Disease trends in a young Chinese cohort according to fecal metagenome and plasma metabolites

Jie, Zhuye; Liang, Suisha; Ding, Qiuxia; Li, Fei; Sun, Xiaohuan; Lin, Yuxiang; Chen, Peishan; Cai, Kaiye; Zhou, Hongcheng; Lu, Haorong; Wang, Xiaohan; Zhang, Tao; Xiao, Liang; Yang, Huanming; Wang, Jian; Hou, Yong; Kristiansen, Karsten; Jia, Huijue; Xu, Xun

Published in:
Medicine in Microecology

DOI:
10.1016/j.medmic.2021.100037

Publication date:
2021

Document version
Publisher's PDF, also known as Version of record

Document license:
CC BY-NC-ND

Citation for published version (APA):
Perspective

Disease trends in a young Chinese cohort according to fecal metagenome and plasma metabolomes

Zhuye Jie\textsuperscript{a,b,c,*,1}, Suisha Liang\textsuperscript{a,1}, Qiuxia Ding\textsuperscript{a,1}, Fei Li\textsuperscript{a}, Xiaohuan Sun\textsuperscript{a}, Yuxiang Lin\textsuperscript{a}, Peishan Chen\textsuperscript{a}, Kaiye Cai\textsuperscript{a}, Hongcheng Zhou\textsuperscript{a}, Haorong Lu\textsuperscript{a}, Xiaohan Wang\textsuperscript{a}, Tao Zhang\textsuperscript{a,c}, Liang Xiao\textsuperscript{a,f,g}, Huamning Yang\textsuperscript{a,h}, Jian Wang\textsuperscript{a,h}, Yong Hou\textsuperscript{a}, Karsten Kristiansen\textsuperscript{a,c,g}, Huijue Jia\textsuperscript{a,b,c,**}, Xun Xu\textsuperscript{a}

\textsuperscript{a} BGI-Shenzhen, Shenzhen, China
\textsuperscript{b} Shenzhen Key Laboratory of Human Commensal Microorganisms and Health Research, BGI-Shenzhen, Shenzhen, China
\textsuperscript{c} Laboratory of Genomics and Molecular Biomedicine, Department of Biology, Universitetsparken 13, University of Copenhagen, Copenhagen, Denmark
\textsuperscript{e} China National Genebank, BGI-Shenzhen, Shenzhen, 518120, China
\textsuperscript{f} Shenzhen Engineering Laboratory of Detection and Intervention of Human Intestinal Microbiome, BGI-Shenzhen, Shenzhen, China
\textsuperscript{g} BGI-Qingdao, BGI-Shenzhen, Qingdao, 266555, China
\textsuperscript{h} James D. Watson Institute of Genome Sciences, Hangzhou, China

ARTICLE INFO

Keywords:
Gut microbiome
Shotgun sequencing
Multi-omic
Young individuals
Uric acid
Disease

ABSTRACT

Most of the disease studies for the gut microbiome have collected cases and control samples from the elderly or the middle-aged. Despite general interest in microbiome health, it is not known how microbial biomarkers from metagenome-wide association studies (MWAS) would perform in a cohort of young individuals, who would be largely free of chronic diseases, as well as medication. Here we analyze high-depth fecal metagenomic shotgun sequencing for 2183 healthy adults with clinical parameters, diet, lifestyle, and metabolite measurements. We provide the first set of large-scale evidence for gut microbiome dysbiosis in hyperuricemia, which relates to meat intake. We build a cardiometabolic disease risk model based on gut microbes for initial screening in a young population and confirm the validity using external cohorts. Fecal bacteria that have been reported to be enriched in colorectal cancer (CRC) are found to correlate with methylhistidines, branched-chain amino acids (BCAA), aromatic amino acids and glutamic acid in these young individuals, which were validated by an additional cohort of 1404 individuals. Our comprehensive data suggest that the gut microbiome could show trends towards diseases years before onset, and the results lay the foundation for the design of larger screens for cardiometabolic diseases and CRC with clinically meaningful cutoffs.

1. Introduction

The gut microbiome has been implicated in a growing list of complex diseases, showing great potential for the diagnosis and treatment of cardiovascular [1], metabolic [2], autoimmune [3] and neurological diseases [4] and cancer [5].

