Supplementary methods

ASD cohort

A total of 773 participants with clinically diagnosed ASD were recruited (aged between 16 months and 19 years) (online supplemental tables S1 and S2). ASD was diagnosed based on the Diagnostic and Statistical Manual of Mental Disorders Fifth Edition (DSM-V) criteria by two or three psychiatrists in a face-to-face observation and interviewing the guardian(s). In the case of children < 2 years of age, the Childhood Autism Rating Scale (CARS) and Gesell Developmental Schedules (GDS) schedules were conducted. For children > 2 years of age, the Autism Diagnostic Observation Schedule (ADOS), Autism Diagnostic Interview-Revised (ADI), or ADOS+ADI were used. Most participant diagnoses were confirmed with ADI-R (197/773), ADOS (354/773), or both (ADOS + ADI, 14/773), natively or locally. All participants will be reassessed in our outpatient or ward and obtained continuous clinical observations based on diagnostic criteria in the follow-up clinical intervention. Given the economic acceptability, convenience, and consistency in clinical practice, the severity of ASD was scored by accumulating the severity scores of clinical manifestations by psychiatrists. The patients' clinical manifestations were clinically evaluated as mild (score 1), moderate (score 2), and severe (score 3) according to DSM-IV and DSM-V[1]; the online supplemental table S3 presents the detailed clinical evaluation standard.

Control cohort

A total of 429 neurotypical (NT) children (aged between 11 months and 15 years) and 20 unrelated healthy adults (aged 16–24 years) were recruited (online supplemental tables S1 and S2). All healthy participants were from schools or companies that cooperated with our hospital for routine examinations. The 429 relatively young participants were mainly from the provinces of Hunan, Shandong, Zhejiang, Shanghai, and Beijing (online supplemental table S4) and were employed in the prospective, multiregional, and observational cohort study. The metadata used in this study are detailed in online supplemental table S1. For the information registration, each sample was scored as 1 (yes) or 0 (no) for each factor. Comorbidities, such as gastrointestinal (GI) problems, sleep complaints, and immune abnormalities, indicated the body conditions during the past two weeks before sampling. The summaries of age, demographic, clinical, and district characteristics are provided in online supplemental tables S2 and S4. No participant took any drug, such as antibiotics, opioids, metformin and statins, or dietary supplements, such as probiotics or prebiotics, which have been proven to impact gut microbiota, in the month before sampling. The use of antibiotics in

the three months prior to sampling was recorded in detail (online supplemental table S1).

Validation cohort

Full age brackets involving the appended cohort were conducted as validation cohort 1, which was recruited by our outpatient department from 2019-2021. This cohort included 73 subjects with ASD and 32 age-matched NT subjects (online supplemental tables S1 and S2).

The unrelated cohort reported by Dan et al. 2020[2] was utilized as validation cohort 2 and included 143 subjects with a clinical diagnosis of ASD (average age, 4.937 ± 0.155) and 143 age- and sex-matched NT individuals (average age, 5.189 ± 0.170) in China.

The unrelated cohort reported by Cao et al. 2021[3] was employed as validation cohort 3 and included 45 subjects with ASD (average age 6.80 ± 3.79) and 41 NT subjects (average age 5.16 ± 0.99) in China.

Stool sample collection and DNA extraction

All of the stool samples from participants were collected by themselves or their trained guardians at home/outpatient/ward within 3 minutes after defecation. After sampling, the container was labeled and transferred (<20 °C) to GUHE Laboratories (Hangzhou, China) within 3 days and stored at -80 °C until further processing. Stool samples were excluded if any organic changes were detected. Total bacterial genomic DNA was extracted using the GHFDE100 DNA isolation kit (Zhejiang Hangzhou Equipment Preparation: 20190952) in accordance with the manufacturer's instructions. Incidentally, the genomic extraction method employed by the kit has obtained a Chinese national invention patent (NO: ZL201511009389.7). The quantity and quality of the extracted DNA were measured using NanoDrop ND-1000 spectrophotometer (Thermo Fisher Scientific, USA) and agarose gel electrophoresis, respectively.

16S rRNA gene sequencing

For each sample, we amplified variable V4 regions of the 16S rRNA gene using the forward primer 515F (5'-GTGCCAGCMGCCGCGGTAA-3') and the reverse primer 806R (5'-GGACTACHVGGGTWTCTAAT-3'). The PCR components contained 25 μl of Phusion High-Fidelity PCR Master Mix, 3 μl (10 uM) of each Forward and Reverse primer, 10 μl of DNA Template, 3μl of DMSO, and 6 μl of ddH2O. The following cycling conditions were used: initial denaturation at 98 °C for 30s followed by 25 cycles of

denaturation at 98 °C for 15 s, annealing at 58 °C for 15 s, and extension at 72°C for 15 s and a final extension of 1 min at 72°C. PCR amplicons were purified with Agencourt AMPure XP Beads (Beckman Coulter, Indianapolis, IN) and quantified using the PicoGreen dsDNA Assay Kit (Invitrogen, Carlsbad, CA, USA).

