







Leveraging Existing 16SrRNA Microbial Data to Define a Composite Biomarker for Autism Spectrum Disorder

 YuShuang Xu,^a  YiHua Wang,^b JinShuang Xu,^c Yu Song,^d  BingQiang Liu,^b  ZhiFan Xiong^a

^aDivision of Gastroenterology, Liyuan Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China

^bSchool of Mathematics, Shandong University, Jinan, China

^cDivision of Nephrology, Jiaozhou Hospital of Tongji University DongFang Hospital, Jiaozhou, China

^dDivision of Gastroenterology, Jiaozhou Hospital of Tongji University DongFang Hospital, Jiaozhou, China

YuShuang Xu and YiHua Wang contributed equally to this work. The order was determined by the corresponding author after negotiation.

ABSTRACT Cumulative studies have utilized high-throughput sequencing of the 16SrRNA gene to characterize the composition and structure of the microbiota in autism spectrum disorder (ASD). However, they do not always obtain consistent results; thus, conducting cross-study comparisons is necessary. This study sought to analyze the alteration of fecal microbiota and the diagnostic capabilities of gut microbiota biomarkers in individuals with ASD using the existing 16SrRNA microbial data and explore heterogeneity among studies. The raw sequence and metadata from 10 studies, including 1,019 samples, were reanalyzed. Results showed no significant difference in alpha diversity of fecal microbiota between ASD and the control group. However, a significant difference in the composition structure of fecal microbiota was observed. Given the large differences in sample selection and technical differences, the separation of fecal microbiota between ASD and controls was not observed. Subgroup analysis was performed on the basis of different country of origin, hypervariable regions, and sequencing platforms, and the dominant genera in ASD and healthy control groups were determined by linear discriminant analysis (LDA) of the effect size (LEfSe) algorithm and Wilcoxon rank-sum test. Machine learning analyses were carried out to determine the diagnostic capabilities of potential microbial biomarkers. A total of 12 genera were identified to distinguish ASD from control, and the AUC of the training set and verification set was 0.757 and 0.761, respectively. Despite cohort heterogeneity, gut microbial dysbiosis of ASD has been proven to be a widespread phenomenon. Therefore, fecal microbial markers are of great significance in diagnosing ASD diseases and possible candidates for further mechanistic study of the role of intestinal microbiota in ASD.

IMPORTANCE This study provides an updated analysis to characterize the gut microbiota in ASD using 16SrRNA gene high-throughput sequencing data from 10 publicly available studies. Our analysis suggests an association between the fecal microbiota and ASD. Sample selection and technical differences between studies may interfere with the species composition analysis of the ASD group and control group. By summarizing the results of 16SrRNA gene sequencing from multiple fecal samples, we can provide evidence to support the use of microbial biomarkers to diagnose the occurrence of ASD. Our study provides a new perspective for further revealing the correlation between gut microbiota and ASD from the perspective of 16SrRNA sequencing in larger samples.

KEYWORDS autism spectrum disorder, fecal microbiota, 16SrRNA, biomarker

Autism spectrum disorder (ASD) is a group of neurodevelopmental disorders characterized by deficits in social interactions, communication, and restricted and repetitive behavior (1, 2). Patients with ASD have a poor coping ability (3), low-quality of life (4), high rates of psychiatric comorbidity (5), and increased risk of suicidal behavior,

Editor Kristen M. DeAngelis, University of Massachusetts Amherst

Copyright © 2022 Xu et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to ZhiFan Xiong, xiongzhi@126.com, or BingQiang Liu, bingqiang@sdu.edu.cn.

The authors declare no conflict of interest.

Received 26 January 2022

Accepted 3 June 2022

Published 28 June 2022

which brings heavy psychological and economic burdens to families and society. The World Health Organization (WHO) reported that one in 160 children worldwide has ASD, and the prevalence rate shows an increasing trend (6, 7). At present, the clinical diagnosis of ASD primarily based on the information gathered from a detailed history, physical examination, and the observation of specific characteristic behaviors, thereby limiting the remaining diagnostic biomarkers (7). The uncertainty of the potential etiology and unique pathogenesis of ASD and the sporadic effectiveness of existing treatment methods promote the exploration of specific diagnostic biomarkers and effective treatment strategies.

Gut microbiota plays a role in guiding and promoting brain development, and it has a long-term effect on health (8, 9). Gut microbiota and hosts cooperate to regulate immunity, metabolism, and nervous system development and function through the dynamic two-way communication of the “gut–brain axis” (10). Germ-free mice transplanted with gut microbiota from human donors with ASD exhibit hallmark autistic behavior, and treatment of ASD mouse models with candidate microbial metabolites can improve behavioral abnormalities and regulate neuronal excitability in the brain (11). In the ASD susceptibility gene *Cntnap2*^{-/-} model for neurodevelopmental disorders, the researchers have found that different maladaptive behaviors are regulated by the interdependence of microbiota and host genes (12). These results indicate a potential link between the gut microbiota (or specific microbiota) and the brain in patients with ASD (10–12). The differences in species richness, diversity, composition, and structure of the gut microbiota between ASD patients and controls using 16S sequencing have been repeatedly reported, and they have received intense attention; however, a consensus among studies is rarely reported. For example, Zurita et al. found that *Bacteroides*, *Akkermansia*, *Coprococcus*, and different species of *Ruminococcus* increased in children with ASD relative to the controls (13). However, another study has found greater abundance of *Bacteroides* and *Prevotella* in ASD children and lower abundance of *Clostridium XIVa*, *Eisenbergiella*, *Escherichia/Shigella*, and *Akkermansia* (14). Disputes have also been considered to identify ASD-related microbial-based biomarkers. Large-scale multicenter studies using standardized methods may contribute to addressing these questions.

This study provided updated analysis to characterize the gut microbiota diversity in patients with ASD using 16SrRNA gene high-throughput sequencing data from 10 publicly available studies. Then, we evaluated the effect of study projects, country of origin, sequencing platform, and hypervariable region sequenced on the gut microbiota of patients with ASD. Finally, we identified different genera between the two groups using the LEfSe algorithm and Wilcoxon test and investigated whether these specific members of the community can be used as biomarkers to classify individuals as ASD or healthy controls accurately.

