npj | science of food

Article

Published in partnership with Beijing Technology and Business University & International Union of Food Science and Technology

https://doi.org/10.1038/s41538-025-00575-9

Comprehensive multi omics explore the microbial function in metabolic pathway flow during altered diet

Check for updates

Yiming Zhao[1,2,4], Siqi Yao[1,4], Abdulrahim Umar[1], Jing Huang[1], Shuijiao Chen[2,3] ✉ & Zheng Yu[1] ✉

High-protein diet (HPD) and high-fiber diet (HFiD) are two common types of diet, and the transition from HPD to HFiD frequently happens in people who perform diet control. Gut microbiota provide a bridge between diet and health. However, most of the research left out the metabolic landscape change and gut microbial influence under this kind of altered diet. Here, we examined gut microbiome and metabolites change using multi-omics (16S rRNA gene sequencing, shotgun metagenomics, LC-MS/MS metabolomics), and machine-learning methods were introduced to capture the response of gut microbiota to altered diet and associate the variable metabolites and microbial factors. The results suggest that the conversion from HPD to HFiD improves the gut microbial diversity and promotes the pathways of tryptophan, galactose, fructose, and mannose metabolism, which are associated with different gut microbiota and special metabolites. Among them, *Faecalibacterium rodentium* and *Akkermansia muciniphila* indicate potential roles in the multi-metabolism. Although these findings are exploratory and hypothesis-generating, they provide a foundation for future mechanistic investigations and emphasize the importance of diverse diets.

With growing attention to health, research on exploring healthy dietary patterns has become a focal point. Diet plays a crucial role in regulating the microbiota that colonize the gastrointestinal tract, which is essential for maintaining host growth, development, and overall health[1]. In recent years, the impact of Western and Mediterranean dietary patterns on host health has attracted significant attention[2,3]. The increased intake of nutrients such as protein in the Western diet has been associated with increased permeability of the intestinal epithelial mucus layer, thereby raising the risk of gastrointestinal diseases[4]. Meanwhile, studies on patients with inflammatory bowel disease (IBD) have found that a Mediterranean diet primarily based on dietary fiber helps reduce the expression of inflammatory factors such as C-reactive protein, fecal calprotectin, TNF-α, and clinical disease scores[5]. However, current research primarily focuses on the transition between Western and Mediterranean diets, while studies on the transition between specific, single macronutrient-based dietary patterns remain relatively insufficient. Given the impact of protein and dietary fiber on the host, in-depth research on the effects of transitioning between the high-protein diet (HPD) and the high-fiber diet (HFiD) patterns on the body is urgently needed. Therefore, we further consider studying the impact of transitioning between these two common nutritional patterns on host health. HPD is characterized by an increased intake of protein-rich foods such as meat, legumes, and dairy products[6,7]. while HFiD emphasizes the consumption of plant-based foods rich in dietary fiber to promote digestive health and reduce the risk of chronic diseases[8,9]. Notably, the impact of diet on human health and disease is often mediated or modified by the gut microbiota.

The dynamic interplay between diet and gut microbiota composition has emerged as a focal point in contemporary microbiome research[10–12]. The gastrointestinal tract harbors a diverse ecosystem of microbes, collectively known as the gut microbiota, which play pivotal roles in host metabolism[13], immune regulation[14], and overall well-being. Dietary components serve as substrates for microbial fermentation[15], shaping the composition and metabolic activity of gut microbes[16–18]. Studies on the impact of factors such as nutrients on the composition of the gut microbiota have shown that the gut microbiota influences cellular metabolism in the liver and adipose tissue, thereby regulating lipid homeostasis[19]. The metabolites produced by the gut microbiota also play an important role as signaling molecules and substrates in metabolic reactions within the host. Gut microbiota-derived metabolites can promote host metabolism, such as food digestion, absorption, and energy expenditure[20,21], and can also serve as ligands for G-protein-coupled receptors and hormone receptors, influencing the physiological functions of

[1]Human Microbiome and Health Group, Department of Microbiology, Xiangya School of Basic Medical Science, Central South University, Changsha, Hunan, China. [2]Department of Gastroenterology, Xiangya Hospital Central South University, Changsha, China. [3]National Clinical Research Center for Geriatric Disorders, Xiangya Hospital Central South University, Changsha, China. [4]These authors contributed equally: Yiming Zhao, Siqi Yao. ✉e-mail: shuijiaoxy@csu.edu.cn; yuzheng@csu.edu.cn

mammals[22]. Consequently, alterations in dietary patterns can induce profound changes in the gut microbiota landscape, with potential implications for host health and disease susceptibility.

Understanding the impact of transitioning from HPD to HFiD on gut microbiota is of paramount importance. It provides insights into the plasticity of the gut microbiota in response to dietary interventions, elucidating the mechanisms underlying microbial adaptation to varying nutrient substrates. Although previous studies have primarily focused on the effects of single dietary interventions on gut microbiota composition and metabolite profiles, the specific changes in gut microbiota structure, functional potential, and metabolic pathways in response to dietary changes remain poorly characterized. Therefore, this study seeks to investigate the impact of transitioning from HPD to HFiD on gut microbiota composition, diversity, and metabolic function. By employing amplicon sequencing, shotgun-metagenomic sequencing, and metabolome profiling, we explored the changes occurring within the gut microbial community associated with dietary shifts. In addition, the gut microbe-diet-metabolite relationships from HPD to HFiD were captured through comprehensive analysis. We endeavor to contribute to the growing body of knowledge on the intricate interplay between diet, gut microbiota, and human health.

Results

The resilience of gut microbiota from a high-protein to a high-fiber diet

To explore the differential gut microbial response induced by an altered diet and clarify the relationship between metabolites and altered gut microbiome, two groups (SD and Alter) of mice were monitored over an 8-week period. The mice in the Alter group underwent a diet transition from a high-protein diet (HPD) to a high-fiber diet (HFiD) (Fig. 1). Firstly, the 16S RNA gene sequencing was performed on the feces of mice to explore the composition of the microbial community. There was no statistically significant difference in the gut microbial composition ($P > 0.05$, Fig. S1A) and alpha-diversity ($P > 0.05$, Fig. S1B) before dietary intervention. However, after dietary intervention, the overall gut microbiota of mice became significantly different, as illustrated by NMDS (PERMANOVA, $P = 0.001$, $R^2 = 0.298$, Fig. 2A). In detail, Proteobacteria was decreased with Actinobacteriota increasing after the intervention of HPD (Fig. 2B). After the transformation of HFiD, the relative abundance of Firmicutes decreased; however, Verrucomicrobiota enriched in the gut (Fig. 2B). These results emphasized the alteration of gut microbial composition during dietary alteration.

In addition, we found the beta-diversity of gut microbiota decreased significantly when fed with HPD (Bray–Curtis: $P < 0.0001$, Cliff's $\delta = -0.83$; weighted-unifrac: $P < 0.0001$, Cliff's $\delta = -0.70$, Fig. 2C). However, it began to significantly increase after switching to HFiD (Bray–Curtis: $P < 0.0001$, Cliff's $\delta = -0.53$; weighted-unifrac: $P < 0.0001$, Cliff's $\delta = 0.50$, Fig. 2C). Alpha diversity significantly decreased under HPD compared to SD (Shannon–Wiener: $P = 0.038$, Cliff's $\delta = -0.63$; Gini-Simpson: $P = 0.038$, Cliff's $\delta = -0.67$). The subsequent HFiD intervention significantly increased alpha diversity compared to HPD (Shannon–Wiener: $P = 0.038$, Cliff's $\delta = 0.89$; Gini–Simpson: $P = 0.038$, Cliff's $\delta = 1.00$), and the diversity level in HFiD-fed mice was statistically indistinguishable from that of the SD group ($P > 0.05$, Fig. 2D). The data of shotgun metagenomics also supported the resilience of gut microbial alpha diversity during the altered diet (Fig. S2).

Specific diet is associated with different gut microbiota alteration

To further clarify the associated gut microbiota during the dietary alteration, we determined the diet-related microbial taxonomic markers by MaAsLin2[23], which incorporated different diets and experimental time as fixed effects, and identification information of mice as a random effect. The result showed that different gut microbiota responded in the process of an altered diet (Fig. 3A). Afterwards, we utilized the data from metagenomics sequencing to further clarify the food-related microbes at the species level. Several microbes were found to be significantly enriched under HPD, including *Faecalibaculum rodentium* and *Muribaculum gordoncarteri*, the core microbes in the healthy people and mice gut. Moreover, two probiotics, *Lactobacillus johnsonii* and *Lactobacillus intestinalis*, were enriched under HPD (Fig. 3B and Fig. S3). After switching to HFiD, the relative abundance of *Akkermansia muciniphila*, which belonged to the Verrucomicrobiota phylum colonized in the intestinal mucosa (Fig. 3C and Fig. S4). We further validated the variations in the relative abundance of these bacteria in fecal samples from mice subjected to different dietary interventions using qPCR (Fig. S5).