Bioinformatics and statistical analysis

Briefly, raw sequences with exactly matched barcodes were assigned to their unique corresponding samples and identified as valid sequences. Inferior sequences were filtered out according to the following criteria: (i) sequences with <150 bp length or <20 average Phred score, (ii) sequences that contained ambiguous bases or >8 bp mononucleotide repeats[4 5] and the average of clean reads from each sample was 126340. Qualified paired-end reads were blasted, dereplicated (--derep_fulllength), clustered (--cluster_unoise), and chimera detected (--uchime3_denovo) using VSEARCH (V2.4.4)[6] against the SILVA138 database[7] followed by assembly into operational taxonomic units (OTUs) with sequence similarity ≥97% using the Quantitative Insights Into Microbial Ecology (QIIME2, v2020.6) pipeline. OTUs containing <0.001% of the total sequences across all samples were discarded. To minimize the difference in sequencing depth across samples, an average, rounded rarefied OTU table was generated by averaging 100 evenly resampled OTU subsets under 90% of the minimum sequencing depth for further analysis. OTU-level alpha diversity, such as the Chao1, richness, abundance-based coverage estimator, Shannon, and Simpson index of each sample were calculated using the OTU table in QIIME. Beta diversity analysis was performed using UniFrac distance metrics[8 9] and visualized via principal component analysis (PCA) based on the OTU-level compositional profiles[10]. The significance in the differentiation of microbiota structure among groups was assessed by permutational multivariate analysis of variance (PERMANOVA) using the R package “vegan” For taxa, the relative abundance changes between groups were statistically analyzed using the Kruskal-Wallis test from the R stats package at the phylum, class, order, family, genus, and species levels. Based on the normalized OTU tables, functional modules predicted by PICRUSt2.3 were used to predict metagenomic functions (MetaCYC) [11]. Human gut-brain modules (GBMs) were profiled using the Omixer-RPM version 1.0 (https://github.com/raeslab/omixer-rpm) with default parameters. Co-occurrence analysis was performed by calculating Spearman’s rank correlations between microbial taxa/function and clinical phenotype. The *Veillonella* correlation network was performed using the Pearson correlation. Correlations with $p < 0.05$, were validated as pre-significant co-features. Microbial taxa pre-co-feature are visualized in Figure 4B.

For each functional pre-co-feature, we calculated the perturbed score, defined as

$$S = -log10(P) * |RHO|$$

where p-value was the paired t-test between functional 'pre-co-feature' and clinical phenotype. Values of the perturbed score of each functional 'pre-co-feature' (mainly to METACYC) at different age brackets were written into the same box and then averaged as

$$Rs = \frac{\sum_{t=1}^{T} -\log_{10} P * |RHO|}{T}$$

All functional pre-co-features were ranked by Rsscore for further analysis. The correlation network was drawn using Cytoscape according to the perturbed score, and the circos plot was plotted using the R package.

The multivariate linear modeling system was used to calculate the association between selected microbial features and factors for the fixed effects of potentially confounding covariates (other factors from metadata). Associations from MaAslin2[12 13] generally represented the causal relationship between the abundance change of microbial features and metadata (usually the categorical data). Significance values across all associations were then adjusted using the Benjamini–Hochberg False Discovery Rate (FDR) method.

EnvFit[14 15] was performed using the 'vegan' R package. The covariates and significance of each factor were determined using EnvFit based on NMDS with Bray-Curtis dissimilarity. A total of 33 factors were included in the effect size calculation (Online Supplemental Table S6). The significance value of each factor was determined based on 10,000 permutations and adjusted using FDR adjustment (Benjamini–Hochberg procedure).

Random forest analysis was performed to discriminate samples from different groups using the R package "randomForest" with 1,000 trees, and all default settings turned off [16 17]. The generalization error was estimated using 10-fold cross-validation. The expected "baseline" error was also included, which was obtained using a classifier that simply predicts the most common category label. The SHAP (SHapley Additive exPlanations) value was regarded as the sum of all quantitative impacts of potential influencing factors on the feature. The feature with equal importance in the ASD or NT group typically indicated that individual factors induced the feature with the lowest bias.

Deep neural network for microbiota age quantification.

