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

Zhuye Jie[1,2,3,*,†], Suisha Liang[1,*], Qiuxia Ding[1,*], Shanmei Tang[1], Fei Li[1], Dan Wang[1,3], Yuxiang Lin[1,3], Peishan Chen[1,3], Kaiye Cai[1,3], Weiqiao Rao[1], Yunli Liao[1], Dongsheng Zhou[1], Xuemei Qiu[1], Xiaomin Chen[1], Jin Zi[1], Xiaohan Wang[1],Wei Zhang[1], Tao Zhang[1,4,5], Liang Xiao[1,6,7], Huanming Yang[1,8], Jian Wang[1,8], Yong Hou[1,3], Xiao Liu[1], Karsten Kristiansen[1,4], Huanzi Zhong[1,4], Huijue Jia[1,2,3,9,†], Xun Xu[1,3,†]

1. BGI-Shenzhen, Shenzhen, China;
2. Shenzhen Key Laboratory of Human Commensal Microorganisms and Health Research, BGI-Shenzhen, Shenzhen, China;
3. China National Genebank, BGI-Shenzhen, Shenzhen 518120, China;
4. Department of Biology, Ole MaalØes Vej 5, University of Copenhagen, Copenhagen, Denmark;
5. Shenzhen Key Laboratory of Cognition and Gene Research, BGI-Shenzhen, Shenzhen 518083, China;
6. Shenzhen Engineering Laboratory of Detection and Intervention of human intestinal microbiome, BGI-Shenzhen, Shenzhen, China;
7. BGI-Qingdao, BGI-Shenzhen, Qingdao, 266555, China;
8. James D. Watson Institute of Genome Sciences, Hangzhou, China;
9. Macau University of Science and Technology, Avenida Wai long, Taipa, Macau, China.

* These authors contributed equally to this work.

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

Summary

More than a decade of gut microbiome studies has a common goal for human health. As most of the disease studies sample the elderly or the middle-aged, a reference cohort for young healthy 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 ~2183 adults. A number of vitamins, amino acids and trace elements leave their marks on the gut microbiome and cluster with T cell receptor types. Gender and hormones, especially aldosterone have their impacts. Effects of physical fitness, sleeping habits and dairy consumption are also demonstrated in this large multi-omic cohort. Thus, our comprehensive data are poised to advise future study designs and offer support for exercising and yogurt consumption.

INTRODUCTION

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 (Wang and Jia, 2016), recently published studies

have emphasized difficulty in extrapolating to natural cohorts due to location and ethnicity (Deschasaux et al., 2018; He et al., 2018). So far only the LifeLines Deep cohort (n=1135) from the Netherlands made use of low-depth metagenomic sequencing (7.5 million reads per sample) instead of 16S rRNA gene amplicon sequencing (Zhernakova et al., 2016). 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 (FDA; USGS), but are unexplored in the microbiome field. Biological sex is a strong determent for the gut microbiome in mice and livestock (Xiao et al., 2015, 2016; Zhao et al., 2013), but the impact of hormones, especially non-gonadal hormones on the human gut microbiome remains unclear.

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 2183 adults, along with questionnaire data, physical fitness tests, facial skin features, plasma metabolome and immune repertoire. Trans-omics analyses uncover previously overlooked measurements such as aldosterone, trace elements, vitamin A that heavily influence the gut microbiome. This is also to our knowledge the largest cohort with facial skin data and immune repertoire data.

A recent study casted doubt over the health benefits of yogurt consumption, concluding that colonization of the yogurt strains was highly variable between individuals (Zmora et al., 2018). Our large cohort unequivocally showed yogurt strains, especially *Streptococcus thermophilus* in feces, along with beneficial effects in cardiometabolic and mental health.

RESULTS

The gut microbiome as a relatively independent dimension for health

Fecal samples were collected during a physical examination, and 2183 samples were subjected to metagenomic shotgun sequencing, yielding 81.20 ± 24.59 million high-quality non-human reads per sample (Table S1), ensuring accurate taxonomic and functional profilings. 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 1507 Metagenomic Species (MGSs) (Li et al., 2014; Moll1 et al., 2019; Nielsen et al., 2014) and 2981 metagenomic linkage groups (MLGs) (Qin et al, 2012, Jie et al, 2017).

Other omics data, including 104 plasma metabolites (3980 samples), 634 immune indices (VDJ usage and diversity, 4120 samples) from buffy coat, 72 basic medical data (body measurements and routine blood test, 2715 samples), 49 facial skin imaging scores (2049 samples), 18 entries from psychological

questionnaire (2039 samples), and 56 entries from lifestyle questionnaire (3820 samples) were collected from the same individuals (Table S2; S3).

An inter-omics prediction value between omics data was calculated using a 5-fold cross-validated random forest model (Figure 1A). Basic medical data showed the highest global systematic association with other omics data, the accuracy of prediction from basic medical data to physical fitness and metabolites to basic medical data reaching 75% quantile RFCV R = 0.461 and 0.399, respectively (Figure 1A, B, Figure S1). Medical test and immune indices showed high prediction accuracy to metabolites, especially serum creatinine, BMI, waist to hip ratio, hematocrit, triglyceride in medical test data (Figure 1A, B, C). Metabolites constituted the most important factor to immune indices (R= 0.292), in particular, vitamin K made the greatest contribution to prediction, followed by selenium, argininosuccinic acid, N5-methyltetrahydrofolate and glutamic acid (Figure 1 B, C). Life style also showed high prediction accuracy to physical fitness conditions (R= 0.255 Figure 1A, B, C). Most facial skin features can be predicted by or to predict many omics data type, especially medical test, metabolites and physical fitness factors (Figure 1B, C). Among the life style questionnaire, smoking, drinking (especially low concentration alcohol), sports habits (especially resistance training), high sugar and high fat food diet habit, going to bed late past 0.00 am can be predicted by multiple omics data type (Figure 1B).

A number of factors have been reported to explain gut microbial composition, while the total percentage of variance explained remained in single digits (Falony et al., 2016). In our cohort, defecation frequency and gender were the strongest factors to explain gut microbiome composition, followed by serum uric acid level, triglyceride, testosterone, estradiol, bust circumference, grip strength, as well as psychological symptoms and pores and speckles on the cheek, each explaining more than 0.002 of the variances (PERMANOVA test Figure S2). However, none of these data explained more than 0.1 (10%) of the variances to the gut microbiome composition (Figure S2**)**. Gut microbiome was not a strong factor for other omics either.

These results were confirmed by another RFCV predict model which was a more efficient model than global PERMANOVA in capturing the specific gut microbes association with omics data factors (Figure 2). Gut microbe showed prediction power for plasma levels of vitamins (VA, folic acid, VB5, VD), hormones (testosterone, aldosterone) and trace elements (mercury, selenium, arsenic) and amino acids (BCAA, glutamic acid, tryptophan, histidine) in metabolites. Grip strength, vital capacity, facial skin porphyrins and staying up late past 0.00 am can also be predicted by the gut microbiome. Influence from other lifestyle questionnaire factors such as defecation, yogurt, age, gender, smoking, milk, drinking alcohol, fruit and vegetables were also detected by the RFCV model. Influence from ABO blood group was also detected (Figure 2), 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).

Defecation, hormone and gender

Perhaps due to population differences and a younger age of this cohort (29.6 ± 5.5, average ± sev, Table S1), the gender (female 1016, male 1007) effect on the gut microbiome is clearly evident (Benjamin Hochberg-adjusted p value < 0.05, Figure S3). The females showed a greater microbial diversity (BH-adjusted pvalue =1.52e-12). *Fusobacterium mortiferum, Eubacterium dolichum, Blautia hansenii, Prevotella stercorea* and *Ruminococcus mgs0922* were significantly more abundant in males (Figure S3), despite the absence of colorectal cancer in this young cohort. *Bifidobacterium* and its five species (*B. longum*, *B. adolescentis*, *B. breve*, *B. bifidum*, *B. animalis* and *B.catenulatum-B.pseudocatenulatum* complex) were all significantly enriched in females (Figure S3). GABA (γ-aminobutyric acid) production by bacteria including the female-enriched *Bifidobacterium* and *Bacteroides* could impact on the brain and immune functions in the gut (Strandwitz et al., 2018). The greater gut microbiome gene richness in women could also approximate a larger number of cells per gram of feces (Vandeputte et al., 2017), and influence defecation frequency and regularity.

In the unstratified cohort, testosterone negatively correlated with *Clostridia*. Aldosterone, one of the major adrenal gland mineralcorticoid, positively correlated with three *Bacteroides* species (*B. intestinalis, B. cellulosilyticus, B. stercorirosoris)* and *Eubacterium eligens*. Estradiol/Estrogen positively correlated with *Alistipes obesi*; pyridoxine (Vitamin B6) positively correlated with *Eubacterium siraeum* and *Eubacterium eligens* (Table S5). In addition, plasma levels of the hormones progesterone, estradiol, and aldosterone positively associated with gut microbial enzymes for secondary bile acids (Figure S4). Genes for producing secondary bile acids have been found in bacteria such as *Lactobacillus*, *Streptococcus* and *Bifidobacterium* (Gu et al., 2017), and could affect defecation as well as metabolism (Figure S4).

Facial skin features reflect systemic health

Facial skin imaging data (percentile score relative to VISIA database population of the same gender and age) could be predicted by other omics such as basic medical test and lifestyle (Figure 1). We found that individuals with a higher level of HDLC and water content scored younger than one's actual age. Distinct patterns were seen for males and females (Figure 3). In males, four of the TCR VDJ usage TRBV6−5:TRBJ2−6,TRBV12−3:TRBJ1-1/TRBJ2−1, and TRBV11−1:TRBJ1−4, together with immune diversity score and serum dehydroepiandrosteron correlated with more porphyrins yet less UV stains and brown spots on the face. In females, dehydroepiandrosteron associated with not only porphyrins but also frontal and lateral pores. For females but not males, a

better score from the sleeping and diet assessment corresponded to less porphyrins (Figure 3), providing a strong incentive for keeping healthy habits. Both males and females appeared older than actual age with increased BMI, systolic blood pressure and metabolites associated with cardiometabolic risk, especially lead, glutamate, vitamin D, B1 and B5. However, higher selenium and mercury corresponded to less UV stains in females. In general, the gut microbiome did not have a lot of associations with facial skin features, except for one association between an MGS from Oscillospiraceae and less red zone on the cheeks for females.