Metagenome-wide association studies (MWAS) for colorectal cancer (CRC) and autoimmune diseases are beginning to provide a converging list of fecal biomarkers that distinguish between healthy and diseased individuals in cohorts from different countries [6,7]. Metabolic disease [2] is the first type of disease associated with the microbiota studied by MWAS. SCFA-producing bacteria, such as Clostridiales sp. SS3/4 [8], have been found enriched in type 2 diabetes compared with control individuals. Clostridioides hathewayi [8] and Clostridioides bolteae [9] enriched on the control individuals. Obesity is another metabolic disease frequently reported associated with gut microbiota. The ratio of Bacteroidetes and Firmicutes increased in obesity [10] has been found in many animal and human studies. Altered gut microbiome in gout patients also has been reported. Antibiotic treatment has shown the potential influence of gut microbiome on uric acid metabolism in mice. Furthermore,
fetal microbiota transplantation has been shown to improve gout symptoms. Biomarkers for cardiometabolic diseases and neuropsychiatric disorders are more controversial both due to the larger sample size required for weaker signals, and due to medication use.

Most of the disease association have involved participants who were elderly or middle-aged. But the result in a young reference cohort is still lacking. Here we analyze multi-omic data, including the data from fecal microbiome, plasma metabolites, medical test, physical fitness, and psychological and lifestyle questionnaire, from a discovery cohort of 2183 volunteers and a validation cohort of 1404 volunteers from different regions. Both discovery and validation cohorts have an average age of no more than 30 years old [11]. From microbiome and metabolome data, we captured associations for hyperuricemia, cardiometabolic diseases, and colorectal cancer. In addition, amino acids including methylhistidines, branched-chain amino acids (BCAA), aromatic amino acids (AAA) and glutamate associated with intake of meat were correlated with CRC-enriched bacteria in these young individuals based on gut microbiome data.

2. Materials and methods

2.1. Study cohort

As part of 4D-SZ, all the >2000 volunteers for the first cohort were recruited between May 2017 and July 2017 during a physical examination in Shenzhen. The 1400 volunteers for the second cohort were also recruited in 2017 until late September in multiple cities in China, including Wuhan, Qingdao, Tianjin. Detailed sample information regarding the cohorts has been reported [11]. The omics use of samples is shown in Supplementary Table 1c [11]. Baseline characteristics of the cohort are shown in Supplementary Table 1b [11], 1d [11].

The study was approved by the Institutional Review Boards (IRB) at BGI-Shenzhen (BGI-IRB19121), and all participants provided written informed consent at enrolment.

2.2. Demographic data collection

The lifestyle questionnaire contained 56 entries involving age, marital status, disease history of the volunteer and his/her family, eating and exercise habits [Supplementary Table 1b [11], 1d [11]]. The psychological questionnaire contained 18 entries for the evaluation of irritability, dizziness, frustration, fear, appetite, self-confidence, resilience [Supplementary Table 1b [11]].

2.3. Samples collection

Fecal samples were self-collected by the volunteers preferably on the same morning, using a kit containing a room temperature stabilizing reagent to preserve the metagenome [12], which preserves the fecal metagenome for at least two weeks at room temperature. The samples were frozen at −80 °C on the same day and only thawed for DNA extraction. The overnight fasting blood samples were drawn from a cubital vein of volunteers by medical doctors.

2.4. DNA extraction and metagenomics shotgun sequencing

DNA extraction of the stored fecal samples within a few months after collection was performed as previously described [5]. This manual extraction protocol has been described in detail [13], and used in many studies [14–17]. Briefly, a frozen aliquot of fecal sample was suspended in 250 μl of guanidine thiocyanate, 0.1 M Tris (pH 7.5) and 40 μl of 10% N-lauroyl sarcosine. Then 500 μl 5% N-lauroyl sarcosine was added. After 1 h incubation, 500 μl of glass beads (0.1 mm) and 500 μl of TENP were added to the tube for vortexing followed by centrifugation. The supernatant was transferred to a new tube and DNA was precipitated by isopropanol [12]. Metagenomic sequencing was performed on the BGISEQ-500 platform (100bp of single-end reads for fecal samples, and four libraries were constructed for each lane), and quality-controlled as previously reported [18]. The library construction and sequencing protocol for BGISEQ-500 were PCR-free, and size selection was not necessary.

2.5. Amino acid measurements

40 μl plasma was deproteinized with 20 μl 10% (w/v) sulfoalicylic acid (Sigma) containing internal standards, then 120 μl aqueous solution (distilled water) was added. After centrifugation, 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 0.5% and 0.1% formic acid (v/v), (B) acetonitrile containing 0.5% 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).