The gentle change of microbial function under an altered diet

Although the obvious taxonomic remodeling was detected under altered diet (Fig. 2B), the main functional pathways seemed to be relatively stable (Fig. 4A). However, the functional composition of gut microbial community at the gene level under different diets shows significant diet-related differences (PERMANOVA, $P = 0.003$, $R^2 = 0.385$, Fig. 4B). On the one hand, these results revealed that an altered diet won't cause severe gut microbial disorder like large-scale function change. The predominant microbial functions were still composed of carbohydrate and amino acid metabolism, involved in microbial basic life metabolism. On the other hand, it suggested that food-related microbial altered functions may be mapped on the upstream or downstream specific metabolites.

Specifically, there were 2006 under-represented and 7169 over-represented genes in Alter group (Fig. 4C). These genes enriched in carbohydrate metabolism (e.g., starch, sucrose, fructose, mannose, and galactose metabolism) and amino acid metabolism (including tryptophan,

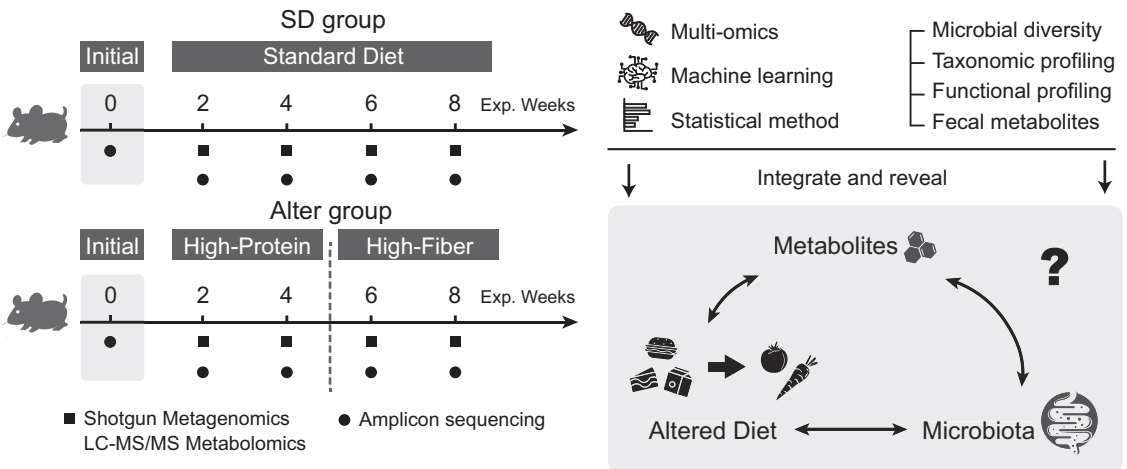


Fig. 1 | The scheme of the total research design.

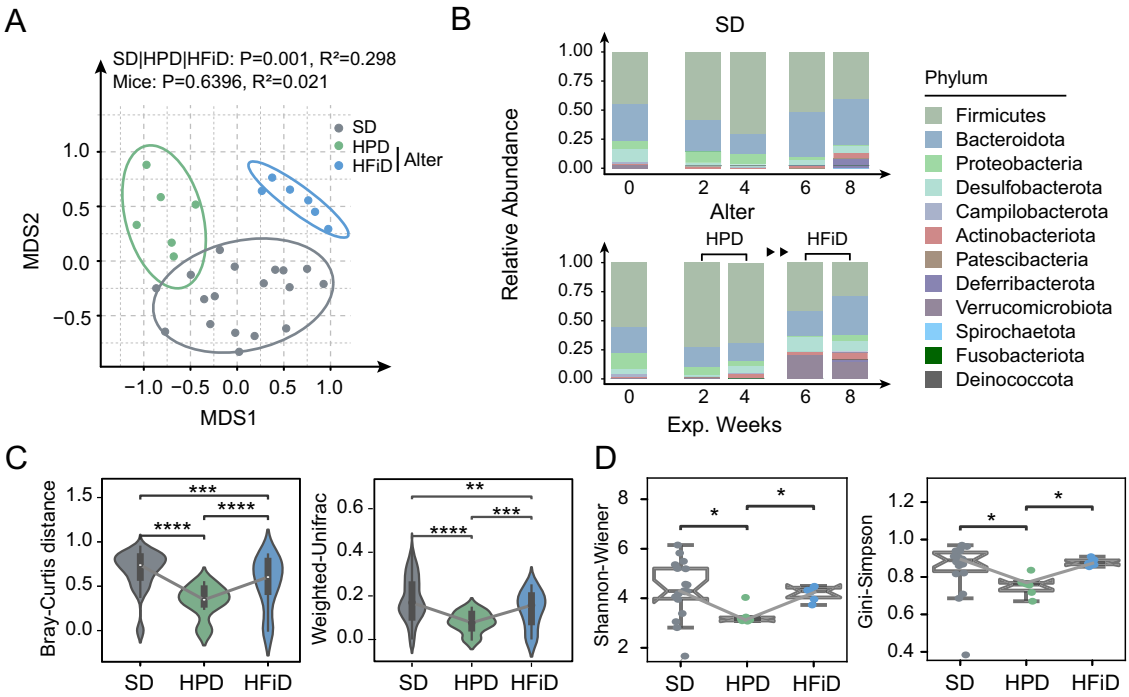


Fig. 2 | Altered gut microbiota in various diet. A The NMDS based on intestinal microbial composition and the PERMANOVA were applied to test the difference among features. **B** The gut microbial composition in different experimental weeks and groups, with the average relative abundance of phylum. **C** The Bray-Curtis and weighted-unifrac distances were used to estimate the beta-diversity of gut microbiota under three diets. **D** The Shannon–Wiener and Gini–Simpson index were calculated to estimate the alpha-diversity. *P* values were calculated by two-sided Mann–Whitney–Wilcoxon test, and adjusted by BH method (*0.01 < adj. $P < 0.05$; ** adj. $P < 0.01$).

phenylalanine, and tyrosine metabolism) were found to be differentially abundant between the dietary conditions with over two-fold changes (Fisher exact test with BH adjusted). Moreover, some paramount vitamin metabolism pathways were detected, such as riboflavin and thiamine metabolism, which were possibly implicated in host health (Fig. 4D). Through the traceability analysis of genes, we found that these differential genes enriched came from a variety of microbes, among which *Lachnospiraceae*, *Akkermansiaceae*, *Erysipelotrichaceae*, *Muribaculaceae*, *Clostridiaceae,* and *Desulfovibrionaceae*. Not surprisingly, these microbes coincided with the associated microbiota on diets (Fig. 3A).

Transition to high-fiber diet linked to gut microbial metabolic pathways

Next, we detected the fecal metabolites by LC-MS/MS untargeted metabolomics to study the metabolic spectrum under an altered diet. The results showed that the composition of the intestinal metabolic spectrum was significantly different under various diets (PERMANOVA, $P = 0.03$, $R^2 = 0.266$, Fig. 5A). There were 545 metabolites detected as differential metabolites between HPD and HFiD using an integrated pipeline (Fig. S6). Afterwards, these differential metabolites were mapped to the MetOrigin database[24]. A total of 177 metabolites were determined to be associated with food, whereas 52 metabolites were attributed to microbiota. More importantly, metabolites linked to microbiota have been found to exhibit a significant overlap with those associated with food (Fisher exact test, $P < 0.0001$, Fig. 5B). Equally, Procrustes and partial mantel analysis of amplicon sequencing, metagenomics, and metabolomics further underscores the intricate interplay between the microbiome, functional potential, and the metabolic landscape (Fig. 5C, Fig. S7, and Table S1). These results indicated that the changes in intestinal metabolism caused by diet changes were closely related to gut microbes.

