alzheimer's [84,85], may be no less important than endurance or resistance training; and our results suggest that different types of exercise could have differential impacts on the gut microbiome and the microbiome changes could be a readout for monitoring effects of training. Endurance training actually lowers testosterone [86] and could lead to hyperuricemia, especially if combined with high-fructose food and drinks and lack of dairy consumption [41].

Our large-scale analyses provide substantial support for health benefits of yogurt consumption. The universally present species were *Streptococcus thermophilus* and *Bifidobacterium animalis* instead of commonly tested probiotics from *Lactobacillus*. An

orally administered strain of *B. longum* has been shown to persist in 30% of individuals for at least 6 months [87], while we failed to detect in feces an *L. casei* strain gavaged to rats [88,89], suggesting general differences between *Bifidobacterium* and *Lactobacillus*. The strains used by Zmora et al. included a number of *Lactobacillus*, *Bifidobacterium*, as well as *Streptococcus* and *Lactococcus*, all detectable in various gastrointestinal sites despite laxative and colonoscopy [11]. One potential explanation for the association with desirable cardiometabolic and psychological scores observed in our study for yogurt or milk is the production of metabolites such as folate and GABA by *S. thermophilus*, *Bifidobacterium* and *Lactobacillus* [90,91]. Moreover, Lactobacilli have been reported to sequester heavy metals including lead and cadmium [92]. All of these live or dead probiotics could potentially exert functions on the immune system or even the brain. The positive association with endogenous *Bifidobacterium* species with milk intake is more likely due to live bacteria which help metabolize the lactose in this largely lactose-intolerant population. It remains to be seen whether and how diary consumption affects the gut microbiome in other cohorts, and there appears to be regional differences in China already.

Thus, this study provides a young reference for the gut microbiome with physical fitness test and questionnaire data, and reveals interrelationship with other omics such as trace elements, hormones and immune repertoire that have so far not been included in other study designs. There is a lot more to investigate both in vitro and in vivo by researchers across disciplines. Interventional as well as mechanistic studies will be needed to see how physical activity, well-timed sleeping and dietary interventions such as yogurt, milk and

vegetables might improve the gut microbiome, hormone levels, cardiometabolic and mental health.

**Data and materials availability:** Metagenomic sequencing data for all samples have been deposited to the CNSA (https://db.cngb.org/cnsa/) of (CNGB) database under the accession code CNP00 00426, CNP0000289.

**Acknowledgments**

The authors are very grateful to colleagues at BGI-Shenzhen and China National Genebank (CNGB), Shenzhen for sample collection, DNA extraction, library construction, sequencing, and discussions. We thank Dr. Qiang Sun (University of Toronto), our colleagues Chen Chen and Yanmei Ju, Jinghua Wu for helpful comments.

**Author contributions:**

J.W. initiated the overall health project. Y.Z., H.Z., K.C., P.C., X.X. organized the sample collection and processing, with immune repertoire from X.L., W.Z, metabolomics from X.Q., Q.L., Y.L., D.Z., H.Lian, Y.Z., X.C., W.R., Y.R., Y.W., J.Z., R.W., raw metagenomic profile from Q.D, X.W., and J.Z., Q.D., S.T., Y.L., D.W. checked the host metadata and matched the omics data. H.Zhou, H.Lu led the DNA extraction and sequencing, respectively. Z.J. led the bioinformatic analyses, including S.L. and F.L. H.Zhong, Q.D., S.T., D.W. performed early analyses which are not included in the

manuscript. Z.J., H.J. interpreted the data. H.J., S.L., Z.J., F.L. wrote the manuscript and rendered the display items. All authors contributed to finalizing this manuscript.

**Competing interests:** The authors declare no competing financial interest.

**References:**

1. Wang, J. & Jia, H. Metagenome-wide association studies: fine-mining the microbiome. *Nat. Rev. Microbiol.* **14**, 508–522 (2016).
2. He, Y. *et al.* Regional variation limits applications of healthy gut microbiome reference ranges and disease models. *Nat. Med.* **24**, 1532–1535 (2018).
3. Deschasaux, M. *et al.* Depicting the composition of gut microbiota in a population with varied ethnic origins but shared geography. *Nat. Med.* (2018) doi:10.1038/s41591-018-0160-1.
4. Zhernakova, A. *et al.* Population-based metagenomics analysis reveals markers for gut microbiome composition and diversity. *Science* **352**, 565–569 (2016).
5. Xie, H. *et al.* Shotgun Metagenomics of 250 Adult Twins Reveals Genetic and Environmental Impacts on the Gut Microbiome. *Cell Syst.* **3**, 572-584.e3 (2016).
6. Rothschild, D. *et al.* Environment dominates over host genetics in shaping human gut microbiota. *Nature* **555**, 210–215 (2018).
7. Mehta, R. S. *et al.* Stability of the human faecal microbiome in a cohort of adult men. *Nat. Microbiol.* **3**, 347–355 (2018).
8. Xiao, L. *et al.* A catalog of the mouse gut metagenome. *Nat. Biotechnol.* **33**, 1103–1108 (2015).
9. Xiao, L. *et al.* A reference gene catalogue of the pig gut microbiome. *Nat. Microbiol.* **1**, 16161 (2016).
10. Zhao, L. *et al.* Quantitative genetic background of the host influences gut microbiomes in chickens. *Sci. Rep.* **3**, 1163 (2013).
11. Zmora, N. *et al.* Personalized Gut Mucosal Colonization Resistance to Empiric Probiotics Is Associated with Unique Host and Microbiome Features. *Cell* **174**, 1388-1405.e21 (2018).
12. Li, J. *et al.* An integrated catalog of reference genes in the human gut microbiome. *Nat. Biotechnol.* **32**, 834–841 (2014).
13. Nielsen, H. B. *et al.* Identification and assembly of genomes and genetic elements in complex metagenomic samples without using reference genomes. *Nat. Biotechnol.* **32**, 822–828 (2014).
14. Moll1, J. M. *et al.* Pathway-based approach to classify unannotated gut bacteria associated with insulin sensitivity. *Revis. Nat Microbiol* (2019).

15. Falony, G. *et al.* Population-level analysis of gut microbiome variation. *Science* **352**, 560–564 (2016).
16. Jie, Z. *et al.* The gut microbiome in atherosclerotic cardiovascular disease. *Nat. Commun.* **8**, 845 (2017).
17. Wilck, N. *et al.* Salt-responsive gut commensal modulates TH17 axis and disease. *Nature* **551**, 585–589 (2017).
18. Veldhoen, M. & Ferreira, C. Influence of nutrient-derived metabolites on lymphocyte immunity. *Nat. Med.* **21**, 709–718 (2015).
19. Grizotte-Lake, M. *et al.* Commensals Suppress Intestinal Epithelial Cell Retinoic Acid Synthesis to Regulate Interleukin-22 Activity and Prevent Microbial Dysbiosis. *Immunity* **49**, 1103-1115.e6 (2018).
20. Lawson, P. A. *et al.* Reclassification of Eubacterium formicigenerans Holdeman andMoore 1974 as Dorea formicigenerans gen. nov., comb. nov., and description of Dorea longicatena sp. nov., isolated from human faeces. *Int. J. Syst. Evol. Microbiol.* **52**, 423–428 (2002).
21. Liu, R. *et al.* Gut microbiome and serum metabolome alterations in obesity and after weight-loss intervention. *Nat. Med.* **23**, 859–868 (2017).
22. Beaumont, M. *et al.* Heritable components of the human fecal microbiome are associated with visceral fat. *Genome Biol.* **17**, 189 (2016).
23. Stolz, J. F. & Oremland, R. S. Bacterial respiration of arsenic and selenium. *FEMS Microbiol. Rev.* **23**, 615–27 (1999).
24. Coryell, M., McAlpine, M., Pinkham, N. V., McDermott, T. R. & Walk, S. T. The gut microbiome is required for full protection against acute arsenic toxicity in mouse models. *Nat. Commun.* **9**, 5424 (2018).
25. Cheng, S. *et al.* Metabolite profiling identifies pathways associated with metabolic risk in humans. *Circulation* **125**, 2222–31 (2012).
26. Vieira-Silva, S. *et al.* Species–function relationships shape ecological properties of the human gut microbiome. *Nat. Microbiol.* **29**, 16088 (2016).
27. Frankl, J. A., Thearle, M. S., Desmarais, C., Bogardus, C. & Krakoff, J. T-cell receptor repertoire variation may be associated with type 2 diabetes mellitus in humans. *Diabetes Metab. Res. Rev.* **32**, 297–307 (2016).
28. Sakamoto, M. *et al.* Butyricimonas synergistica gen. nov., sp. nov. and Butyricimonas virosa sp. nov., butyric acid-producing bacteria in the family 'Porphyromonadaceae' isolated from rat faeces. *Int. J. Syst. Evol. Microbiol.* **59**, 1748–1753 (2009).
29. Vital, M., Howe, A. C. & Tiedje, J. M. Revealing the bacterial butyrate synthesis pathways by analyzing (meta)genomic data. *mBio* **5**, e00889 (2014).
30. Mitry, P. *et al.* Plasma concentrations of anserine, carnosine and pi-methylhistidine as biomarkers of habitual meat consumption. *Eur. J. Clin. Nutr.* **73**, 692–702 (2019).
31. He, Q. *et al.* Two distinct metacommunities characterize the gut microbiota in Crohn's disease patients. *GigaScience* **6**, 1–11 (2017).
32. Qin, N. *et al.* Alterations of the human gut microbiome in liver cirrhosis. *Nature* **513**, 59–64 (2014).
33. Yu, J. *et al.* Metagenomic analysis of faecal microbiome as a tool towards targeted non-invasive biomarkers for colorectal cancer. *Gut* **66**, 70–78 (2017).