RESULTS

Grouping of ASD microbiota data sets. The flowchart of study selection is shown in Fig. 1. The systematic searches obtained 792 records from the PubMed database and 10 projects from the GMrepo database. After preliminary screening of titles and abstracts, 51 records and one project were identified for full-text review. In addition, the 16SrRNA gene sequencing data from 10 microbiota studies met the criteria for further analysis. The 10 data sets were labeled as S1 to S10 (13–22). This combined data set consisted of 1,019 participants (569 ASD and 450 healthy controls), and the study sizes varied from 12 to 286 subjects. Two patients with ASD and four control samples in the S9 project were excluded because their number of valid tags was less than 6,000. The average age of two of the 10 studies was unclear, but all participants were minors. The sequencing platforms of all studies are Illumina. The sequencing fragments include V3 to 4, V4, and V4 to 5, and the research countries include China, Ecuador, Italy, and Korean. Detailed information on the data sets regarding demographic

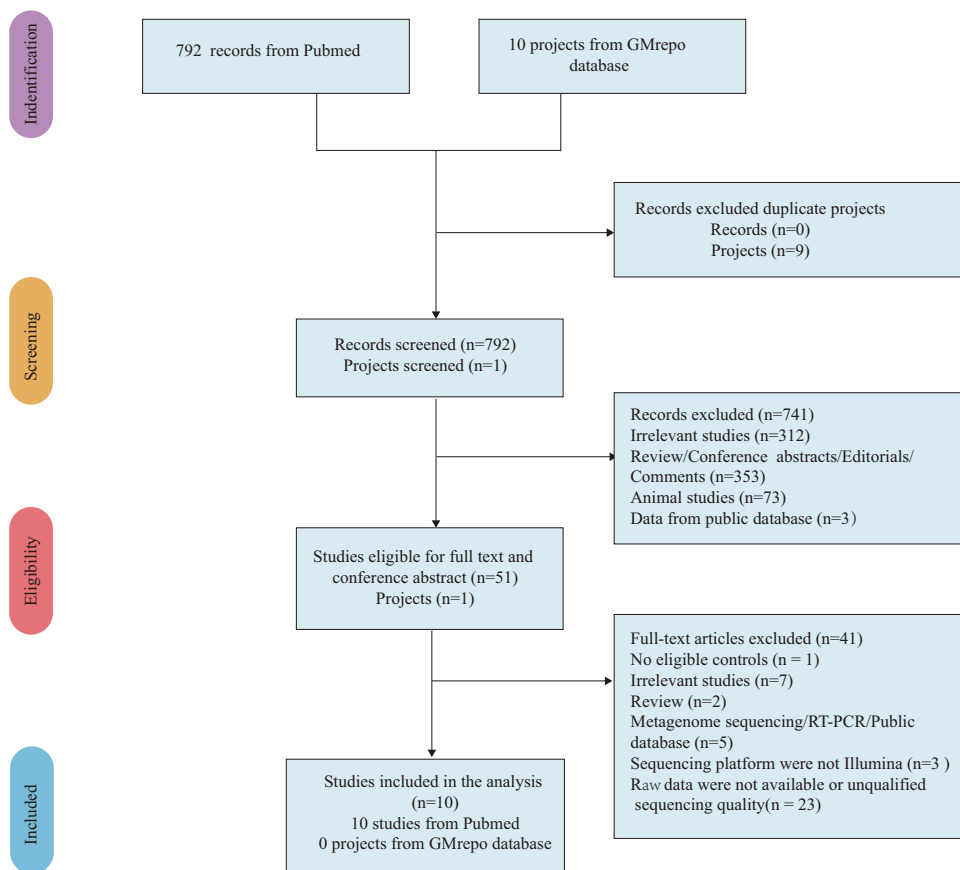


FIG 1 Flowchart of the study selection.

characteristics, diagnostic methods of ASD, PCR primer, hypervariable regions, and informatics technology of studies included is depicted in Table 1.

Alpha diversity analysis in patients with ASD and controls. First, we assessed whether variation in alpha diversity was associated with the occurrence of disease. Analysis of fecal samples showed that no statistical differences were observed using the Shannon diversity index ($P = 0.782$), observed operational taxonomic units (OTUs) ($P = 0.356$), and Pielou's evenness index ($P = 0.822$), indicating that no significant difference in richness and uniformity of the gut microbiota was observed between ASD and control groups. Next, subgroup analysis was performed in accordance with different studies. In two of the 10 studies, the alpha diversity of the ASD group was lower than that of the healthy control group (all $P < 0.05$). The random model was selected for meta-analysis, and the results showed that no significant difference in Shannon diversity index (SMD = -0.113 , 95% CI = -0.324 to 0.097 , $P = 0.290$; $I^2 = 56.5\%$), observed OTUs (SMD = -0.120 , 95% CI = -0.357 to 0.116 , $P = 0.318$; $I^2 = 65.4\%$), and Pielou's evenness index (SMD = -0.231 , 95% CI = -0.559 to 0.098 , $P = 0.169$; $I^2 = 82.9\%$) was observed between ASD and control groups. The results are in Fig. 2 and Table S1. Further subgroup analysis was conducted on the basis of the country of origin, variable regions, and sequencing platform. and found that no significant difference in species evenness and richness was observed between ASD and healthy control groups. The results are shown in Tables S2 to S4.

Cluster pattern of PCoA. Ordination analysis based on Bray-Curtis dissimilarity (PERMANOVA, pseudo-F = 4.440, $P = 0.001$) and unweighted uniFrac distances (PERMANOVA, pseudo-F = 4.735, $P = 0.001$) revealed substantial variation among ASD and control groups regarding microbial community composition. The difference among groups was more significant than within groups (ANOSIM, $R = 0.023$, $P = 0.001$). However, the PCoA revealed apparent overlap distinguishing distributions between the two groups, in which the center points were close (Fig. 3a). Considering the significant differences in

TABLE 1 Characteristics of included projects in the analysis

ID	Study yr	ASD ^a n/age (mean)	Control n/age (mean)	Diagnostic methods of ASD	PCR primer	Sequencing region	Country	Seq plat	Data availability (accession no.)
S1	Chiappori et al. (19) 2022	6/NA	6/NA	DSM-5	341F, 805R	V3 to 4	Italy	Illumina MiSeq	SRA: PRJNA813424
S2	Chen et al. (20) 2021	138/6.1	60/6.7	ADOS, CARS	NA	V3 to 4	China	Illumina MiSeq	SRA: PRJNA769228
S3	Ha et al. (21) 2021	54/7.0	38/6.0	DSM-5	341F, 805R	V3 to 4	Korean	Illumina MiSeq	ENA: PRJEB45948
S4	Huang et al. (22) 2021	39/4.7	44/5.1	DSM-5	515F, 926R	V4 to 5	China	Illumina MiSeq	SRA: PRJNA687773
S5	Zou et al. (14) 2020	48/5.0	48/4.0	DSM-4	338F, 806R	V3 to 4	China	Illumina MiSeq	SRA: PRJNA578223
S6	Dan et al. (23) 2020	143/4.9	143/5.2	DSM-5	515F, 806R	V4	China	Illumina HiSeq 2500	SRA: PRJNA453621
S7	Zurita et al. (13) 2020	25/8.9	31/8.5	ADI-R	NA	V4	Ecuador	Illumina MiSeq	SRA: PRJEB27306
S8	Zhao et al. (24) 2019	30/4.3	20/4.4	ADOS, ADI-R	338F, 806R	V3 to 4	China	Illumina MiSeq	SRA: PRJNA624252
S9	Ding et al. (25) (2020)	75/NA	46/NA	DSM-5	515F, 806R	V4	China	Illumina HiSeq 4000	SRA: PRJNA589343
S10	Coretti et al. (26) 2018	11/2.9	14/2.9	DSM-5, ADOS	NA	V3 to 4	Italy	Illumina MiSeq	SRA: PRJEB29421

^aASD, autism spectrum disorder; DSM-5, diagnostic and statistical manual of mental disorders, 5th Edition; DSM-4, diagnostic and statistical manual of mental disorders, 4th Edition; ADOS, Autism Diagnostic Observation Schedule; ADI-R, Autism diagnostics interview-revised (ADI-R) questionnaire; CARS, Childhood Autism Rating Scale; Seq Plat, sequencing platform; SRA, Sequence Read Archive; ENA, European Nucleotide Archive; V1, V3, V4, V5, variable regions of the 16S rRNA gene.