To further clarify the pathways involved in the altered diet from HPD to HFiD, the enrichment analysis was carried out. The results showed that the metabolites about tryptophan (amino acid metabolism), galactose, fructose, and mannose metabolism (carbohydrate metabolism) were altered during the diet alteration (Fisher exact test, $P < 0.05$ and BH <0.2, Fig. 5D). We observed that the transition from HPD to HFiD led to an enrichment of specific metabolites, including those within the same metabolic pathway (Fig. 5E and Fig. S8). Afterwards, Mmvec[25], a bioinformatic tool utilizing deep-learning model to model microbe-metabolite co-occurrence probabilities through gradient-based optimization of latent representation, was performed and trained to a highly accurate level on our data for investigating metabolite-microbe associations (Fig. S9). We found the five previously identified food-related microbes, which are closely associated with the metabolic composition (with an importance score from MMvec of more than five). Specifically, *A. muciniphila* showed the highest consistency with the HFiD metabolic composition, while the other four taxa (*F. rodentium*, *L. johnsonii*, *M. gordoncarteri*, and *L. intestinalis*) were consistent with the HPD. Moreover, these co-occurrence probabilities derived from MMvec analysis aligned with the earlier observed changes in the relative abundances of these five microbes, which suggested the consistency between the enrichment of metabolites and the altered microbes (Fig. 5F). To identify associations between metabolites and microbial species that changed during the dietary transition, we further restricted the threshold of detection by decreasing the MMvec co-occurrence probability and limitation of the consistent changes in microbes and metabolites with the dietary intervention. We identified nine potential microbial-related metabolic reactions associated with tryptophan metabolism. Among them, *A. muciniphila* was associated with five metabolic reactions during the HPD to HFiD transition and three metabolic reactions within the HFiD. Furthermore, one metabolic reaction within the high-protein diet was associated with *F. rodentium* (Fig. 5G).

Discussion

Here, we present the findings of a study of mice with an altered diet (from HPD to HFiD), investigating alterations in gut microbiome composition and function associated with a specific diet. The metabolic spectrum we

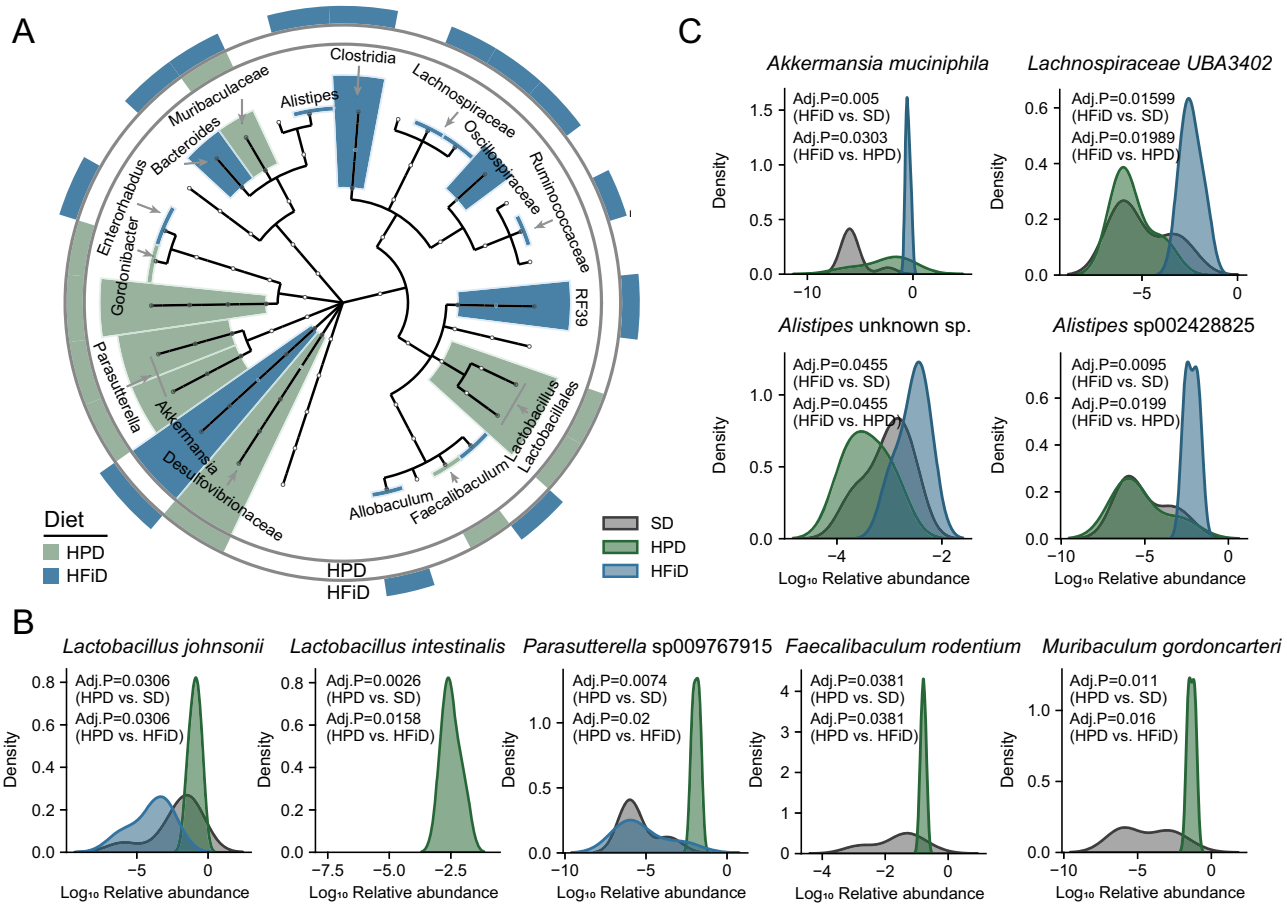


Fig. 3 | Identification of related gut microbiota under different diets. A The clade-specific taxonomic alterations of the gut microbiota by diets (HPD and HFiD). The green color represents the association between the HPD and microbes. And the blue represents HFiD and microbes. The outer ring summarizes all associations between different diets and microbial taxa, and internal branches were colored by strongest association. All associations were identified using MaAsLin2. The relative abundances of species associated with HPD and HFiD were shown in **B** and **C**, respectively. The adjusted *P* values in the figure were calculated by a two-sided Mann–Whitney–Wilcoxon test, and corrected by the BH method.

detected was associated with this altered diet and largely paralleled changes identified in gut microbiota, such as microbial profiling and functions. Some microbial taxa (such as *A.muciniphila* and *F. rodentium*) were detected as associated taxa for specific diets. These changes were consistent with the probability of differential metabolites co-occurrence. In addition, we found that the dietary transition from HPD to HFiD was related to the flow of multi-metabolic pathways.

HPD and HFiD showed significant specificity for intestinal microbes; this reflects the differences in the regulatory mechanisms of diet on the gut microbiota[18,26,27]. We found that the diversity of intestinal microbes decreased under HPD, which is similar to previous studies[3,28]. However, some studies have reported either neutral or even positive effects of high-protein diets on gut microbiota, depending on the protein source, fiber content, and feeding duration[29–31]. Although some previous studies have indicated that for obese patients, high protein intake does not affect the diversity of the gut microbiota and has a protective effect on the gut[32]. The discrepancy in these results may be related to the protein content and source[33,34]. The microbial shifts observed may reflect functionally relevant adaptations. Proteobacteria are often associated with gut dysbiosis, and their reduction may suggest improved intestinal health[35]. The rise in *Akkermansia* a key member of Verrucomicrobiota, has been linked to enhanced mucosal integrity and metabolic benefits[36], while certain Firmicutes are associated with obesity and low-grade inflammation[37]. These compositional changes thus imply potential functional improvements in host–microbiota interactions. This may be related to the metabolism of microbes, as the gut microbiota maintains intestinal barrier function and regulates immune responses in a fiber-rich environment[38,39]. After considering the covariate influence caused by feeding time and individual deviation, the predominant microbial community in each sample under different diets was analyzed. The results showed that there was an enrichment of different microbes in HPD and HFiD. *Lactobacillus* was significantly enriched under HPD, which was consistent with the previous study[40]. Additionally, *Parasutterella* was found to be enriched under HPD, which was found to utilize protein substrate before[41]. Besides, *Muribaculaceae*[42] and *Desulfovibrionaceae*[30] enriched under HPD were also found to be related to protein intake. Conversely, the abundance of these microbes decreased after switching to HFiD. Some other microbes were detected to respond to HFiD in our study, such as *A.muciniphila*[43], which contrasts with a previous study[44] reporting a decrease in this species under high-fat conditions. One possible explanation is the fiber content in our HFiD formulation, which may have mitigated the inhibitory effects typically seen with fat-rich, fiber-poor diets.