2.6. Hormone measurements

250 μl plasma was diluted with 205 μl aqueous solution (distilled water). For SPE experiments, HLB (Waters, USA) was activated with 1.0 ml of dichloromethane, acetonitrile, methanol, respectively and equilibrated with 1.0 ml of water. The pretreated plasma sample was loaded onto the cartridge 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 Phenomenex 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 1 mM Ammonium acetate, (B) Methanol containing 1 mM Ammonium acetate. The gradient elution was 25% B kept for 0.9 min, then changed linearly to 40% B during 0.9 min, 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.

2.7. 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) HNO3,2 mg/L AU, and internal standards (20 μg/L). The mixture was sonicated for 10 min before ICP-MS analysis. Multi-element determination was performed on an Agilent 7700× ICP-MS (Agilent Technologies, Tokyo, Japan) equipped with an octopole
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 mm diameter injector with a Shield Torch system, a Scott double-pass spray chamber and nickel cones (Agilent Technologies, Tokyo, Japan). A glass concentric Micro-Mist nebulizer (Agilent Technologies, Tokyo, Japan) was used for the analysis of diluted samples.

2.8. 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-y1-13C3) hydrochloride (Sigma-Aldrich), levomefolic acid-13C, d₃, riboflavin-13C, 15N2, 4-pyridoxic acid-d₃ and pantoteninic acid-13C3,15 N hemi calcium salt (Toronto Research Chemicals). 500 μl supernatant was dried by nitrogen flow. 60 μl water 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 a Waters Xevo TQ-S Triple quadrupole mass spectrometry (Waters, USA) with the electrospray ionization (ESI) source in positive ion mode. A Waters ACQUITY UPLC HSS T3 column (1.7 μm, 1.5 × 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, 1.0%–1.0% B; 1–1.5 min, 1.0%–3.0% B; 1.5–2 min, 3.0%–30.0% B; 2–3.5 min, 30%–70% B; 3.5–4.0 min, 70%–90% B; 4.0–4.8 min, 90%–90% B; 4.9–6.0 min, 1.0%–1.0% B. The factors in each type of omics were regressed against the relative abundances of MGS and MLGs. We also compute the Metaphlan2 [20] profile (v2.6.0) with default parameter.

2.11. Physical fitness test

Eight kinds of tests were performed to evaluate volunteers’ physical fitness condition (Supplementary Table 1b [11]). 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-FT. 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 measured value from each test. Then each measure value score was assigned 1 through 5 based on its corresponding age-matched national standards (Supplementary Table 5 [11]). Both the direct measurements and the scores were used for analyses (Supplementary Table 2 [11], Supplementary Table 3 [11]).

2.12. Quality control, taxonomic annotation and abundance calculation

The sequencing reads were quality-controlled as described previously [18] and implementation of the pipeline is available at https://github.com/jiezhuyu/cOGM. Briefly, adapter trimming/filtering is automatically processed by the BGISEQ-500 sequencing platform. The raw sequences were quality filtered and trimmed by with overall accuracy (OA) control strategy [18] using OAs2 (-Qys = 33, -minLen = 30, -Scut = 0.9, -Qcut = 0.8), and high-quality reads aligned to hg19 using SOAP2.22 (identity ≥ 0.9) to remove human reads by removeHost (-D 4 -s 30 -r 1 -v 7 -i 0.9). The retained clean reads were aligned to the integrated gene catalog (IGC) by using SOAP2.22 (identity ≥ 0.95). Adapter trimming/filtering is automatically processed by the BGISEQ-500 sequencing platform. Taxonomic assignment of the high-quality fecal metagenomic data was performed using the reference gene catalog comprising 9,879, 896 gene [13] by using SOAP2.22 (identity ≥ 0.95). Taxonomy of the fecal MGSs/MLGs were then determined from their constituent genes, as previously described [6,18,19]. That is trimmed mean (remove the highest and lowest 5% abundance genes) of genes abundance those belong to the same MGSs/MLGs. We also compute the Metaphlan2 [20] profile (v2.6.0) with default parameter.

2.13. The factors predicted by gut microbiome

The factors in each type of omics were regressed against the relative abundances of MGS profiles (found in at least 10% of the samples) in the fecal samples using default parameters in the RFCV function from randomForest package in R. Dichotomous variables (such as gender) and unordered categorical variable (such as region) were recoded into dummy variables. Frequency items such as yogurt eating habit were assigned integers. RFCV R defined as Spearman’s correlation between the measured value and 5-fold cross-validation predicted value was calculated.