552 34. Zhang, X. *et al.* The oral and gut microbiomes are perturbed in rheumatoid
553 arthritis and partly normalized after treatment. *Nat. Med.* **21**, 895–905 (2015).
554 35. Qin, J. *et al.* A metagenome-wide association study of gut microbiota in type 2
555 diabetes. *Nature* **490**, 55–60 (2012).
556 36. Zeller, G. *et al.* Potential of fecal microbiota for early-stage detection of colorectal
557 cancer. *Mol. Syst. Biol.* **10**, 766 (2014).
558 37. Feng, Q. *et al.* Gut microbiome development along the colorectal adenoma–
559 carcinoma sequence. *Nat. Commun.* **6**, 6528 (2015).
560 38. Sreekumar, A. *et al.* Metabolomic profiles delineate potential role for sarcosine in
561 prostate cancer progression. *Nature* **457**, 910–914 (2009).
562 39. Hai, Y., Huang, A. M. & Tang, Y. Structure-guided function discovery of an
563 NRPS-like glycine betaine reductase for choline biosynthesis in fungi. *Proc. Natl. Acad.*
564 *Sci.* **116**, 10348–10353 (2019).
565 40. Tikkanen, E., Gustafsson, S. & Ingelsson, E. Associations of Fitness, Physical
566 Activity, Strength, and Genetic Risk With Cardiovascular Disease: Longitudinal
567 Analyses in the UK Biobank Study. *Circulation* **137**, 2583–2591 (2018).
568 41. Kuo, C.-F., Grainge, M. J., Zhang, W. & Doherty, M. Global epidemiology of
569 gout: prevalence, incidence and risk factors. *Nat. Rev. Rheumatol.* **11**, 649–662 (2015).
570 42. Chen, L. *et al.* Trans-provincial health impacts of atmospheric mercury emissions
571 in China. *Nat. Commun.* **10**, 1484 (2019).
572 43. Freeman, J. D., Warren, R. L., Webb, J. R., Nelson, B. H. & Holt, R. A. Profiling
573 the T-cell receptor beta-chain repertoire by massively parallel sequencing. *Genome Res.*
574 **19**, 1817–1824 (2009).
575 44. Ahmadzadeh, M. *et al.* Tumor-infiltrating human CD4 + regulatory T cells
576 display a distinct TCR repertoire and exhibit tumor and neoantigen reactivity. *Sci.*
577 *Immunol.* **4**, eaao4310 (2019).
578 45. Wang, L. *et al.* Connecting blood and intratumoral Treg cell activity in predicting
579 future relapse in breast cancer. *Nat. Immunol.* (2019) doi:10.1038/s41590-019-0429-7.
580 46. Zhang, C. *et al.* Impact of a 3-Months Vegetarian Diet on the Gut Microbiota and
581 Immune Repertoire. *Front. Immunol.* **9**, (2018).
582 47. Davenport, E. R. *et al.* ABO antigen and secretor statuses are not associated with
583 gut microbiota composition in 1,500 twins. *BMC Genomics* **17**, 941 (2016).
584 48. Zeevi, D. *et al.* Structural variation in the gut microbiome associates with host
585 health. *Nature* (2019) doi:10.1038/s41586-019-1065-y.
586 49. Png, C. W. *et al.* Mucolytic bacteria with increased prevalence in IBD mucosa
587 augment in vitro utilization of mucin by other bacteria. *Am. J. Gastroenterol.* **105**, 2420–
588 8 (2010).
589 50. Marchandin, H. *et al.* Phylogenetic analysis of some Sporomusa sub-branch
590 members isolated from human clinical specimens: description of Megasphaera
591 micronuciformis sp. nov. *Int. J. Syst. Evol. Microbiol.* **53**, 547–53 (2003).
592 51. Wang, J. *et al.* Genome-wide association analysis identifies variation in vitamin D
593 receptor and other host factors influencing the gut microbiota. *Nat. Genet.* **48**, 1396–1406
594 (2016).
595 52. Bonder, M. J. *et al.* The effect of host genetics on the gut microbiome. *Nat. Genet.*
596 **48**, 1407–1412 (2016).

53. Turpin, W. *et al.* Association of host genome with intestinal microbial composition in a large healthy cohort. *Nat. Genet.* (2016) doi:10.1038/ng.3693.
54. Liu, X. *et al.* M-GWAS for the Gut Microbiome in Chinese Adults Illuminates on Complex Diseases. *Cell* (2019) doi:10.2139/ssrn3383800.
55. Rausch, P. *et al.* Colonic mucosa-associated microbiota is influenced by an interaction of Crohn disease and FUT2 (Secretor) genotype. *Proc. Natl. Acad. Sci.* **108**, 19030–19035 (2011).
56. Rausch, P. *et al.* Multigenerational Influences of the Fut2 Gene on the Dynamics of the Gut Microbiota in Mice. *Front. Microbiol.* **8**, 991 (2017).
57. Zhernakova, D. V. *et al.* Individual variations in cardiovascular-disease-related protein levels are driven by genetics and gut microbiome. *Nat. Genet.* (2018) doi:10.1038/s41588-018-0224-7.
58. Liu, Q. P. *et al.* Bacterial glycosidases for the production of universal red blood cells. *Nat. Biotechnol.* **25**, 454–464 (2007).
59. Rahfeld, P. *et al.* An enzymatic pathway in the human gut microbiome that converts A to universal O type blood. *Nat. Microbiol.* (2019) doi:10.1038/s41564-019-0469-7.
60. Sinha, T. *et al.* Analysis of 1135 gut metagenomes identifies sex-specific resistome profiles. *Gut Microbes* 1–9 (2018) doi:10.1080/19490976.2018.1528822.
61. Schnorr, S. L. *et al.* Gut microbiome of the Hadza hunter-gatherers. *Nat. Commun.* **5**, 3654 (2014).
62. Schloissnig, S. *et al.* Genomic variation landscape of the human gut microbiome. *Nature* **493**, 45–50 (2012).
63. Schüssler-Fiorenza Rose, S. M. *et al.* A longitudinal big data approach for precision health. *Nat. Med.* **25**, 792–804 (2019).
64. Zhou, W. *et al.* Longitudinal multi-omics of host–microbe dynamics in prediabetes. *Nature* **569**, 663–671 (2019).
65. Dankers, A. C. A. *et al.* Hyperuricemia influences tryptophan metabolism via inhibition of multidrug resistance protein 4 (MRP4) and breast cancer resistance protein (BCRP). *Biochim. Biophys. Acta* **1832**, 1715–22 (2013).
66. Song, Y. *et al.* Alistipes onderdonkii sp. nov. and Alistipes shahii sp. nov., of human origin. *Int. J. Syst. Evol. Microbiol.* **56**, 1985–90 (2006).
67. Cervenka, I., Agudelo, L. Z. & Ruas, J. L. Kynurenines: Tryptophan's metabolites in exercise, inflammation, and mental health. *Science* **357**, eaaf9794 (2017).
68. So, A. K. & Martinon, F. Inflammation in gout: mechanisms and therapeutic targets. *Nat. Rev. Rheumatol.* (2017) doi:10.1038/nrrheum.2017.155.
69. Parvez, F. *et al.* Assessment of arsenic and polycyclic aromatic hydrocarbon (PAH) exposures on immune function among males in Bangladesh. *PloS One* **14**, e0216662 (2019).
70. Cri an, T. O. *et al.* Soluble uric acid primes TLR-induced proinflammatory cytokine production by human primary cells via inhibition of IL-1Ra. *Ann. Rheum. Dis.* **75**, 755–62 (2016).
71. Thomas, A. M. *et al.* Metagenomic analysis of colorectal cancer datasets identifies cross-cohort microbial diagnostic signatures and a link with choline degradation. *Nat. Med.* **25**, 667–678 (2019).

642 72. Manna, S. K. *et al.* Biomarkers of Coordinate Metabolic Reprogramming in
643 Colorectal Tumors in Mice and Humans. *Gastroenterology* **146**, 1313–1324 (2014).
644 73. Qiu, Y. *et al.* Serum metabolite profiling of human colorectal cancer using GC-
645 TOFMS and UPLC-QTOFMS. *J Proteome Res* **8**, 4844–4850 (2009).
646 74. Yachida, S. *et al.* Metagenomic and metabolomic analyses reveal distinct stage-
647 specific phenotypes of the gut microbiota in colorectal cancer. *Nat. Med.* (2019)
648 doi:10.1038/s41591-019-0458-7.
649 75. Jie, Z. *et al.* The vagino-cervical microbiome as a woman's life history. *Revis.*
650 *Cell* (2019) doi:10.1101/533588.
651 76. van Kessel, S. P. *et al.* Gut bacterial tyrosine decarboxylases restrict levels of
652 levodopa in the treatment of Parkinson's disease. *Nat. Commun.* **10**, 310 (2019).
653 77. Maini Rekdal, V., Bess, E. N., Bisanz, J. E., Turnbaugh, P. J. & Balskus, E. P.
654 Discovery and inhibition of an interspecies gut bacterial pathway for Levodopa
655 metabolism. *Science* **364**, eaau6323 (2019).
656 78. McAlpine, C. S. *et al.* Sleep modulates haematopoiesis and protects against
657 atherosclerosis. *Nature* (2019) doi:10.1038/s41586-019-0948-2.
658 79. Lane, J. M. *et al.* Genome-wide association analyses of sleep disturbance traits
659 identify new loci and highlight shared genetics with neuropsychiatric and metabolic traits.
660 *Nat. Genet.* **49**, 274–281 (2016).
661 80. Cappuccio, F. P. *et al.* Meta-analysis of short sleep duration and obesity in
662 children and adults. *Sleep* **31**, 619–26 (2008).
663 81. Cedernaes, J. *et al.* Acute sleep loss results in tissue-specific alterations in
664 genome-wide DNA methylation state and metabolic fuel utilization in humans. *Sci. Adv.*
665 **4**, eaar8590 (2018).
666 82. Barton, W. *et al.* The microbiome of professional athletes differs from that of
667 more sedentary subjects in composition and particularly at the functional metabolic level.
668 *Gut* gutjnl-2016-313627 (2017) doi:10.1136/gutjnl-2016-313627.
669 83. Price, N. D. *et al.* A wellness study of 108 individuals using personal, dense,
670 dynamic data clouds. *Nat. Biotechnol.* **35**, 747–756 (2017).
671 84. Jensen, C. S. *et al.* Exercise as a potential modulator of inflammation in patients
672 with Alzheimer's disease measured in cerebrospinal fluid and plasma. *Exp. Gerontol.* **121**,
673 91–98 (2019).
674 85. da Silva, D. E. *et al.* High-Intensity Interval Training in Patients with Type 2
675 Diabetes Mellitus: a Systematic Review. *Curr. Atheroscler. Rep.* **21**, 8 (2019).
676 86. Popovic, B. *et al.* Acute Response to Endurance Exercise Stress: Focus on
677 Catabolic/anabolic Interplay Between Cortisol, Testosterone, and Sex Hormone Binding
678 Globulin in Professional Athletes. *J. Med. Biochem.* **38**, 6–12 (2019).
679 87. Maldonado-Gómez, M. X. *et al.* Stable Engraftment of Bifidobacterium longum
680 AH1206 in the Human Gut Depends on Individualized Features of the Resident
681 Microbiome. *Cell Host Microbe* **20**, 515–526 (2016).
682 88. Pan, H. *et al.* A gene catalogue of the Sprague-Dawley rat gut metagenome.
683 *GigaScience* **7**, (2018).
684 89. Pan, H. *et al.* A single bacterium resurrects the microbiome-immune balance to
685 protect bones from destruction in a rat model of rheumatoid arthritis. *Microbiome* **7**, 107
686 (2019).

