Biomedicine: 2024; 44(2):190-203 *April-June 2024*
Research article

Autism Biomarker Identification using an Integrative Systems Biology and Machine Learning Approach Highlighting TC.FEV.OM, acrA, and ABCB-BAC Genes in Gut Microbiome Analysis

Bhumika Narang[1], Pradhumn Sharma[1], Sudeepti Kulshrestha[1], Ronit Gandotra[1], Somnath S. Pai[2], Muskan Syed[1], Priyanka Narad[3], Abhishek Sengupta[1]

[1]Systems Biology and Data Analytics Research Lab, Centre for Computational Biology and Bioinformatics, Amity Institute of Biotechnology, Amity University Uttar Pradesh, Noida, India
[2]Amity Institute of Virology & Immunology, Amity University, Noida, India
[3]Previous Affiliation :Amity Institute of Biotechnology, Amity University Uttar Pradesh, Noida, India
Current Affiliation:Indian Council of Medical Research, Ansari Nagar, New Delhi, India

(Received : 24.04.2024 Revised : 03.06.2024 Accepted : 06.06.2024)

Corresponding author: **Abhishek Sengupta[1]** Email : asengupta@amity.edu

ABSTRACT

Introduction and Aim: Autism Spectrum Disorder (ASD) is a multifaceted neurodevelopmental disorder with complicated origins, and recent research points to a possible connection between dysbiosis of the gut microbiome and the pathophysiology of ASD.

Materials and Methods: In the present study, gut microbiome samples from public repositories (NCBI BioProject IDs: PRJNA815491 and PRJNA642975) were meta-analyzed using an integrated computational methodology. The gut microbiome 16S rRNA samples (n = 98) were subjected to taxonomic classification, functional profiling, statistical analysis as well as LEfSE and T-test analysis to find microbial biomarkers. Lastly, Machine Learning (ML) was employed to find the important features related to ASD.

Results: The results indicated nine significant features namely Sutterella, Prevotella, Blautia, Substance dependence pathway, Circulatory system pathway, Parasitic infectious disease, K02014 (TC.FEV.OM) gene, K03585 (acrA) gene, and K06147 (ABCB-BAC) gene. Moreover, complex relationships between microbial taxa, functional pathways, and genetic components were discovered by network analysis utilizing Cytoscape, which provided insight into possible microbial-host interactions and their relevance to the pathophysiology of ASD.

Conclusion: Overall, our research sheds light on potential microbial biomarkers, pathways, and genes dysregulated in ASD, as well as the gut microbiome and functional changes linked to the disorder. These findings suggest interesting directions for future research and therapeutic approaches targeting the gut-brain axis in the management of ASD. They also add to a fuller knowledge of the intricate interactions between the gut microbiome, host genetics, and ASD pathogenesis.

Keywords: Gut Microbiome, Autism, 16S rRNA analysis, Systems Biology, Machine Learning.

1. Introduction

Disorders pertaining to the growth and development of the brain or central nervous system (CNS) are known as neurodevelopmental disorders. A child diagnosed with autism spectrum disorder (ASD) has severe neurodevelopmental impairments that affect their capacity to connect and communicate with others (1). ASD frequently co-occurs with other clinical symptoms, such as gastrointestinal abnormalities (up to 70%), motor deficits (79%), sleep issues (50–80%), and intellectual incapacity (45%), in addition to these important conditions (2). In the past several years, the prevalence of autism has increased significantly over the world to 1 in 132 people, with boys experiencing a notably higher incidence of the disorder than girls (3-4). From 1 in 150 children in 2000 to 1 in 54 in 2016, the frequency of ASD has increased in the US (CDC, 2024). ASD children have a greater frequency of gastrointestinal (GI) problem comorbidity than neurotypical (NT) children, and there is evidence of a relationship between the gut microbiome and ASD (5-8). Abdominal pain, bloating, constipation, or loose stools are some of the symptoms of these GI issues (6). By making it easier to integrate multi-omics data, several research have shown how successful machine learning (ML) is in characterizing gut dysbiosis in ASD. Additionally, these methods have been used to

investigate the connection between gut microbiome and the intensity of ASD symptoms (9).