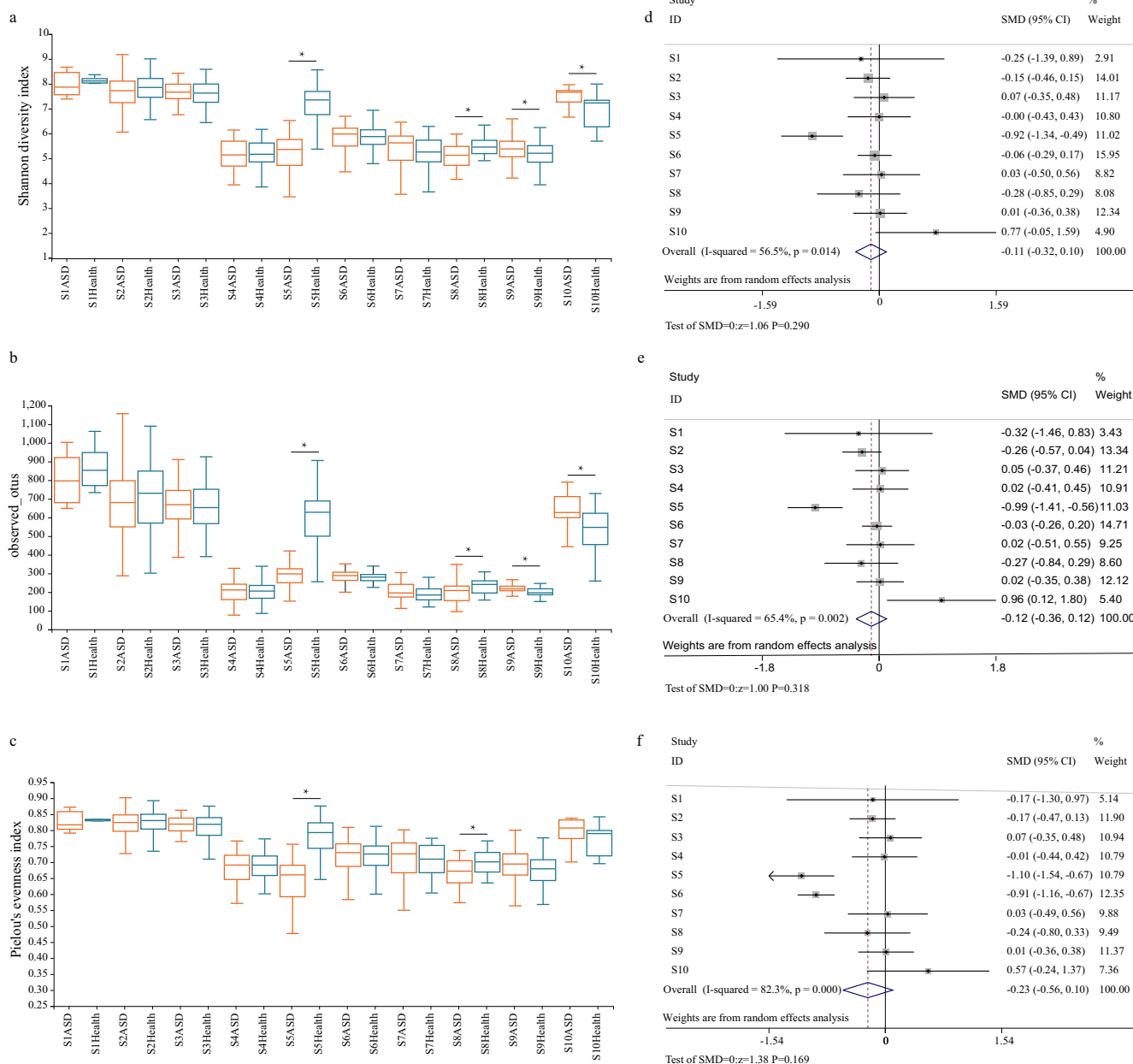


FIG 2 Alpha diversity comparison between ASD and control groups using the Shannon diversity index (a), observed OTUs (b), and Pielou's evenness index (c) in different studies. The meta-analysis results of the alpha diversity in ASD and control groups. Forest plot displaying SMD and 95% CIs for the Shannon diversity index (d), observed OTUs (e), and Pielou's evenness index (f). *, $P < 0.05$ was characterized as significant difference.

some parameters, such as country of origin, variable regions, and sequencing platform, the batch effect in each study is too large. Therefore, further subgroup analysis was performed. The difference between the two groups based on different countries was less significant than that within groups (ANOSIM, $R = -0.032$, $P = 0.997$). However, the difference between groups based on different studies (ANOSIM, $R = 0.426$, $P = 0.001$), different variable regions (ANOSIM, $R = 0.178$, $P = 0.001$), and different sequencing platforms (ANOSIM, $R = 0.214$, $P = 0.001$) was more significant than that within groups.

The microbiota composition across ASD and healthy controls in subgroups was compared using PERMANOVA based on Bray-Curtis dissimilarities and unweighted unFrac distance. Significant intergroup differences were observed in five out of 10 studies (S1, S5, S6, S9, and S10, all $P < 0.05$, Table S1). The gut microbiota composition between patients with ASD and healthy control sequenced using primers that targeted the V3 to 4 and V4 to 5