The important functions of altered microbes in diets and the conjecture that functional genes drive microbial abundance both prompt us to carry out further metagenomic sequencing and function analysis. The results of the functional composition at the pathway level seem not to fluctuate, which coincides with the study of large-scale people in the Human Microbiome Project[45]. This kind of substitutability from microbiota provides functional redundancy, and the main microbial functional genes are provided by

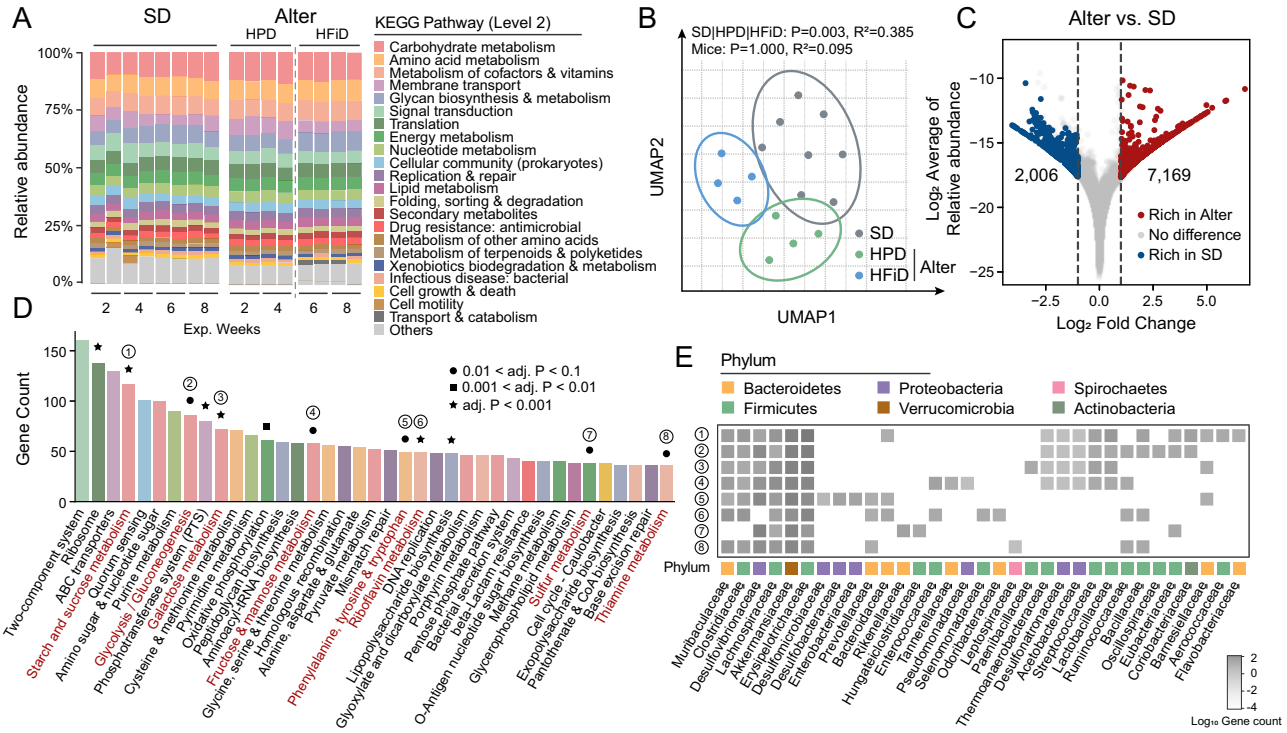


Fig. 4 | Gut microbial functions under an altered diet. A The functional proportion of the 2nd level pathway of KEGG. Only the top 20 pathways are shown, and the remaining ones are grouped as "Others". **B** The composition of genes in different groups with UMAP. The top of the figure shows the *P* value and R^2-value of gene composition under different variable using PERMANOVA test. **C** Compared with the SD group, the increasing or decreasing of intestinal microbial gene abundance in the Alter group. Only genes with $P < 0.05$ and adjusted $P < 0.2$ (BH method) were considered as the "rich" genes, and genes were colored according to the $\log_2$ fold change of them (blue: <−1, and red: >1). **D** The number of differential genes contained in KEGG's level 3 pathway. The color of the column is based on KEGG's level 2 pathway. For visualization, only the pathways were shown with the number of differential genes more than 35. The BH method was used to adjust the *P* value. **E** Traceability analysis of differential genes in eight KEGG pathways from Figure **D**.

protein and carbohydrate metabolic pathways related to the basic life activities of microbes in our study. However, nearly ten thousand differential genes were identified when compared at detailed gene levels, which were enriched in the tryptophan, galactose, fructose and mannose metabolism pathways that might be related to the metabolism of the host. Most striking was the difference in metabolites in the same pathway and consistent microbial co-occurrence during the altered diet (HPD to HFiD). For example, the reactions of indole pyruvate and tryptamine to indole-3-acetaldehyde (IAA) were direct and catalyzed by *A.muciniphila*, which was previously found to be the key contributor in the tryptophan metabolism[46]. We also checked this kind of metabolic alteration using ELISA (Fig. S10). Equally, indole pyruvate and tryptamine as substrates were enriched in HPD, whereas IAA was enriched in HFiD. Although IAA was usually regarded as one of the phytohormones to control plant development[47,48], one recent study supported the idea that IAA could be produced from gut microbiota[49].

The changes in tryptophan and its related metabolites under HFiD are particularly prominent, and they may play an important role in restoring microbial homeostasis and host health after being threatened by HPD. As an essential amino acid, tryptophan is provided by a large amount of dietary protein and serves as a precursor for the synthesis of most metabolites[50]. Its metabolites play a key role in maintaining the structure and function of intestinal epithelial cells, gastrointestinal motility, and insulin secretion by pancreatic β-cells[51,52]. The degradation of protein in the intestine leads to the release of tryptophan, which is further converted by the gut microbiota into various metabolites[53]. Therefore, indole-3-acetate and tryptamine are enriched as substrates under HPD, while indole-3-acetate and IAA are further enriched under HFiD, and this response is significantly associated with *A. muciniphila*. *A. muciniphila* has previously been identified as a key microbe in tryptophan metabolism, promoting the accumulation of indole pathway-related substances[46]. Although only a small portion of tryptophan is converted into indole compounds, indole and its derivatives contribute to gastrointestinal function and immune system regulation. Furthermore, studies have shown that indole promotes the secretion of glucagon-like peptide-1, delays gastric emptying, and suppresses appetite[54,55]. Indole and isopropanol can also improve mucosal barrier function and reduce intestinal mucosal permeability by interacting with the pregnane X receptor[55]. In summary, after the dietary switch to HFiD, the gut microbiota promotes significant changes in these substances, and these alterations may be crucial factors for the host in restoring physiological functions.

A key limitation of this study lies in its modest sample size across some omics layers and the absence of direct mechanistic experiments. While multi-omics associations are suggestive, the use of different animals for different omics layers limits the validity of cross-omics integration. Although we introduced procrustes analysis and an additional test (Fig. S11) to strengthen the feasibility of overall analysis, the associations should not be interpreted as conclusive evidence of causality. We recommend it as exploratory and hypothesis-generating, laying the groundwork for future validation studies or functional interventions. The exclusive use of female mice, while simplifying hormonal variability, overlooks potential sex-specific effects of diet on gut microbiota and metabolic outcomes. Given the influence of sex hormones on immune responses and microbial composition, future investigations should include both male and female subjects to elucidate sex-dependent mechanisms. While machine learning methods like MMvec were employed to infer microbe-metabolite associations, the analysis could benefit from integrating advanced algorithms to better capture nonlinear relationships and improve predictive accuracy.

In summary, this study provides exploratory insights into the complex interactions among diet, gut microbiota, and host metabolism, using a multi-omics framework. By examining changes in microbial composition