Despite significant attempts, the precise mechanisms behind ASD remain unclear. ASD presents a complex neurodevelopmental profile with diverse symptoms and biomarkers, underscoring the importance of understanding its underlying mechanisms for early diagnosis and effective treatment. Our goal was to devise a computational method for taxonomic and functional profiling and correlational analysis within the gut microbiome, emphasizing identifying patterns linked to ASD. We compared the taxonomic and functional profiles of the control and ASD samples to understand the differences in the microbiomes and their functional makeup and to investigate the gut microbiome settings under the two disease conditions. With these initiatives, we hope to understand better the complex interactions and mechanisms unique to dysbiotic conditions in the gut microbiome that are linked to a higher risk of autism in humans. This research may clarify how the gut microbiome contributes to autism pathophysiology, opening the door to more focused treatments and interventions.

2. MATERIALS AND METHODS

2.1. Data acquisition and retrieval

We examined two publicly available datasets employing 16S rRNA gut microbiome data from the NCBI Bioproject. The studies were based on the gut microbiomes of children aged 2 to 10 years old, stratified by disease condition (Control/Autistic), and were obtained for our work using NCBI BioProject IDs PRJNA815491 and PRJNA642975. A total of 98 raw samples were downloaded in Ubuntu 20.04.4 in the FASTQ file format and categorized by disease condition.

2.2. 16S rRNA data bioinformatics analysis

The *Parallel-Meta 3* software (version 3.3.2) was utilized to classify the pre-processed data taxonomically. The genome sequences were aligned to the reference database using Bowtie2, which produced an Operational Taxonomic Units (OTUs) table exhibiting relative abundances. Using *SILVA version 138* and *Greengenes version 13–8*, as reference databases, the "*PM-select-taxa*" command made it easier to create taxa feature tables. Then, log10(1 + x) normalization of absolute counts was applied for transformed abundances visualization in R using the *ggplot2* (version 4.1.2) and *ggpubr* packages. Additionally, using the *KEGG* database, the "*PM-select-func*" tool created functional characteristic tables for each sample, normalizing absolute counts and utilizing the *ggplot2* library in R to visualize pathway abundances.

2.3. Diversity analysis

The *vegan* library's "*diversity"* command, which is centered on Simpson's and Shanon's index, was used to calculate the Alpha-diversity for both taxonomic classification and functional profiling data in R. The *ggplot2* and ggpubr libraries were used to show the results. The beta-diversity calculation was performed using the Manhattan distance measure using a custom function based on the *vegan* and *ape* library, and visualized using the *ggplot2* packages in R.

2.4. Variation analysis

Using the *vegan* library's "*adonis2*" command in R, PERMANOVA analysis based on the Manhattan distance method was carried out for the findings of both taxonomic classification and functional profiling. The R *ggplot2* library was used to visualize the R2 values showing variations in the makeup of microbes and pathways.

2.5. Biomarker Identification using *LEfSE* analysis, T-test, and Systems Biology approaches

Using the MicrobiomeAnalyst tool, *LEfSE* analysis was performed based on results for taxonomic classification. To identify microbial biomarkers across different disease conditions, the minimum count threshold was assigned as 4, features with low counts and variations were eliminated based on sample prevalence set at 20% and coefficient of variation, respectively, and was normalized using the Total sum scaling (TSS) method. The FDR-adjusted p-value cut-off was changed to 0.05.

T-test was conducted using the *dplyr* and *rstatix* libraries in the R to check the statistical difference of each microbiological taxonomy or functional profile based on "group"(disease condition) and its significance. The effect size was computed using *Cohen's D* to standardize the difference between the two groups (control and autistic). The functional and microbiological biomarkers that were found, together with the associated P-values and effect sizes, were shown using the *ggplot2* and ggpubr libraries.