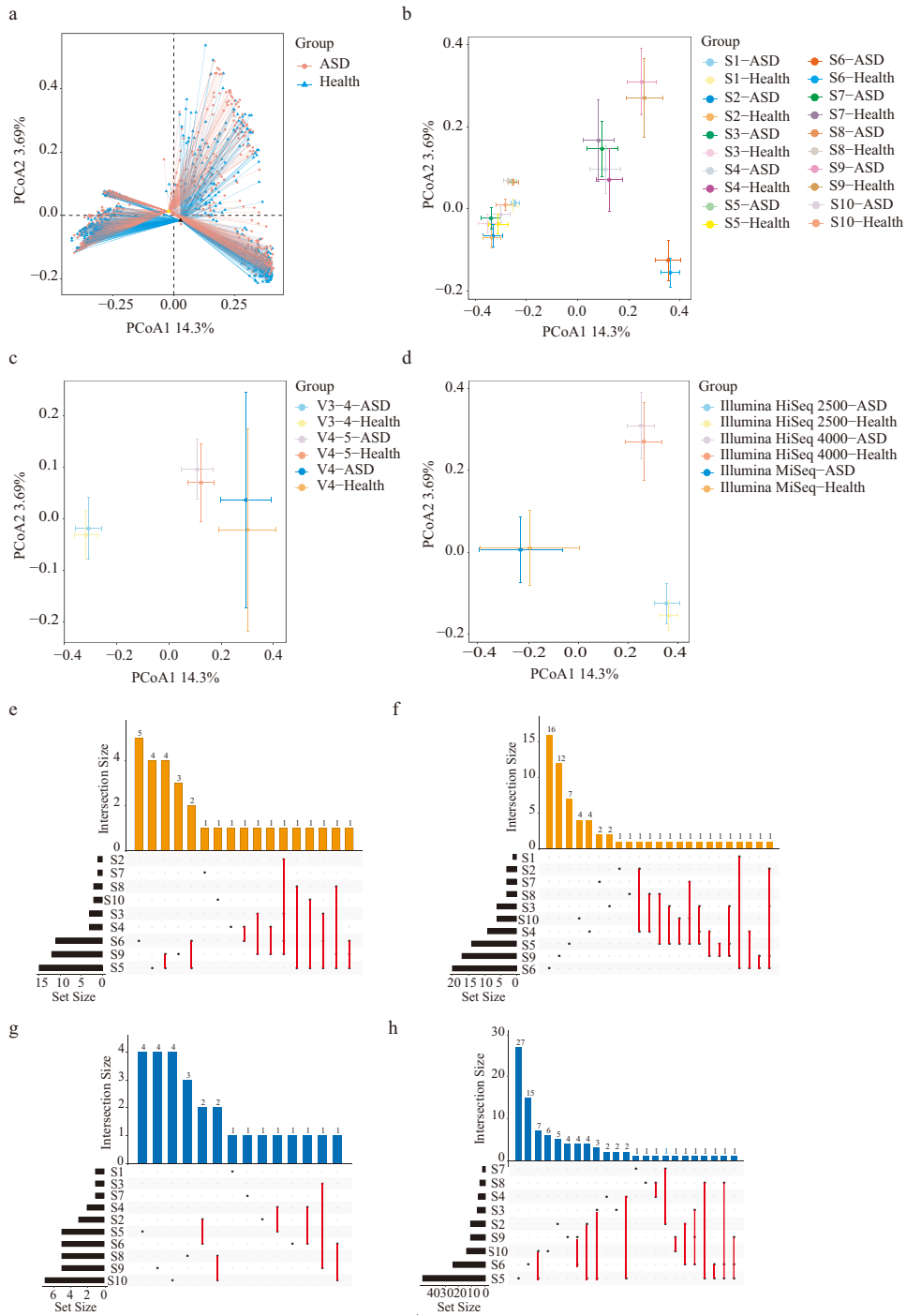


FIG 3 Beta-diversity comparison between ASD and healthy control using Bray–Curtis dissimilarity (a). The PCoA based on Bray–Curtis dissimilarity shows the distribution of the microbiota composition structure of ASD and healthy control group grouped by different studies (b), variable regions (c), and sequencing platform (d). The overlapping feature of dominant genera in the ASD group between studies S1 to 10 using LefSe algorithm (e) and Wilcoxon rank-sum test (f). The overlapping feature of dominant genera in the control group between studies S1 to 10 using LefSe algorithm (g) and Wilcoxon rank-sum test (h).

regions differed significantly (all $P < 0.05$, Table S2). In addition, the gut microbiota composition between patients with ASD and healthy control sequenced using sequencing platform Illumina MiSeq, Illumina HiSeq 2500, and Illumina HiSeq 4000 differ significantly (all $P < 0.05$, Table S3). The PCoA plots based on Bray–Curtis dissimilarities between the ASD group and control group classified by projects, variable regions, and sequencing platform are shown in Fig. 3b to d.

Identifying the fecal composite biomarker for ASD. We identified the dominant genera in the ASD and control groups using LEfSe and Wilcoxon rank-sum test and included the intersection of the results in the two groups. The results showed that no significant dominant genera were found in the ASD group, whereas the dominant genera in the control group included *Ruminococcus 2*, *Oscillibacter*, and *Veillonella*. Next, subgroup analysis was performed on the basis of different studies, variable regions, and sequencing platforms. The dominant genera in ASD and control groups in studies S1 to 10 using LEfSe and Wilcoxon rank-sum test were screened (Tables S8 to S11). The overlapping feature of the dominant genera of ASD and control groups in different studies is shown in Fig. 3e to h. The dominant genera of the ASD group included *Lachnospiraceae_incertae_sedis*, *Clostridium_XVIII*, *Eubacterium*, *Anaerostipes*, *Clostridium_sensu_stricto*, *Coprococcus*, *Dorea*, and *Faecalibacterium*. The dominant genera in the control group included *Gemmiger*, *Bacteroides*, *Roseburia*, *Dialister*, *Akkermansia*, *Haemophilus*, *Megamonas*, *Parabacteroides*, and *Streptococcus*. The results of dominant genera in ASD and control group subgrouped by three variable regions using LEfSe and Wilcoxon rank-sum test are summarized in Tables S12 to S15. The dominant genera of the ASD group included *Lachnospiraceae_incertae_sedis*, and *Eubacterium*. The dominant bacteria in the control group included *Dialister*, *Prevotella*, *Megamonas*, *Parabacteroides*, *Clostridium_XVIII*, and *Roseburia*. Then, subgroup analysis was performed on the basis of different sequencing platforms, and the dominant genera in ASD and control groups were screened (Tables S15 to S19). The dominant genera of the ASD group included *Eubacterium*, *Bifidobacterium*, *Blautia*, *Dialister*, *Coprococcus*, and *Lachnospiraceae_incertae_sedis*. The dominant bacteria in the control group included *Parabacteroides*, *Prevotella*, *Ruminococcus2*, *Romboutsia*, *Megamonas*, and *Clostridium_XIVa*.

Next, we used the dominant genera identified in different subgroups to build prediction models based on the random forest model. The top 12 important genera in the ASD and healthy control groups determined by subgroup analysis of different studies can distinguish the two groups, and the area under the receiver operating characteristic curve (AUC) of the training set and verification set was 0.688 and 0.706, respectively (Fig. 4a). The eight genera in the ASD and healthy control groups determined by subgroup analysis of different variable regions also can distinguish the two groups, and the AUC of the training set and verification set was 0.725 and 0.658, respectively (Fig. 4b). In contrast, the 12 genera in the ASD and healthy control groups determined by subgroup analysis of different sequencing platforms could distinguish ASD from control, and the AUC of the training set and verification set was 0.757 and 0.761, respectively (Fig. 4c).

DISCUSSION

In this study, we found significant differences in fecal microbiota composition between ASD and healthy controls. However, the batch effect between the two groups was too large. Study projects, sequencing platform, and hypervariable region sequenced were important interference factors for intergroup differences among fecal sample groups, and the influence of the country of origin was relatively small. Then, we selected 12 dominant genera in ASD and healthy control groups by LEfSe and Wilcoxon rank-sum test, and these genera were analyzed to identify and validate microbiota-based biomarkers that could be used to classify individuals as ASD or healthy controls. Random forest classification models constructed to differentiate individuals with ASD from healthy controls using fecal samples showed good performance.