90. Yatsunenko, T. *et al.* Human gut microbiome viewed across age and geography. *Nature* **486**, 222–7 (2012).
91. Valenzuela, J. A., Flórez, A. B., Vázquez, L., Vasek, O. M. & Mayo, B. Production of γ-aminobutyric acid (GABA) by lactic acid bacteria strains isolated from traditional, starter-free dairy products made of raw milk. *Benef. Microbes* **10**, 579–587 (2019).
92. Daisley, B. A. *et al.* Immobilization of cadmium and lead by Lactobacillus rhamnosus GR-1 mitigates apical-to-basolateral heavy metal translocation in a Caco-2 model of the intestinal epithelium. *Gut Microbes* 1–13 (2018) doi:10.1080/19490976.2018.1526581.
93. Han, M. *et al.* A novel affordable reagent for room temperature storage and transport of fecal samples for metagenomic analyses. *Microbiome* **6**, 43 (2018).
94. Fang, C. *et al.* Assessment of the cPAS-based BGISEQ-500 platform for metagenomic sequencing. *GigaScience* **7**, 1–8 (2018).
95. Zhang, W. *et al.* IMonitor: A Robust Pipeline for TCR and BCR Repertoire Analysis. *Genetics* **201**, 459–472 (2015).
96. Marbach, D. *et al.* Wisdom of crowds for robust gene network inference. *Nat. Methods* **9**, 796–804 (2012).
97. Louppe, G., Wehenkel, L., Sutera, A. & Geurts, P. Understanding variable importances in forests of randomized trees. in *Advances in Neural Information Processing Systems 26* (eds. Burges, C. J. C., Bottou, L., Welling, M., Ghahramani, Z. & Weinberger, K. Q.) 431–439 (Curran Associates, Inc., 2013).
98. Morgan, X. C. *et al.* Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. *Genome Biol.* **13**, R79 (2012).

720

721

722

723

724

725 **Online Methods：**

726 **Study Cohort**

727 As part of 4D-SZ, all the >2000 volunteers for the first cohort were recruited
728 between May 2017 and July 2017 during a physical examination. The 1400 volunteers for
729 the second cohort were also recruited in 2017, with no overlaps. The samples in each
730 omics are shown in Supplementary Table 1c. Baseline characteristics of the cohort are
731 shown in Supplementary Table 1b, 1d.

732 The study was approved by the Institutional Review Boards (IRB) at BGI-Shenzhen,
733 and all participants provided written informed consent at enrolment.

734

735 **Demographic Data Collection**

736 The lifestyle questionnaire contained 56 entries involving age, marital status, disease
737 history of the volunteer and his/her family, eating and exercise habits (Supplementary
738 Table 1b, 1d). The psychological questionnaire contained 18 entries for the evaluation of
739 irritability, dizziness, frustration, fear, appetite, self-confidence, resilience
740 (Supplementary Table 1b).

741

742 **Samples Collection**

743 Fecal samples were self-collected by volunteers, using a kit containing a room
744 temperature stabilizing reagent to preserve the metagenome[93]. The samples were frozen
745 on the same day. The overnight fasting blood samples were drawn from a cubital vein of
746 volunteers by the doctors.

747

748 **DNA extraction and metagenomics shotgun sequencing**

749 DNA extraction of the stored fecal samples within the next few months was
750 performed as previously described (Qin et al., 2012). Metagenomic sequencing was done
751 on the BGISEQ-500 platform (100bp of singled-end reads for fecal samples and four
752 libraries were constructed for each lane) [94].

753

754 **Amino Acid Measurements**

755 40 µl plasma was deproteinized with 20 µl 10% (w/v) sulfosalicylic acid (Sigma)
756 containing internal standards, then 120 µl aqueous solution was added. After centrifuged,
757 the supernatant was used for analysis. The analysis was performed by ultra high pressure
758 liquid chromatography (UHPLC) coupled to an AB Sciex Qtrap 5500 mass spectrometry
759 (AB Sciex, US) with the electrospray ionization (ESI) source in positive ion mode. A
760 Waters ACQUITY UPLC HSS T3 column (1.8 µm, 2.1 × 100 mm) was used for amino
761 compound separation with a flow rate at 0.5 ml/min and column temperature of 55 °C.
762 The mobile phases were (A) water containing 0.05% and 0.1% formic acid (v/v), (B)
763 acetonitrile containing 0.05% and 0.1% formic acid (v/v). The gradient elution was 2% B
764 kept for 0.5 min, then changed linearly to 10% B during 1 min, continued up to 35% B in

2 min, increased to 95% B in 0.1 min and maintained for 1.4 min. Multiple Reaction Monitoring (MRM) was used to monitor all amino compounds. The mass parameters were as follows, Curtain gas flow 35 L/min, Collision Gas (CAD) was medium, Ion Source Gas 1 (GS 1) flow 60 l/min, Ion Source Gas 2 (GS 2) flow 60 l/min, IonSpray Voltage (IS) 5500V, temperature 600 °C. All amino compound standards were purchased from sigma and Toronto research chemical (TRC).

**Hormone Measurements**

250 µl plasma was diluted with 205 µl aqueous solution, For SPE experiments, HLB (Waters, USA) was activated with 1.0 ml of dichloromethane, acetonitrile, methanol, respectively and was equilibrated with 1.0 ml of water. The pretreated plasma sample was loaded onto the cartridge and was extracted using gravity. Clean up was accomplished by washing the cartridges with 1.0 ml of 25% methanol in water. After drying under vacuum, samples on the cartridges were eluted with 1.0 ml of dichloromethane. The eluted extract was dried under nitrogen and the residual was dissolved with 25% methanol in water and was transferred to an autosampler vial prior to LC–MS/MS analysis. The analysis was performed by UHPLC coupled to an AB Sciex Qtrap 5500 mass spectrometry (AB Sciex, US) with the atmospheric pressure chemical ionization (APCI) source in positive ion mode. A Phenomone Kinetex C18 column (2.6 µm, 2.1 × 50 mm) was used for steroid hormone separation with a flow rate at 0.8 ml/min and column temperature of 55 °C. The mobile phases were (A) water containing 1mM Ammonium acetate, (B) Methanol containing 1mM Ammonium acetate. The gradient elution was 25% B kept for 0.9min, then changed linearly to 40% B during 0.9min,

continued up to 70% B in 2 min, increased to 95% B in 0.1 min and maintained for 1.6 min. Multiple Reaction Monitoring (MRM) was used to monitor all steroid hormone compounds. The mass parameters were as follows, Curtain gas flow 35 l/min, Collision Gas (CAD) was medium, Ion Source Gas 1 (GS 1) flow 60 l/min, Ion Source Gas 2 (GS 2) flow 60 l/min, Nebulizer Current (NC) 5, temperature 500 °C. All steroid hormone profiling compound standards were purchased from sigma, Toronto research chemical (TRC), Cerilliant and DR. Ehrenstorfer.

**Trace element Measurements**

200 µl of whole blood were transferred into a 15 mL polyethylene tube and diluted 1:25 with a diluent solution consisting of 0.1% (v/v) Triton X-100, 0.1% (v/v) HNO3,2mg/L AU, and internal standards (20 µg/L). The mixture was sonicated for 10min before ICP-MS analysis. Multi-element determination was performed on an Agilent 7700x ICP-MS (Agilent Technologies, Tokyo, Japan) equipped with an octupole reaction system (ORS) collision/reaction cell technology to minimize spectral interferences. The continuous sample introduction system consisted of an autosampler, a quartz torch with a 2.5-mmdiameter injector with a Shield Torch system, a Scott double-pass spray chamber and nickel cones (Agilent Technologies, Tokyo, Japan). A glass concentric MicroMist nebuliser (Agilent Technologies, Tokyo, Japan) was used for the analysis of diluted samples.

**Water-soluble Vitamins Measurements**

200 µl plasma were deproteinized with 600 µl methanol (Merck), water, acetic acid (9:1:0.1) containing internal standards, thiamine-(4-methyl-13C-thiazol-5-yl-13C3) hydrochloride (Sigma-Aldrich), levomefolic acid-13C, d3, riboflavin-13C,15N2, 4-pyridoxic acid-d3 and pantothenic acid-13C3,15N hemi calcium salt (Toronto Research Chemicals). 500 µl supernatant were dried by nitrogen flow. 60 µl water were added to the residues, vortexed, the mixture was centrifuged and the supernatant was for analysis. The analysis was performed by UPLC coupled to a Waters Xevo TQ-S Triple Quad mass spectrometry (Waters, USA) with the electrospray ionization (ESI) source in positive ion mode. A Waters ACQUITY UPLC HSS T3 column (1.7 µm, 2.1 × 50 mm) was used for water-soluble vitamins separation with a flow rate at 0.45 ml/min and column temperature of 45 °C. The mobile phases were (A) 0.1 % formic acid in water, (B) 0.1% formic acid in methanol. The following elution gradient was used: 0–1 min,99.0%–99.0% A; 1–1.5 min, 99.0% A–97.0% A; 1.5–2 min, 97.0% A–70.0% A,2–3.5 min, 70% A–30% A; 3.5–4.0 min, 30%A–10.0%A; 4.0–4.8 min, 10%A–10.0%A; 4.9–6.0 min, 99.0%A–99.0%A. Multiple Reaction Monitoring (MRM) was used to monitor all water-soluble vitamins. The mass parameters were as follows, the capillary voltages of 3000V and source temperature of 150°C were adopted. The desolvation temperature was 500°C. The collision gas flow was set at 0.10 ml/min. The cone gas and desolvation gas flow were 150 l/h and 1000 l/h, respectively. All water-soluble vitamins standards were purchased from Sigma-Aldrich (USA).

**Fat-soluble Vitamins Measurements**

250 µl plasma were deproteinized with 1000 µl methanol and acetonitrile, (v/v,1:1) (Fisher Chemical) containing internal standards, all-trans-Retinol-d5, 25-HydroxyVitamin-D2-d6, 25-HydroxyVitamin-D3-d6, vitamin K1-d7, α-Tocopherol-d6 (Toronto Research Chemicals). 900 µl supernatant were dried by nitrogen flow. 80 µl 80% acetonitrile were added to the residues, vortexed, the mixture was centrifuged, and the supernatant was used for analysis. The analysis was performed by UPLC coupled to an AB Sciex Qtrap 4500 mass spectrometry (AB Sciex, USA) with the atmospheric pressure chemical ionization (APCI) source in positive ion mode. A Waters ACQUITY UPLC BEH C18 column (1.7 µm, 2.1 × 50 mm) was used for fat-soluble vitamins separation with a flow rate at 0.50 ml/min and column temperature of 45 °C. The mobile phases were (A) 0.1 % formic acid in water, (B) 0.1% formic acid in acetonitrile. The following elution gradient was used: 0–0.5 min,60.0%–60.0% A; 0.5–1.5 min, 60.0% A–20.0% A; 1.5–2.5 min, 20.0% A–0% A,2.5–4.5 min, 0% A–0% A; 4.5–4.6 min, 0%A–60.0%A; 4.6–5.0 min, 60.0%A–60.0%A. Multiple Reaction Monitoring (MRM) was used to monitor all fat-soluble vitamins. The mass parameters were as follows, Curtain gas flow 30 l/min, Collision Gas (CAD) was medium, Ion Source Gas 1 (GS 1) flow 40 l/min, Ion Source Gas 2 (GS 2) flow 50 l/min, Nebulizer Current (NC) 5, temperature 400 °C. All fat-soluble vitamins standards were purchased from Sigma-Aldrich (USA), Toronto research chemical (TRC).