2.14. Statistical adjustment for factors that may influence the gut microbiome

Associations between gut microbiome MGSs, functional modules, Shannon diversity, and variance explained and other type omics data were all adjusted for factors that probably influence the gut microbiome by linear model or partial Spearman correlation, 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) as in recent studies [21]. 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.

2.15. Mathematical transformation for compositional data

We use mmvec [22,23] to re-rank and highlight associations picked up by average ranks methods (see Materials and methods 2.17), which account for the composition effect. Gut microbiome relative abundance profile is compositional data (sum to 1). “mmvec” (microbe-metabolite vectors) was a compositional technique that can handle multi-omics datasets. It reports co-occurrence probabilities between microbes and metabolites by two-layer neural networks. Softmax transform enforces scale invariance to remove the composition effect. It has several users defined hyperparameter. In our study, we set “–latent-dim 3 –min-feature-count 10 –learning-rate 1e-5 –beta1 0.9 –beta2 0.99”. Both the logloss and cross-entropy curves decaying, and plateau is close to zero, which suggest a good model. Other hyperparameter can set the same good fitting too, such as “–latent-dim 3 –min-feature-count 10 –learning-rate 1e-5 –beta1 0.9 –beta2 0.95” and “–latent-dim 5 –min-feature-count 10 –learning-rate 1e-5 –beta1 0.85 –beta2 0.9”. But the ranks will vary small across different hyperparameter set. And also the ranks vary by the log (x+1) or mm transform of metabolite profile in R package edgeR [24]. So the best rank for a paired metabolite and microbe among different hyper parameter set and transform was reported in Supplementary Table 4b [11].

2.16. Benjamini-Hochberg multiple hypothesis testing correction

The multiple hypotheses testing Benjamini-Hochberg corrections were done for one source-target omics pair each time. We show two versions of Benjamini-Hochberg correction for Shannon and other type omics in Supplementary Table 2a [11]. One of the BH adjustments was done within one omics each time. Another adjustment was done overall on all omics.

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

A rank average method [25] 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.

2.17.1. Step 1: data preprocessing

Dichotomous variables (such as gender) and unordered categorial variable (such as region) were re-coded 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 were removed

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

For each factor in one omics, we performed regression using RFCV function with default parameter based on all factors in one other omics and calculated RFCV R [26]. 5-fold average variable importance was output for step 3. Partial Spearman's correlations (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 [27]) with default parameters after adjusting for the above confounders.

2.17.3. Step 3: Robust networks construction

To get the robust and strongest association between factors from any two type of omics, in other words, to filter predictor factors and target factors, we performed two steps. First, to choose the target factors, we just kept the top 20 target factors with the highest RFCV R. Then to choose microbe factors (MGSs) for every selected target factor, we kept MGSs with top 30 average ranks and retained edges with partial Spearman's correlation BH-adjusted pvalue <0.05. The average rank was computed as the 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 computed the 5-fold cross-validation predict power (RFCV R) for each metabolite and partial Spearman correlation. 20 metabolites with the 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 30th and passed the partial Spearman BH-adjusted pvalue <0.05.

2.17.4. Step 4: Network visualization

For each target factor, the top 5–10 average ranks MGSs in each source omics type were selected as representative factors to make barplots using ggplot2 package or pheatmap package. The ComplexHeatmap package in R was used to plot omics triadic relation.

2.18. Microbial metabolic syndrome risk index validation in cardiometabolic cohort

Using multi-omics analyses method described above after controlling for the potential confounders, we picked 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 (mmol/L), γ-glutamyl transpeptidase (U/L), serum alanine aminotransferase (U/L), fasting blood glucose (mmol/L)) (Fig. 2, Supplementary Fig. 2). We linked the MGSs to the BCAA metabolites (valine/leucine/alanine), tryptophan, glutamic acid (q < 0.1, Supplementary Table 3a [11]). For the published disease studies from China, all the MGS 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 [8,15,16,28,29] (Supplementary Fig. 2). The microbial metabolic syndrome risk index is similar to the T2D index [8]. For each individual validation sample, the microbial metabolic syndrome risk index of sample j that denoted by MMSRj was computed by the formula below:

\[
PR_j = \frac{\text{Count}(A_i > R)}{N}
\]

\[
J^0 = \sum_{i \in B} PR_j
\]

\[
J^1 = \sum_{i \in G} PR_j
\]

\[
MMSR_j = \frac{J^0}{|B|} - \frac{J^1}{|G|}
\]

Where \(A_i\) is a scalar representing the relative abundance of MGS i in validation sample j, R is a vector representing the relative abundance of MGS i of all samples in this cohort which served as healthy reference. Nis the sample size of this cohort that is 2183. Percentile rank \(PR_j\) is the percentage of test sample’s MGS i relative abundance in its reference cohort frequency distribution that is equal to or lower than it. B is 12 out of 80 MGSs that were positively correlated with BMI and triglyceride. G is 68 out of 80 MGSs that were negatively correlated with BMI and triglyceride. Finally, \(|B|\) 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.