Compared with a single study, the data set of multiple studies can comprehensively detect the changes in ASD-related microbiota composition by increasing the sample size and evaluating interference factors. In our results, PCoA showed that the ASD group was closer to the control group in the same research project, indicating that the factors such as sample selection and technical differences greatly affected microbiota analysis. Previous studies have shown that geographical and ethnic groups are crucial in forming specific microbial communities (23). Chen et al. (24) found that the amplification of different hypervariable regions of the bacterial 16S rRNA gene (V1 to V2, V3 to

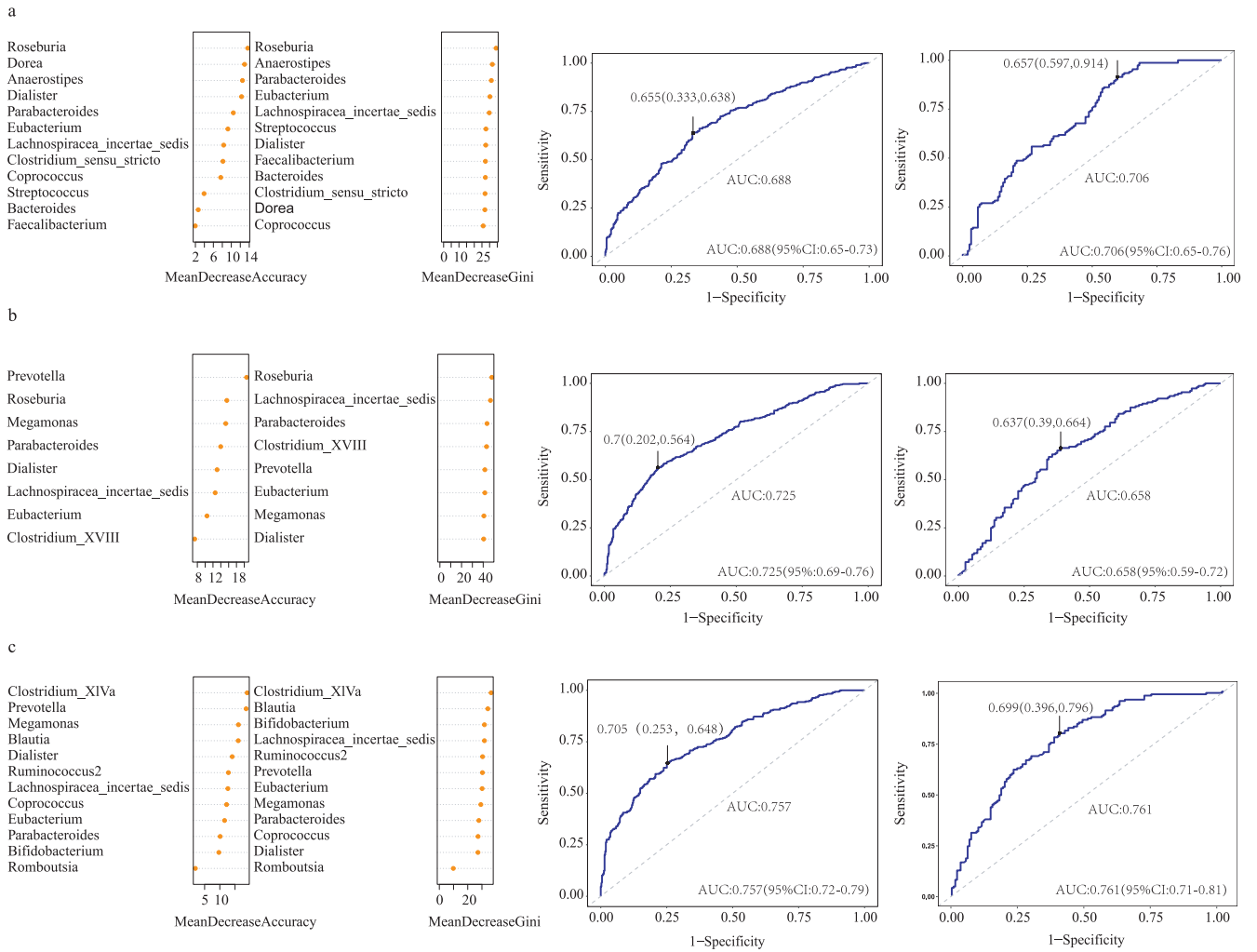


FIG 4 RF model was used to build a predictive model of genus-level abundant genera. The relative importance of each genus in the predictive model was evaluated using the mean decreasing accuracy and Gini coefficient. ROC curve generated using genera determined by subgroup analysis of different studies (a), variable regions (b) and sequencing platforms (c).

V4, and V4) affected the identification of the entire microbial community and some bacterial groups. This result is consistent with our results, this is, the difference between groups based on different variable region targeted was more significant than that within groups. Batch bias caused by experimental protocols such as sample collection, sequencing platform, and bioinformatics analysis is important. Therefore, a consistent approach must be used to minimize batch deviations and facilitate comparison among data sets.

In a clinical treatment study, fecal bacteria transplantation can reduce the gastrointestinal and behavioral symptoms of autistic patients and change the composition and structure of intestinal flora, which is characterized by a significant reduction in the relative abundance of *Eubacterium coprostanogenes* (25). The offspring of LPS induced maternal immune activation showed an abnormal brain-gut-microbiota axis, accompanied by social behavior defects, anxiety-like and repetitive behavior, and ASD-like microbiota characteristics. The abundance of *Coprococcus* was relevant to the anxiety-like and repetitive behavior (26). The relative abundance of *Lachnospiraceae incertae sedis* was significantly increased in children and adults (14, 27). In previous research, Sharon et al. (11) systematically analyzed the intestinal flora of ASD patients, ASD mice, and their offspring. They found significant differences among the gut microbiota of ASD patients, ASD mice, and the control group, and this difference can be vertically

transmitted to the offspring mice. *Verrucomicrobia*, *Proteobacteria phyla*, *Bacteroides ovatus*, and *Parabacteroides merdae* were significantly enriched in normal individuals/mice. Spearman correlation analysis showed that *Bacteroides* and *Parabacteroids* were positively correlated with the decrease of repetitive behavior and increased social behavior. The use of microbiota transfer therapy can significantly improve the relative abundance of *Bifidobacterium*, *Prevotella*, and *Desulfovibrio* in patient with ASD, and these changes persisted even after the cessation of treatment. Other studies have shown that the ASD group has lower levels of *Romboutsia* and *Megamonas* (27, 28). These results partially overlaps with our results. Based on previous physiological studies, the biomarkers that we identified, likely play a pathogenic or therapeutic role in the occurrence of ASD. However, several studies have concluded that the relative abundance of *Bifidobacterium*, *Blautia*, and *Dialist* in patients with ASD is significantly reduced, which is different from our results (29). This is a controversial issue that needs further study.