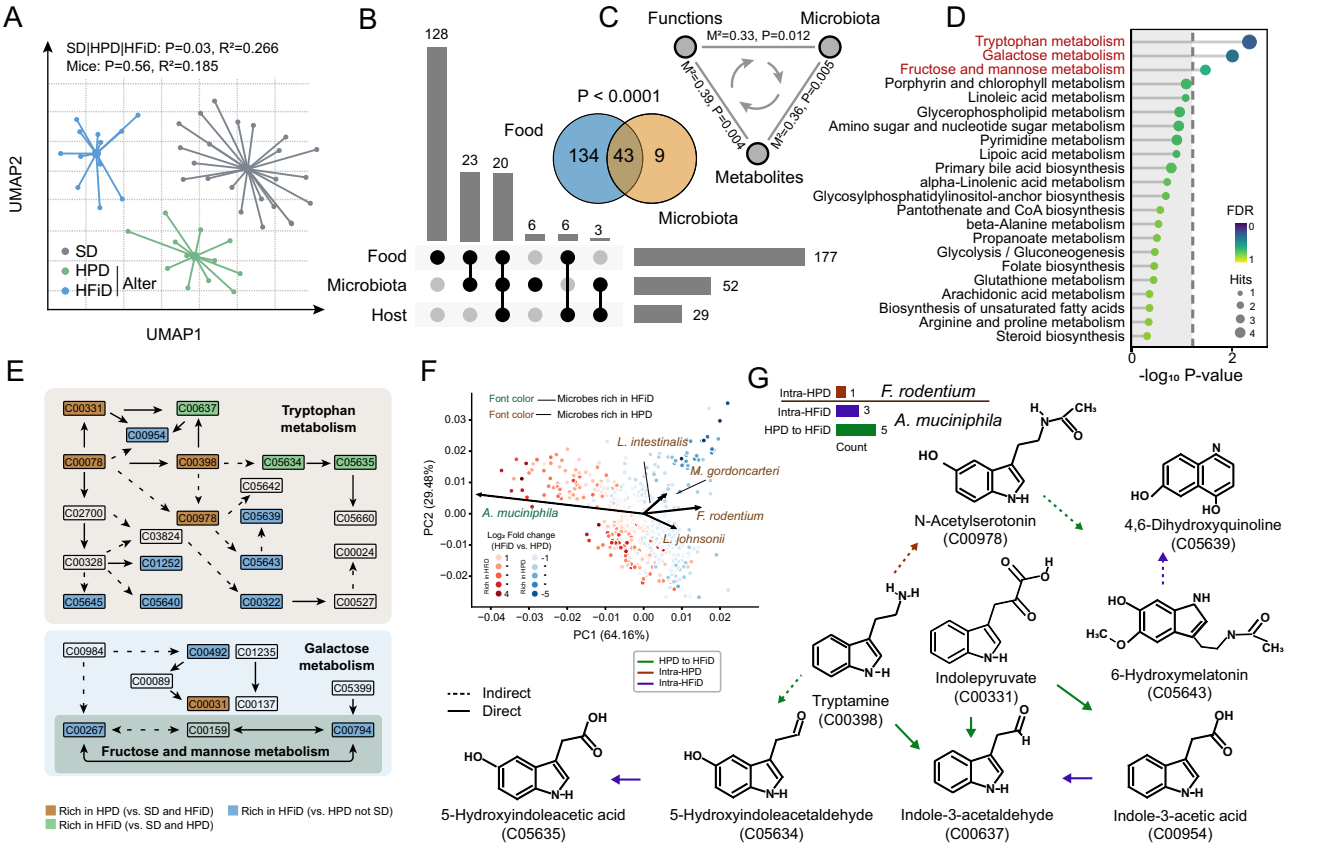


Fig. 5 | Association of fecal metabolite changes with the gut microbiome. A The composition of metabolites in different groups with UMAP. The centroid points were shown in the middle of each group. The top of the figure showed the *P* value and R^2-value of metabolite composition under different variable using PERMANOVA test. **B** The Upset diagram and Venn diagram showed the number of metabolites related with different variables. The columns on the right of the Upset graph represent the total number of metabolites. The Fisher exact test was used to calculate the *P* value of overlap in Venn graph. **C** Overview of Procrustes analysis among microbiota, functions and metabolites. **D** The enrichment analysis of differential metabolites, and the red font represent the pathways with $P < 0.05$ and BH <0.2. **E** The metabolic maps of the tryptophan, galactose, fructose and mannose metabolism that contain multiple differential metabolites. **F** Biplot of metagenomics and metabolic data sets using mmvec. Metabolic features (dots) are labeled according to their strength of change with respect to food change size (red is associated with HFiD, blue with HPD). The differential microbes were represented by arrows indicating their cooccurrences with the metabolites. Only the microbes with MMvec importance score <5 showed. **G** The predicted microbially mediated reactions are shown, with the top panel indicating the microbes inferred to be responsible for these reactions under different dietary conditions. A color-coded, solid-line or dashed-line arrow represented a direct or indirect reaction, respectively.

and metabolites across dietary interventions, our findings generate hypotheses regarding the potential resilience of the gut ecosystem under fiber reintroduction. While these results are preliminary and limited by small sample sizes and a lack of mechanistic validation, they highlight key microbial taxa and pathways that warrant further investigation. This work serves as a conceptual foundation for future mechanistic studies aimed at understanding how dietary modulation may influence host physiology through the microbiota.

Methods

Animals and sample selection

Female BALB/c mice (3 weeks old, 12–15 g) were purchased from Hunan Sleek Jingda (SLAC), Changsha, China. All mice were raised in plastic cages covered with metal fences under specific pathogen-free conditions. The mice were housed in individually ventilated cages (IVCs) with free access to food and water. The environmental conditions of housing were maintained as follows: temperature (25 ± 5 °C), humidity (60–70%), and light (12/12 h light/dark cycle). Before commencing the experiment, mice were provided with a period of 5 days to adapt to the experimental environment.

After adaptive feeding, the mice were randomly divided into two groups, with eight mice in each group. According to the experimental design, the two groups were fed different types of diets. The two groups are as follows: the standard diet (SD) group, healthy mice with an AIN-93G diet. The Alter group, healthy mice with a high protein diet (HPD) for 4 weeks, followed by a high fiber diet (HFiD) for four weeks. HPD and HFiD are both modified based on the AIN-93G standard diet. The diet information of the two groups of mice was shown in Table S2. And the content of nutrients and energy in different diets were shown in Table S3. All mice were allowed to freely eat during the feeding phase. Dietary intervention was performed on mice after adapting to the environment. Collect fecal samples from 0 weeks before dietary intervention, and then collect fecal samples every 2 weeks after dietary intervention.

After immobilizing the mice, lifting their tail and gently pressing the lower abdomen to collect fresh feces into sterile EP tubes, which were then stored at −80 °C for further analysis. We randomly selected six mice in each group and performed LC-MS/MS non-targeted metabolomics on fecal samples from 2, 4, 6, and 8 weeks after dietary intervention.

Three mice were randomly selected in each group, and a total of 30 fecal samples from 0, 2, 4, 6, and 8 weeks of dietary intervention were subjected to 16S rRNA gene sequencing. And the other two mice were selected, and a total of 16 fecal samples from 2, 4, 6, and 8 weeks of dietary intervention were subjected to shotgun metagenomic sequencing. Five and four mice were selected from the mice with different dietary interventions for quantitative

detection of microorganisms in feces by qPCR and ELISA analysis of specific metabolites.

All food was produced by Beijing HFK Bioscience Co., Ltd., Beijing, China. The production numbers were 20220249 and 20220274. The diet formulation for the SD group was complied with the American Institute of Nutrition's food-grade ingredient formulation standard AIN-93G[56]. High protein and high fiber diets were modified according to AIN-93G criteria. The study was approved by the Laboratory Animal Ethics Committee of Xiangya Hospital of Central South University (No. 2022030735).

DNA extraction and 16S rRNA gene sequencing

The fecal samples were accessed using 16S rRNA gene sequencing. In detail, DNA extraction total genomic DNA samples were extracted using the OMEGA Soil DNA Kit (M5635-01) (Omega Bio-Tek, Norcross, GA, USA), following the manufacturer's instructions. PCR amplification of the bacterial 16S rRNA genes V3–V4 region was performed using the forward primer 338F (5'-ACTCCTACGGGAGGCAGCA-3') and the reverse primer 806R (5'-GGACTACHVGGGTWTCTAAT-3'). Sample-specific 7 bp barcodes were incorporated into the primers for multiplex sequencing. The PCR components contained 5 μl of buffer (5×), 0.25 μl of Fast pfu DNA Polymerase (5 U/μl), 2 μl (2.5 mM) of dNTPs, 1 μl (10 μM) of each forward and reverse primer, 1 μl of DNA template, and 14.75 μl of ddH_2O. Thermal cycling consisted of initial denaturation at 98 °C for 5 min, followed by 25 cycles consisting of denaturation at 98 °C for 30 s, annealing at 53 °C for 30 s, and extension at 72 °C for 45 s, with a final extension of 5 min at 72 °C. PCR amplicons were purified with Vazyme VAHTSTM DNA Clean Beads (Vazyme, Nanjing, China) and quantified using the Quant-iT PicoGreen dsDNA Assay Kit (Invitrogen, Carlsbad, CA, USA). After the individual quantification step, amplicons were pooled in equal amounts, and pair-end 2 × 250 bp sequencing was performed using the Illumina NovaSeq platform with NovaSeq 6000 SP Reagent Kit (500 cycles).