Feature selection models were trained using a full list of OTU-level features, which

included 7,573 microbial taxa. Training and validation sets were separated to contain 90% and 10% of the profiles in all cases. The regressors were built using taxonomic profiles derived from individual samples (sample-based models). The model was trained as a regressor using five-fold cross-validation. After completing the grid search for various model configurations, the best-performing model was selected based on the maximal R^2 score. The model contained three hidden layers with 512 nodes in each, with a PReLU activation function, Adam optimizer, dropout fraction of 0.5, and a learning rate of 0.001. To verify the importance of features derived from the sample-based DNN model, gradient boosting was used, as implemented in the XGBoost Python library. The best-performing XGBoost model was trained using the following parameters: linear_nthread = 35, max_depth = 6, max_delta_step = 2, lambda= 0, gamma = 0.1, eta=0.1, and alpha = 0.5. The performance of the XGBoost models was evaluated using the MAE.

Microbial relationship alteration analysis

The PM score was used to quantify the relationship alteration to remove the potential impact of changes on microbial abundance. A threshold FDR q-value < 0.05, p-value < 0.05, PM score > 0.25, and taxa detection rate > 0.25 were used to filter significant relationship alterations.

In addition to the PM2RA analysis, changes in the occurrence network were calculated. Pearson's correlation was used to describe the relationships between taxa and build the occurrence network. After obtaining the two occurrence networks for NT and ASD cohorts, Pearson correlation differences were calculated pairwise. Pearson correlation alterations with difference > 0.7 or difference < -0.7 and taxa detection rate > 0.25 were used to build the final altered occurrence network.

Identification of microbial relationship alteration with increasing ASD score

The clinical symptoms of children with ASD were clustered into four groups: group 1 with ASD scores 1-3 (included), group 2 with ASD scores 4-5 (included), group 3 with ASD score 6, and group 4 with ASD scores 7-9 (included). The paired microbial relationship alteration between these four groups and the NT cohort was quantified using PM2RA. We performed linear regressions on each microbial relationship to identify changes in the microbial relationship with increasing ASD scores. Several thresholds were used to filter the expanded relationship alterations. The linear regression R-square should be >0.7, and the regression coefficients should be > 0.05. Compared with NT, the PM score in groups 3 and 4 should be >0.2. The corresponding

detection rate of taxa should be >0.25. The information regarding the 54 identified microbial relationship alterations is shown in online supplemental table S11.

References

1. (APA) APA. Diagnostic and statistical manual of mental disorders (DSM-5®). American Psychiatric Pub2013.
2. Dan Z, Mao X, Liu Q, et al. Altered gut microbial profile is associated with abnormal metabolism activity of Autism Spectrum Disorder. *Gut Microbes* 2020;11(5):1246-67.
3. Cao X, Liu K, Liu J, et al. Dysbiotic Gut Microbiota and Dysregulation of Cytokine Profile in Children and Teens With Autism Spectrum Disorder. *Front Neurosci* 2021;15:635925.
4. Chen H, Jiang W. Application of high-throughput sequencing in understanding human oral microbiome related with health and disease. *Frontiers in Microbiology* 2014;5 doi: ARTN 508.
5. Gill SR, Pop M, DeBoy RT, et al. Metagenomic analysis of the human distal gut microbiome. *Science* 2006;312(5778):1355-59. doi: 10.1126/science.1124234
6. Rognes T, Flouri T, Nichols B, et al. VSEARCH: a versatile open source tool for metagenomics. *PeerJ* 2016;4:e2584.
7. Quast C, Pruesse E, Yilmaz P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* 2013;41(Database issue):D590-6.
8. Lozupone C, Knight R. UniFrac: a new phylogenetic method for comparing microbial communities. *Appl Environ Microb* 2005;71(12):8228-35.
9. Lozupone CA, Hamady M, Kelley ST, et al. Quantitative and qualitative beta diversity measures lead to different insights into factors that structure microbial communities. *Appl Environ Microb* 2007;73(5):1576-85.
10. Ramette A. Multivariate analyses in microbial ecology. *Fems Microbiol Ecol* 2007;62(2):142-60.
11. Douglas GM, Maffei VJ, Zaneveld JR, et al. PICRUSt2 for prediction of metagenome functions. *Nat Biotechnol* 2020;38(6):685-88.
12. Morgan XC, Tickle TL, Sokol H, et al. Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. *Genome Biol* 2012;13(9):R79.
13. Mallick H, Rahnavard A, McIver LJ, et al. Multivariable Association Discovery in Population-scale Meta-omics Studies. *bioRxiv* 2021 doi: https://doi.org/10.1101/2021.01.20.427420
14. He Y, Wu W, Zheng HM, et al. Regional variation limits applications of healthy gut microbiome reference ranges and disease models. *Nat Med* 2018;24(10):1532-35.
15. Stewart CJ, Ajami NJ, O'Brien JL, et al. Temporal development of the gut microbiome in early childhood from the TEDDY study. *Nature* 2018;562(7728):583-88.
16. Breiman L. Random forests. *Mach Learn* 2001;45(1):5-32.

17. Svetnik V, Liaw A, Tong C, et al. Random forest: a classification and regression tool for compound classification and QSAR modeling. *J Chem Inf Comput Sci* 2003;43(6):1947-58.