The metabolome-immune-gut axis

Among the strongest associations between different omics is that between immune repertoire and plasma metabolites (Figure 1). More strikingly, when we plot 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) (Figure 4). Vitamin A, N5-methyltetrahydrofolicacid, selenium and mercury and serum aldosterone all showed negative associations with a few TRBJ1-4, TRBJ2-1 and positive associations with TRBJ2-4. Vitamin B5, Vitamin E, phosphoserine, arginosuccinic acid and arsenic showed these associations with TRBJ1-4, TRBJ2, TRBJ2-4, as well as positive associations with TRBV20-1 and TRBV3-1. Glutamic acid and serine showed a largely opposite pattern, except for positive associations with TRBV20-1.

When we looked into the gut microbiome, a number of associations could also be identified with the immune repertoire data (Figure 4). Vitamin A is central to a healthy immune system but is typically studied for its role in early development (Veldhoen and Ferreira, 2015). The Clostridiales species likely reduce intestinal level of vitamin A (Figure S5**)** through modulating expression of retinol dehydrogenase 7 and dampen antimicrobial response in the gut (Grizotte-Lake et al., 2018). Vitamin D derivatives did not show significant associations with fecal MGSs, yet largely followed the pattern of N5-methyltetrahydrofolicacid, which positively associated with the butyrate-producing and atherosclerotic cardiovascular disease-depleted bacterium *Eubacterium eligens* (Jie et al., 2017). N5-methyltetrahydrofolicacid also negatively associated with the formate or hydrogen-producing and obesity-related bacteria including *Dorea longicatena* and *Blautia* (Beaumont et al., 2016; Liu et al., 2017; Taras et al., 2002) (Figure 4), implying that folic acid supplementation might modulate fecal levels of these bacteria. Associations between the gut microbiome and plasma levels of trace elements including mercury, selenium and arsenic might be surprising (Figure 4). Selenium-containing rice is commercially promoted as anti-cancer (Stolz and Oremland, 1999), and we found that the association pattern largely followed arsenic, consistent with these two trace elements' similar function in anaerobic respiration. Selenium and mercury also correlated with disease-associated species such as *Clostridium bolteae*, *Streptococcus oralis* and *Ruminococcus gnavus* in the gut microbiome. Mercury, selenium and arsenic also associated with facial skin features, favoring UV spots versus porphyrins, while lead and

magnisium, along with glutamic acid, aminoadipic acid and vitamin B5 correlated with red areas (Figure 3**)**.

The cluster represented by vitamin B5 associated with the butyrate-producing and arsenic-metabolizing bacterium, *Faecalibacterium prausnitzii* (Coryell et al., 2018). *Bacteroides coprophilus*, a prevalent but not very abundant species from the *Bacteroides* genus, showed the opposite pattern (Figure 4). The third cluster represented by glutamic acid showed negative associations with previously reported bacteria implicated with lower BMI such as *Alistipes shahii*, *Bacteroides cellulosilyticus*, *Ruminococcus lactis* and *Eubacterium eligens* (Jie et al., 2017) in this large cohort (Figure 4), consistent with higher glutamic acid in obese individuals (Liu et al., 2017).

Biomarkers for cardiometabolic diseases and colorectal cancer

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 positively correlated with vitamin A, B5, D3 and E, high sugar/high fat diet, amino acids (glutamic acid and alanine) and trace elements (arsenic, mercury and selenium), while negatively associated with defecation frequency, testosterone and aldosterone, and beneficial gut bacteria such as *Faecalibacterium prausnitzii*, *Alistipes shahii*, *Oscillospiraceae* and *Bacteroides intestinalis* (Figure 5A, Figure S6A).

We next defined a score according to 8 blood parameters and 80 fecal microbiome features for cardiometabolic disease risk (see methods) and tested it in previously published case-control samples. Fecal metagenomic samples from Chinese patients with atherosclerotic cardiovascular disease (ACVD), liver cirrhosis, obesity and inflammatory bowel diseases all scored higher compared to control samples without the disease, while those from diseases such as colorectal cancer, rheumatoid arthritis and unstratified T2D did not (Figure 5B) (He et al., 2017; Jie et al., 2017; Liu et al., 2017; Qin et al., 2012, 2014; Yu et al., 2017; Zhang et al., 2015b).

This young multi-omic cohort also helps understand colorectal cancer (CRC). The butyrate-producing *E. eligens* positively associated with fruit and vegetable intake, plasma hydroxyproline, while negatively associated with plasma alanine (Figure 5A). Previously reported CRC-enriched bacteria (Feng et al., 2015; Wang and Jia, 2016; Yu et al., 2017; Zeller et al., 2014) *Peptostreptococcus stomatis* associated with leucine, phenylalanine, alanine, tyrosine in the plasma, *Solobacterium moorei* with leucine and tryptophan, *Fusobacterium mortiferum* with glutamic acid; Enterobactericeae including *Escherichia coli, Klebsiella pneumoniae, Enterobacter cloacae, Citrobacter freundii* positively associated with sarcosine, hydroxylysine, branched chain amino acids, tyrosine, tryptophan, 1−methylhistidine, hydroxyproline, and argininosuccinic acid (Figure 5A, Figure

S6B). *Flavonifractor plautii* and *Bilophila wadsworthia* also showed positive correlation with leucine, 1-methylhistidine and aminobutyric acid (Figure S6B).

Physical fitness, exercising and sleeping reflected in the gut microbiome

Vital capacity, a commonly taken measure for lung function, explained a considerable fraction of variance in the gut microbiome and associated with bacteria such as *A. shahii*, *F. prausnitzii* and *Bifidobacterium adolescentis*, while negatively correlated with cardiometabolic disease-related bacterium *Clostridium clostridioforme*, *Ruminococcus gnavus* and *E. coli* (Figure 2, Figure 6). Grip strength, a known protective factor for cardiovascular casualty, negatively associated with *E. coli* (Figure 6). Vertical jump score negatively associated with *E. coli*, while positively associated with *B. cellulosilyticus*, *B. intestinalis*, *Butyrisimonas virosa*. Reassuringly, sports activities, intensity and time of exercising correlated with better scores in such physical fitness tests. Gut microbiome diversity (Shannon index) associated with favorable scores in most of the fitness tests (Table S4). *Bacteroides cellulosilyticus* and *B. stercorirosoris*, which associated with exercise intensity, even correlated with a faster reaction time (Figure 6).

Besides, those who often stay up after midnight also showed negative correlations with *Haemophilus parainfluenzae, Veillonella tobetsu*ensis and hydroxy vitamin D3/D, while positively correlated with *Clostridium phoceensis*,

mercury, selenium, arsenic, vitamin A, hydroxyproline and phosphoserine (Figure S3, 7). And selenium positively correlated with high level serum uric acid (Figure S6A). Thus, we recommend going to bed earlier from an omics point of view.

A presumably benign species from yogurt in the healthy microbiome

Besides defecation frequency and gender, yogurt consumption explained a considerable portion of variances in the gut microbiome (Figure 2). A recent study casted doubt over the health benefits of yogurt consumption, concluding that colonization of the yogurt strains was highly variable between individuals (Zmora et al., 2018). In our large cohort, *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 (Figure 7A, D). *Bifidobacterium animalis*, probably representing the star strain from CHR HANSEN, *B. animalis* subsp. lactis BB-12, was also enriched in yogurt eaters (Figure 7A). In contrast, there was no significant increase in any *Lactobacillus* strains. The increase in *S. thermophilus* and *B. animalis* by taking yogurt at least reduced the amount of *Clostridium bolteae*, which was known to be elevated in a number of cardiometabolic diseases (Wang and Jia, 2016).

A closer look into the host metadata revealed that yogurt consumption was associated with a number of favorable measurements such as a higher HDL

(high-density lipoprotein) cholesterols, lower triglycerides, better elastic thinking, less depression and lower serum vitamin A (Figure 7A). Consistently, fecal relative abundance of *S. thermophilus* associated with a lower diastolic blood pressure, lower indirect bilirubin, lower hematocrit (Figure 7C). Other intriguing effects included facial skin improvements seen as less redness at the front and less porphyrins; and metabolomic changes such as higher estradiol, proline, lower glycine and ornithine (Figure 7A).

Regarding increasing *Bifidobacterium* in the gut microbiome, however, taking milk was more effective than yogurt in this Chinese cohort (Figure 7A, B), implying that some of the beneficial effects come from the exogenous strains in yogurt. The changes in *Blautia* species due to milk consumption may affect methane metabolism in the gut and susceptibility to inflammatory bowel diseases or irritable bowel syndrome (Halkjær et al., 2018; Manor and Borenstein, 2017).

DISCUSSION

In summary, our trans-omic investigation of thousands volunteers establish an unprecedented reference data set for the healthy gut microbiome. A number of factors in circulation crosstalk with the gut microbiome, and then manifest on the face, in the heads and in fitness tests. Levels of trace elements, such as heavy metals, as important cofactors for bacteria respiration and other functions, should be measured even in uncontaminated regions. The TCR usage in such a large

cohort is a great resource for discovering microbial antigens other than those from traditional pathogens.

Potential influence of physical activity on the gut microbiota has been analyzed in rugby athletes (Barton et al., 2017) and colorectal cancer (Feng et al., 2015). Compliance to recordings such as Fitbit is notoriously bad in healthy individuals (Price et al., 2017). 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.

Our large-scale analyses provide substantial support for health benefits of yogurt consumption. The universally present species were *Streptococcus thermophilus* and to a lesser extend *Bifidobacterium animalis* instead of commonly tested probiotics from *Lactobacillus*. Yet, all of these live or dead probiotics could nonetheless exert functions on the immune system or even the brain. Faced with skyrocketing incidences of metabolic diseases, we have yogurt as an inexpensive option.

ACKNOWLEDGMENTS

The authors are very grateful to colleagues at BGI-Shenzhen for sample collection, DNA extraction, library construction, sequencing, and discussions. We thank Chenchen for helpful comments.

AUTHOR CONTRIBUTIONS

J.W. initiated the overall health project. Y.Z., H.Z., X.X. organized the sample collection, with immune repertoire from X.L., W.Z, serum metabolomics from W.R., Y.L., D.Z., X.Q., X.C., 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. Z.J. led the bioinformatic analyses, including S.L. and F.L. H.Z., Q.D., S.T., D.W. performed early analyses which did not make it into the manuscript. Z.J., H.J. interpreted the data. H.J., S.L., Z.J. wrote the manuscript. All authors contributed to finalizing this manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interest.

REFERENCES

Barton, W., Penney, N.C., Cronin, O., Garcia-Perez, I., Molloy, M.G., Holmes, E., Shanahan, F., Cotter, P.D., and O'Sullivan, O. (2017). The microbiome of professional athletes differs from that of more sedentary subjects in composition and particularly at the functional metabolic level. Gut gutjnl-2016-313627.