**Immune indices Measurements**

10 ml whole blood was centrifuged at 3,000 r/min for 10 min, then 165 µl buffy coat were obtained to extract DNA using MagPure Buffy Coat DNA Midi KF Kit (Magen,

855 China). The DNA was sequenced on the BGISEQ- 500 platform using 200 bp singled-
856 end reads. The data processing was performed using Immune IMonitor [95]. VJ Gene use
857 diversity is shannon index of VJ gene usage profile. Immune cell diversity is Shannon
858 index of CDR3. Immune cell species result is unique CDR3 number. Immunity
859 uniformity is CDR3 pielou index. Score of above index is the sample rank in population.
860

861 **Medical Parameters**

862 All the volunteers were recruited during the physical examination. The medical test
863 including blood tests, urinalysis, routine examination of cervical secretion. All the
864 medical parameters were measured by the physical examination center and shown in
865 Supplementary Table 1b, 1d.
866

867 **Facial Skin feature**

868 The volunteer's frontal face without makeup was photographed by VISIA-CRTM
869 imaging system (Canfield Scientific, Fairfield, NJ, USA) equipped with chin supports
870 and forehead clamps that fix the face during the photographing process and maintain a
871 fixed distance between the volunteers and the camera at all times. Eight indices were
872 obtained including spots, pores, wrinkles, texture, UV spots, porphyrins, brown spots and
873 red area from the cheek and forehead, respectively (Supplementary Table 1b). The
874 percentile of index was calculated based on the index value ranked in the age-matched
875 database, the higher the better (Supplementary Table 1b).
876

877 **Physical fitness test**

878 8 kinds of tests were performed to evaluate volunteers' physical fitness condition
879 (Supplementary Table 1b). Vital capacity was measured by HK6800-FH (Hengkangjiaye,
880 China). Eye-closed and single-legged standing was measured by HK6800-ZL. Choice
881 reaction time was measured by HK6800-FY. Grip strength was measured by HK6800-
882 WL. Sit and reach was measured by HK6800-TQ. Sit-ups was measured by HK6800-YW.
883 Step index was measured by HK6800-TJ. Vertical jump was measured by HK6800-ZT.
884 We got a measure value from each test. Then each measure value score was assigned 1
885 through 5 based on its corresponding age-matched national standards (Supplementary
886 Table 4). Both the direct measurements and the scores were used for analyses
887 (Supplementary Table 2, Supplementary Table 3).
888
889 **Quality control, taxonomic annotation and abundance calculation**
890 The sequencing reads were quality-controlled as described previously [94]. Taxonomic
891 assignment of the high-quality fecal metagenomic data was performed using the reference
892 gene catalog comprising 9,879,896 gene[12]. Taxonomy of the fecal MGSs/MLGs were
893 then determined from their constituent genes, as previously described[1,13,14,35].
894
895 **The factors in each type of omics predicted by other type omics**
896 The factors in each type of omics were regressed against the relative abundances of
897 mgs profile (found in at least 10% of the samples) in the fecal samples using default
898 parameters in the RFCV function from randomForest package in R. Dichotomous
899 variables (such as gender ) and unordered categorical variable (such as region) were re-
900 coding into dummy variables. Frequency items such as yogurt eating habit were assigned

integers. RFCV R defined as spearman's correlation between measured value and 5-fold cross-validation predicted value was calculated, and then rank the top 5 predictable factors in each omics type. The same prediction process was done between any two types of omics. Then ggplot2 package in R was used to boxplot predict power of target omics factors by all kinds of other predictor omics (Fig. 2b). 75% quantile RFCV R between any two types omics (from a to b and from b to a) was used to construct the bi-direction global omics correlation network using CytoScape (Fig. 2a). R pheatmap and barplot was used to make heatmap plot for some representative factors (Fig. 2c, Supplementary Fig. 1).

**Adjusting for potential confounders**

Associations between gut microbiome MGSs, functional modules, Shannon diversity, and variance explained and other omics data were all adjusted for factors that probably influence the gut microbiome, including gender, age, BMI, health products (amino acid, vitamin, calcium), antivirus, antibiotics, drugs (currently using antihypertensive drugs, hyperglycemic drugs, lipid lowering drugs), days since last menstrual bleeding, pregnant, lactation, bowel problem (defecation). Besides the above basic set of confounders, we also show the results adjusting for more potential confounders including dietary (dietary taste spicy, sweet, salty, oil, or light, high sugar and high-fat diet habit, fruit and vegetable intake, favors fat meat), exercise (exercise frequency, exercise intensity, average time per exercise), drinking, smoking and Bristol's stool score.

**Benjamini-Hochberg multiple hypothesis testing correction**

The multiple hypothesis testing Benjamini-Hochberg corrections are done for one source target omics pair each time for Fig. 4-6, except immune index and gut microbe pair which BH-adjust was done on one immune index each time. We show two versions of Benjamini-Hochberg correction for Shannon and other omics in Supplementary Table 2a. One of the BH adjust was done within one omics each time. Another adjust was done overall on all omics.

**Robust association network construction between any two omics data type including fecal microbial MGSs**

An rank average method [96] was used to combine the results of multiple inference methods to make a robust omics association network. We combined two non-linear models, one-to-many randomforest and one-to-one partial spearman's correlation, to test the association between factor from any two types omics.

*Step 1: Data preprocessing.*

Dichotomous variables (such as gender) and unordered categorical variable (such as region) were re-coding into dummy variables. Frequency items such as yogurt eating habit were assigned integers. We removed variables following these rules: (i) The microbial species less than 10% in all the samples. (ii) Near zero variance. (iii) With more than 70% missing value. Missing values were filled with median. Outliers were defined as outside of the 95% quartiles and outliers samples are removed.

*Step 2: Computation of associations using multiple inference methods.*

For each factor in one omics, we did regression using RFCV function with default parameter based on all factors in one other omics and calculated RFCV R. [97]. 5-fold average variable importance was output for step3. Partial spearman's correlation (ppcor R package) between factors from any two types of omics were also output. Potential confounders were considered as described above. We also show generalized linear model results from MaAslin R package[98]) with default parameters after adjusting above confounders.

*Step 3: Robust networks construction.*

To get the robust and strongest association between factors from any two type omics, in other words, to filter predictor factors and target factors, we did it in two steps. First to choose the target factors, we just kept the top 20 target factors with highest RFCV R. Then to choose predictor factors for every selected target factor, we kept predictor factors with top 30 average ranks and retained edges with partial spearman's correlation BH-adjusted pvalue <0.05. The average rank was computed as sum of the ranks across the RFCV importance and absolute partial spearman rho. For example, metabolites as target and gut microbe as source. We regressed gut microbes against the metabolites and compute the 5-fold cross validation predict power (RFCV R) for each metabolites and partial spearman correlation. 20 metabolites with highest RFCV were kept. For each of the 20 select metabolites such as VA, average ranks across RFCV and partial spearman were done. Gut microbe biomarker for VA was found with average rank top 30$^{th}$ and passed the partial spearman BH-adjusted pvalue <0.05.

*Step 4: Network visualization.*

For each target factor, top 5-10 average ranks source factor in each source omics type were selected as representative factors to make barplots using ggplot2 package (Fig. 6). The pheatmap package was used to plot the common representative factors that could be strongest predicted by multiple omics data type (Fig. 2c). All the source-target factors pair RFCVR (a as source, b as target and b as source, a as target) was boxplot (Fig. 2b) using ggplot2. The ComplexHeatmap package in R was used to plot omics triadic relation (Fig. 4-6). CytoScape was also used to visualize the global omics network (Fig. 2a).

**Microbial metabolic syndrome risk index validation in cardiometabolic cohort.**

Using multi-omics analyses method described above after controlling for the potential confounders above, we picked up 80 MGSs that significantly correlated with one of the eight cardiometabolic risk factors (waist Hip Ratio, BMI, triglyceride (mmol/L), High-Density-Lipoprotein (mmol/L), serum Uric Acid ( $\mu$ mol/L),γ-glutamyl transpeptidase (U/L), serum alanine aminotransferase(U/L), fasting blood glucose (mmol/L) ) (Supplementary Fig. 5, 6). And they are link to the BCAA metabolites (valine / leucine / alanine), tryptophan, glutamic acid ($q<0.1$, Supplementary Table 3a). For the published disease studies from China, all the MGSs abundances were derived from metagenomic shotgun data, while the 8 clinical measurements could be missing, e.g. liver cirrhosis and Crohn's disease only had BMI available [21,31,32,34,35] (Supplementary Fig. 5, 6). The microbial metabolic syndrome risk index is similar with the T2D index (Qin et al, 2012). For each individual validation sample, the microbial metabolic syndrome risk index of sample *j* that denoted by *MMSR j* was computed by the formula below:

$$PR_{ij} = Count(A_{ij} > R_i)/N$$

$$J^B = \sum_{i \in B} PR_{ij}$$

$$J^G = \sum_{i \in G} PR_{ij}$$

$$MMSR_j = \frac{J^B}{|B|} - \frac{J^G}{|G|}$$

Where $Aij$ is a scalar represents the relative abundance of MGS $i$ in validation sample $j$. $Ri$ is a vector represents the relative abundance of MGS $i$ of all samples in this cohort which served as healthy reference. $N$ is the sample size of this cohort that is 2183. Percentile rank PR$ij$ is the percentage of test sample $j$'s MGS $i$ relative abundance in its reference cohort frequency distribution that are equal to or lower than it. $B$ is 12 out of 80 MGS that were positively correlated with BMI and Triglyceride. $G$ is 68 out of 80 MGS that were negatively correlated with BMI and Triglyceride. And $|B|$ and $|G|$ are the sizes of these two sets. We used percentile rank instead of relative abundance to avoid that the index was influenced too much by the dominant species.

**Figure legends:**

**Fig. 1 | Overview of the multi-omic cohorts.** Diagram for features available from the main cohort of 2,183 individuals and validation cohort of 1,404 volunteers. Details are available in Supplementary Table 1.

**Fig. 2 | Overview of the interrelationship between omics in the main cohort. a**, Global association strength between omics datasets. Each arrow is a 5-fold cross-validation random forest (RFCV) prediction. The direction of the arrow indicated the

direction of prediction, used the source omics dataset to predict the target dataset. The darkness and size of the arrow lines indicated 75% quantile of spearman's correlation between measured value and 5-fold cross-validation RFCV predicted value (RFCV R). **b**, Detailed predict power of source omics for each target omics. Tick label in x-axis is target omics. Title in top is source omics. Each node in box is a target factor. The color of the node and box line indicated the target omics data type. Y-axis is the target factor RFCV R predicted from source omics. **c**, Common representative factors that could be strongest predicted by multiple omics data type. Y-axis tick label is source omics. Title is target omics. X-axis tick label is common representative factors (target factors). The cell color in heat map indicated the RFCV R using the omics data in y-axis to predict each factor in x-axis.

**Fig. 3 | Factors associated with gut microbiome in both cohorts.** Top 45 factors with RFCV R > 0.1 in each type of omics that are predicted by gut microbiome. Factors with R ≤ 0.1 in main cohort are not shown. The length of the bar indicated the rank RFCV R using all samples and the color indicated the rank of max of RFCV R using male or female samples only, the darker the greater. Due to missing medical data in the validation cohort (Fig. 1, Supplementary Table 1), only red blood cell count can be validated.