Furthermore, correlations among the microbes, pathways, and genes were calculated using the "*corr"* function in Python based on Spearman's correlational scores. Separate correlation files were generated for each pair (pathways-microbes, pathways-genes, microbes-genes). These correlation files were then imported into *Cytoscape version 3.10.2.* using the *MetScape* plugin. Three individual networks depicting correlations between pathways-microbes, pathways-genes, and microbes-genes were created. The "*Merge Network*" function in *Cytoscape* was then utilized to combine the paired networks into one integrated network. The network was then enriched by adding annotations. Further, we utilized the *CytoHubba* plugin to identify the top 25 nodes based on the "*betweenness*" measure of centrality (since it indicates the extent of a node's influence on the interactions of other nodes within the network) including the microbes, pathways, and genes contributing towards ASD.

2.6. Predictive Modeling using Machine Learning (ML) algorithms

Based on the *LEfSE* analysis and correlation analysis, two distinct datasets were compiled based on taxonomic classification data (Top microbes) and functional profiling data (top pathways and top genes) for developing a predictive model for ASD. Subsequently, missing values were addressed by removing them to uphold data integrity. Standardization of datasets was then carried out using the '*StandardScaler'* from the '*sklearn.preprocessing*' package to ensure uniform scaling across all variables and mitigate bias towards larger range variables. Skewness assessment and normalization were performed using the '*PowerTransformer'* with the Yeo-Johnson method. Various classification models, namely Random Forest, Gradient Boosting, AdaBoost, K-Nearest Neighbors, Support Vector Machine, Decision Tree, Logistic Regression, and Gaussian Naive Bayes, were employed to explore the relationship between features and ASD. These models, sourced from the '*sklearn.ensemble', 'sklearn.neighbors', 'sklearn.svm', 'sklearn.tree'*, and '*sklearn.linear_model*' packages in *scikit-learn* library, were evaluated using metrics such as accuracy, precision, F1 score, recall, and confusion matrix. Evaluation results identified the Gaussian Naive Bayes model as most suitable for the combined dataset of "Top pathways and Top genes", while logistic regression emerged as the preferred model for the "Top microbes" dataset. Further analysis involved the use of a random forest classifier to determine feature importance, selecting the top three features in each case. Visualization techniques including heatmaps and bar graphs, facilitated by the '*seaborn'* and '*matplotlib.pyplot*' packages, were employed to present the findings.

3. Results

3.1. Data retrieval

As shown in Table 1, samples from two publicly available studies were taken into consideration for this investigation. Two subcategories were created from the samples: Control and Autistic.

	Title	Publication
PRJNA815491	A preliminary investigation on therelationship between gut microbiome and gene expressions in peripheral mononuclear cells of infants with autism spectrum disorders	https://doi.org/10.1080/09168451.2016.1222267
PRJNA642975	Altered Gut Microbiota in Chinese Children with Autism Spectrum	https://doi.org/10.3389/fcimb.2019.00040

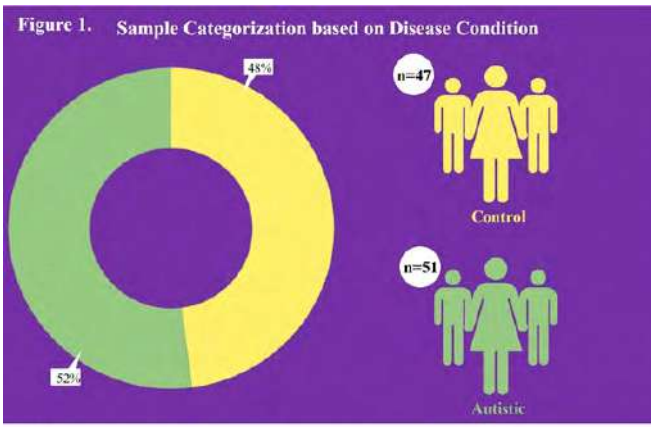

Fig.1 - Depicts the sample distribution for each of the subcategories.

3.2. 16S rRNA data bioinformatics analysis

The microbial, pathway, and gene abundances of the gut microbiome were revealed by taxonomic classification and functional profiling using Parallel-Meta 3 for both the control and autistic disorder conditions was obtained through. A total of 38 genera were found in our samples. When microbial abundances were compared between Control vs Autistic samples, an increase in microbial abundances including *Blautia*, *Bacteroides*, *Ruminococcus*, *Faecalibacterium*, *Lachnospiraceae*, *Prevotella*, *Subdoligranulum* and others, were found to be more abundant in autistic samples was noted in autistic circumstances, indicating instability in the gut milieu.