16S rRNA gene sequencing data processing and analysis

The sequenced 16S rRNA gene sequence is screened with low quality (Discards bases with Phred score <4, and removes sequences with any ambiguous bases), merged, truncated to 400 bp, and then the chimera was removed, as referred to our previous article[18]. Finally, the remaining sequences were clustered into operational taxonomic unit (OTU). To reduce false positives of obtaining data, only the OTUs mapped reads more than ten were retained and rarefied to the minimum read counts using R package "picante"[57]. By rarefication, each sample has 21,506 effective sequences (reads), and the sequencing depth reached saturation (Fig. S12). Next, the taxonomic classification was performed according to the Silva database (version 138)[58]. To investigate the variation of microbial communities across samples, the Shannon–Wiener and Gini–Simpson indexes were calculated. At the same time, the Bray–Curtis distance metrics were calculated based relative abundance of OTUs and visualized via nonmetric multidimensional scaling (NMDS) by R package "vegan"[59].

Metagenomic sequencing and functional analysis. To study the microbial function alteration, the shotgun metagenomic sequencing was carried out. Total microbial genomic DNA was extracted using the QIAGEN DNeasy PowerWater Kit (14900-100-NF). The extracted microbial DNA was processed to construct metagenome shotgun sequencing libraries with insert sizes of 400 bp by using Illumina TruSeq Nano DNA LT Library Preparation Kit. Each library was sequenced by the Illumina HiSeq X-ten platform (Illumina, USA) with a PE150 strategy.

To ensure the data accuracy and obtain the clean reads, raw sequences were processed to remove low-quality (average quality <35 or contains more than 16 unclear bases) reads using Fastp[60] (version 0.21.0) and FastUniq[61] (version 1.1.0) to eliminate duplicates in paired short DNA sequence reads in a FASTQ format. The sequences from humans were filtered out by mapping to the mice reference genome (mm39) using Bowtie2[62] (version 2.3.5) with sensitive mode. Afterwards, the clean reads were assembled into contigs using Megahit[63] (version 1.2.9) with default settings. The contigs greater than 500 bp were used to perform the gene prediction with MetaGeneMark[64] (version 3.38). After that, we used hmmer[65] (version 3.3.1) based on the AntiFam database[66] to remove spurious genes. A gene catalog was established using mmseq2[67] (version 13.45111) based on minimum sequence identity >95% and clustering threshold >90%. The function annotation for the gene catalog was performed by kofam-scan[68] (version 1.3.0), and only the annotation with e-value less than 0.00001 was remained. For each gene, reads from that sample were mapped using BWA[69] (version 0.7.17), and calculated the gene abundance according to the mapped read number divided by gene length. Functional composition analysis was performed based on gene abundance. And then we performed Permutational multivariate analysis of variance (PERMANOVA) using the R package "vegan" and uniform manifold approximation and projection (UMAP) using the Python package "UMAP" to analyze the functional composition of gut microbial community at the gene level.

For traceability analysis of differential genes, we extracted all original gene sequences first. Secondly, the differential gene sequences were mapped to the NCBI non-redundant database[70] by DIAMOND[71] (version 2.0.2). At last, we utilized MEGAN[72] (version 6.20.11) software to determine its belonging classification of microbes.

Binning and taxonomic profiling of metagenomic sequencing data

Metagenome assembled genomes (MAGs) were generated using MetaBAT2[73] (version 2.12.1). The quality of MAGs was evaluated with CheckM[74] (version 1.2.0) for completeness and contamination as a genome. The high or median quality MAGs (HM-MAGs) with >50% completeness and <5% contamination were remained. The remained HM-MAGs were clustered to species-level genome bins (SGBs) based on average nucleic acid identity ≥95% using dRep[75] (version 3.4.0). The GTDB-tk[76] or a set of five linoleic and isomers of conjugated linoleic acids discernible in the data (as informed by the metabolomic analysis) (version 1.7.0) was used for taxonomic identification with Genome Taxonomy Database (r202). The abundance of each species was calculated using CoverM (version 0.6.1).

Bacteria quantification using qPCR

Genomic DNA was isolated from mouse feces TIANGEN Fecal DNA Extraction Kit (#DP328–02, TIANGEN, China). Quantitative real-time PCR (RT-qPCR) was conducted using Universal SYBR Green Fast qPCR Mix (#RK21203, ABclonal, China) in combination with specific primers listed in Table S4 and the extracted DNA templates. Amplification was carried out on the ROCHE LightCycler® 480 System (Rotor-Gene 6000 Software, Sydney, Australia). The relative bacterial abundance was determined using the −ΔCt method, with universal 16S rRNA gene primers for Eubacteria serving as the internal reference.

Fecal metabolomic assays

The sample was stored at −80 °C and thawed on ice prior to analysis. A 400 μL solution of methanol and water (7:3, v/v), which included an internal standard, was combined with 20 mg of the sample and vortexed for 3 min. Subsequently, the mixture underwent sonication in an ice bath for 10 min, followed by an additional vortex for 1 min. It was then placed at −20 °C for 30 min. The sample was centrifuged at 12,000 rpm for 10 min at 4 °C, after which the sediment was discarded. The supernatant was centrifuged again at the same speed for 3 min, and a 200 μL aliquot was collected for LC-MS analysis.

For LC-MS, we employed a Waters ACQUITY UPLC BEH C18 column (1.8 μm, 2.1 mm × 100 mm) maintained at 40 °C, with a flow rate of 0.4 mL/min and an injection volume of 2 μL. The solvent system consisted of water (0.1% formic acid) and acetonitrile (0.1% formic acid). The elution process began with 5% mobile phase B, followed by a linear gradient to 90% mobile phase B over 11 min, maintained for 1 min, and then quickly reverted to 5% within 0.1 min, followed by a 1.9-min hold to restore initial conditions.

For TOF MS scans, the mass range was established at 50–1000 Da, with an accumulation time of 200 ms and dynamic background subtraction activated. Product ion scan parameters included a mass range of 25–1000 Da, an accumulation time of 40 ms, collision energy set to 30 or −30 V, and a collision energy spread of 15. Additional specifications comprised a resolution of UNIT, a charge state of 1 to 1, intensity at 100 cps, exclusion of isotopes within 4 Da, a mass tolerance of 50 mDa, and monitoring a maximum of 12 candidate ions per cycle.

The quality control (QC) samples were prepared by pooling extracts from all study samples to evaluate the reproducibility of sample processing. Multiple QC samples were interspersed during instrumental analysis to monitor analytical repeatability. The results showed high overlap in total ion chromatogram (TIC) curves of metabolite detection, with consistent retention times and peak intensities, indicating good signal stability for the same sample analyzed at different times (Fig. S13A). Additionally, Pearson correlation analysis of QC samples demonstrated strong consistency, confirming the overall stability of the detection process (Fig. S13B).

Identification of metabolite

The original data of mass spectrometry was converted into mzML format by ProteoWizard, and the peak extraction, alignment and retention time correction were carried out by the XCMS program. The "SVR" method is used to correct the peak area, and the peaks with a missing rate >50% in each group of samples are filtered. The screened peaks were corrected, and the identification information of metabolites was obtained by searching the self-built database in the laboratory, integrating the public database and the metDNA method.

The analysis software used for the identification of non-target metabolites is MetDNA. Priority of data using different databases: Metlin (http://metlin.scripps.edu/index.php), HMDB (https://hmdb.ca/) and Kegg (https://www.kegg.jp/). Mona (https://mona.fiehnlab.ucdavis.edu/), MassBank (http://www.massbank.jp/). followed by MetDNA construction and derivation, and finally Pubchem, ChemSpider. Non-target qualitative provides substances with a score of more than 0.5. Detailed matching rules are as follows: mz error of first-level and second-level matching is 25 ppm, rt error of first-level and second-level matching is 6 s, allowable error of parent ion Q1 searching is 25 ppm, and allowable error of MS2 searching is 50 ppm. rt carries out searching and filtering, with an allowable error of 60 s and the lowest second-level score of 0.3. The final score consists of three parts: fragment score and forward searching. The weight of the reverse search score is: the proportion of debris score = 0.1, the proportion of forward search score = 0.3, and the proportion of reverse search score = 0.6.

Quantification of tryptamine, indolepyruvate, and indole-3-acetaldehyde in feces

Levels of tryptamine, indolepyruvate, and indole-3-acetaldehyde in mouse fecal samples were measured using commercial ELISA kits (#F0233, F0442, F0450; FANNKEW Bioscience, China) following the manufacturer's instructions. In this format, antigen in the sample competes with a fixed amount of enzyme-labeled antigen (HRP-conjugated) for binding to a limited amount of specific antibody precoated on the 96-well plate. Briefly, standards and samples were diluted as required and added to the plate along with the HRP-conjugated antigen. After incubation at 37 °C for 1 h, the plate was washed five times with washing buffer to remove unbound components. Then, the TMB substrate solution was added and incubated in the dark for 15 min, followed by the addition of the stop solution. Absorbance was measured at 450 nm. The optical density is inversely proportional to the concentration of the target analyte in the sample. A standard curve was generated, and sample concentrations were interpolated accordingly. All samples were run in triplicate.