**Fig. 4 | Association map of the four-tiered analyses integrating the metabolites, clinical indices, life style and the fecal microbiome.** The color of heat map show the partial spearman correlation adjusted for factors that probably influence the gut microbiome, as shown in Supplementary Fig. 3. BH-adjusted p-value is denoted: +, q-value<0.1; *, q-value<0.05; **, q-value<0.01.

**Fig. 5 | Gut microbiome associated with physical fitness and exercise in the main cohort.** The color of heat map shows the partial spearman correlation adjusted factors that probably influence the gut microbiome, as shown in Supplementary Fig. 3. BH adjusted p-value is denoted: +, q-value<0.1; *, q-value<0.05; **, q-value<0.01

**Fig. 6 | Influence of yogurt and milk intake on omics in the main cohort. a-e**, The top 5 factor in each omics data associated with yogurt, milk intake habit and the *Streptococcus thermophiles, Bifidobacterium animalis and Clostridium bolteae* abundance. The length of the bars represents partial Spearman's correlation coefficient adjusted for factors that probably influence the gut microbiome, as shown in Supplementary Fig. 3. BH adjusted p-value is denoted: +, q-value<0.1; *, q-value<0.05; **, q-value<0.01; ***, q-value<0.001; ****, q-value<0.0001. **f**, Fecal relative abundance of *S. thermophilus* in volunteers with increasing frequency of yogurt consumption.

## 49 facial skin imaging indices (only in the main cohort)

Spots, Pores, Wrinkles, Porphyrins, Texture, UV spots, Porphyrins, Brown spots and red area from the cheek and forehead, respectively.

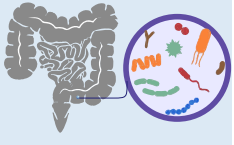
## Psychological questionnaire (18 entries)

The evaluation of irritability, Appetite, Dizziness, Frustration, Resilience, Fear, Self-confidence...

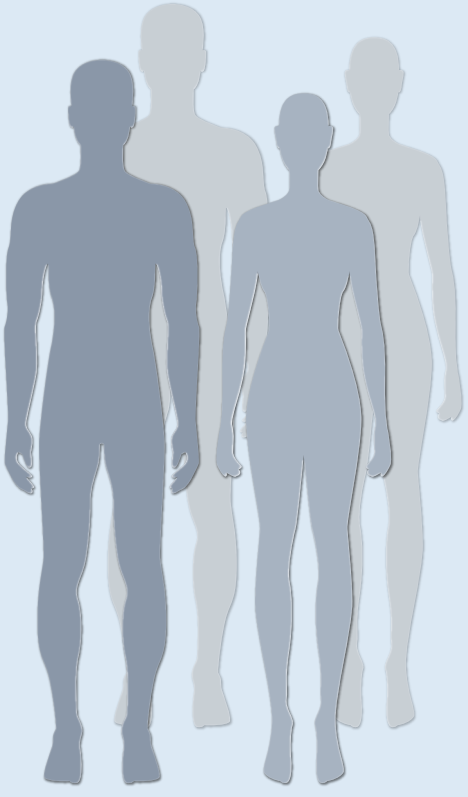
## Lifestyle questionnaire (56 entries)

Age, Gender, Marital status, BMI, Smoking, Drinking, Disease history, Eating habits, Exercise habits...

## Fecal microbiome



Shotgun metagenomics (1,507 MGSs and 2,981 MLGs)



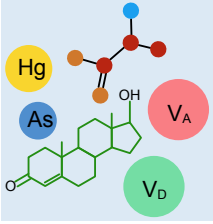
**3,587 Healthy adults**

The main cohort: 2,183 individuals
The validation cohort: 1,404 individuals

## 24 physical fitness data (only in the main cohort)

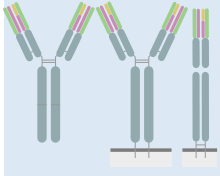
Vital capacity, Grip strength, Sit-ups, Choice reaction time, Sit and reach, One-leg stand with eyes closed, Step index, Vertical jump...

## 104 plasma metabolites



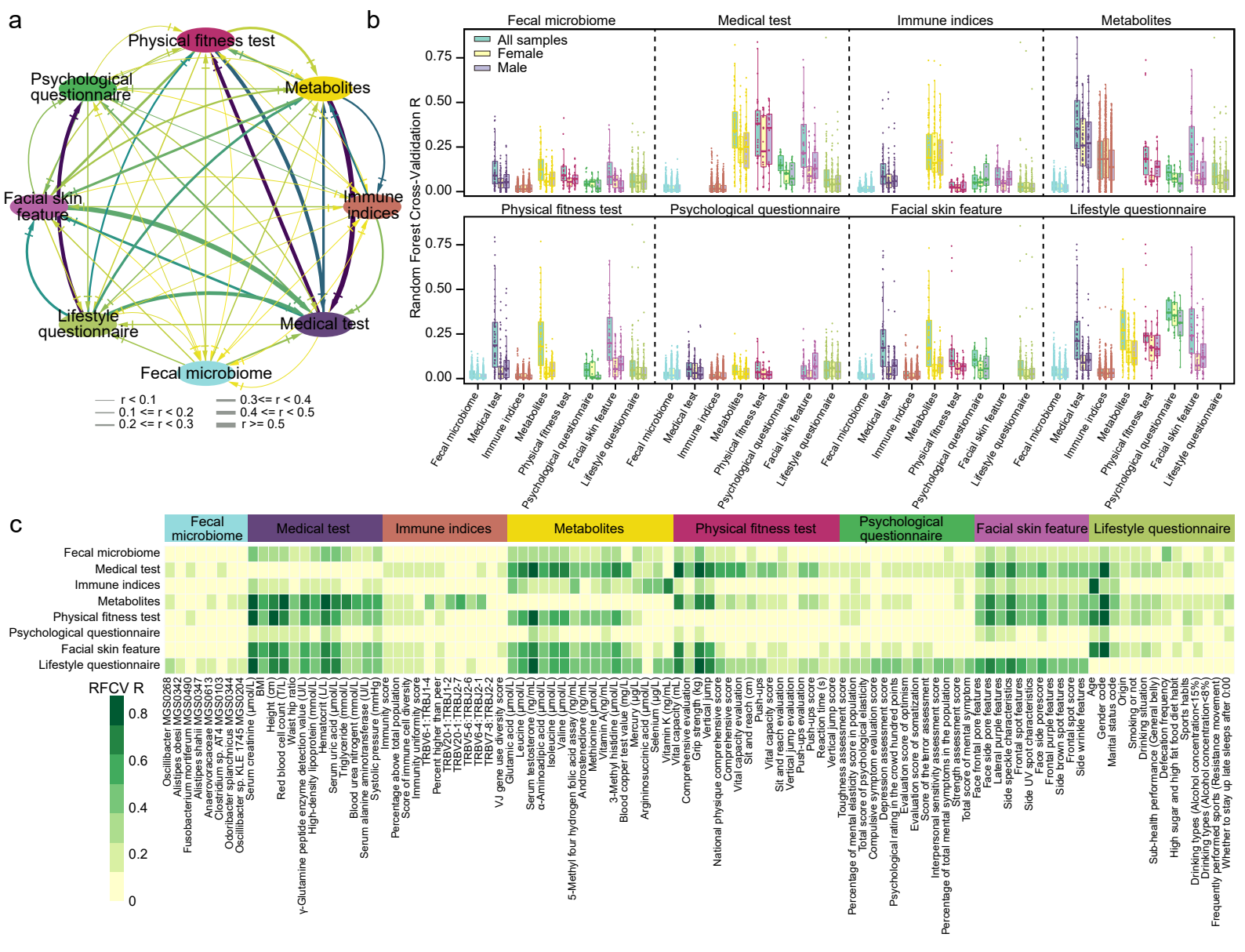

Amino acids (plasma)
Hormones (plasma)
Vitamins (plasma)
Trace elements (whole blood)

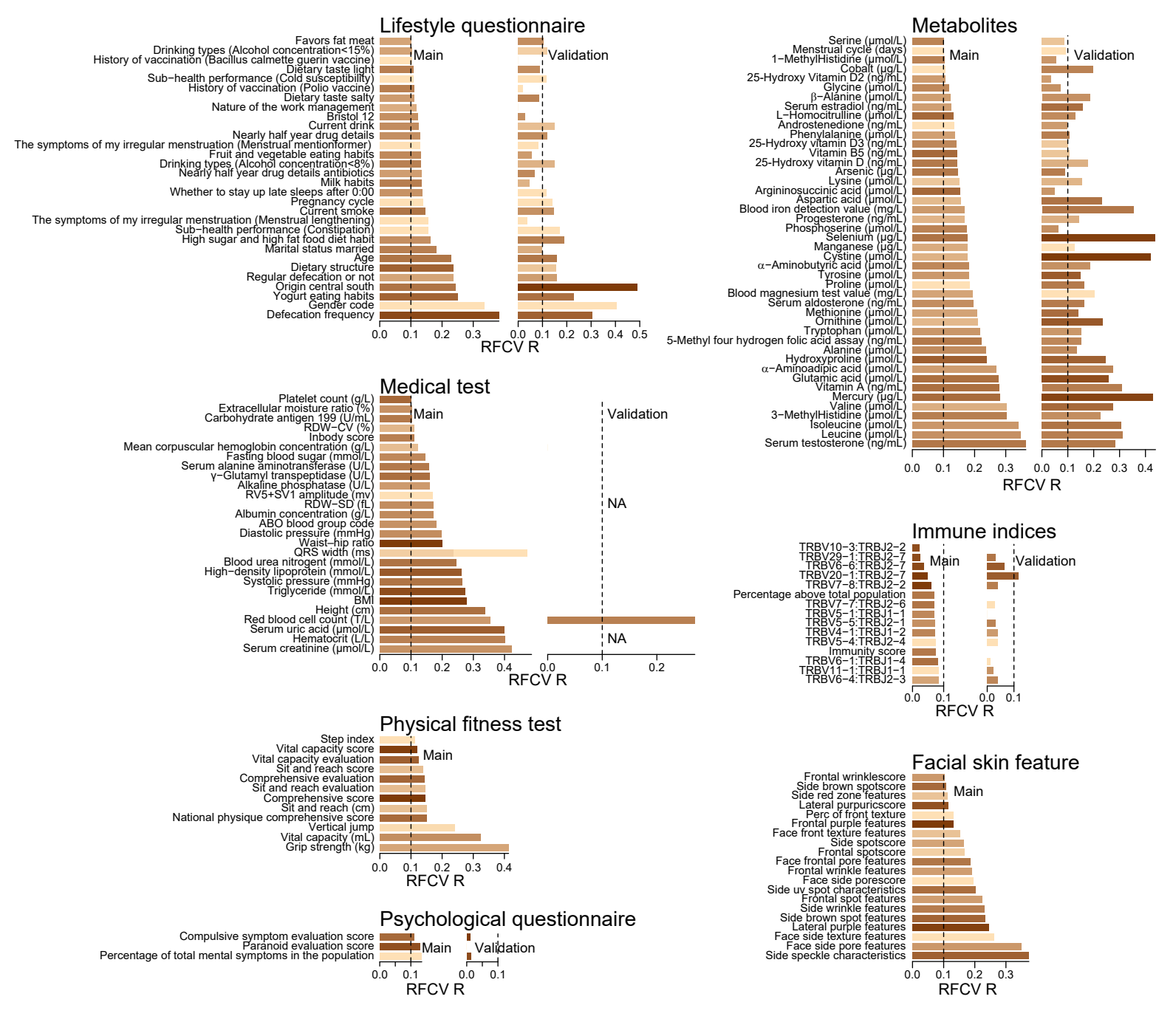
## 634 immune indices (buffy coat)