2.19. CRC probability computation

We built a randomForest model based on CRC-enriched bacteria [7, 17, 30–35] (Supplementary Table 4a [11]) for HK study [30]. Since our cohort was geographically connected to HK, it was effective to use the model directly. We applied the model to predict the probability of CRC in our cohort.

3. Results

Metagenomic and metabolomic data were collected from 2183 Chinese subjects from whom we had collected extensive clinical, physical fitness, psychological questionnaire data (Materials and methods 2.17, Supplementary Tables 1b-d). A detailed summary of the cohort has been reported in other paper [11]. This Asian cohort is characterized by young individuals (Age: 29.6 ± 5.5, average ± stdev), lean (BMI: 21.7 ± 3.8, average ± stdev), sex (46.5% female), rare chronic medication usage and free of cardiometabolic diseases (e.g. fasting blood glucose: 5.2 ± 0.5 mmol/L, low-density lipoprotein cholesterol: 2.7 ± 0.6 mmol/L). Another geographically distant cohort is composed of mostly young (Age (29.5 ± 5.2)), lean (BMI (21.5 ± 0.2)), sex (40.6% female) individuals, with rare chronic medication usage and free of cardiometabolic diseases (Supplementary Tables 1a and 1b) for scientific validation.

3.1. Novel associations with hyperuricemia and meat metabolites in the gut microbiome

Hyperuricemia is a common disease in the East Asian population. The main symptom of hyperuricemia is the excretion of urate in urine or through the gastrointestinal tract. Hyperuricemia was defined as serum uric acid (SUA) > 416 μmol/L (7 mg/dL) for men and >357 μmol/L (6 mg/dL) for women. In our cohort, 406 of the 970 males had urate content >416 μmol/L and 125 of 991 females had a content >357 μmol/L. Levels of uric acid was the fifth largest explanatory factor on microbiome composition (P = 0.0002, q = 4.66E-03) among the 88 microbiome covariates (Supplementary Table 1b). The uric acid level was also negatively correlated with alpha-diversity (q = 6.39E-03, Supplementary Table 1a). We further found that uric acid was the best predictive microbiome covariate (R = 0.398, Supplementary Table 3c).

We next used average rank (Materials and methods 2.17) to infer the
drivers of prediction. At FDR < 0.05 and after adjustment for medication and dietary supplements (Materials and methods 2.14), we identified 654 associations between microbiome features and the level of uric acid (Supplementary Table 3c). Notably, for example, higher level of uric acid was characterized by higher relative abundance of Flavonifractor plautii (mlg, q = 2.851E-05; metaplan2, q = 0.098), a member of the bacteria Lachnospiraceae family (Ruminococcus gravis (MLG, q = 1.770E-05; metaplan2, q = 0.001). In addition, abundances of Lachnospiraceae bacterium 1_4_56FAA (MLG, q = 0.207), Lachnospiraceae bacterium 2_1_58FAA (MLG, q = 0.068), and Lachnospiraceae bacterium 1_4_56FAA was associated with high salt diet (q = 0.016). We confirmed the presence of xanthine dehydrogenase/oxidase (XO; 1.17.1.4) and xanthine phosphoribosyltransferase (XPRT) to catalyze the conversion of xanthine to hypoxanthine to uric acid [36], but their Kyoto Encyclopedia of Genes and Genomes (KEGG) ontology (KO) term abundances were not correlated with uric acid. Low level of uric acid was characterized by high abundance of Faecalibacterium prausnitzii (MLG, q = 6.595E-3), Alistipes shahii (MLG, q = 0.003; metaplan2, q = 0.073), Oscillospiraceae (MLG, q = 0.003), Butyrivibrio fibrisolvens (MLG, q = 0.005) and Bacteroides intestinales (MLG, q = 0.0284) (Fig. 1).