Associations between microbes and metabolites

In order to identify microbial features associated with diets and the metabolites, we measured probabilities of co-occurrence between species (based on metagenomic data) and either all metabolites. The Mmvec[25], a neural network solution inspired by deep-learning processing, to build a log-transformed conditional probability matrix from each cross-omics feature pair and apply singular value decomposition in order to represent co-occurrence in the form of biplots. We chose the model where accuracy was highest for different initialization conditions for the gradient descent algorithm (batch-size of 1000, 2500, and 5000, learn-rate of 0.001, epochs 9999), with low cross-validation error and model likelihood.

After obtaining the co-occurrence probabilities between metabolites and microbes significantly enriched under specific dietary conditions, we retained, for each metabolite, the microbes with the highest and lowest co-occurrence probabilities. To ensure biological relevance, we further filtered these pairs by requiring that the microbe with the highest co-occurrence probability be enriched under the same dietary condition as the metabolite, while the microbe with the lowest co-occurrence probability be enriched under the opposite dietary condition.

Metabolites origin and pathway analysis

To clarify where metabolites among diets come from, we utilized MetOrigin[24], an online tool for the integrative analysis of the metabolites and discriminating between host, food, microbiota, and co-metabolism activities associated with microbial metabolites. MetOrigin obtains the source information of metabolites by integrating multiple databases and cross-validating database IDs, chemical formulas, and compound names. In addition, the enrichment analysis of metabolite data was performed using MetaboAnalyst[77], which is a comprehensive platform dedicated to metabolomics data analysis. The hypergeometric test was used to calculate *p* values and corrected with the Benjamini and Hochberg (BH) method.

Other statistical analysis. For parametric feature-wise multivariable testing, we used MaAsLin2[23], which finds associations between microbial features and metadata of interest (including responsive microbes and metabolites on diets). To focus on the leading microbes and decrease the false positive, we only consider the microbes with relative abundance more than 0.1% in each sample. MaAsLin2 can utilize a transformed generalized linear model to associate each feature iteratively with covariates.

The Procrustes Analysis applied by the R package vegan of the correlation among OTUs, genes and metabolites was performed based on the principal component analysis (PCA) matrix obtained from the package sklearn (version 1.1.1) in Python. Using PCA for each data, we kept ten dimensions, which in addition can keep more than 95% information of each dataset. All statistical tests were using "scipy" or "statsmodels" package in Python. Most data visualization was applied using "ggplot2" in R or "matplotlib" in Python. In addition, we used GraPhlAn[78] to construct the cladogram. To evaluate whether the sample size was sufficient to detect meaningful biological effects, we performed a power analysis using the "effsize" R package[79]. And the detailed results are in Table S5.

Data availability

The raw sequence data reported in this paper have been deposited in the Genome Sequence Archive in the National Genomics Data Center, China National Center for Bioinformation/Beijing Institute of Genomics, Chinese Academy of Sciences, under accession number CRA013779 that are publicly accessible at https://bigd.big.ac.cn/gsa.

Received: 12 November 2024; Accepted: 8 September 2025;
Published online: 29 October 2025

References

1. Agus, A., Clément, K. & Sokol, H. Gut microbiota-derived metabolites as central regulators in metabolic disorders. *Gut* **70**, 1174–1182 (2021).
2. Chassaing, B. et al. Randomized controlled-feeding study of dietary emulsifier carboxymethylcellulose reveals detrimental impacts on the gut microbiota and metabolome. *Gastroenterology* **162**, 743–756 (2022).

3. Li, Z. et al. Effects of dietary protein on gut development, microbial compositions and mucin expressions in mice. *J. Appl. Microbiol.* **132**, 2262–2269 (2022).
4. Schroeder, B. O. et al. Bifidobacteria or fiber protects against diet-induced microbiota-mediated colonic mucus deterioration. *Cell Host Microbe* **23**, 27–40.e27 (2018).
5. Murros, K. E. Hydrogen sulfide produced by gut bacteria may induce Parkinson's disease. *Cells* **11**, 978 (2022).
6. Clifton, P. Effects of a high protein diet on body weight and comorbidities associated with obesity. *Br. J. Nutr.* **108**, S122–S129 (2012).
7. Nunes, E. A. et al. Systematic review and meta-analysis of protein intake to support muscle mass and function in healthy adults. *J. Cachexia Sarcopenia Muscle* **13**, 795–810 (2022).
8. Lattimer, J. M. & Haub, M. D. Effects of dietary fiber and its components on metabolic health. *Nutrients* **2**, 1266–1289 (2010).
9. Waddell, I. S. & Orfila, C. Dietary fiber in the prevention of obesity and obesity-related chronic diseases: from epidemiological evidence to potential molecular mechanisms. *Crit. Rev. Food Sci. Nutr.* **63**, 8752–8767 (2023).
10. Zmora, N., Suez, J. & Elinav, E. You are what you eat: diet, health and the gut microbiota. *Nat. Rev. Gastroenterol. Hepatol.* **16**, 35–56 (2019).
11. Albenberg, L. G. & Wu, G. D. Diet and the intestinal microbiome: associations, functions, and implications for health and disease. *Gastroenterology* **146**, 1564–1572 (2014).
12. Gentile, C. L. & Weir, T. L. The gut microbiota at the intersection of diet and human health. *Science* **362**, 776–780 (2018).
13. Tremaroli, V. & Bäckhed, F. Functional interactions between the gut microbiota and host metabolism. *Nature* **489**, 242–249 (2012).
14. Geuking, M. B., Köller, Y., Rupp, S. & McCoy, K. D. The interplay between the gut microbiota and the immune system. *Gut Microbes* **5**, 411–418 (2014).
15. Sugihara, K. & Kamada, N. Diet-microbiota interactions in inflammatory bowel disease. *Nutrients* https://doi.org/10.3390/nu13051533 (2021).
16. Russell, W. R. et al. High-protein, reduced-carbohydrate weight-loss diets promote metabolite profiles likely to be detrimental to colonic health. *Am. J. Clin. Nutr.* **93**, 1062–1072 (2011).
17. Huang, M.-C. et al. Inadequate energy and excess protein intakes may be associated with worsening renal function in chronic kidney disease. *J. Ren. Nutr.* **18**, 187–194 (2008).
18. Zhao, Y. et al. Genome-centric investigation of the potential succession pattern in gut microbiota and altered functions under high-protein diet. *Curr. Res. Food Sci.* **7**, 100600 (2023).
19. Moreno-Navarrete, J. M. & Fernandez-Real, J. M. The gut microbiota modulates both browning of white adipose tissue and the activity of brown adipose tissue. *Rev. Endocr. Metab. Disord.* **20**, 387–397 (2019).
20. Krautkramer, K. A., Fan, J. & Bäckhed, F. Gut microbial metabolites as multi-kingdom intermediates. *Nat. Rev. Microbiol.* **19**, 77–94 (2021).
21. Kolodziejczyk, A. A., Zheng, D. & Elinav, E. Diet–microbiota interactions and personalized nutrition. *Nat. Rev. Microbiol.* **17**, 742–753 (2019).
22. Chen, H. et al. A forward chemical genetic screen reveals gut microbiota metabolites that modulate host physiology. *Cell* **177**, 1217–1231.e1218 (2019).
23. Mallick, H. et al. Multivariable association discovery in population-scale meta-omics studies. *PLoS Comput. Biol.* **17**, e1009442 (2021).
24. Yu, G., Xu, C., Zhang, D., Ju, F. & Ni, Y. MetOrigin: discriminating the origins of microbial metabolites for integrative analysis of the gut microbiome and metabolome. *IMeta* **1**, e10 (2022).
25. Morton, J. T. et al. Learning representations of microbe–metabolite interactions. *Nat. Methods* **16**, 1306–1314 (2019).
26. Dolan, K. T. & Chang, E. B. Diet, gut microbes, and the pathogenesis of inflammatory bowel diseases. *Mol. Nutr. food Res.* **61**, 1600129 (2017).
27. Makki, K., Deehan, E. C., Walter, J. & Bäckhed, F. The impact of dietary fiber on gut microbiota in host health and disease. *Cell Host Microbe* **23**, 705–715 (2018).
28. Oliver, A. et al. High-fiber, whole-food dietary intervention alters the human gut microbiome but not fecal short-chain fatty acids. *Msystems* **6**, 00115–00121 (2021).
29. Kiilerich, P. et al. Effect of a long-term high-protein diet on survival, obesity development, and gut microbiota in mice. *Am. J. Physiol. Endocrinol. Metab.* **310**, E886–E899 (2016).
30. Li, Z. et al. The gut microbiome on a periodized low-protein diet is associated with improved metabolic health. *Front. Microbiol.* **10**, 709 (2019).
31. Peled, S. & Livney, Y. D. The role of dietary proteins and carbohydrates in gut microbiome composition and activity: a review. *Food Hydrocoll.* **120**, 106911 (2021).
32. Dong, T. S. et al. A high protein calorie restriction diet alters the gut microbiome in obesity. *Nutrients* **12**, 3221 (2020).
33. La Rosa, S. L. et al. The human gut Firmicute Roseburia intestinalis is a primary degrader of dietary β-mannans. *Nat. Commun.* **10**, 905 (2019).
34. David, L. A. et al. Diet rapidly and reproducibly alters the human gut microbiome. *Nature* **505**, 559–563 (2014).
35. del Campo, R. et al. Improvement of digestive health and reduction in proteobacterial populations in the gut microbiota of cystic fibrosis patients using a *Lactobacillus reuteri* probiotic preparation: a double blind prospective study. *J. Cyst. Fibros.* **13**, 716–722 (2014).
36. Everard, A. et al. Cross-talk between Akkermansia muciniphila and intestinal epithelium controls diet-induced obesity. *Proc. Natl Acad. Sci. USA* **110**, 9066–9071 (2013).
37. Turnbaugh, P. J. et al. An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* **444**, 1027–1031 (2006).
38. Roager, H. M. et al. Whole grain-rich diet reduces body weight and systemic low-grade inflammation without inducing major changes of the gut microbiome: a randomised cross-over trial. *Gut* **68**, 83–93 (2019).
39. Wastyk, H. C. et al. Gut-microbiota-targeted diets modulate human immune status. *Cell* **184**, 4137–4153.e4114 (2021).
40. Zhu, Y. et al. Intake of meat proteins substantially increased the relative abundance of genus *Lactobacillus* in rat feces. *PLoS ONE* **11**, e0152678 (2016).
41. Amaretti, A. et al. Profiling of protein degraders in cultures of human gut microbiota. *Front. Microbiol.* **10**, 2614 (2019).
42. Nakamura, S., Kuda, T., Midorikawa, Y., Takahashi, H. & Kimura, B. Typical gut indigenous bacteria in ICR mice fed a soy protein-based normal or low-protein diet. *Curr. Res. Food Sci.* **4**, 295–300 (2021).
43. Richard, C. et al. 679 High fiber diet modifies gut microbiome, propionate production, intratumor immune response and is associated with outcome in patients with lung cancer treated with immune checkpoint inhibitors. *J. Immunothe. Cancer* **8**, 3 (2020).
44. Everard, A. & Cani, P. D. Diabetes, obesity and gut microbiota. *Best. Pract. Res. Clin. Gastroenterol.* **27**, 73–83 (2013).
45. Structure function and diversity of the healthy human microbiome. *Nature* **486**, 207–214 (2012).
46. Yang, F. et al. Effect of diet and intestinal AhR expression on fecal microbiome and metabolomic profiles. *Microb. Cell Fact.* **19**, 1–18 (2020).
47. Duca, D., Lorv, J., Patten, C. L., Rose, D. & Glick, B. R. Indole-3-acetic acid in plant–microbe interactions. *Antonie Van Leeuwenhoek* **106**, 85–125 (2014).
48. Spaepen, S., Vanderleyden, J. & Remans, R. Indole-3-acetic acid in microbial and microorganism-plant signaling. *FEMS Microbiol. Rev.* **31**, 425–448 (2007).
49. Tintelnot, J. et al. Microbiota-derived 3-IAA influences chemotherapy efficacy in pancreatic cancer. *Nature* **615**, 168–174 (2023).