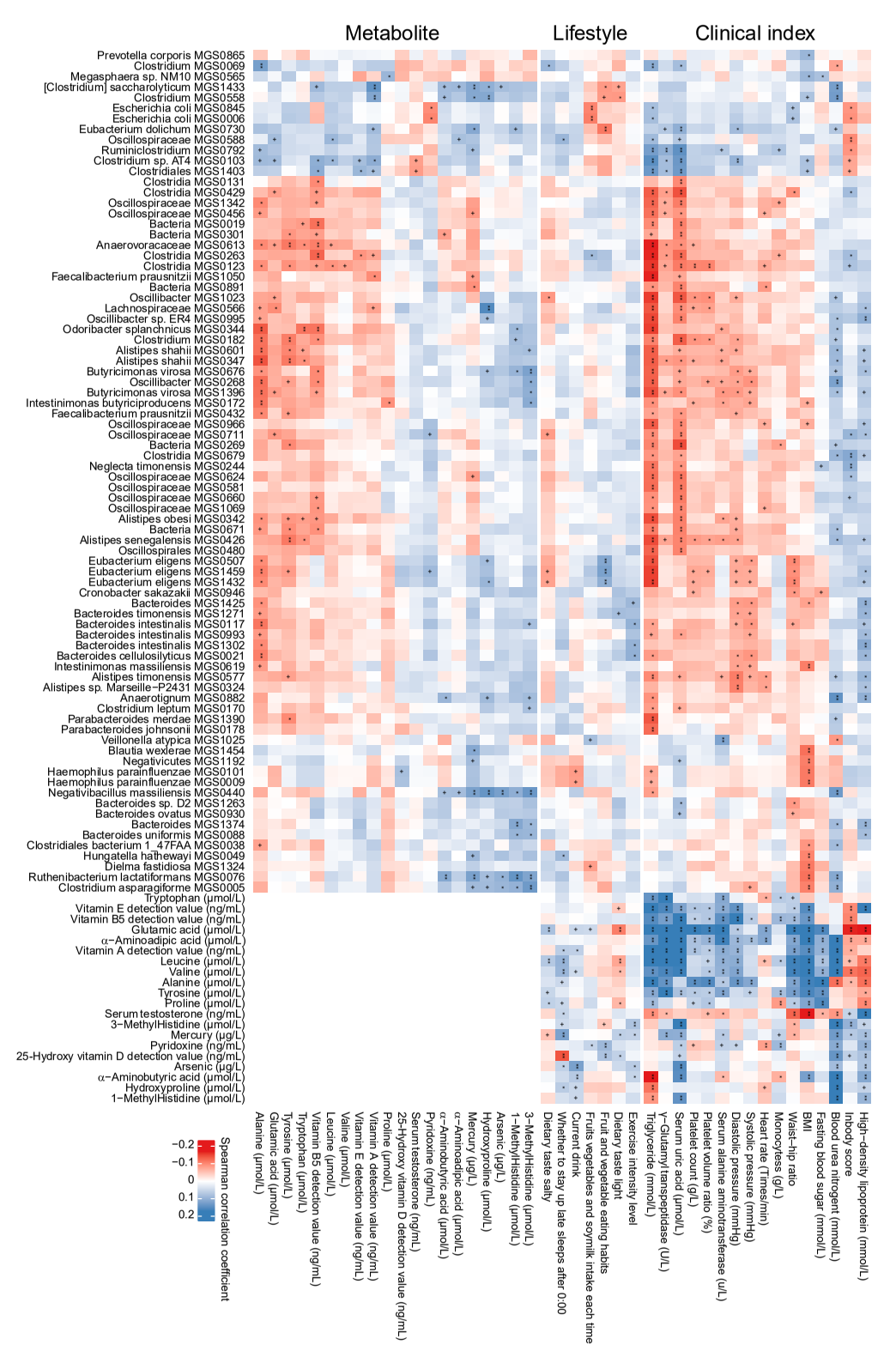


## 72 basic medical data (9 in validation cohort)

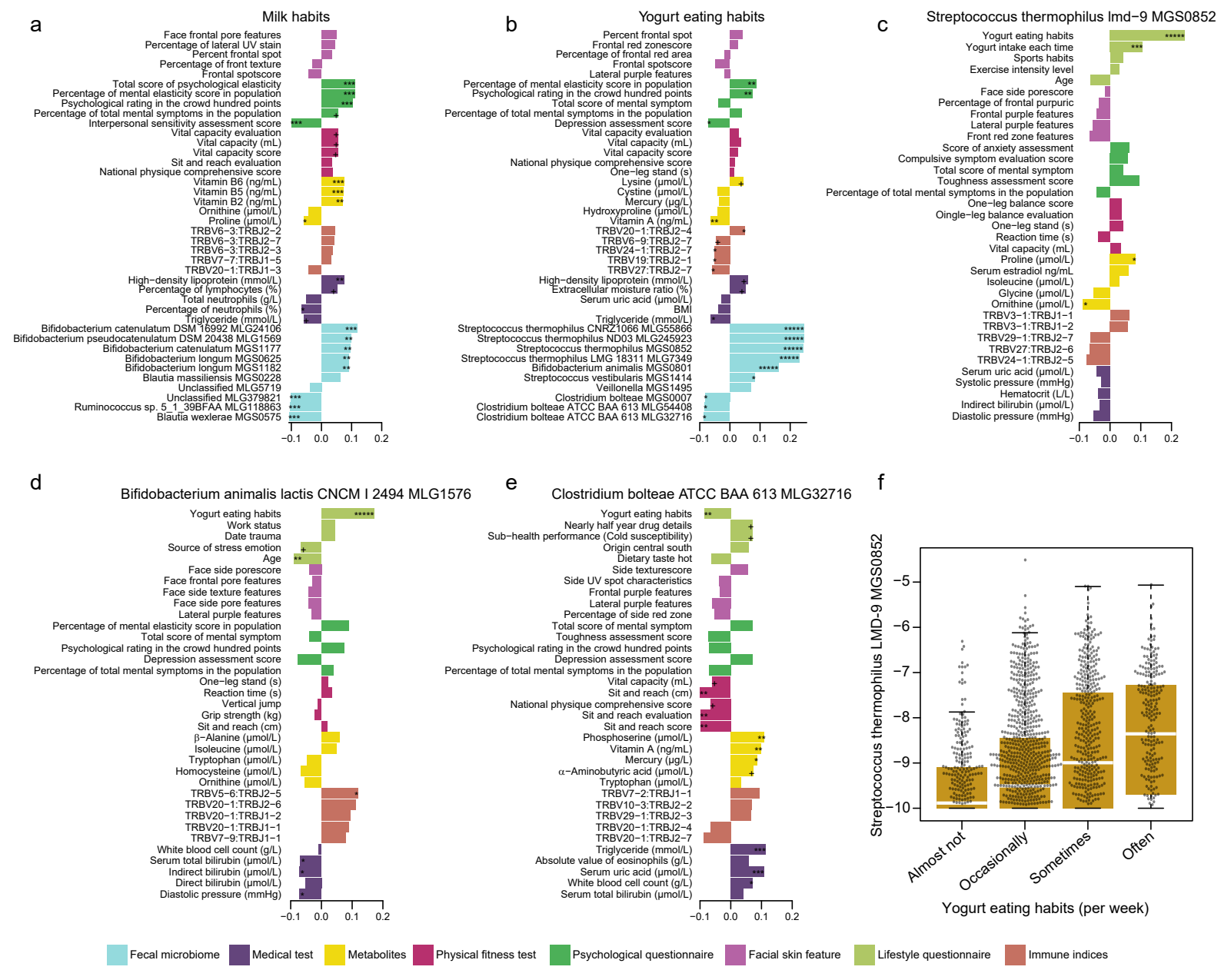
Body measurements:
BMI, Chest circumference, Uric acid
Routine blood test:
Alkaline phosphatase, Bilirubin, HDL, LDL, Globulin, Creatinine (serum)


a
Physical fitness test
Psychological questionnaire
Metabolites
Facial skin feature
Immune indices
Lifestyle questionnaire
Medical test
Fecal microbiome
r < 0.1
0.1 <= r < 0.2
0.2 <= r < 0.3
0.3<= r < 0.4
0.4 <= r < 0.5
r >= 0.5
b
Fecal microbiome
Medical test
Immune indices
Metabolites
Physical fitness test
Psychological questionnaire
Facial skin feature
Lifestyle questionnaire
All samples
Female
Male
Random Forest Cross-Validation R
0.00
0.25
0.50
0.75
c
RFCV R
0
0.2
0.4
0.6
0.8
Oscillibacter MGS0268
Alistipes obesi MGS0342
Fusobacterium mortiferum MGS0490
Alistipes shahii MGS0347
Anaerovoracaceae MGS0013
Clostridium sp. CAG:81 MGS0103
Odoribacter splanchnicus MGS0344
Oscillibacter sp. KLE 1745 MGS0204
Serum creatinine (μmol/L)
BMI
Height (cm)
Red blood cell count (T/L)
Waist hip ratio
γ-Glutamine peptide enzyme detection value (U/L)
High-density lipoprotein (mmol/L)
Hematocrit (L/L)
Serum uric acid (μmol/L)
Triglyceride (mmol/L)
Blood urea nitrogen (mmol/L)
Serum alanine aminotransferase (U/L)
Systolic pressure (mmHg)
Immunity score
Percentage above total population
Score of immune cell diversity
Immune cell diversity score
TRBV6-1:TRBJ1-4
Percent higher than peer
TRBV20-1::TRBJ1-2
TRBV20-1::TRBJ2-7
TRBV5-6:TRBJ2-6
TRBV6-4:TRBJ2-1
TRBV7-8:TRBJ2-2
VJ gene use diversity score
Glutamic acid (μmol/L)
Leucine (μmol/L)
Serum testosterone (ng/mL)
α-Aminoadipic acid (μmol/L)
Isoleucine (μmol/L)
Valine (μmol/L)
5-Methyl four hydrogen folic acid assay (ng/mL)
Androstenedione (ng/mL)
Methionine (μmol/L)
Methionine A (mg/mL)
3-Methyl histidine A (mg/mL)
Blood copper test value (mg/L)
Mercury (μg/L)
Argininosuccinic acid (μmol/L)
Selenium (μg/L)
Vitamin K (ng/mL)
Vital capacity (mL)
Comprehensive evaluation
Grip strength (kg)
Vertical jump
National physique comprehensive score
Comprehensive score
Vital capacity evaluation
Sit and reach (cm)
Push-ups
Vital capacity score
Sit and reach evaluation
Vertical jump evaluation
Push-ups evaluation
Push-ups score
Reaction time (s)
Vertical jump score
Toughness assessment score
Percentage of mental elasticity score in population
Total score of psychological elasticity
Compulsive symptom evaluation score
Depression assessment score
Psychological rating in the crowd hundred points
Evaluation score of optimism
Evaluation score of somatization
Score of the terror assessment
Interpersonal sensitivity assessment score
Percentage of total mental symptoms in the population
Strength assessment score
Total score of mental symptom
Face frontal pore features
Face side pore features
Face speckle features
Side speckle characteristics
Frontal spot features
Side UV spot characteristics
Face side porescore
Frontal purple features
Side brown spot features
Frontal spot score
Side wrinkle features
Age
Gender code
Marital status code
Origin
Smoking or not
Drinking situation
Sub-health performance (General belly)
Defecation frequency
High sugar and high fat food diet habit
Drinking types (Alcohol concentration<15%)
Drinking types (Alcohol concentration>=8%)
Frequently performed sports (Resistance movement)
Whether to stay up late sleeps after 0:00