To better understand the mechanisms by which the gut microbiota could drive hyperuricemia or lower uric acid, we focused on microbiota-prediction metabolites (Supplementary Table 3a). Compared to the lower uric acid samples, we found significant increases in the concentration of microbiota-prediction metabolites in higher uric acid abundance samples, including vitamins (vitamin A (r = 0.447, q = 5.30E-126), B5 (r = 0.238, q = 9.63E-34), D3 (r = 0.195, q = 7.59E-23) and E (r = 0.0969, q = 2.37E-06)), amino acids (glutamic acid (r = 0.439, q = 1.01E-120), tryptophan (r = 0.278, q = 3.34E-46) and alanine (r = 0.274, q = 7.98E-45)), trace elements (arsenic (r = 0.106, q = 5.24E-07) and mercury (r = 0.168, q = 6.87E-16)) (Supplementary Fig. 1). Testosterone was negatively associated with high uric acid (r = 0.560, q = 1.47E-211) (Supplementary Fig. 1). Potential microbial drivers of differences in these metabolites were further identified (Fig. 2). For example, the negative associations between fecal Butyricimonas virosa, Odoribacter splanchicus and plasma alanine (q = 0.003; q = 1.66E-3) were consistent with these bacteria containing butyrate production from amino acids (Supplementary Table 3e) [37,38], which together with positive associations with methylhistidines (q = 0.003, q = 0.944), hinted at a meal-excess diet (Fig. 2). Self-reported dietary structures showed association with serum uric acid (q = 0.0266, Supplementary Table 3e). Alistipes shahii abundance was negatively associated with plasma tryptophan (P = 7.43E-05, q = 0.020, Fig. 2, Supplementary Table 3e). Hyperuricemia has been reported to skew tryptophan metabolism towards kynurenine production in mice models [40], instead of indole production by A. shahii [41], potentially modulating signalling through aryl hydrocarbon receptors (AhR) [42]. This novel microbial metabolic pathway may shift hyperuricemia towards healthy status. Our study provided the first set of large-scale evidence for gut microbiome dysbiosis in hyperuricemia, together with hormonal, metabolic and potentially immunological differences.

3.2. Consistent biomarkers for cardiometabolic diseases in young individuals cohort

Hyperuricemia is accompanied by cardiometabolic disease risk. Early studies on type 2 diabetes, obesity, and cardiometabolic diseases contributed to most of the fecal metagenomic association studies [6]. Through this cohort healthy and young cohort, we defined a score according to 8 routine blood parameters and 80 fecal microbiome features for cardiometabolic disease risk (see Materials and methods 2.18, Fig. 2). This disease score was further tested in previously published case-control samples from older individuals. Using species markers in the fecal microbiota alone, metagenomic samples from Chinese patients with atherosclerotic cardiovascular disease (ACVD), liver cirrhosis, obesity and Crohn’s disease all obtained higher scores compared to control samples without the disease (P < 0.05) (Supplementary Fig. 2a). Rather
than diseases such as colorectal cancer, rheumatoid arthritis and medication-unstratified T2D (Supplementary Fig. 2a) [8,14–16,28–30]. Inclusion of the eight clinical parameters into the cardiometabolic disease score helped to distinguish T2D cases and controls, and excluded Crohn’s disease (Supplementary Fig. 2b). Although regional differences and misidentifications remain a concern, we illustrated the potential for population-wide screens of cardiometabolic diseases using the fecal microbiome.

3.3. Relationships of CRC-associated biomarkers and plasma metabolites

Both the CRC-enriched bacteria [7,17,30–35] and the plasma metabolome have been actively studied for CRC biomarkers. But to our knowledge they have not been investigated in the same cohort (Supplementary Table 3a). Although nobody was diagnosed with CRC, 1.288% of the samples in our cohort were associated with high CRC risk (>85%) based on the prediction model of the HK study [30] (Supplementary Fig. 3a, Materials and methods 2.19), which was geographically connected to our cohort. Notably, a few of the samples contained very high relative abundance of the CRC-enriched bacteria (Supplementary Fig. 3b). Furthermore, previously reported CRC-enriched bacteria [7,17,30–35] were found to be associated with plasma metabolites in this young cohort, regardless of statistical adjustment for covariates (Supplementary Table 3a). More specific, Peptostreptococcus stomatis positively associated with plasma branched-chain amino acids (BCAA) (leucine (MLG, q = 0.020; metaplan2, q = 0.009) and alanine (MLG, q = 0.045; metaplan2, q = 0.001)), aromatic amino acids (AAA)(phenylalanine (MLG, q = 0.041; metaplan2, q = 0.00) and tyrosine (MLG, q = 0.023; metaplan2, q = 0.001)), and sarcosine (MLG, q = 0.046), a metabolite studied for prostate cancer and a degradation intermediate of betaine [43,44] (Fig. 3, Supplementary Table 4b). We confirmed the presence of the alanine biosynthesis asdA gene in P. stomatis genomes using eggNOG mapper.