Beaumont, M., Goodrich, J.K., Jackson, M.A., Yet, I., Davenport, E.R., Vieira-Silva, S., Debelius, J., Pallister, T., Mangino, M., Raes, J., et al. (2016). Heritable components of the human fecal microbiome are associated with visceral fat. Genome Biol. *17*, 189.

Coryell, M., McAlpine, M., Pinkham, N. V., McDermott, T.R., and Walk, S.T. (2018). The gut microbiome is required for full protection against acute arsenic toxicity in mouse models. Nat. Commun. *9*, 5424.

Deschasaux, M., Bouter, K.E., Prodan, A., Levin, E., Groen, A.K., Herrema, H., Tremaroli, V., Bakker, G.J., Attaye, I., Pinto-Sietsma, S.-J., et al. (2018). Depicting the composition of gut microbiota in a population with varied ethnic origins but shared geography. Nat. Med.

Falony, G., Joossens, M., Vieira-Silva, S., Wang, J., Darzi, Y., Faust, K., Kurilshikov, A., Bonder, M.J., Valles-Colomer, M., Vandeputte, D., et al. (2016). Population-level analysis of gut microbiome variation. Science (80-.). *352*, 560–564.

Fang, C., Zhong, H., Lin, Y., Chen, B., Han, M., Ren, H., Lu, H., Luber, J.M., Xia, M., Li, W., et al. (2018). Assessment of the cPAS-based BGISEQ-500 platform for metagenomic sequencing. Gigascience *7*.

FDA Arsenic in Rice and Rice Products.

Feng, Q., Liang, S., Jia, H., Stadlmayr, A., Tang, L., Lan, Z., Zhang, D., Xia, H., Xu, X.X., Jie, Z., et al. (2015). Gut microbiome development along the colorectal adenoma–carcinoma sequence. Nat. Commun. *6*, 6528.

Grizotte-Lake, M., Zhong, G., Duncan, K., Kirkwood, J., Iyer, N., Smolenski, I., Isoherranen, N., and Vaishnava, S. (2018). Commensals Suppress Intestinal Epithelial Cell Retinoic Acid Synthesis to Regulate Interleukin-22 Activity and Prevent Microbial Dysbiosis. Immunity *49*, 1103–1115.e6.

Halkjær, S.I., Christensen, A.H., Lo, B.Z.S., Browne, P.D., Günther, S., Hansen, L.H., and Petersen, A.M. (2018). Faecal microbiota transplantation alters gut microbiota in patients with irritable bowel syndrome: results from a randomised, double-blind placebo-controlled study. Gut gutjnl-2018-316434.

He, Q., Gao, Y., Jie, Z., Yu, X., Laursen, J.M.J.M., Xiao, L., Li, Y., Li, L., Zhang, F., Feng, Q., et al. (2017). Two distinct metacommunities characterize the gut microbiota in Crohn's disease patients. Gigascience *6*, 1–11.

He, Y., Wu, W., Zheng, H.-M., Li, P., McDonald, D., Sheng, H.-F., Chen, M.-X., Chen, Z.-H., Ji, G.-Y., Zheng, Z.-D.-X., et al. (2018). Regional variation limits applications of healthy gut microbiome reference ranges and disease models. Nat. Med. *24*, 1532–1535.

Jie, Z., Xia, H., Zhong, S.-L., Feng, Q., Li, S., Liang, S., Zhong, H., Liu, Z., Gao, Y., Zhao, H., et al. (2017). The gut microbiome in atherosclerotic cardiovascular disease. Nat. Commun. *8*, 845.

Jie, Z., Chen, C., Hao, L., Li, F., Tian, L., and Tong, X. (2019). The vagino-cervical microbiome as a woman's life history. Under Consid. Nature.

Li, J., Jia, H., Cai, X., Zhong, H., Feng, Q., Sunagawa, S., Arumugam, M., Kultima, J.R.J.R., Prifti, E., Nielsen, T., et al. (2014). An integrated catalog of reference genes in the human gut microbiome. Nat. Biotechnol. *32*, 834–841.

Liu, R., Hong, J., Xu, X.X., Feng, Q., Zhang, D., Gu, Y., Shi, J., Zhao, S., Liu, W., Wang, X.X., et al. (2017). Gut microbiome and serum metabolome alterations in obesity and after weight-loss intervention. Nat. Med. *23*, 859–868.

Louppe, G., Wehenkel, L., Sutera, A., and Geurts, P. (2013). Understanding variable importances in forests of randomized trees. In Advances in Neural Information Processing Systems 26, C.J.C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K.Q. Weinberger, eds. (Curran Associates, Inc.), pp. 431–439.

Manor, O., and Borenstein, E. (2017). Systematic Characterization and Analysis of the Taxonomic Drivers of Functional Shifts in the Human Microbiome. Cell Host Microbe *21*, 254–267.

Marbach, D., Costello, J.C., Küffner, R., Vega, N.M., Prill, R.J., Camacho, D.M., Allison, K.R., DREAM5 Consortium, Kellis, M., Collins, J.J., et al. (2012). Wisdom of crowds for robust gene network inference. Nat. Methods *9*, 796–804.

Moll1, J.M., Rausch, P., Eriksen, C., Nielsen, H.B., Feng, Q., Xia, H., and Brix, S. (2019). Pathway-based approach to classify unannotated gut bacteria associated with insulin sensitivity. Under Revis. Nat Microbiol.

Nielsen, H.B., Almeida, M., Juncker, A.S., Rasmussen, S., Li, J., Sunagawa, S., Plichta, D.R., Gautier, L., Pedersen, A.G., Le Chatelier, E., et al. (2014). Identification and assembly of genomes and genetic elements in complex metagenomic samples without using reference genomes. Nat. Biotechnol. *32*, 822–828.

Price, N.D., Magis, A.T., Earls, J.C., Glusman, G., Levy, R., Lausted, C., McDonald, D.T., Kusebauch, U., Moss, C.L., Zhou, Y., et al. (2017). A wellness study of 108 individuals using personal, dense, dynamic data clouds. Nat. Biotechnol. *35*, 747–756.

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

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

Stolz, J.F., and Oremland, R.S. (1999). Bacterial respiration of arsenic and selenium. FEMS Microbiol. Rev. *23*, 615–627.

Strandwitz, P., Kim, K.H., Terekhova, D., Liu, J.K., Sharma, A., Levering, J., McDonald, D., Dietrich, D., Ramadhar, T.R., Lekbua, A., et al. (2018). GABA-modulating bacteria of the human gut microbiota. Nat. Microbiol.

Taras, D., Simmering, R., Collins, M.D., Lawson, P.A., and Blaut, M. (2002). 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.

USGS Arsenic and Drinking Water.

Vandeputte, D., Kathagen, G., D'hoe, K., Vieira-Silva, S., Valles-Colomer, M., Sabino, J., Wang, J., Tito, R.Y., De Commer, L., Darzi, Y., et al. (2017). Quantitative microbiome profiling links gut community variation to microbial load. Nature *551*, 507–511.

Veldhoen, M., and Ferreira, C. (2015). Influence of nutrient-derived metabolites on lymphocyte immunity. Nat. Med. *21*, 709–718.

Wang, J., and Jia, H. (2016). Metagenome-wide association studies: fine-mining the microbiome. Nat. Rev. Microbiol. *14*, 508–522.

Xiao, L., Feng, Q., Liang, S., Sonne, S.B., Xia, Z., Qiu, X., Li, X., Long, H., Zhang, J., Zhang, D., et al. (2015). A catalog of the mouse gut metagenome. Nat. Biotechnol. *33*, 1103–1108.

Xiao, L., Estellé, J., Kiilerich, P., Ramayo-Caldas, Y., Xia, Z., Feng, Q., Liang, S., Pedersen, A.Ø., Kjeldsen, N.J.N.J., Liu, C., et al. (2016). A reference gene catalogue of the pig gut microbiome. Nat. Microbiol. *1*, 16161.

Yu, J., Feng, Q., Wong, S.H.S.H., Zhang, D., Liang, Q.Y., Qin, Y., Tang, L., Zhao, H., Stenvang, J., Li, Y., et al. (2017). Metagenomic analysis of faecal microbiome as a tool towards targeted non-invasive biomarkers for colorectal cancer. Gut *66*, 70–78.

Zeller, G., Tap, J., Voigt, A.Y., Sunagawa, S., Kultima, J.R., Costea, P.I., Amiot, A., Böhm, J., Brunetti, F., Habermann, N., et al. (2014). Potential of fecal microbiota for early-stage detection of colorectal cancer. Mol. Syst. Biol. *10*, 766.

Zhang, W., Du, Y., Su, Z., Wang, C., Zeng, X., Zhang, R., Hong, X., Nie, C., Wu, J., Cao, H., et al. (2015a). IMonitor: A Robust Pipeline for TCR and BCR Repertoire Analysis. Genetics *201*, 459–472.

Zhang, X., Zhang, D., Jia, H., Feng, Q., Wang, D., Liang, D. Di, Wu, X., Li, J.J.J., Tang, L., Li, Y.Y.Y., et al. (2015b). The oral and gut microbiomes are perturbed in rheumatoid arthritis and partly normalized after treatment. Nat. Med. *21*, 895–905.

Zhao, L., Wang, G., Siegel, P., He, C., Wang, H., Zhao, W., Zhai, Z., Tian, F., Zhao, J., Zhang, H., et al. (2013). Quantitative genetic background of the host influences gut microbiomes in chickens. Sci. Rep. *3*, 1163.

Zhernakova, A., Kurilshikov, A., Bonder, M.J., Tigchelaar, E.F., Schirmer, M., Vatanen, T., Mujagic, Z., Vila, A.V., Falony, G., Vieira-Silva, S., et al. (2016). Population-based metagenomics analysis reveals markers for gut microbiome composition and diversity. Science (80-.). *352*, 565–569.

Zmora, N., Zilberman-Schapira, G., Suez, J., Mor, U., Dori-Bachash, M., Bashiardes, S., Kotler, E., Zur, M., Regev-Lehavi, D., Brik, R.B.-Z., et al. (2018). Personalized Gut Mucosal Colonization Resistance to Empiric Probiotics Is Associated with Unique Host and Microbiome Features. Cell *174*, 1388–1405.e21.

FIGURE LEGENDS

Figure 1. Overview of the interrelationship between omics.

(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 source omics. Title in top is target omics. Each node in box is a target factor. The color of the node and box line indicated the source omics data type. The fill color of box indicated blue means predict using all samples, yellow means used female samples only, and purple means used male samples only. Y-axis is the target factor RFCV R predicted from source omics in x-axis tick label. (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.

