Supplementary Methods

Clinical trial registration

This study was registered in the Chinese Clinical Trial Registry (ChiCTR-OOC-17012768) and ClinicalTrials.gov (NCT03390582).

Exclusion criteria

The exclusion criteria were applied to all groups: pregnancy; lactation; cigarette smoking; alcohol addiction; hypertension; diabetes mellitus; body mass index (BMI)<18.5; BMI>27; recent (<3 months prior) use of antibiotics, probiotics, prebiotics, symbiotics, hormonal medication, laxatives, proton pump inhibitors, insulin sensitizers or traditional Chinese medicine; nodular or cystic thyroid disease; a known history of a disease with an autoimmune component such as multiple sclerosis (MS), rheumatoid arthritis (RA), irritable bowel syndrome (IBS), or inflammatory bowel disease (IBD); and a history of malignancy or any gastrointestinal tract surgery (e.g., gastrectomy, bariatric surgery, colectomy, ileectomy, cholecystectomy or appendectomy).

Sample collection and clinical parameters

All subjects were examined in the morning after an overnight fast (≥8 h). Peripheral blood (17 mL) was collected from all participants and stored in ethylenediaminetetraacetic acid (EDTA) anticoagulation tubes or separation gel coagulation tubes for routine blood, hepatic function, blood lipid, thyroid function and thyroid autoantibody examinations. All subjects were provided with commode specimen collection kits for stool collection, and the samples were sent to our laboratory on an ice pack within two hours. Each fecal sample was immediately divided into aliquots, frozen in liquid nitrogen, and stored at -80°C. In addition, sodium azide was added to the samples for metabolomics analysis.

Routine blood and biochemical analyses of clinical parameters

Except for thyrotropin receptor antibody (TR-Ab), the clinical parameters in this study were examined in the clinical laboratory of the First Affiliated Hospital of Harbin Medical University. TR-Ab was examined in the clinical laboratory of

Heilongjiang Provincial Hospital. Serum thyroid-stimulating hormone (TSH), free thyroxine (fT4), and free triiodothyronine (fT3) levels were measured by a chemiluminescent immunoassay (Abbott Diagnostics, Tokyo, Japan). TSH receptor antibody (TR_Ab) was also measured by a chemiluminescence immunoassay (Snibe Diagnostic, Shenzhen, China). Routine blood indices were measured by flow cytometry (SYSMEX, Tokyo, Japan). Blood biochemical indices (hepatic function and blood lipids) were assessed using an automated biochemistry analyzer (Beckman Coulter, California, USA) and auxiliary reagents.

The normal reference ranges for the clinical parameters are as follows: **(1) Thyroid function:** TSH, 0.3-4.94 μIU/L; fT4, 0.70-1.48 ng/dL; fT3, 1.71-3.71 pg/mL; and TR-Ab, 0.00-1.50 IU/L. **(2) Routine blood parameters:** WBC, 3.69-9.16 10^9/L; RBC, 3.68-5.13 10^{12}/L; PLT, 98.00-300.20 10^9/L; NEUT, 2.00-7.00 10^9/L; LYMP, 0.84-4.00 10^9/L; MONO, 0.12-1.00 10^9/L; EO, 0.02-0.50 10^9/L; BASO, 0.00-1.00 10^9/L; HGB, 113.00-151.00 g/L; HCT, 33.50-45.00%; MCV, 82.60-99.10 fL; MCH, 26.90-33.30 pg; and MCHC, 322.00-363.00 g/L. **(3) Hepatic function:** ALT, 5.00-35.00 U/L; AST, 8.00-40.00 U/L; TP, 60.00-83.00 g/L; ALB, 3400-54.00 g/L; GLB, 16.00-35.00 g/L; GGT, 7.0-45.00 U/L; AKP/ALP, 40.00-150.00 U/L; TBIL, 3.40-21.00 umol/L; DBIL, 0.00-3.40 umol/L; IBIL, 0.00-17.10 umol/L; and TBA, 0.00-10.00 umol/L. **(4) Blood lipids:** CHOL, 3.35-5.71 mmol/L; TG, 0.48-2.25 mmol/L; HDL, 1.03-1.55 mmol/L; LDL, 0.26-4.11 mmol/L; ApoA, 1.20-1.60 g/L; ApoB, 0.80-1.05 g/L; and LPa, 0.000-400.00 mg/L.

Because clinical parameters have normal reference ranges, they are not only count data but also measurement data. In this study, the count data were used to calculate the incidence of complications, and the measurement data were used for correlation analyses.

gDNA extraction

DNA extraction was performed within one month after sample collection. Bacterial DNA was extracted from fecal samples at Novogene Bioinformatics Technology Co., Ltd. (Beijing, China) using a TIANGEN kit according to the manufacturer's recommendations. The DNA concentration and purity were monitored

on 1% agarose gels. After concentration determination, the DNA was diluted to 1 ng/μL using sterile water. The extracted DNA was stored at -20°C. The concentration of bacterial DNA was measured using a Nanodrop 2000 (Thermo Scientific, USA).