Lifestyle questionnaire
Favors fat meat
Drinking types (Alcohol concentration<15%)
History of vaccination (Bacillus calmette guerin vaccine)
Dietary taste light
Sub−health performance (Cold susceptibility)
History of vaccination (Polio vaccine)
Dietary taste salty
Nature of the work management
Bristol 12
Current drink
Nearly half year drug details
The symptoms of my irregular menstruation (Menstrual mentionformer)
Fruit and vegetable eating habits
Drinking types (Alcohol concentration<8%)
Nearly half year drug details antibiotics
Milk habits
Whether to stay up late sleeps after 0:00
Pregnancy cycle
Current smoke
The symptoms of my irregular menstruation (Menstrual lengthening)
Sub−health performance (Constipation)
High sugar and high fat food diet habit
Marital status married
Age
Dietary structure
Regular defecation or not
Origin central south
Yogurt eating habits
Gender code
Defecation frequency
Main
Validation
0.0 0.1 0.2 0.3
0.0 0.1 0.2 0.3 0.4 0.5
RFCV R
Metabolites
Serine (μmol/L)
Menstrual cycle (days)
1−MethylHistidine (μmol/L)
Cobalt (μg/L)
25-Hydroxy Vitamin D2 (ng/mL)
Glycine (μmol/L)
β−Alanine (μmol/L)
Serum estradiol (ng/mL)
L−Homocitrulline (μmol/L)
Androstenedione (ng/mL)
Phenylalanine (μmol/L)
25-Hydroxy vitamin D3 (ng/mL)
Vitamin B5 (ng/mL)
25-Hydroxy vitamin D (ng/mL)
Arsenic (μg/L)
Lysine (μmol/L)
Argininosuccinic acid (μmol/L)
Aspartic acid (μmol/L)
Blood iron detection value (mg/L)
Progesterone (ng/mL)
Phosphoserine (μmol/L)
Selenium (μg/L)
Manganese (μg/L)
Cystine (μmol/L)
α−Aminobutyric acid (μmol/L)
Tyrosine (μmol/L)
Proline (μmol/L)
Blood magnesium test value (mg/L)
Serum aldosterone (ng/mL)
Methionine (μmol/L)
Ornithine (μmol/L)
Tryptophan (μmol/L)
5-Methyl four hydrogen folic acid assay (ng/mL)
Alanine (μmol/L)
Hydroxyproline (μmol/L)
α−Aminoadipic acid (μmol/L)
Glutamic acid (μmol/L)
Vitamin A (ng/mL)
Mercury (μg/L)
Valine (μmol/L)
3−MethylHistidine (μmol/L)
Isoleucine (μmol/L)
Leucine (μmol/L)
Serum testosterone (ng/mL)
Main
Validation
0.0 0.1 0.2 0.3 0.4
0.0 0.1 0.2 0.3 0.4
RFCV R
Medical test
Platelet count (g/L)
Extracellular moisture ratio (%)
Carbohydrate antigen 199 (U/mL)
RDW−CV (%)
Inbody score
Mean corpuscular hemoglobin concentration (g/L)
Fasting blood sugar (mmol/L)
Serum alanine aminotransferase (U/L)
γ−Glutamyl transpeptidase (U/L)
Alkaline phosphatase (U/L)
RV5+SV1 amplitude (mv)
RDW−SD (fL)
Albumin concentration (g/L)
ABO blood group code
Diastolic pressure (mmHg)
Waist−hip ratio
QRS width (ms)
Blood urea nitrogent (mmol/L)
High−density lipoprotein (mmol/L)
Systolic pressure (mmHg)
Triglyceride (mmol/L)
BMI
Height (cm)
Red blood cell count (T/L)
Serum uric acid (μmol/L)
Hematocrit (L/L)
Serum creatinine (μmol/L)
Main
Validation
NA
NA
0.0 0.1 0.2 0.3 0.4
0.0 0.1 0.2
RFCV R
Immune indices
TRBV10−3:TRBJ2−2
TRBV29−1:TRBJ2−7
TRBV6−6:TRBJ2−7
TRBV20−1:TRBJ2−7
TRBV7−8:TRBJ2−2
Percentage above total population
TRBV7−7:TRBJ2−6
TRBV5−1:TRBJ1−1
TRBV5−5:TRBJ2−1
TRBV4−1:TRBJ1−2
TRBV5−4:TRBJ2−4
Immunity score
TRBV6−1:TRBJ1−4
TRBV11−1:TRBJ1−1
TRBV6−4:TRBJ2−3
Main
Validation
0.0 0.1
0.0 0.1
RFCV R
Physical fitness test
Step index
Vital capacity score
Vital capacity evaluation
Sit and reach score
Comprehensive evaluation
Sit and reach evaluation
Comprehensive score
Sit and reach (cm)
National physique comprehensive score
Vertical jump
Vital capacity (mL)
Grip strength (kg)
Main
0.0 0.1 0.2 0.3 0.4
RFCV R
Facial skin feature
Frontal wrinklescore
Side brown spotscore
Side red zone features
Lateral purpuricscore
Perc of front texture
Frontal purple features
Face front texture features
Side spotscore
Frontal spotscore
Face frontal pore features
Frontal wrinkle features
Face side porescore
Side uv spot characteristics
Frontal spot features
Side wrinkle features
Side brown spot features
Lateral purple features
Face side texture features
Face side pore features
Side speckle characteristics
Main
0.0 0.1 0.2 0.3
RFCV R
Psychological questionnaire
Compulsive symptom evaluation score
Paranoid evaluation score
Percentage of total mental symptoms in the population
Main
Validation
0.0 0.1
0.0 0.1
RFCV R


Metabolite
Lifestyle
Clinical index
Prevotella corporis MGS0865
Clostridium MGS0069
Megasphaera sp. NM10 MGS0565
[Clostridium] saccharolyticum MGS1433
Clostridium MGS0558
Escherichia coli MGS0845
Escherichia coli MGS0006
Eubacterium dolichum MGS0730
Oscillospiraceae MGS0588
Ruminiclostridium MGS0792
Clostridium sp. AT4 MGS0103
Clostridiales MGS1403
Clostridia MGS0131
Clostridia MGS0429
Oscillospiraceae MGS1342
Oscillospiraceae MGS0456
Bacteria MGS0019
Bacteria MGS0301
Anaerovoracaceae MGS0613
Clostridia MGS0263
Clostridia MGS0123
Faecalibacterium prausnitzii MGS1050
Bacteria MGS0891
Oscillibacter MGS1023
Lachnospiraceae MGS0566
Oscillibacter sp. ER4 MGS0995
Odoribacter splanchnicus MGS0344
Clostridium MGS0182
Alistipes shahii MGS0601
Alistipes shahii MGS0347
Butyricimonas virosa MGS0676
Oscillibacter MGS0268
Butyricimonas virosa MGS1396
Intestinimonas butyriciproducens MGS0172
Faecalibacterium prausnitzii MGS0432
Oscillospiraceae MGS0966
Oscillospiraceae MGS0711
Bacteria MGS0269
Clostridia MGS0679
Neglecta timonensis MGS0244
Oscillospiraceae MGS0624
Oscillospiraceae MGS0581
Oscillospiraceae MGS0660
Oscillospiraceae MGS1069
Alistipes obesi MGS0342
Bacteria MGS0671
Alistipes senegalensis MGS0426
Oscillospirales MGS0480
Eubacterium eligens MGS0507
Eubacterium eligens MGS1459
Eubacterium eligens MGS1432
Cronobacter sakazakii MGS0946
Bacteroides MGS1425
Bacteroides timonensis MGS1271
Bacteroides intestinalis MGS0117
Bacteroides intestinalis MGS0993
Bacteroides intestinalis MGS1302
Bacteroides cellulosilyticus MGS0021
Intestinimonas massiliensis MGS0619
Alistipes timonensis MGS0577
Alistipes sp. Marseille−P2431 MGS0324
Anaerotignum MGS0882
Clostridium leptum MGS0170
Parabacteroides merdae MGS1390
Parabacteroides johnsonii MGS0178
Veillonella atypica MGS1025
Blautia wexlerae MGS1454
Negativicutes MGS1192
Haemophilus parainfluenzae MGS0101
Haemophilus parainfluenzae MGS0009
Negativibacillus massiliensis MGS0040
Bacteroides sp. D2 MGS1263
Bacteroides ovatus MGS0930
Bacteroides MGS1374
Bacteroides uniformis MGS0088
Clostridiales bacterium 1_47FAA MGS0038
Hungatella hathewayi MGS0049
Dielma fastidiosa MGS1324
Ruthenibacterium lactatiformans MGS0076
Clostridium asparagiforme MGS0005
Tryptophan (μmol/L)
Vitamin E detection value (ng/mL)
Vitamin B5 detection value (ng/mL)
Glutamic acid (μmol/L)
α−Aminoadipic acid (μmol/L)
Vitamin A detection value (ng/mL)
Leucine (μmol/L)
Valine (μmol/L)
Alanine (μmol/L)
Tyrosine (μmol/L)
Proline (μmol/L)
Serum testosterone (ng/mL)
3−MethylHistidine (μmol/L)
Mercury (μg/L)
Pyridoxine (ng/mL)
25-Hydroxy vitamin D detection value (ng/mL)
Arsenic (μg/L)
α−Aminobutyric acid (μmol/L)
Hydroxyproline (μmol/L)
1−MethylHistidine (μmol/L)
Alanine (μmol/L)
Glutamic acid (μmol/L)
Tyrosine (μmol/L)
Tryptophan (μmol/L)
Vitamin B5 detection value (ng/mL)
Leucine (μmol/L)
Valine (μmol/L)
Vitamin E detection value (ng/mL)
Vitamin A detection value (ng/mL)
Proline (μmol/L)
25-Hydroxy vitamin D detection value (ng/mL)
Serum testosterone (ng/mL)
Pyridoxine (ng/mL)
α−Aminobutyric acid (μmol/L)
α−Aminoadipic acid (μmol/L)
Mercury (μg/L)
Hydroxyproline (μmol/L)
Arsenic (μg/L)
1−MethylHistidine (μmol/L)
3−MethylHistidine (μmol/L)
Dietary taste salty
Whether to stay up late sleeps after 0:00
Current drink
Fruits vegetables and soymilk intake each time
Fruit and vegetable eating habits
Dietary taste light
Exercise intensity level
Triglyceride (mmol/L)
γ−Glutamyltranspeptidase (U/L)
Serum uric acid (μmol/L)
Platelet count (g/L)
Platelet volume ratio (%)
Serum alanine aminotransferase (u/L)
Diastolic pressure (mmHg)
Systolic pressure (mmHg)
Heart rate (Times/min)
Monocytes (g/L)
Waist−hip ratio
BMI
Fasting blood sugar (mmol/L)
Blood urea nitrogent (mmol/L)
Inbody score
High−density lipoprotein (mmol/L)
Spearman correlation coefficient
−0.2
−0.1
0
0.1
0.2