50. Agus, A., Planchais, J. & Sokol, H. Gut microbiota regulation of tryptophan metabolism in health and disease. *Cell Host Microbe* **23**, 716–724 (2018).
51. Mawe, G. M. & Hoffman, J. M. Serotonin signalling in the gut—functions, dysfunctions and therapeutic targets. *Nat. Rev. Gastroenterol. Hepatol.* **10**, 473–486 (2013).
52. Gribble, F. M. & Reimann, F. Metabolic messengers: glucagon-like peptide 1. *Nat. Metab.* **3**, 142–148 (2021).
53. Roager, H. M. & Licht, T. R. Microbial tryptophan catabolites in health and disease. *Nat. Commun.* **9**, 3294 (2018).
54. Dong, F. et al. Intestinal microbiota-derived tryptophan metabolites are predictive of Ah receptor activity. *Gut Microbes* **12**, 1788899 (2020).
55. Venkatesh, M. et al. Symbiotic bacterial metabolites regulate gastrointestinal barrier function via the xenobiotic sensor PXR and Toll-like receptor 4. *Immunity* **41**, 296–310 (2014).
56. Reeves, P. G., Rossow, K. L. & Lindlauf, J. Development and testing of the AIN-93 purified diets for rodents: results on growth, kidney calcification and bone mineralization in rats and mice. *J. Nutr.* **123**, 1923–1931 (1993).
57. Kembel, S. W. et al. Picante: R tools for integrating phylogenies and ecology. *Bioinformatics* **26**, 1463–1464 (2010).
58. Quast, C. et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* **41**, D590–D596 (2012).
59. Dixon, P. VEGAN, a package of R functions for community ecology. *J. Veg. Sci.* **14**, 927–930 (2003).
60. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).
61. Xu, H. et al. FastUniq: a fast de novo duplicates removal tool for paired short reads. *PLoS ONE* **7**, e52249 (2012).
62. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
63. Li, D., Liu, C. M., Luo, R., Sadakane, K. & Lam, T. W. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* **31**, 1674–1676 (2015).
64. Zhu, W., Lomsadze, A. & Borodovsky, M. Ab initio gene identification in metagenomic sequences. *Nucleic Acids Res.* **38**, e132–e132 (2010).
65. Wheeler, T. J. & Eddy, S. R. nhmmer: DNA homology search with profile HMMs. *Bioinformatics* **29**, 2487–2489 (2013).
66. Eberhardt, R. Y. et al. AntiFam: a tool to help identify spurious ORFs in protein annotation. *Database* **2012**, bas003 (2012).
67. Steinegger, M. & Soding, J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat. Biotechnol.* **35**, 1026–1028 (2017).
68. Aramaki, T. et al. KofamKOALA: KEGG Ortholog assignment based on profile HMM and adaptive score threshold. *Bioinformatics* **36**, 2251–2252 (2020).
69. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
70. Pruitt, K. D., Tatusova, T. & Maglott, D. R. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.* **35**, D61–D65 (2007).
71. Buchfink, B., Xie, C. & Huson, D. H. Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* **12**, 59–60 (2015).
72. Huson, D. H., Auch, A. F., Qi, J. & Schuster, S. C. MEGAN analysis of metagenomic data. *Genome Res.* **17**, 377–386 (2007).
73. Kang, D. D. et al. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ* **7**, e7359 (2019).
74. Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* **25**, 1043–1055 (2015).
75. Olm, M. R., Brown, C. T., Brooks, B. & Banfield, J. F. dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *ISME J.* **11**, 2864–2868 (2017).
76. Zhou, C. et al. GTDB: an integrated resource for glycosyltransferase sequences and annotations. *Database* https://doi.org/10.1093/database/baaa047 (2020).
77. Pang, Z. et al. MetaboAnalyst 5.0: narrowing the gap between raw spectra and functional insights. *Nucleic Acids Res.* **49**, W388–W396 (2021).
78. Asnicar, F., Weingart, G., Tickle, T. L., Huttenhower, C. & Segata, N. Compact graphical representation of phylogenetic data and metadata with GraPhlAn. *PeerJ* **3**, e1029 (2015).
79. Meissel, K. & Yao, E. S. Using Cliff's delta as a non-parametric effect size measure: an accessible web app and R tutorial. *Pract. Assess. Res. Eval.* **29**, 1 (2024).

Acknowledgements

We express our gratitude to Syeda Sundas Batool for the valuable assistance in the preparation of this manuscript. This work was funded by the National Natural Science Foundation of China (32170071 and 82270564), Natural Science Foundation of Hunan Province (2022JJ30916), China Postdoctoral Science Foundation (2022M713521), and Central South University Innovation-Driven Research Program (No. 2023CXQD059).

Author contributions

Z.Y. and S.J.C. conceived the study. Y.M.Z. analyzed the data. Y.S.Q. performed the experiments. Y.Z., J.H., and S.Y. wrote the manuscript. All authors contributed to the article and approved the submitted version.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41538-025-00575-9.

Correspondence and requests for materials should be addressed to Shuijiao Chen or Zheng Yu.

Reprints and permissions information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

© The Author(s) 2025