Amplicon generation and purification

Bacterial genomic DNA was amplified with the primers 341F (CCTAYGGGRBGCASCAG) and 806R (GGACTACNNGGGTATCTAAT), which are specific to the V3-V4 hypervariable regions of the 16S rRNA gene. All PCRs were carried out with Phusion® High-Fidelity PCR Master Mix (New England Biolabs, Ipswich, Massachusetts, USA). An equal volume of 1× loading buffer (containing SYB green) was mixed with the PCR products and subjected to electrophoresis on 2% agarose gels for detection. Samples with a bright, primary band between 400-450 bp were selected for further experiments. The products of the same sample were combined and subjected to electrophoresis. DNA of the correct size was purified using a gel extraction kit (Qiagen, Hilden, Germany) and quantified using a Qubit instrument (Life Technologies, Carlsbad, CA).

Library preparation and sequencing

Sequencing libraries were generated using a TruSeq® DNA PCR-Free Sample Preparation Kit (Illumina, San Diego, CA, USA) following the manufacturer's recommendations, and index codes were added. Library quality was assessed using a Qubit@ 2.0 Fluorometer (Thermo Scientific, USA) and an Agilent Bioanalyzer 2100 system (Agilent Technologies, Santa Clara, CA). Finally, the library was sequenced on an Illumina HiSeq 2500 platform (Illumina), and 250-bp paired-end reads were generated.

Paired-end read assembly and quality control

Paired-end reads were assigned to samples based on their unique barcodes and were truncated by cleaving the barcode and primer sequence. Paired-end reads were merged using FLASH (v.1.2.7; San Jose, California, USA), a rapid and highly accurate analysis tool designed to merge paired-end reads when at least some of the reads overlap the reads generated from the opposite end of the same DNA fragment. The splicing sequences were termed raw tags.

Quality filtering of the raw tags was performed under specific filtering conditions to obtain high-quality clean tags according to the QIIME (v.1.7.0) quality control (QC) process. The tags were compared with a reference database (the Gold database) using the UCHIME algorithm to detect chimera sequences, and the chimera sequences were then removed. The effective tags were finally obtained.

OTU cluster and species annotation

Sequence analysis was performed using Uparse software (v.7.0.1001; Edgar, Tiburon, California, USA, http://drive5.com/uparse/). Sequences with ≥97% similarity were assigned to the same operational taxonomic units (OTUs). Representative sequences for each OTU were screened for further annotation. For each representative sequence, the SILVA128/16S database was used based on the Ribosomal Database Project (RDP) classifier (v.2.2) algorithm to annotate taxonomic information. To study the phylogenetic relationships between different OTUs and differences in dominant species in different samples (groups), we conducted multiple sequence alignment using MUSCLE software (v.3.8.31). OTU abundance information was normalized using the sequence number corresponding to the sample with the fewest sequences.

Ultra-performance liquid chromatography quadrupole time-of-flight tandem mass spectrometry (UPLC-Q-TOF/MS) metabonomic profiling of fecal samples

Fecal water was extracted by mixing a weighed sample of thawed stool with methanol at a ratio of 3 mL/g. Then, 20 μL of internal standard (0.3 mg/mL 2-chloro-D-phenylalanine) was added. All samples were ground, subjected to ultrasonic extraction in ice water for 30 min, and centrifuged at 10,000 rpm at 4°C for 10 min. For UPLC-Q-TOF/MS analysis, 200 μL of the supernatant was harvested. Metabonomic analysis was performed using a Waters UPLC system equipped with a binary solvent delivery manager and a sample manager coupled with a Waters Q-TOF Mass Spectrometer and equipped with an electrospray ionization (ESI) source that can operate in both positive and negative ion modes (Waters Corporation, Milford, MA, USA). Standardized samples were prepared for quality control (QC) by pooling aliquots of all samples and then analyzed using the same method. The pooled QC

sample was injected two times at the beginning and one time at the end of the run, and QC samples were injected at regular intervals (every 6 samples) throughout the analytical run to obtain a set of data from which repeatability could be assessed.

LC conditions

Column: acquity BEH C18 column (100 mm × 2.1 mm i.d., 1.7 μm; Waters, Milford, USA). Solvent: the column was maintained at 45°C, and separation was achieved using the following gradient: 5% B-80% B over 0-10 min, 80% B-100% B over 10.0-12.5 min, 100% B-5% B over 12.5-12.6 min, and holding for 12.6-14.0 min at 5% B at a flow rate of 0.40 mL/min, where B is acetonitrile (0.1% (v/v) formic acid) and A is aqueous formic acid (0.1% (v/v) formic acid). The injection volume was 3 μL, and the column temperature was set at 45℃. The source temperature and desolvation temperature were set at 120℃ and 500℃, respectively. Nitrogen was used as both the desolvation gas (900 L/h) and the cone gas (50 L/h).