Figure 2. Factors associated with gut microbiome.

Top 45 factors with RFCV R > 0.1 in each type of omics that are predicted by gut microbiome. 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 independently.

Figure 3. The association of facial skin imaging.

The omics factors association with facial skin is sex specific using multi-omics analyses method (see methods). The cell color of heat map indicated the spearman correlation. +, BH-adjusted pvalue<0.1; *, BH-adjusted pvalue <0.05; **, BH-adjusted pvalue <0.01.

Figure 4. The metabolome-immune-gut axis

Immune repertoire is clustered into 3 groups by fecal microbiome and plasma metabolites. Factors are selected using multi-omics analyses method (see methods). The cell color of triadic heatmap shows the partial spearman correlation adjust age and gender. BH-adjusted p-value is denoted: +, q-value<0.1; *, q-value<0.05; **, q-value<0.01.

Figure 5. Gut microbial markers for hyperuricemia and other cardiometabolic diseases.

(A) Association map of the four-tiered analyses integrating the metabolites, clinical indices, life style and the fecal microbiome. The color of heat map shows the partial spearman correlation adjusted age and gender. BH-adjusted p-value is denoted: +, q-value<0.1; *, q-value<0.05; **, q-value<0.01. (B) Microbial metabolic syndrome risk index was computed for eight additional cardiometabolic cohort. The box depicts the interquartile range (IQR) between the first and third quartiles (25th and 75th percentiles, respectively) and the line inside denotes the median, whereas the points represent the microbial metabolic syndrome risk index in each sample

Figure 6. Gut microbiome associated with physical fitness and exercise.

The color of heat map shows the partial spearman correlation adjusted age, gender and BMI. BH adjusted p-value is denoted: +, q-value<0.1; *, q-value<0.05; **, q-value<0.01

Figure 7. Influence of yogurt and milk intake on omic.

(A-C) The top 5 factor in each omics data associated with yogurt, milk intake habit and the *Streptococcus thermophiles* abundance. The length of the bars represents partial spearman's correlation coefficient adjusted age, gender.

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. (D) Fecal relative abundance of *S. thermophilus* in volunteers with increasing frequency of yogurt consumption.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Study Cohort

All the >2000 volunteers were recruited between May 2017 and July 2017 during physical examination in BGI-Shenzhen, China, the samples in each omics are shown in Table S3. Baseline characteristics of the cohort are shown in Table S2.

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

METHODS DETAILS

Demographic Data Collection

The lifestyle questionnaire contained age, disease history, eating and exercise habits and so on. The psychological questionnaire contained the evaluation of irritability, dizziness, frustration, fear, appetite, self-confidence, resilience and so on (Table S2).

Samples Collection

Fecal samples were self-collected by volunteers, using a kit containing a room temperature stabilizing reagent to preserve the metagenome (Han et al., 2018). The samples were frozen on the same day. The overnight fasting blood samples were drawn from a cubital vein of volunteers by the doctors.

DNA extraction and metagenomics shotgun sequencing

DNA extraction of the stored fecal samples within the next few months was performed as previously described (Qin et al., 2012). Metagenomic sequencing

was done on the BGISEQ-500 platform (100bp of singled-end reads for fecal samples and four libraries were constructed for each lane) (Fang et al., 2018).

Amino Acid Measurements

40 µl plasma was deproteinized with 20 µl 10% (w/v) sulfosalicylic acid (Sigma) containing internal standards, then 120 µl aqueous solution was added. After centrifuged, the supernatant was used for analysis. The analysis was performed by ultra high pressure liquid chromatography (UHPLC) coupled to an AB Sciex Qtrap 5500 mass spectrometry (AB Sciex, US) with the electrospray ionization (ESI) source in positive ion mode. A Waters ACQUITY UPLC HSS T3 column (1.8 µm, 2.1 × 100 mm) was used for amino compound separation with a flow rate at 0.5 ml/min and column temperature of 55 °C. The mobile phases were (A) water containing 0.05% and 0.1% formic acid (v/v), (B) acetonitrile containing 0.05% and 0.1% formic acid (v/v). The gradient elution was 2% B 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) HNO_3,2mg/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, China). The DNA was sequenced on the BGISEQ- 500 platform using 200 bp singled-end reads. The data processing was performed using Immune IMonitor(Zhang et al., 2015a).

Medical Parameters

All the volunteers were recruited during the physical examination. The medical test including blood tests, urinalysis, routine examination of cervical secretion and so on. All the medical parameters were measured by the physical examination center and shown in Table S2.

Facial Skin feature

The volunteer's frontal face without makeup was photographed by VISIA-CRTM imaging system (Canfield Scientific, Fairfield, NJ, USA) equipped with chin supports and forehead clamps that fix the face during the photographing process and maintain a fixed distance between the volunteers and the camera at all times. Eight indices were obtained including spots, pores, wrinkles, texture, UV spots, porphyrins, brown spots and red area from the cheek and forehead, respectively

(Table S2). The percentile of index was calculated based on the index value ranked in the age-matched database, the higher the better (Table S2).

Physical fitness test

8 kinds of tests were performed to evaluate volunteers' physical fitness condition (Table S2). Vital capacity was measured by HK6800-FH (Hengkangjiaye, China). Eye-closed and single-legged standing was measured by HK6800-ZL. Choice reaction time was measured by HK6800-FY. Grip strength was measured by HK6800-WL. Sit and reach was measured by HK6800-TQ. Sit-ups was measured by HK6800-YW. Step index was measured by HK6800-TJ. Vertical jump was measured by HK6800-ZT. We got a measure value from each test. Then each measure value score was assigned grades A through E based on its corresponding age-matched database.

QUANTIFICATION AND STATISTICAL ANALYSIS

Quality control, Taxonomic annotation and abundance calculation

The sequencing reads were quality-controlled as described previously(Fang et al., 2018).

Taxonomic assignment of the high-quality fecal metagenomic data was performed using the reference gene catalog comprising 9,879,896 genes(Li et al., 2014). Taxonomy of the fecal MGSs/MLGs were then determined from their constituent genes, as previously described(Moll1 et al., 2019; Nielsen et al., 2014; Qin et al., 2012; Wang and Jia, 2016).

Gut microbiota composition explained by factors in each type of omics.

To assess the effect of each factor in each type of omics on gut microbiome, we performed PERMANOVA (Adonis) test on the MGS/MLG abundance profiles using euclidean distance and 999 permutations in the vegan R package. Factors with BH adjusted P<0.05 were considered to associate with MGS/MLG (Figure S2B).

The global effect size between fecal microbiome and omics data.

The cumulative contribution of metadata variables/omics data lists was determined by forward stepwise redundancy analysis on MGS profile with the forward.sel function in packfor R package with variables that was sub-selected by envfit function in vegan (Figure S2A).

The factors in each type of omics predicted by fecal microbiome

The factors in each type of omics were regressed against the relative abundances of mgs profile (found in at least 10% of the samples) in the fecal samples using default parameters in the RFCV function from randomForest package in R. All the categorical variables were converted into continue variables, and nominal variables were converted into dummy variables. RFCV R defined as spearman's correlation between measured value and 5-fold cross-validation predicted value was calculated, then rank the top 5 predictable factors in each omics type. The same prediction process was done between any two types of omics. Then ggplot package in R was used to boxplot predict power of target omics factors by all kinds of other predictor omics (Figure 1B). 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 (Figure 1A). R package pheatmap and barplot was used to make heatmap plot for some representative factors (Figure 1C, Figure 2, Figure S1).

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

An rank average method(Marbach et al., 2012) was used to combine the results of multiple inference methods to make a robust omics association network. In our study, we consider to combine variance importance from randomForest that was an all vs one non-linear regression model and spearman's correlation that was a one vs one non-linear model.

Step 1: data prepare.

All categorical variables in other omics data were converted into continuous variables, and nominal variables were converted into dummy variables. Missing values were filled with median and variables with more than 70% missing were dropped. The microbial species less than10% in all the samples were also removed. Near zero variance variables was removed. Outliers were defined as outside of the 95% quartiles and removed.

Step 2: multiple inference methods perform.

Random forest variable importance was used to identify the most important predictor variables(Louppe et al., 2013). RFCV regression function from randomForest package in R with default parameter was used to get the 5-fold average variable importance. Spearman's correlation with the cor.test function in base R software. Partial spearman's correlation with ppcor package.

Step 3: Robust networks construction.

The average rank (source factors -target factors edge from any two different omics) was computed as sum of the ranks across the RFCV importance and spearman rho. For the scales of variables, the RFCV importance was rank within each target factor in each omics. For each omics type, we just kept the top 20 target factors with highest RFCV R. Second we kept top 30 average ranks for each target variable and retained edges with spearman's correlation BH-adjusted pvalue <0.05. 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 spearman correlation between all gut microbes and all metabolites. 20 of 107 metabolites with highest RFCV were kept. For each of the 20 select metabolites such as VA, average ranks across RFCV and spearman were done. At last gut microbe biomarker for VA was found with average rank top 30th and passed the spearman BH-adjusted pvalue <0.05.

Step 4: network visualize.