BCAA and AAA were previously be linked to increasing CRC risk [7]. Here we found that microbiome data could better predict metabolites that were branched-chain amino acids (BCAA) (isoleucine, R = 0.342; valine, R = 0.303; leucine, R = 0.348; alanine, R = 0.235), as opposed to aromatic amino acids (AAA) (phenylalanine, R = 0.136; tyrosine, R = 0.181), serine (R = 0.100) and glycine (R = 0.029). Enterobacteriaceae including Escherichia coli, Klebsiella pneumoniae, Enterobacter cloacae and Citrobacter freundii were significantly positively associated (q < 0.05) with sarcosine, hydroxylysine, branched chain amino acids, tyrosine, tryptophan, 1–methylhistidine, hydroxyproline, and argininosuccinic acid (Fig. 3, Supplementary Table 4b). 1-methylhistidine is a marker for habitual meat intake, especially red meat [35]. Enterobacteriae genomes had the complete set of genes for glutamate family amino acids (arginine, proline and histidine) metabolism. Functional analysis also showed that KO term abundance from glutamate family amino acids metabolism pathway (hyuB(K01474), q = 0.009; amiE (K01426), q = 0.004) correlated with levels of glutamate family amino acids (hydroxyproline and 3-methylhistidine).

Microbiome data better predicted levels of 3-methylhistidine (R = 0.303), tryptophan (R = 0.217) and weekly argininosuccinic acid (R = 0.152), 1-methylhistidine (R = 0.104), sarcosine (R = 0.088), hydroxylysine (R = 0.078). Bacteria such as Bacteroides massiliensis, Bacteroides thetaiotaomicron, Clostridium asparagiforme, and Butyricimonas virosa were associated with 3-methylhistidine (q = 4.388E-3, q = 0.023, q = 4.381E-3, q = 0.003) (Supplementary Tables 3a and 4b), a

![Fig. 3. Association of plasma metabolites and previously reported CRC-enriched bacteria. The MGS related to the plasma metabolites are listed in the heat map. The top panel indicates significant partial associations (q value < 0.05) between gut microbiome and the indicated metabolite adjusted factors that potentially influence the gut microbiome. BH-adjusted p-value is denoted: +, q-value<0.1; *, q-value<0.05; **, q-value<0.01.](https://example.com/fig3)
 metabolite that could indicate muscle catabolism without consuming meat [45]. Clostridium asparagiforme, and Butyrivibrio virosa contained genes encoding enzymes to convert histidine into 5-Phosphohypoxilase 1-pyrophosphate (PRPP), glutamate and 3-methylhistidinase. In addition, the butyrate-producing E. eligs positively associated with fruit and vegetable intake, folic acid (q = 0.006), while negatively associated with alanine (q = 0.006) (Fig. 3). The E. eligs genome harboured the dfrd and dfrA genes that are involved in folic acid usage. B. longum, which has been reported to be depleted in advanced adenoma [7,17] and could stimulate a renewal of intestinal stem cells [46], was seen here to negatively associated with plasma hydroxyproline (q = 0.018). A number of these associations were also observed in the geographically independent cohort, e.g. Enterobacter cloacae and hydroxylysine (q = 0.022), B. wadsworthia and arsenic (q = 0.079) after adjusting for confounders. Without adjustment for confounders, E. eligs, folic acid (q = 0.0294) and alanine (q = 0.022), B. longum and hydroxyproline (q = 0.009) could also be replicated (Supplementary Table 3a). These results corroborated fecal markers of CRC with plasma metabolites (Supplementary Table 4a), suggesting further studies on the long-term interplay between dietary metabolites and bacteria for CRC etiology, and the population threshold for early intervention.

4. Discussion

Here we provided the first set of large-scale evidence for gut microbiome dysbiosis in hyperuricemia. We confirmed the depletion of Faecalbacteria prausnitzii in gout reported in a small 16S rRNA gene amplicon-based Chinese study [47]. The uric acid production pathway from xanthine or hypoxanthine by Lachnospiraceae bacterium 1_4_56FCAA reported in an irritable bowel syndrome study [36] has also been confirmed. In addition, we identified novel species associated with hyperuricemia, of which two member of the Lachnospiraceae family (Ruminococcus grauus and Lachnospiraceae bacterium 2_1_58FCAA), Clostridium boltae, Flavonifractor plautii, and Eubacterium doliichum were significantly elevated in individuals with hyperuricemia. Alstipes shahii, Oscillospiraceae, Butyrivibrio virosa and Bacteroides intestinalis were depleted individuals with hyperuricemia.