Centroid data were collected from 50 to 1,000 m/z with a scan time of 0.1 s and an interscan delay of 0.02 s over a 13-min analysis time using the following parameters: retention time (RT) range, 0.5-14.0 min; mass range, 50-1000 Da; and mass tolerance, 0.01 Da. Isotopic peaks were excluded for analysis, the noise elimination level was set at 10, and the minimum intensity was set to 15% of the base peak intensity. Finally, RT tolerance was set to 0.01 min. An excel file with three-dimensional datasets including the m/z, peak RT and peak intensities was obtained, and RT-m/z pairs were used as the identifier for each ion. The resulting matrix was further reduced by removing any peaks with missing values (ion intensity=0) in more than 60% of the samples. The internal standard was used for data quality control (reproducibility).

Microbiome analysis

Alpha diversity was applied to analyze the complexity of the species diversity in each sample based on 6 indices, including observed species richness, the Chao1 index, the Shannon diversity index, the Invsimpson index, the abundance-based coverage estimator (ACE), and Good's coverage. These indices were calculated for our samples using Quantitative Insights Into Microbial Ecology (QIIME, v.1.7.0) based on rarefied

OTU counts and displayed using R software (v.2.15.3). The *Firmicutes*/*Bacteroidetes* (F/B) ratio and microbial dysbiosis index (MDI) were used to evaluate gut microbiota dysbiosis. The MDI is defined as the log of the total abundance levels of species that are increased in patients divided by the total abundance levels of species that are decreased at the genus level. Beta-diversity analysis was used to evaluate similarities and differences in species complexity among the samples, and the beta diversity (Bray-Curtis distance) was calculated using QIIME software (v.1.7.0) based on the rarefied OTU counts. Principal coordinate analysis (PCoA) was performed to obtain principal coordinates and visualize complex multidimensional data. PCoA results were displayed using the Weighted Correlation Network Analysis (WGCNA) package, stats package and ggplot2 package in R software (v.2.15.3). Permutational multivariate analysis of variance (PERMANOVA) of the distance matrices was implemented in the vegan package in R to determine whether the case/control status explained the variation in the gut microbial community composition. A Venn diagram was used to illustrate the overlap of different species or Kyoto Encyclopedia of Genes and Genome (KEGG) pathways.

Based on the normalized relative abundance matrix, LEfSe uses the Mann-Whitney U test to detect features with significantly different abundance levels between assigned taxa and performs linear discriminant analysis (LDA) to estimate the effect size of each feature, emphasizing both statistical significance and biological relevance. The metagenomes of the gut microbiota were imputed from 16S rRNA gene sequences using PICRUSt. The predicted functional composition profiles were respectively collapsed into level 1, 2 and 3 KEGG pathways, which represent pathways that were present in <10% of the samples but were not included in the comparison analysis.

Metabolome analysis

UPLC-Q-TOF/MS raw data were analyzed using progenesis QI software (Waters Corporation, Milford, USA). Positive and negative data were respectively imported into SIMCA software (v.14.1, Umetrics, Umeå, Sweden) for multivariate statistical analysis. Principal component analysis (PCA) and supervised partial least

squares-discriminant analysis (PLS-DA) were performed to visualize metabolic alterations among the groups. Orthogonal partial least squares-discriminate analysis (OPLS-DA) was applied to build a classification model to identify differential metabolites based on the variable importance in projection (VIP) value, fold change (FC) value and P value. The goodness of fit was quantified by R^2Y, while the predictive ability was indicated by Q^2, and models with R^2Y and Q^2 values greater than or equal to 0.5 were considered suitable for recognition analysis. A cross-validation procedure and testing with 200 random permutations were performed using SIMCA software (v.14.1) to avoid overfitting of the OPLS-DA model. Metabolites were identified by comparing the observed accurate mass and mass spectrometry (MS/MS) spectra; the accurate mass data and spectra were available in online databases, including the METLIN (http://metlin.scripps.edu), human metabolite (HMDB; http://hmdb.ca) and progenesis QI (Waters Corporation, Milford, USA) databases. An independent-samples t-test (Mann-Whitney U test) was applied to measure the significance of each metabolite between two groups. Different metabolites with the greatest VIP values and statistical significance ($VIP>1.0$, $P<0.05$ and $FC>1.5$) were selected, and volcano plots were generated in SIMCA software (v.14.1) to show differential metabolites.

A pathway analysis of the differential metabolites was first performed with MetaboAnalyst 3.0 (http://www.metaboanalyst.ca/), which is based on database sources, including the KEGG database (http://www.genome.jp/kegg/) and the HMDB, to identify the top altered pathways for analysis and visualization. Then, other implicated biomarker pathways were further interpreted using references and databases. Enrichment analyses are usually performed to determine whether a group of metabolites appears in a functional node. Compared with all background metabolites, significant pathway enrichment analyses enhance the reliability of a study and identify biological processes that are most relevant to biological phenomena. KOBAS 2.0 (http://kobas.cbi.pku.edu.cn/home.do) was used to identify significantly enriched pathways with Fisher's exact test. Pathways with $P<0.05$ were considered significantly enriched. The KEGG enrichment pathways are displayed as a bubble

plot generated using the ggplot2 package in R (v.2.15.3).