Top N average ranks for each target as representative factors was used to make barplots using ggplot2 package (Figure 2,7). The pheatmap package was used to plot the common representative factors that could be strongest predicted by multiple omics data type(Figure 1C). All the source-target factors pair RFCVR (a as source, b as target and b as source, a as target) was boxplot (Figure1B) using ggplot2. The ComplexHeatmap package in R was used to plot omics triadic relation (Figure 3-6). CytoScape was also used to visualized the global omics

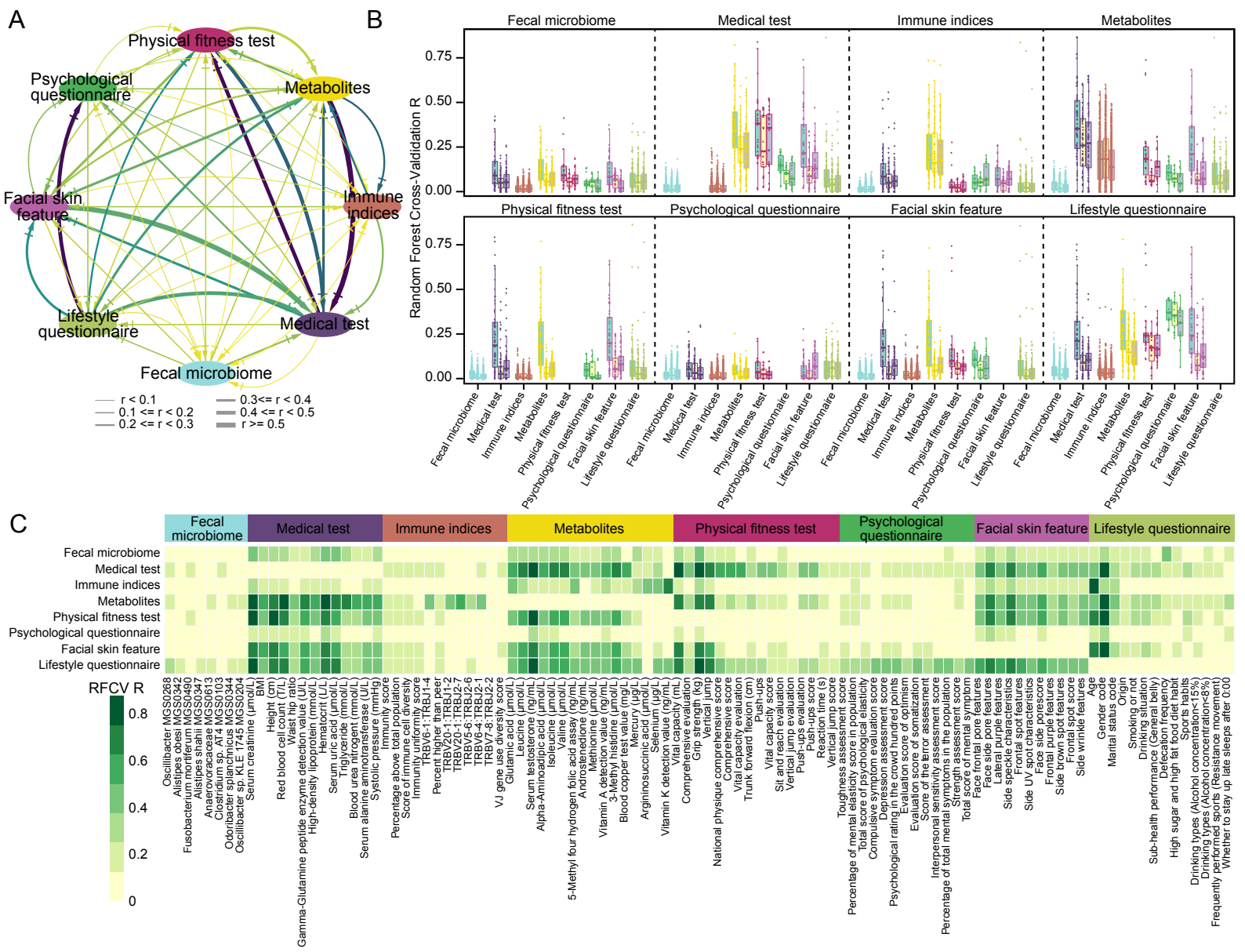
network (Figure 1A). R package igraph was used to plot omics factors network that are all significant correlate to serum uric acid (Figure S6A).

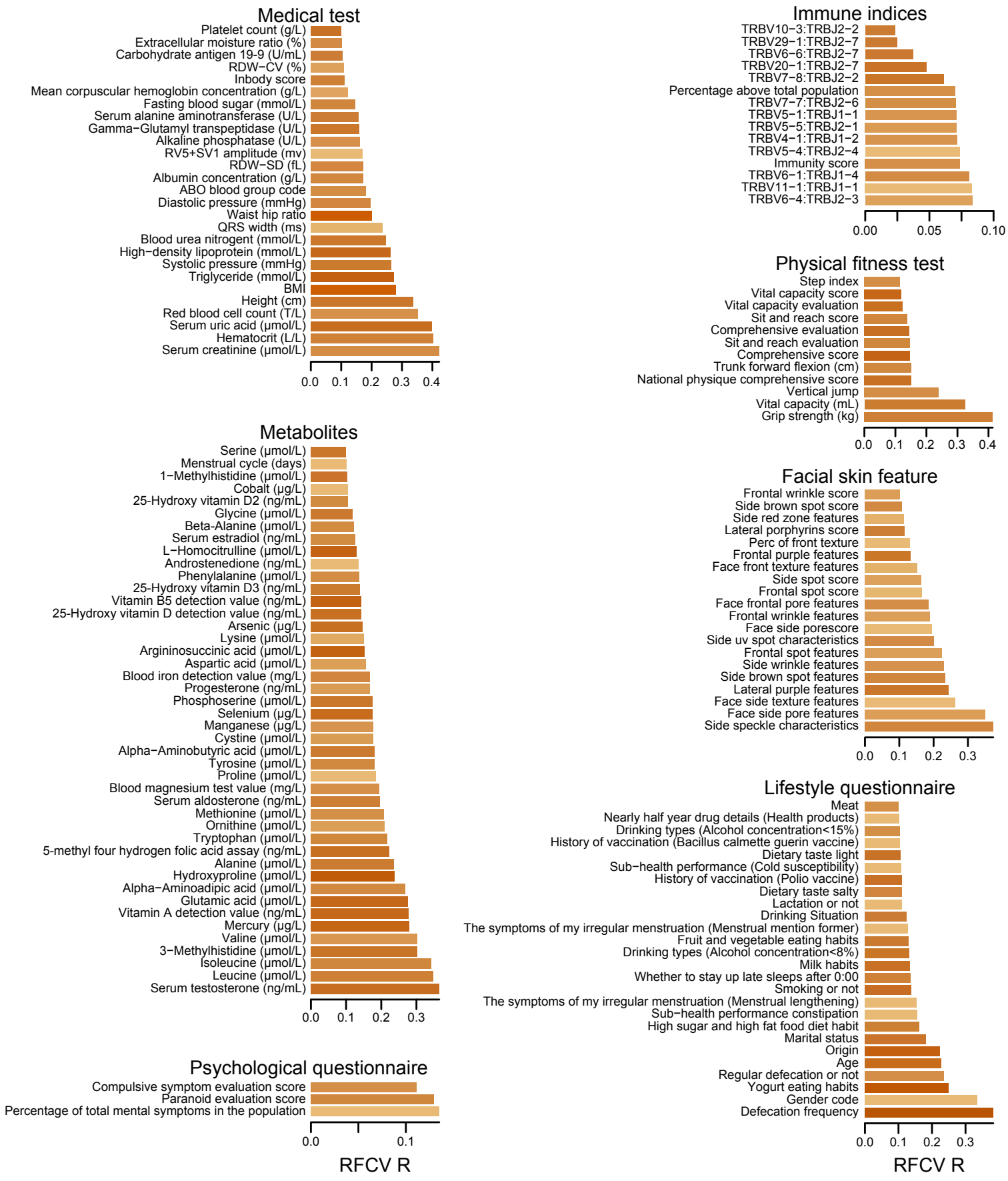
Microbial metabolic syndrome risk index validation in cardiometabolic cohort.

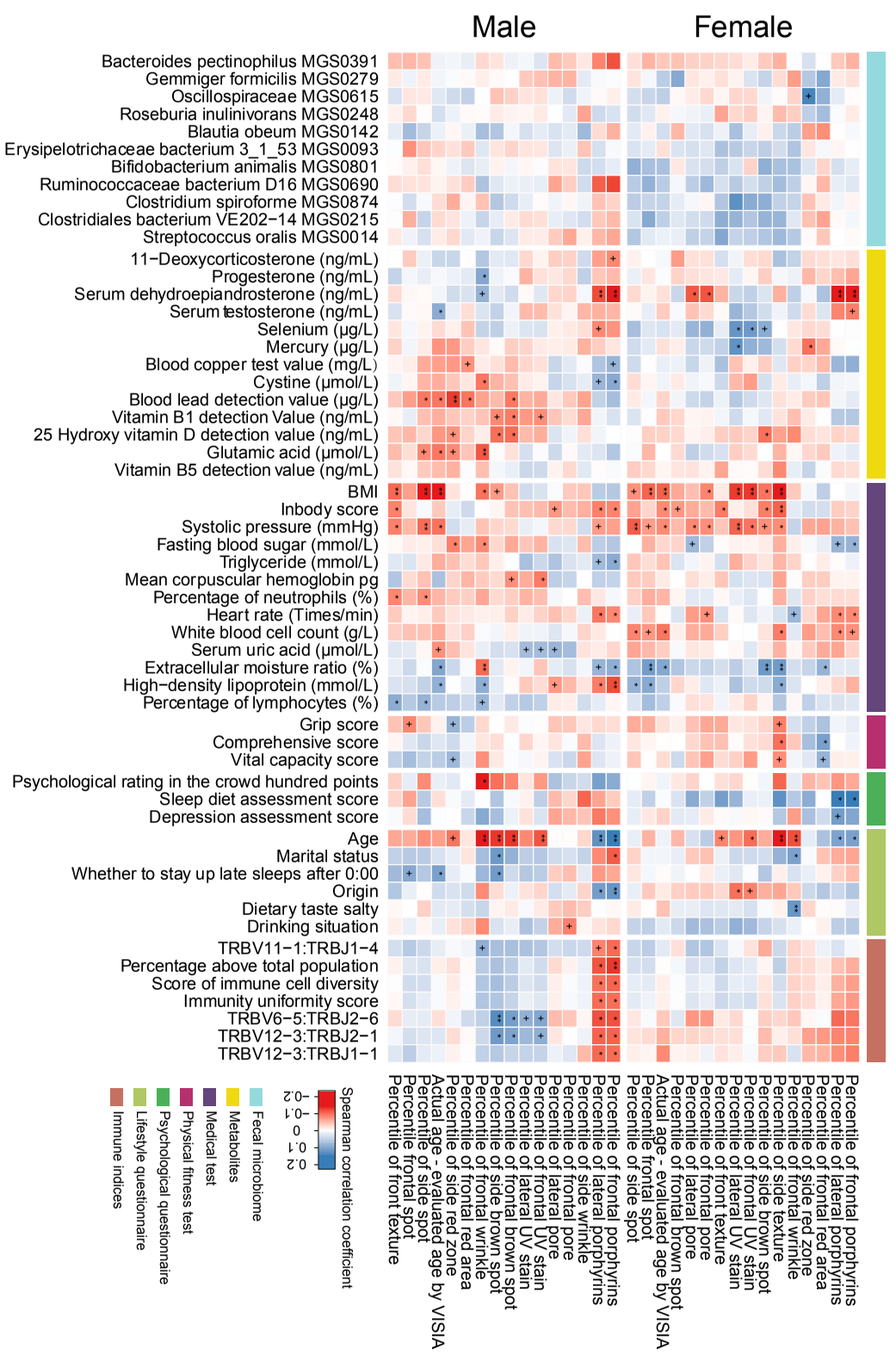
Using multi-omics analyses method, based on the current cohort we pick up 80 MGS significant correlate with one of the eight cardiometabolic risk factors (Waist Hip Ratio, BMI, Triglyceride (mmol/L),High-Density-Lipoprotein (mmol/L),Serum Uric Acid (μ mol/L), γ -Glutamyl Transpeptidase (U/L),Serum Alanine Aminotransferase(U/L),Fasting Blood Sugar (mmol/L)). 12 of 80 MGS was defined as bad cardiometabolic microbe for they were positively correlate with BMI and Triglyceride. The rest 68 MGS was defined as good with cardiometabolic microbe who are negatively correlate with BMI and Triglyceride. For each MGS and each validation chinese disease cohort samples (ACVD, CD, UC, HongKong CRC, Cirrhosis, Obesity, RA, T2D) we get the percentage rank based onthe current cohort as reference. Then microbial metabolic syndrome risk index for new sample computed as average bad cardiometabolic microbe percentage rank minus average good cardiometabolic microbe percentage rank. Rank sum test was done between the disease and healthy group to test the microbial metabolic syndrome risk index performance in validation disease cohort.