Previous hyperuricemia studies based on analyses of the gut microbiome have not extensively included fecal or plasma metabolites. Here we identified potential microbe-metabolite pathways associated with hyperuricemia. A. shahii negatively associated with plasma tryptophan. Hyperuricemia has been reported to skew tryptophan metabolism towards kynurenine production in mice models [40], instead of indole production by A. shahii [41], potentially modulating signals through aryl hydrocarbon receptors (AHR) [42]. One bacteria that was negatively associated with serum uric acid, F. prausnitzii, has been reported to encode a methyltransferase for arsenic detoxification [48]. IL-1β, the major cytokine responsible for gout [49], has been associated with urinary level of arsenic [50]. Co-stimulation of patient-derived PBMCs with monosodium urate crystals and TLR2 or TLR4 (toll-like receptors) ligands have been shown to disrupt IL-1β/IL-1Ra (IL1 receptor antagonist) balance [51], consistent with involvement of microbes in gout.

Metagenomic-wide association studies (MWAS) have documented gut microbial perturbations in a growing list of diseases by comparing cases versus controls [8,14–16,28–30]. Hyperuricemia has been related to cardiovascular diseases, which is the leading cause of death in China. We expected to observe trends of cardiometabolic illnesses in this deep-sequencing metagenomic cohort, with mean age below 30 years. Alarmingly enough, trends for cardiometabolic diseases, type 2 diabetes [52] and colorectal cancer can already be seen from the fecal microbiome and a few parameters in the blood. The importance of a set of healthy gut microbes, such as A. shahii, F. prausnitzii, E. eligs and B. cellulosilyticus for leanness has been demonstrated previously [14]. Interestingly, few associations have been found with Akkermansia, reflecting a healthy gut in our cohort. But this result may differ between individuals [53,54] or require mucosal sampling rather than fecal sample. Future experiments are needed to elucidate the potential harmful gut microbes listed above. And more intervention study about whether decrease the abundance of potentially harmful bacteria, including R. grauus with exercise and diet, or fend off C. boltae with yoghurt, also in need.

Genetic potential for histidine degradation instead of synthesis has been observed to increase in CRC relative to healthy controls according to metagenomic studies [17,55]. 1-methylhistidinase, a marker for habitual meat intake [39], can be metabolized into histidine. The plasma level of the amino acid proline was reported to increase in a mouse model of CRC [33], but found in another study to decrease in human CRC in another study [35]. In this young cohort of young individuals from China, we did not observe a significant associations between proline and known gut microbiome markers of CRC. Hydroxyproline, on the other hand, predicted gut microbiome composition better compared to proline. Enterobacteriaceae such as Escherichia coli and Klebsiella pneumoniae positively associated with hydroxyproline in this cohort. A recent study that analyzed fecal metabolites together with the fecal microbiome and reported an increase in branched chain amino acids and aromatic amino acids in CRC [7,17,30–35] (Supplementary Table 4b). Here we observed 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 [7,17], implying that leucine was normally not in excess.

We also find it intriguing that decarboxylases generally appear important for bacterial stress response in the microbiome, i.e. maintaining a balanced pH for themselves. The top one for gut microbes may be glutamate decarboxylase (which produces GABA (γ-aminobutyr acid) from glutamate), while histidine decarboxylase in the female reproductive tract might contribute to menstrual pains [56]. Besides, recent studies identified tyrosine decarboxylases in gut microbes that could digest the drug levodopa used to treat Parkinson’s disease [57,58].

In summary, this study provides clues for baseline gut microbiome with deviations towards diseases. Interventional and mechanistic studies will be needed to determine how to improve the gut microbiota, metabolite levels and cardio-metabolic health.

Credit author statement


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 CNP0000426 [59,60], CNP0000289 [59,60]. Please refer all the Supplementary Tables in our first paper of the same cohort [11].
Declaration of competing interest
The authors declare no competing financial interest.

Acknowledgments
This study was funded by the Guangdong key areas R&D projects, China (No. 2019B020229003). The authors are very grateful to colleagues at BGZ-Genecan 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.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.medmic.2021.100037.

References