Notably, we have not consistently identified bacteria in 10 intestinal microecology studies, that is, bacteria with an intersection. This result may be due to the relatively small number of individuals involved in limited studies. Their sporadic distribution among individuals shows that the role of the gut microbiota in ASD may be overestimated, but many mechanisms can lead to ASD, or a variety of bacteria can mediate a single mechanism (such as inflammation). For example, children with ASD often exhibit an immune response disorder. Using the mouse model of maternal immune activation, Kim et al. (9) found that the increase of IL-17A in maternal inflammation cannot only increase the risk of developing neurodevelopmental disorders but also affect the CD4+T cells of the offspring through the changes of gut microbiota, resulting in an immune sensitization phenotype of the offspring, thereby increasing susceptibility to bacteria-induced gut inflammation. In addition, the interaction between the gut microbiota and mammalian nervous system forms both adaptive and dysfunctional neurological processes through the “gut-brain axis” pathway. Moreover, intestinal inflammation caused by gut microbiota imbalance can cause intestinal barrier dysfunction and induce systemic inflammation, which may lead to structural changes in the brain barrier and induce mental symptoms (30, 31). Gut microbiota and its metabolites can also affect the brain and behavior by acting on the vagus nerve and intestinal nervous system (32). However, sufficient experimental evidence to support this finding is lacking.

Strengths and limitations. This study can identify some studies and pool their results, with a relatively large number of participants. It is also an update and extension of previous similar studies. Nevertheless, this study has several limitations. First, the metadata information of existing research is incomplete. The composition of individual feces varies greatly, which is affected by many confounding factors, such as region, diet, exercise, gender, age, body mass index, and specimen collection method. Therefore, we hope to integrate other possible confounding factors in future research. Second, the projects included in this study involve different sequencing platforms and variable regions targeted. We tried to eliminate the batch effect by using the combat function (33) and removeBatchEffect limma function (34), but it was not successful. Large batch effects can still be observed among different projects. Therefore, the method of eliminating the batch effect must be developed and utilized more. Third, whether intestinal disorders precede the development of ASD or whether restrictive and repetitive behaviors such as diet in ASD drive this disorder remain unclear. The dynamic changes in the relative abundance of individual features in time sequence may be used to identify the driving pathogenic factors in the development of ASD. Large-scale human microbiota research must collect samples from multiple time points to maximize the detection of minor effects in microbial host interactions.

Conclusion. Our analysis suggests an association between the fecal microbiota and ASD. Sample selection and technical differences among studies may interfere with species composition analysis of the ASD group and control group. By summarizing the results of 16SrRNA gene sequencing from multiple fecal samples, we can provide evidence to support the use of microbial biomarkers to diagnose the occurrence of ASD. Our study provides a new perspective to reveal the correlation between gut microbiota

and ASD based on 16S rRNA sequencing in large samples. Nevertheless, future research is necessary to understand the exact contribution of changing intestinal microbiota to the clinical manifestations of ASD, which lead to the development of prevention and treatment methods.

MATERIALS AND METHODS

Study search, selection, and inclusion. We searched within the PubMed database. The search strategy and search terms were constructed as follows: (((Autism Spectrum Disorder [MeSH Terms]) OR (Autism Spectrum Disorder [Title/Abstract]) OR (Autistic Spectrum Disorder [Title/Abstract]) OR (Autism [Title/Abstract]) OR (Autistic Disorder [Title/Abstract])) AND (((Microbiota [MeSH Terms]) OR (Microbiota [Title/Abstract]) OR (Microbiotas [Title/Abstract]) OR (Microbial Community [Title/Abstract]) OR (Microbial Communities [Title/Abstract]) OR (Microbiome [Title/Abstract]) OR (Microbiomes [Title/Abstract])). Studies were limited to those published from inception to April 1, 2022. In addition, we restricted the search to papers written without language restrictions.

The GMrepo (data repository for Gut Microbiota) database is a curated database of human gut metagenomes, which can provide data resources with high reusability and accessibility (35). It includes 253 projects concerning 92 phenotypes. The GMrepo database adopts a two-step quality control process to ensure data quality. Amplicon sequencing samples/runs with <20,000 reads or only a single taxon were marked as “failed QC (QC status = 0).” We searched project data from the GMrepo database (<https://gmrepo.humangut.info>), and the search condition was set as follows: (i) the phenotype was “autism spectrum disorder”; (ii) healthy controls were mandatory; (iii) studies with fecal microbiota analyzed using 16S rRNA sequencing; and (iv) the percentage of failed runs was set to less than 90% to eliminate the projects whose percentage of failed runs exceeds 90%.

Data set collection. Raw sequence data and metadata were retrieved from the NCBI Short Read Archive (SRA), European Nucleotide Archive (ENA), GMrepo database, or article’s contents. Next, we excluded samples if the metadata were not available. Study information from the selected research included demographic information and methodologies, such as the recruiting area, sample size, mean age, diagnostic methods of ASD, PCR primer, sequencing platform, hypervariable region sequence, and whether or not the raw data available were recorded. The samples were grouped before downstream analyses.

Processing of 16S rRNA gene sequences. All downloaded raw sequencing reads of 16S rRNA gene sequences were processed using USEARCH (36). When the sequences provided by the study were paired-end sequence files, the `fastq_mergpairs` command was used to join the reads. `Fastq_minmergelen 250` command was used to filter reads less than 250 bp, and the resulting merged reads were filtered to exclude low-quality reads using a maximum false rate of 5% as the evaluation standard. Finally, 83.5% of the original reads were retained. When building zOTU, the shortest length of reads was set to obtain the absolute abundance of zOTU, and a data normalization technique was used, namely, total sum scaling, to obtain the relative abundance of zOTU for downstream analysis. Samples with less than 6,000 valid tags were deleted. Meanwhile, when calculating alpha and beta diversity, the sequencing depth was set to 6,000. At least eight exact duplicates sequences were screened as representative sequences. zOTU was obtained by denoising with `unoise3` algorithm. After identifying and removing chimeric sequences, these zOTUs were classified to the deepest taxonomic level that had 80% support using the Naive Bayesian classifier trained on the RDP taxonomy outline (version 14) (37). A total of 51,204,546 valid tags were obtained from 1,019 samples, with an average of 50,250 for each sample.

Microbial community profiling. The Shannon diversity index, observed OTUs, and Pielou’s evenness index were used to compare the differences in alpha diversity between ASD patients and healthy controls in different projects. Stata version 16.0 (College Station, TX) was used for meta-analysis. For continuous variables, comparisons between the two groups assessed the weighted standardized mean difference (SMD) and 95% confidence intervals (CIs), and Hedges’ *g* was selected as the final effect size. *P*-value < 0.05 was characterized as significant difference. Cochran’s *Q* test and *I*² statistics were used to assess the extent of heterogeneity among studies. A high degree of heterogeneity was expected; thus, pooled estimates was obtained through random-effect models. Beta diversity (between-sample) was assessed on the basis of Bray–Curtis dissimilarity and unweighted UniFrac distances and visualized by principal coordinate analysis (PCoA). Moreover, analysis of similarities (ANOSIM) was conducted on the basis of Bray–Curtis dissimilarity to test whether the difference between the two groups was significantly more significant than that within groups. We used the permutational multivariate analysis of variance (PERMANOVA) to assess microbiota differences between patients with ASD and healthy controls.