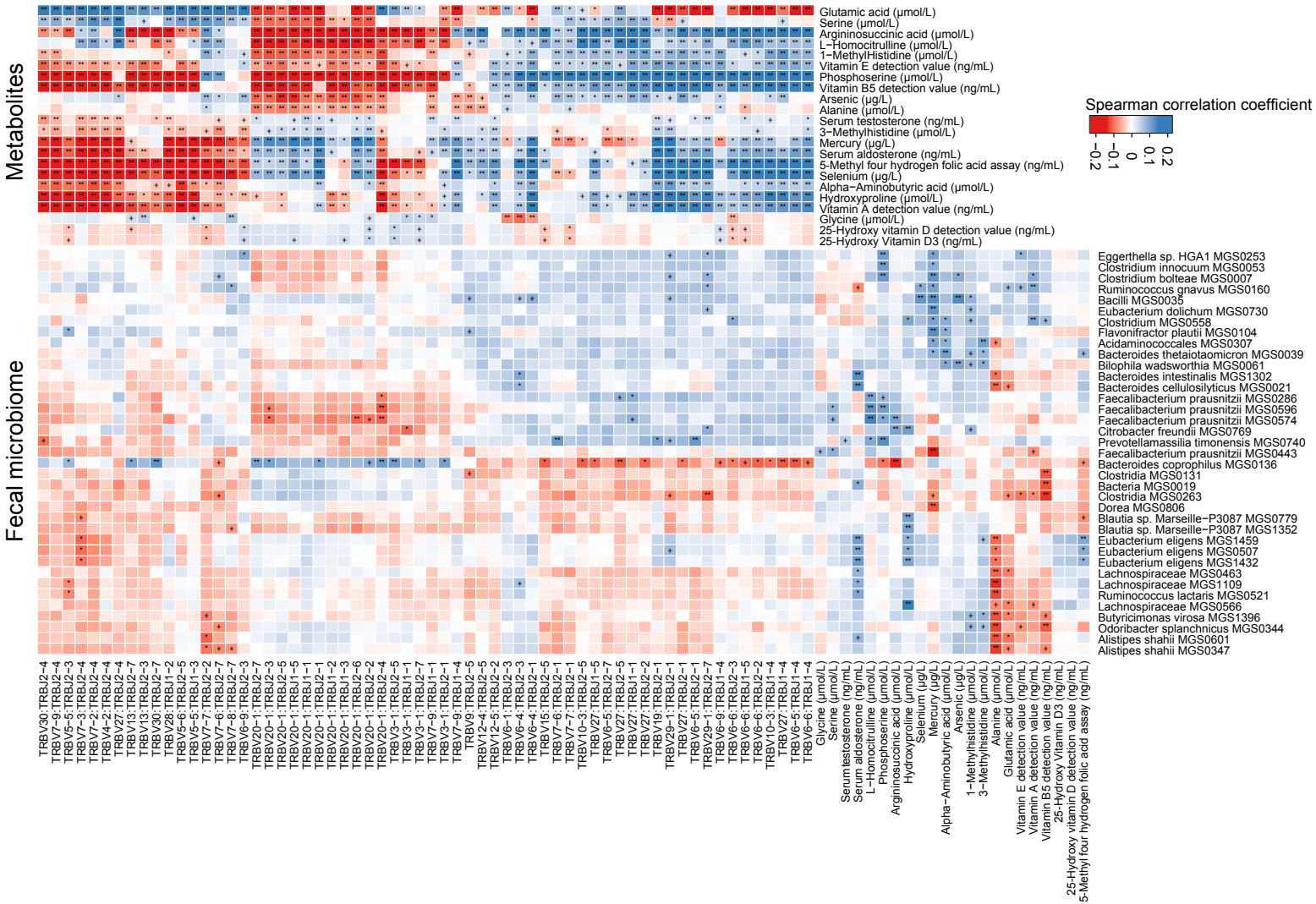
DATA AND SOFTWARE AVAILABILITY

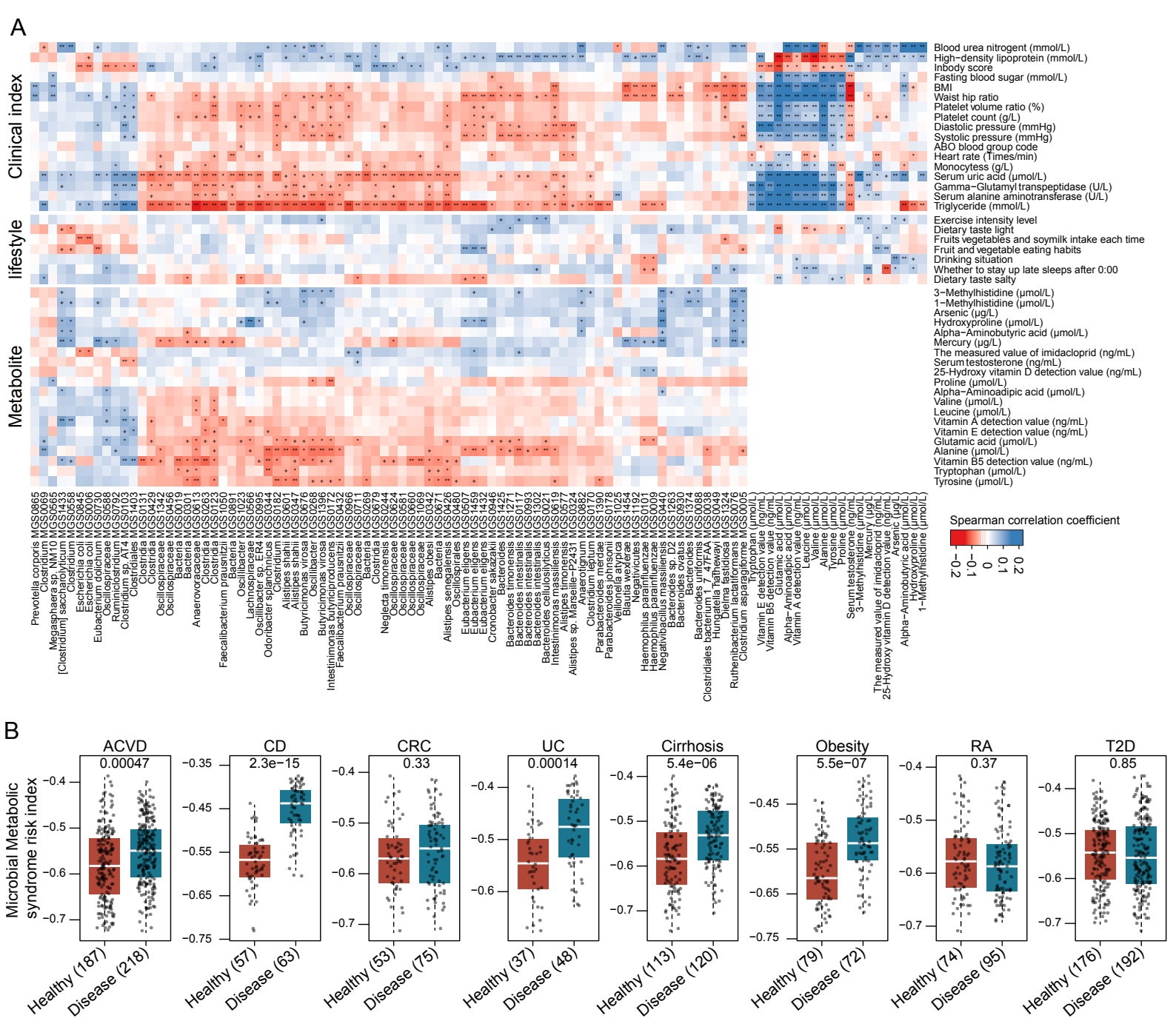
Metagenomic sequencing data for all samples have been deposited to the (CNGB) database under the accession code CNPxx.


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
Random Forest Cross-Validation R
0.00
0.25
0.50
0.75
C
Fecal microbiome
Medical test
Immune indices
Metabolites
Physical fitness test
Psychological questionnaire
Facial skin feature
Lifestyle questionnaire
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. Marseille-P3170 MGS0103
Odoribacter splanchnicus MGS0344
Oscillibacter sp. KLE 1745 MGS0204
Serum creatinine (μmol/L)
BM
Height (cm)
Red blood cell count (T/L)
Waist hip ratio
Gamma-Glutamine peptide enzyme detection value (U/L)
High-density lipoprotein (mmol/L)
Hematocrit (L/L)
Triglyceride (mmol/L)
Serum uric acid (μmol/L)
Blood urea nitrogent (mmol/L)
Serum alanine aminotransferase (U/L)
Systolic pressure (mmHg)
Immunity score
Percentage above total population
Score of immune cell diversity
Immune cell richness 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)
Alpha-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)
Vitamin A detection value (μg/mL)
3-Methyl histidine (μmol/L)
Blood copper test value (mg/L)
Mercury (μg/L)
Argininosuccinic acid (μmol/L)
Selenium (μg/L)
Vitamin K detection value (ng/mL)
Vital capacity (mL)
Comprehensive evaluation
Grip strength (kg)
Vertical jump
National physique comprehensive score
Comprehensive score
Vital capacity evaluation
Trunk forward flexion (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 side pore 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 habits
Drinking types (Alcohol concentration<15%)
Drinking types (Alcohol concentration<8%)
Frequently performed sports (Resistance movement)
Whether to stay up late sleeps after 0:00