Exercise
Physical fitness test
Exercise intensity level
Average time per exercise
Sports habits
Fruit and vegetable eating habits
Sleep quality in the last month
Sub−health performance (General belly)
National physique comprehensive score
One−leg balance score
Reaction time score
Vertical jump score
Step index
Sit and reach score
Grip score
Vital capacity score
Coprobacillus sp. 3 3 56FAA-MGS0084
Escherichia coli-MGS0006
Escherichia coli-MGS0845
Clostridium clostridioforme-MGS0105
Clostridium innocuum-MGS0053
Clostridium clostridioforme-MGS1018
Clostridium bolteae-MGS0007
Ruminococcus gnavus-MGS0160
Clostridia-MGS0123
Lachnospiraceae bacterium 10−1-MGS0614
Oscillospiraceae-MGS0542
Oscillospiraceae-MGS0502
Oscillospiraceae-MGS0794
Oscillospiraceae-MGS1223
Oscillospiraceae-MGS0711
Alistipes senegalensis-MGS0426
Peptoniphilus-MGS1217
Clostridia-MGS0429
Clostridia-MGS0679
Bacteroides-MGS1425
Bacteroides timonensis-MGS1271
Bacteroides cellulosilyticus-MGS0021
Bacteroides stercorirosoris-MGS0066
Alistipes shahii-MGS0347
Bacteroides pectinophilus-MGS0391
Eubacterium rectale-MGS0663
Oscillibacter sp. ER4-MGS0479
Gemmiger formicilis-MGS0279
Eubacterium rectale-MGS0687
Subdoligranulum-MGS0643
Blautia-MGS1320
Blautia obeum-MGS0853
Ruminococcus bicirculans-MGS0166
Lachnospiraceae-MGS0463
Clostridia-MGS0263
Clostridia-MGS0254
Coprobacter fastidiosus-MGS1138
Blautia-MGS0266
Lachnospiraceae-MGS0667
Clostridium-MGS0763
Faecalibacterium prausnitzii-MGS0238
Faecalibacterium prausnitzii-MGS0443
Oscillospiraceae-MGS0630
Bifidobacterium adolescentis-MGS0617
Barnesiella intestinihominis-MGS0192
Anaerovoracaceae-MGS0613
Bacteroides intestinalis-MGS0117
Butyricimonas virosa-MGS0676
Butyricimonas virosa-MGS1396
Oscillibacter-MGS1023
Alistipes shahii-MGS0601
Alistipes putredinis-MGS0563
Oscillospiraceae-MGS0600
Eubacterium ventriosum-MGS0745
Oscillospiraceae-MGS0759
Butyricimonas-MGS0277
Parabacteroides merdae-MGS0068
Parabacteroides sp. Marseille−P3236-MGS1413
Sub−health performance (General belly)
Sleep quality in the last month
Sports habits
Fruit and vegetable eating habits
Exercise intensity level
Average time per exercise
Spearman correlation coefficient
−0.2
−0.1
0
0.1
0.2


a
Milk habits
Face frontal pore features
Percentage of lateral UV stain
Percent frontal spot
Percentage of front texture
Frontal spotscore
Total score of psychological elasticity
Percentage of mental elasticity score in population
Psychological rating in the crowd hundred points
Percentage of total mental symptoms in the population
Interpersonal sensitivity assessment score
Vital capacity evaluation
Vital capacity (mL)
Vital capacity score
Sit and reach evaluation
National physique comprehensive score
Vitamin B6 (ng/mL)
Vitamin B5 (ng/mL)
Vitamin B2 (ng/mL)
Ornithine (μmol/L)
Proline (μmol/L)
TRBV6−3:TRBJ2−2
TRBV6−3:TRBJ2−7
TRBV6−3:TRBJ2−3
TRBV7−7:TRBJ1−5
TRBV20−1:TRBJ1−3
High−density lipoprotein (mmol/L)
Percentage of lymphocytes (%)
Total neutrophils (g/L)
Percentage of neutrophils (%)
Triglyceride (mmol/L)
Bifidobacterium catenulatum DSM 16992 MLG24106
Bifidobacterium pseudocatenulatum DSM 20438 MLG1569
Bifidobacterium catenulatum MGS1177
Bifidobacterium longum MGS0625
Bifidobacterium longum MGS1182
Blautia massiliensis MGS0228
Unclassified MLG5719
Unclassified MLG379821
Ruminococcus sp. 5_1_39BFAA MLG118863
Blautia wexlerae MGS0575
−0.1 0.0 0.1 0.2
b
Yogurt eating habits
Percent frontal spot
Frontal red zonescore
Percentage of frontal red area
Frontal spotscore
Lateral purple features
Percentage of mental elasticity score in population
Psychological rating in the crowd hundred points
Total score of mental symptom
Percentage of total mental symptoms in the population
Depression assessment score
Vital capacity evaluation
Vital capacity (mL)
Vital capacity score
National physique comprehensive score
One−leg stand (s)
Lysine (μmol/L)
Cystine (μmol/L)
Mercury (μg/L)
Hydroxyproline (μmol/L)
Vitamin A (ng/mL)
TRBV20−1:TRBJ2−4
TRBV6−9:TRBJ2−7
TRBV24−1:TRBJ2−7
TRBV19:TRBJ2−1
TRBV27:TRBJ2−7
High−density lipoprotein (mmol/L)
Extracellular moisture ratio (%)
Serum uric acid (μmol/L)
BMI
Triglyceride (mmol/L)
Streptococcus thermophilus CNRZ1066 MLG55866
Streptococcus thermophilus ND03 MLG245923
Streptococcus thermophilus MGS0852
Streptococcus thermophilus LMG 18311 MLG7349
Bifidobacterium animalis MGS0801
Streptococcus vestibularis MGS1414
Veillonella MGS1495
Clostridium bolteae MGS0007
Clostridium bolteae ATCC BAA 613 MLG54408
Clostridium bolteae ATCC BAA 613 MLG32716
−0.1 0.0 0.1 0.2
c
Streptococcus thermophilus lmd−9 MGS0852
Yogurt eating habits
Yogurt intake each time
Sports habits
Exercise intensity level
Age
Face side porescore
Percentage of frontal purpuric
Frontal purple features
Lateral purple features
Front red zone features
Score of anxiety assessment
Compulsive symptom evaluation score
Total score of mental symptom
Toughness assessment score
Percentage of total mental symptoms in the population
One−leg balance score
Oingle−leg balance evaluation
One−leg stand (s)
Reaction time (s)
Vital capacity (mL)
Proline (μmol/L)
Serum estradiol ng/mL
Isoleucine (μmol/L)
Glycine (μmol/L)
Ornithine (μmol/L)
TRBV3−1:TRBJ1−1
TRBV3−1:TRBJ1−2
TRBV29−1:TRBJ2−7
TRBV27:TRBJ2−6
TRBV24−1:TRBJ2−5
Serum uric acid (μmol/L)
Systolic pressure (mmHg)
Hematocrit (L/L)
Indirect bilirubin (μmol/L)
Diastolic pressure (mmHg)
−0.1 0.0 0.1 0.2
d
Bifidobacterium animalis lactis CNCM I 2494 MLG1576
Yogurt eating habits
Work status
Date trauma
Source of stress emotion
Age
Face side porescore
Face frontal pore features
Face side texture features
Face side pore features
Lateral purple features
Percentage of mental elasticity score in population
Total score of mental symptom
Psychological rating in the crowd hundred points
Depression assessment score
Percentage of total mental symptoms in the population
One−leg stand (s)
Reaction time (s)
Vertical jump
Grip strength (kg)
Sit and reach (cm)
β−Alanine (μmol/L)
Isoleucine (μmol/L)
Tryptophan (μmol/L)
Homocysteine (μmol/L)
Ornithine (μmol/L)
TRBV5−6:TRBJ2−5
TRBV20−1:TRBJ2−6
TRBV20−1:TRBJ1−2
TRBV20−1:TRBJ1−1
TRBV7−9:TRBJ1−1
White blood cell count (g/L)
Serum total bilirubin (μmol/L)
Indirect bilirubin (μmol/L)
Direct bilirubin (μmol/L)
Diastolic pressure (mmHg)
−0.1 0.0 0.1 0.2
e
Clostridium bolteae ATCC BAA 613 MLG32716
Yogurt eating habits
Nearly half year drug details
Sub−health performance (Cold susceptibility)
Origin central south
Dietary taste hot
Side texturescore
Side UV spot characteristics
Frontal purple features
Lateral purple features
Percentage of side red zone
Total score of mental symptom
Toughness assessment score
Psychological rating in the crowd hundred points
Depression assessment score
Percentage of total mental symptoms in the population
Vital capacity (mL)
Sit and reach (cm)
National physique comprehensive score
Sit and reach evaluation
Sit and reach score
Phosphoserine (μmol/L)
Vitamin A (ng/mL)
Mercury (μg/L)
α−Aminobutyric acid (μmol/L)
Tryptophan (μmol/L)
TRBV7−2:TRBJ1−1
TRBV10−3:TRBJ2−2
TRBV29−1:TRBJ2−3
TRBV20−1:TRBJ2−4
TRBV20−1:TRBJ2−7
Triglyceride (mmol/L)
Absolute value of eosinophils (g/L)
Serum uric acid (μmol/L)
White blood cell count (g/L)
Serum total bilirubin (μmol/L)
−0.1 0.0 0.1 0.2
f
Streptococcus thermophilus LMD-9 MGS0852
−5
−6
−7
−8
−9
−10
Almost not
Occasionally
Sometimes
Often
Yogurt eating habits (per week)
Fecal microbiome
Medical test
Metabolites
Physical fitness test
Psychological questionnaire
Facial skin feature
Lifestyle questionnaire
Immune indices