Next, the LEfSe method using the Kruskal–Wallis test was performed to identify the microbiota biomarkers. A linear discriminant analysis (LDA) score (\log_{10}) of 2.0 was used as the cutoff (38). The genera that were significantly different in relative abundance between the two groups were determined using the Wilcoxon rank-sum test. The results were considered significant at *P* values of <0.05. LEfSe and Wilcoxon rank sum test were used to obtain the dominant genera of ASD and control group in different subgroups, and the intersection was included. Then, the union of dominant genera of ASD and control group in different subgroups was investigated. After removing the dominant genera that repeatedly existed in ASD and control groups, the remaining genera were used as microbiota biomarkers for subsequent analysis. The random forest model was used to determine whether a composite microbial biomarker could discriminate ASD versus controls using the “RandomForest” package in R (ver. 4.0.3, R Foundation for Statistical Computing). The relative importance of each genus in the predictive model was evaluated using the mean decreasing accuracy and Gini coefficient. If more genera are screened, the top 12 genera of importance were retained for

subsequent analysis. Finally, subject operating characteristic (ROC) curve analysis was used to evaluate the clinical diagnostic ability of microbial biomarkers. Furthermore, 70% data and 30% data were randomly selected as the training set and verification set, respectively.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

SUPPLEMENTAL FILE 1, PDF file, 0.8 MB.

ACKNOWLEDGMENTS

All authors have made substantive contributions to this manuscript. Manuscript writing, sequencing analyses and management: Y.X., Y.W., J.X., Y.S.; project supervision and manuscript revision: Y.W., Z.X., and B.L.; study design, project management, financial support, and manuscript revision: B.L. and Z.X. None of the authors have financial disclosures to report. The authors have no conflicts of interest to declare.

Raw sequence data and metadata were retrieved from the NCBI SRA, ENA, and GMrepo databases. We thank the projects for the availability for data in this study. This study was supported by the National Key Research and Development Program of China (2018YFC2002000).

REFERENCES

- Ma B, Liang J, Dai M, Wang J, Luo J, Zhang Z, Jing J. 2019. Altered gut microbiota in Chinese children with autism spectrum disorders. *Front Cell Infect Microbiol* 9:40. <https://doi.org/10.3389/fcimb.2019.00040>.
- Strati F, Cavalieri D, Albanese D, De Felice C, Donati C, Hayek J, Jousson O, Leoncini S, Renzi D, Calabrò A, De Filippo C. 2017. New evidences on the altered gut microbiota in autism spectrum disorders. *Microbiome* 5:24. <https://doi.org/10.1186/s40168-017-0242-1>.
- Hirvikoski T, Blomqvist M. 2015. High self-perceived stress and poor coping in intellectually able adults with autism spectrum disorder. *Autism* 19: 752–757. <https://doi.org/10.1177/1362361314543530>.
- Jonsson U, Alaie I, Löfgren Wilteus A, Zander E, Marschik PB, Coghill D, Bölte S. 2017. Annual research review: quality of life and childhood mental and behavioural disorders - a critical review of the research. *J Child Psychol Psychiatry* 58:439–469. <https://doi.org/10.1111/jcpp.12645>.
- Russell AJ, Murphy CM, Wilson E, Gillan N, Brown C, Robertson DM, Craig MC, Deeley Q, Zinkstok J, Johnston K, McAlonan GM, Spain D, Murphy DG. 2016. The mental health of individuals referred for assessment of autism spectrum disorder in adulthood: a clinic report. *Autism* 20:623–627. <https://doi.org/10.1177/1362361315604271>.
- Paglia L. 2020. Children diagnosed with “ASD” are first of all . . . children. *Eur J Paediatr Dent* 21:8.
- Hirvikoski T, Boman M, Chen Q, D’Onofrio BM, Mittendorfer-Rutz E, Lichtenstein P, Bölte S, Larsson H. 2020. Individual risk and familial liability for suicide attempt and suicide in autism: a population-based study. *Psychol Med* 50:1463–1474. <https://doi.org/10.1017/S0033291719001405>.
- Mathee K, Cickovski T, Deoraj A, Stollstorff M, Narasimhan G. 2020. The gut microbiome and neuropsychiatric disorders: implications for attention deficit hyperactivity disorder (ADHD). *J Med Microbiol* 69:14–24. <https://doi.org/10.1099/jmm.0.001112>.
- Kim E, Paik D, Ramirez RN, Biggs DG, Park Y, Kwon H-K, Choi GB, Huh JR. 2022. Maternal gut bacteria drive intestinal inflammation in offspring with neurodevelopmental disorders by altering the chromatin landscape of CD4(+) T cells. *Immunity* 55:145–158.e7. <https://doi.org/10.1016/j.immuni.2021.11.005>.
- Morais LH, Schreiber HL, Mazmanian SK. 2021. The gut microbiota-brain axis in behaviour and brain disorders. *Nat Rev Microbiol* 19:241–255. <https://doi.org/10.1038/s41579-020-00460-0>.
- Sharon G, Cruz NJ, Kang D-W, Gandal MJ, Wang B, Kim Y-M, Zink EM, Casey CP, Taylor BC, Lane CJ, Brammer LM, Isern NG, Hoyt DW, Noecker C, Sweredski MJ, Moradian A, Borenstein E, Jansson JK, Knight R, Metz TO, Lois C, Geschwind DH, Krajmalnik-Brown R, Mazmanian SK. 2019. Human gut microbiota from autism spectrum disorder promote behavioral symptoms in mice. *Cell* 177:1600–1618.e17. <https://doi.org/10.1016/j.cell.2019.05.004>.
- Buffington SA, Dooling SW, Sgritta M, Noecker C, Murillo OD, Felice DF, Turnbaugh PJ, Costa-Mattioli M. 2021. Dissecting the contribution of host genetics and the microbiome in complex behaviors. *Cell* 184:1740–1756.e16. <https://doi.org/10.1016/j.cell.2021.02.009>.
- Zurita MF, Cárdenas PA, Sandoval ME, Peña MC, Fornasini M, Flores N, Monaco MH, Berding K, Donovan SM, Kuntz T, Gilbert JA, Baldeón ME. 2020. Analysis of gut microbiome, nutrition and immune status in autism spectrum disorder: a case-control study in Ecuador. *Gut Microbes* 11: 453–464. <https://doi.org/10.1080/19490976.2019.1662260>.
- Zou R, Xu F, Wang Y, Duan M, Guo M, Zhang Q, Zhao H, Zheng H. 2020. Changes in the gut microbiota of children with autism spectrum disorder. *Autism Res* 13:1614–1625. <https://doi.org/10.1002/aur.2358>.
- Wu S, Sun C, Li Y, Wang T, Jia L, Lai S, Yang Y, Luo P, Dai D, Yang YQ, Luo Q, Gao NL, Ning K, He LJ, Zhao X-M, Chen WH. 2020. GMrepo: a database of curated and consistently annotated human gut metagenomes. *Nucleic Acids Res* 48:D545–D553. <https://doi.org/10.1093/nar/gkz764>.
- Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R. 2011. UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* 27:2194–2200. <https://doi.org/10.1093/bioinformatics/btr381>.
- Wang Q, Garrity GM, Tiedje JM, Cole JR. 2007. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol* 73:5261–5267. <https://doi.org/10.1128/AEM.00062-07>.
- Jiang H, Ling Z, Zhang Y, Mao H, Ma Z, Yin Y, Wang W, Tang W, Tan Z, Shi J, Li L, Ruan B. 2015. Altered fecal microbiota composition in patients with major depressive disorder. *Brain Behav Immun* 48:186–194. <https://doi.org/10.1016/j.bbi.2015.03.016>.
- Chiappori F, Cupaioli FA, Consiglio A, Di Nanni N, Mosca E, Licciulli VF, Mezzelani A. 2022. Analysis of faecal microbiota and small ncRNAs in autism: detection of miRNAs and piRNAs with possible implications in host-gut microbiota cross-talk. *Nutrients* 14:1340. <https://doi.org/10.3390/nu14071340>.
- Chen Z, Shi K, Liu X, Dai Y, Liu Y, Zhang L, Du X, Zhu T, Yu J, Fang S, Li F. 2021. Gut microbial profile is associated with the severity of social impairment and IQ performance in children with autism spectrum disorder. *Front Psychiatry* 12:789864. <https://doi.org/10.3389/fpsy.2021.789864>.
- Ha S, Oh D, Lee S, Park J, Ahn J, Choi S, Cheon KA. 2021. Altered gut microbiota in Korean children with autism spectrum disorders. *Nutrients* 13: 3300. <https://doi.org/10.3390/nu13103300>.
- Huang M, Liu K, Wei Z, Feng Z, Chen J, Yang J, Zhong Q, Wan G, Kong XJ. 2021. Serum oxytocin level correlates with gut microbiome dysbiosis in children with autism spectrum disorder. *Front Neurosci* 15:721884. <https://doi.org/10.3389/fnins.2021.721884>.
- Dan Z, Mao X, Liu Q, Guo M, Zhuang Y, Liu Z, Chen K, Chen J, Xu R, Tang J, Qin L, Gu B, Liu K, Su C, Zhang F, Xia Y, Hu Z, Liu X. 2020. Altered gut microbial profile is associated with abnormal metabolism activity of autism spectrum disorder. *Gut Microbes* 11:1246–1267. <https://doi.org/10.1080/19490976.2020.1747329>.
- Zhao RH, Zheng PY, Liu SM, Tang YC, Li EY, Sun ZY, Jiang MM. 2019. Correlation between gut microbiota and behavior symptoms in children with autism spectrum disorder. *Zhongguo Dang Dai Er Ke Za Zhi* 21:663–669. <https://doi.org/10.7499/j.issn.1008-8830.2019.07.009>.