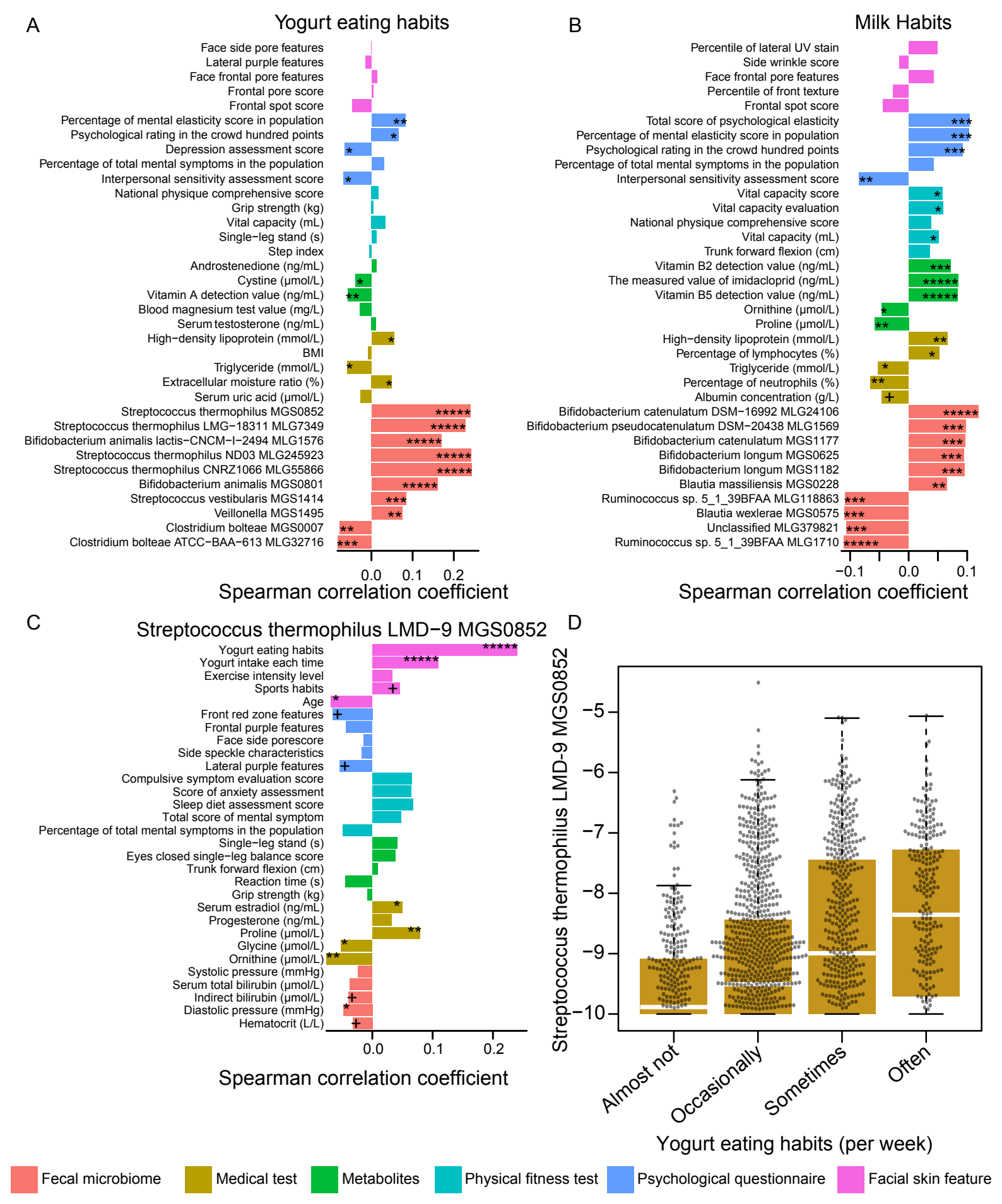

Medical test
Platelet count (g/L)
Extracellular moisture ratio (%)
Carbohydrate antigen 19-9 (U/mL)
RDW−CV (%)
Inbody score
Mean corpuscular hemoglobin concentration (g/L)
Fasting blood sugar (mmol/L)
Serum alanine aminotransferase (U/L)
Gamma−Glutamyl transpeptidase (U/L)
Alkaline phosphatase (U/L)
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)
0.0 0.1 0.2 0.3 0.4
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
0.00 0.05 0.10
Physical fitness test
Step index
Vital capacity score
Vital capacity evaluation
Sit and reach score
Comprehensive evaluation
Sit and reach evaluation
Comprehensive score
Trunk forward flexion (cm)
National physique comprehensive score
Vertical jump
Vital capacity (mL)
Grip strength (kg)
0.0 0.1 0.2 0.3 0.4
Metabolites
Serine (μmol/L)
Menstrual cycle (days)
1−Methylhistidine (μmol/L)
Cobalt (μg/L)
25-Hydroxy vitamin D2 (ng/mL)
Glycine (μmol/L)
Beta-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 detection value (ng/mL)
25-Hydroxy vitamin D detection value (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)
Alpha−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)
Alpha−Aminoadipic acid (μmol/L)
Glutamic acid (μmol/L)
Vitamin A detection value (ng/mL)
Mercury (μg/L)
Valine (μmol/L)
3−Methylhistidine (μmol/L)
Isoleucine (μmol/L)
Leucine (μmol/L)
Serum testosterone (ng/mL)
0.0 0.1 0.2 0.3
Facial skin feature
Frontal wrinkle score
Side brown spot score
Side red zone features
Lateral porphyrins score
Perc of front texture
Frontal purple features
Face front texture features
Side spot score
Frontal spot score
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
0.0 0.1 0.2 0.3
Lifestyle questionnaire
Meat
Nearly half year drug details (Health products)
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
Lactation or not
Drinking Situation
The symptoms of my irregular menstruation (Menstrual mention former)
Fruit and vegetable eating habits
Drinking types (Alcohol concentration<8%)
Milk habits
Whether to stay up late sleeps after 0:00
Smoking or not
The symptoms of my irregular menstruation (Menstrual lengthening)
Sub−health performance constipation
High sugar and high fat food diet habit
Marital status
Origin
Age
Regular defecation or not
Yogurt eating habits
Gender code
Defecation frequency
0.0 0.1 0.2 0.3
Psychological questionnaire
Compulsive symptom evaluation score
Paranoid evaluation score
Percentage of total mental symptoms in the population
0.0 0.1
RFCV R
RFCV R


Male
Female
Bacteroides pectinophilus MGS0391
Gemmiger formicilis MGS0279
Oscillospiraceae MGS0615
Roseburia inulinivorans MGS0248
Blautia obeum MGS0142
Erysipelotrichaceae bacterium 3_1_53 MGS0093
Bifidobacterium animalis MGS0801
Ruminococcaceae bacterium D16 MGS0690
Clostridium spiroforme MGS0874
Clostridiales bacterium VE202−14 MGS0215
Streptococcus oralis MGS0014
11−Deoxycorticosterone (ng/mL)
Progesterone (ng/mL)
Serum dehydroepiandrosterone (ng/mL)
Serum testosterone (ng/mL)
Selenium (μg/L)
Mercury (μg/L)
Blood copper test value (mg/L)
Cystine (μmol/L)
Blood lead detection value (μg/L)
Vitamin B1 detection Value (ng/mL)
25 Hydroxy vitamin D detection value (ng/mL)
Glutamic acid (μmol/L)
Vitamin B5 detection value (ng/mL)
BMI
Inbody score
Systolic pressure (mmHg)
Fasting blood sugar (mmol/L)
Triglyceride (mmol/L)
Mean corpuscular hemoglobin pg
Percentage of neutrophils (%)
Heart rate (Times/min)
White blood cell count (g/L)
Serum uric acid (μmol/L)
Extracellular moisture ratio (%)
High−density lipoprotein (mmol/L)
Percentage of lymphocytes (%)
Grip score
Comprehensive score
Vital capacity score
Psychological rating in the crowd hundred points
Sleep diet assessment score
Depression assessment score
Age
Marital status
Whether to stay up late sleeps after 0:00
Origin
Dietary taste salty
Drinking situation
TRBV11−1:TRBJ1−4
Percentage above total population
Score of immune cell diversity
Immunity uniformity score
TRBV6−5:TRBJ2−6
TRBV12−3:TRBJ2−1
TRBV12−3:TRBJ1−1
Percentile of front texture
Percentile frontal spot
Percentile of side spot
Actual age - evaluated age by VISIA
Percentile of side red zone
Percentile of frontal red area
Percentile of frontal wrinkle
Percentile of side brown spot
Percentile of frontal brown spot
Percentile of lateral UV stain
Percentile of frontal UV stain
Percentile of lateral pore
Percentile of frontal pore
Percentile of side wrinkle
Percentile of lateral porphyrins
Percentile of frontal porphyrins
Percentile of front texture
Percentile frontal spot
Percentile of side spot
Actual age - evaluated age by VISIA
Percentile of frontal brown spot
Percentile of lateral pore
Percentile of frontal pore
Percentile of front texture
Percentile of lateral UV stain
Percentile of frontal UV stain
Percentile of side brown spot
Percentile of side texture
Percentile of frontal wrinkle
Percentile of side red zone
Percentile of frontal red area
Percentile of lateral porphyrins
Percentile of frontal porphyrins
Spearman correlation coefficient
0.2
0.1
0
−0.1
−0.2
Fecal microbiome
Metabolites
Medical test
Physical fitness test
Psychological questionnaire
Lifestyle questionnaire
Immune indices


Metabolites
Fecal microbiome
Glutamic acid (μmol/L)
Serine (μmol/L)
Argininosuccinic acid (μmol/L)
L−Homocitrulline (μmol/L)
1−MethylHistidine (μmol/L)
Vitamin E detection value (ng/mL)
Phosphoserine (μmol/L)
Vitamin B5 detection value (ng/mL)
Arsenic (μg/L)
Alanine (μmol/L)
Serum testosterone (ng/mL)
3−Methylhistidine (μmol/L)
Mercury (μg/L)
Serum aldosterone (ng/mL)
5-Methyl four hydrogen folic acid assay (ng/mL)
Selenium (μg/L)
Alpha−Aminobutyric acid (μmol/L)
Hydroxyproline (μmol/L)
Vitamin A detection value (ng/mL)
Glycine (μmol/L)
25-Hydroxy vitamin D detection value (ng/mL)
25-Hydroxy Vitamin D3 (ng/mL)
Spearman correlation coefficient
−0.2
−0.1
0
0.1
0.2
Eggerthella sp. HGA1 MGS0253
Clostridium innocuum MGS0053
Clostridium bolteae MGS0007
Ruminococcus gnavus MGS0160
Bacilli MGS0035
Eubacterium dolichum MGS0730
Clostridium MGS0558
Flavonifractor plautii MGS0104
Acidaminococcales MGS0307
Bacteroides thetaiotaomicron MGS0039
Bilophila wadsworthia MGS0061
Bacteroides intestinalis MGS1302
Bacteroides cellulosilyticus MGS0021
Faecalibacterium prausnitzii MGS0286
Faecalibacterium prausnitzii MGS0596
Faecalibacterium prausnitzii MGS0574
Citrobacter freundii MGS0769
Prevotellamassilia timonensis MGS0740
Faecalibacterium prausnitzii MGS0443
Bacteroides coprophilus MGS0136
Clostridia MGS0131
Bacteria MGS0019
Clostridia MGS0263
Dorea MGS0806
Blautia sp. Marseille−P3087 MGS0779
Blautia sp. Marseille−P3087 MGS1352
Eubacterium eligens MGS1459
Eubacterium eligens MGS0507
Eubacterium eligens MGS1432
Lachnospiraceae MGS0463
Lachnospiraceae MGS1109
Ruminococcus lactaris MGS0521
Lachnospiraceae MGS0566
Butyricimonas virosa MGS1396
Odoribacter splanchnicus MGS0344
Alistipes shahii MGS0601
Alistipes shahii MGS0347
TRBV30:TRBJ2−4
TRBV7−9:TRBJ2−4
TRBV5−5:TRBJ2−3
TRBV7−3:TRBJ2−4
TRBV7−2:TRBJ2−4
TRBV4−2:TRBJ2−4
TRBV27:TRBJ2−4
TRBV13:TRBJ2−3
TRBV30:TRBJ2−7
TRBV28:TRBJ1−2
TRBV5−6:TRBJ2−5
TRBV5−5:TRBJ1−3
TRBV7−7:TRBJ2−2
TRBV7−6:TRBJ2−7
TRBV7−8:TRBJ2−7
TRBV6−9:TRBJ2−3
TRBV20−1:TRBJ2−7
TRBV20−1:TRBJ2−3
TRBV20−1:TRBJ2−5
TRBV20−1:TRBJ1−5
TRBV20−1:TRBJ1−1
TRBV20−1:TRBJ1−2
TRBV20−1:TRBJ1−3
TRBV20−1:TRBJ2−6
TRBV20−1:TRBJ2−2
TRBV20−1:TRBJ2−4
TRBV3−1:TRBJ2−5
TRBV3−1:TRBJ1−1
TRBV3−1:TRBJ2−7
TRBV7−9:TRBJ1−1
TRBV3−1:TRBJ2−1
TRBV7−9:TRBJ1−4
TRBV9:TRBJ2−5
TRBV12−4:TRBJ2−5
TRBV12−5:TRBJ2−2
TRBV6−1:TRBJ2−3
TRBV6−4:TRBJ2−3
TRBV15:TRBJ2−5
TRBV7−6:TRBJ2−1
TRBV7−7:TRBJ2−1
TRBV10−3:TRBJ2−5
TRBV27:TRBJ1−5
TRBV6−5:TRBJ2−7
TRBV27:TRBJ1−1
TRBV27:TRBJ2−5
TRBV27:TRBJ2−2
TRBV19:TRBJ2−1
TRBV29−1:TRBJ2−1
TRBV27:TRBJ2−1
TRBV6−5:TRBJ1−1
TRBV29−1:TRBJ2−7
TRBV6−9:TRBJ1−4
TRBV6−6:TRBJ2−3
TRBV6−6:TRBJ1−5
TRBV6−6:TRBJ2−2
TRBV10−3:TRBJ1−4
TRBV27:TRBJ1−4
TRBV6−5:TRBJ1−4
TRBV6−6:TRBJ1−4
Glycine (μmol/L)
Serine (μmol/L)
Serum testosterone (ng/mL)
Serum aldosterone (ng/mL)
L−Homocitrulline (μmol/L)
Phosphoserine (μmol/L)
Argininosuccinic acid (μmol/L)
Hydroxyproline (μmol/L)
Selenium (μg/L)
Mercury (μg/L)
Alpha−Aminobutyric acid (μmol/L)
Arsenic (μg/L)
1−Methylhistidine (μmol/L)
3−Methylhistidine (μmol/L)
Alanine (μmol/L)
Glutamic acid (μmol/L)
Vitamin E detection value (ng/mL)
Vitamin A detection value (ng/mL)
Vitamin B5 detection value (ng/mL)
25-Hydroxy Vitamin D3 (ng/mL)
25-Hydroxy vitamin D detection value (ng/mL)
5-Methyl four hydrogen folic acid assay (ng/mL)


A
Clinical index
lifestyle
Metabolite
Blood urea nitrogent (mmol/L)
High−density lipoprotein (mmol/L)
Inbody score
Fasting blood sugar (mmol/L)
BMI
Waist hip ratio
Platelet volume ratio (%)
Platelet count (g/L)
Diastolic pressure (mmHg)
Systolic pressure (mmHg)
ABO blood group code
Heart rate (Times/min)
Monocytess (g/L)
Serum uric acid (μmol/L)
Gamma−Glutamyl transpeptidase (U/L)
Serum alanine aminotransferase (U/L)
Triglyceride (mmol/L)
Exercise intensity level
Dietary taste light
Fruits vegetables and soymilk intake each time
Fruit and vegetable eating habits
Drinking situation
Whether to stay up late sleeps after 0:00
Dietary taste salty
3−Methylhistidine (μmol/L)
1−Methylhistidine (μmol/L)
Arsenic (μg/L)
Hydroxyproline (μmol/L)
Alpha−Aminobutyric acid (μmol/L)
Mercury (μg/L)
The measured value of imidacloprid (ng/mL)
Serum testosterone (ng/mL)
25-Hydroxy vitamin D detection value (ng/mL)
Proline (μmol/L)
Alpha−Aminoadipic acid (μmol/L)
Valine (μmol/L)
Leucine (μmol/L)
Vitamin A detection value (ng/mL)
Vitamin E detection value (ng/mL)
Glutamic acid (μmol/L)
Alanine (μmol/L)
Vitamin B5 detection value (ng/mL)
Tryptophan (μmol/L)
Tyrosine (μmol/L)
Spearman correlation coefficient
−0.2
−0.1
0
0.1
0.2
B
Microbial Metabolic syndrome risk index
ACVD
0.00047
CD
2.3e−15
CRC
0.33
UC
0.00014
Cirrhosis
5.4e−06
Obesity
5.5e−07
RA
0.37
T2D
0.85
Healthy (187)
Disease (218)
Healthy (57)
Disease (63)
Healthy (53)
Disease (75)
Healthy (37)
Disease (48)
Healthy (113)
Disease (120)
Healthy (79)
Disease (72)
Healthy (74)
Disease (95)
Healthy (176)
Disease (192)

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 (Genera belly)
National physique comprehensive score
Eyes closed single−leg balance score
Choice reaction time score
Vertical jump score
Vital capacity score
Step index
Sit and reach score
Grip 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
Yogurt eating habits
Face side pore features
Lateral purple features
Face frontal pore features
Frontal pore score
Frontal spot score
Percentage of mental elasticity score in population
Psychological rating in the crowd hundred points
Depression assessment score
Percentage of total mental symptoms in the population
Interpersonal sensitivity assessment score
National physique comprehensive score
Grip strength (kg)
Vital capacity (mL)
Single−leg stand (s)
Step index
Androstenedione (ng/mL)
Cystine (μmol/L)
Vitamin A detection value (ng/mL)
Blood magnesium test value (mg/L)
Serum testosterone (ng/mL)
High−density lipoprotein (mmol/L)
BMI
Triglyceride (mmol/L)
Extracellular moisture ratio (%)
Serum uric acid (μmol/L)
Streptococcus thermophilus MGS0852
Streptococcus thermophilus LMG−18311 MLG7349
Bifidobacterium animalis lactis−CNCM−I−2494 MLG1576
Streptococcus thermophilus ND03 MLG245923
Streptococcus thermophilus CNRZ1066 MLG55866
Bifidobacterium animalis MGS0801
Streptococcus vestibularis MGS1414
Veillonella MGS1495
Clostridium bolteae MGS0007
Clostridium bolteae ATCC−BAA−613 MLG32716
0.0
0.1
0.2
Spearman correlation coefficient
B
Milk Habits
Percentile of lateral UV stain
Side wrinkle score
Face frontal pore features
Percentile of front texture
Frontal spot score
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 score
Vital capacity evaluation
National physique comprehensive score
Vital capacity (mL)
Trunk forward flexion (cm)
Vitamin B2 detection value (ng/mL)
The measured value of imidacloprid (ng/mL)
Vitamin B5 detection value (ng/mL)
Ornithine (μmol/L)
Proline (μmol/L)
High−density lipoprotein (mmol/L)
Percentage of lymphocytes (%)
Triglyceride (mmol/L)
Percentage of neutrophils (%)
Albumin concentration (g/L)
Bifidobacterium catenulatum DSM−16992 MLG24106
Bifidobacterium pseudocatenulatum DSM−20438 MLG1569
Bifidobacterium catenulatum MGS1177
Bifidobacterium longum MGS0625
Bifidobacterium longum MGS1182
Blautia massiliensis MGS0228
Ruminococcus sp. 5_1_39BFAA MLG118863
Blautia wexlerae MGS0575
Unclassified MLG379821
Ruminococcus sp. 5_1_39BFAA MLG1710
−0.1
0.0
0.1
Spearman correlation coefficient
C
Streptococcus thermophilus LMD−9 MGS0852
Yogurt eating habits
Yogurt intake each time
Exercise intensity level
Sports habits
Age
Front red zone features
Frontal purple features
Face side porescore
Side speckle characteristics
Lateral purple features
Compulsive symptom evaluation score
Score of anxiety assessment
Sleep diet assessment score
Total score of mental symptom
Percentage of total mental symptoms in the population
Single−leg stand (s)
Eyes closed single−leg balance score
Trunk forward flexion (cm)
Reaction time (s)
Grip strength (kg)
Serum estradiol (ng/mL)
Progesterone (ng/mL)
Proline (μmol/L)
Glycine (μmol/L)
Ornithine (μmol/L)
Systolic pressure (mmHg)
Serum total bilirubin (μmol/L)
Indirect bilirubin (μmol/L)
Diastolic pressure (mmHg)
Hematocrit (L/L)
0.0
0.1
0.2
Spearman correlation coefficient
D
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