25. Ding X, Xu Y, Zhang X, Zhang L, Duan G, Song C, Li Z, Yang Y, Wang Y, Wang X, Zhu C. 2020. Gut microbiota changes in patients with autism spectrum disorders. *J Psychiatr Res* 129:149–159. <https://doi.org/10.1016/j.jpsychires.2020.06.032>.
26. Coretti L, Paparo L, Riccio MP, Amato F, Cuomo M, Natale A, Borrelli L, Corrado G, Comegna M, Buommino E, Castaldo G, Bravaccio C, Chiariotti L, Berni Canani R, Lembo F. 2018. Gut microbiota features in young children with autism spectrum disorders. *Front Microbiol* 9:3146. <https://doi.org/10.3389/fmicb.2018.03146>.
27. Mo Z, Huang P, Yang C, Xiao S, Zhang G, Ling F, Li L. 2020. Meta-analysis of 16S rRNA microbial data identified distinctive and predictive microbiota dysbiosis in colorectal carcinoma adjacent tissue. *mSystems* 5:e00138-20. <https://doi.org/10.1128/mSystems.00138-20>.
28. Chen Z, Hui PC, Hui M, Yeoh YK, Wong PY, Chan MCW, Wong MCS, Ng SC, Chan FKL, Chan PKS. 2019. Impact of preservation method and 16S rRNA hypervariable region on gut microbiota profiling. *mSystems* 4:e00271-18. <https://doi.org/10.1128/mSystems.00271-18>.
29. Li N, Chen H, Cheng Y, Xu F, Ruan G, Ying S, Tang W, Chen L, Chen M, Lv L, Ping Y, Chen D, Wei Y. 2021. Fecal microbiota transplantation relieves gastrointestinal and autism symptoms by improving the gut microbiota in an open-label study. *Front Cell Infect Microbiol* 11:759435. <https://doi.org/10.3389/fcimb.2021.759435>.
30. Lee GA, Lin YK, Lai JH, Lo YC, Yang YSH, Ye SY, Lee CJ, Wang CC, Chiang YH, Tseng SH. 2021. Maternal immune activation causes social behavior deficits and hypomyelination in male rat offspring with an autism-like microbiota profile. *Brain Sci* 11:1085. <https://doi.org/10.3390/brainsci11081085>.
31. Zhang Q, Zou R, Guo M, Duan M, Li Q, Zheng H. 2021. Comparison of gut microbiota between adults with autism spectrum disorder and obese adults. *PeerJ* 9:e10946. <https://doi.org/10.7717/peerj.10946>.
32. Deng W, Ke H, Wang S, Li Z, Li S, Lv P, Li F, Chen Y. 2022. Metformin alleviates autistic-like behaviors elicited by high-fat diet consumption and modulates the crosstalk between serotonin and gut microbiota in mice. *Behav Neurol* 2022:6711160. <https://doi.org/10.1155/2022/6711160>.
33. Liu F, Li J, Wu F, Zheng H, Peng Q, Zhou H. 2019. Altered composition and function of intestinal microbiota in autism spectrum disorders: a systematic review. *Transl Psychiatry* 9:43. <https://doi.org/10.1038/s41398-019-0389-6>.
34. Carloni S, Bertocchi A, Mancinelli S, Bellini M, Erreni M, Borreca A, Braga D, Giugliano S, Mozzarelli AM, Manganaro D, Fernandez Perez D, Colombo F, Di Sabatino A, Pasini D, Penna G, Matteoli M, Lodato S, Rescigno M. 2021. Identification of a choroid plexus vascular barrier closing during intestinal inflammation. *Science* 374:439–448. <https://doi.org/10.1126/science.abc6108>.
35. Xu Y, Liu X, Liu X, Chen D, Wang M, Jiang X, Xiong Z. 2021. The roles of the gut microbiota and chronic low-grade inflammation in older adults with frailty. *Front Cell Infect Microbiol* 11:675414. <https://doi.org/10.3389/fcimb.2021.675414>.
36. Wan Y, Zuo T, Xu Z, Zhang F, Zhan H, Chan D, Leung TF, Yeoh YK, Chan FKL, Chan R, Ng SC. 2022. Underdevelopment of the gut microbiota and bacteria species as non-invasive markers of prediction in children with autism spectrum disorder. *Gut* 71:910–918. <https://doi.org/10.1136/gutjnl-2020-324015>.
37. Johnson WE, Li C, Rabinovic A. 2007. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* 8:118–127. <https://doi.org/10.1093/biostatistics/kxj037>.
38. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. 2015. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 43:e47. <https://doi.org/10.1093/nar/gkv007>.