Similarly, a total of 48 pathways were identified in our samples. The most abundant pathways observed in both conditions - control and autistic – were genetic information processing, signaling and cellular processes, carbohydrate metabolism, amino acid metabolism, membrane transport, translation, replication and repair, aging, excretory system, circulatory system, and others. In comparison, the relative abundances of implicated pathways were higher in the autistic condition vs the control condition.

Furthermore, a total of 7695 genes were identified in our samples, among which we used the top 50 genes for further analysis. The most abundant genes observed in both conditions - control and autistic – were the *K03088* (*rpoE*) gene, *K02004* (*ABC.CD.P*) gene, *K01990* (*ABC-2.A*), and others. In comparison, the relative abundances of implicated genes were higher in the Control vs Autistic condition. The comparison of relative abundances for both taxonomic and functional profiling across Control vs Autistic disease conditions has been depicted in Figure 2 (a) and (b).

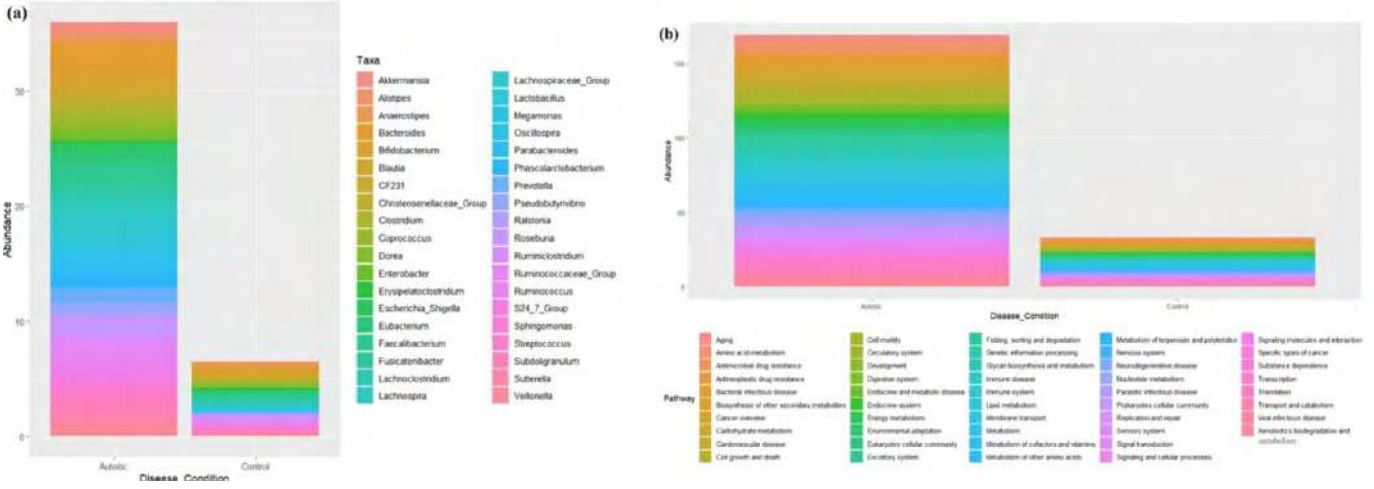

Figure 2. Bar plots representing normalized abundances of the different (a) microorganisms and (b) pathways identified in gut microbiome samples across control and autistic disease conditions.

3.3. Alpha diversity

Based on the taxonomic classification of the gut microbiome samples (Shannon and Simpson indices), Figure 3(a) shows the alpha diversity of those samples. According to the disease condition, the control condition had the lowest alpha diversity, and the autistic condition had the greatest variability. The alpha diversity of gut microbiome samples according to functional profiling (Shannon and Simpson indices) is shown in Figure 3(b) and 3(c) for pathways and genes respectively. According to the state of the disease, controls had the lowest alpha diversity and autistics the most.

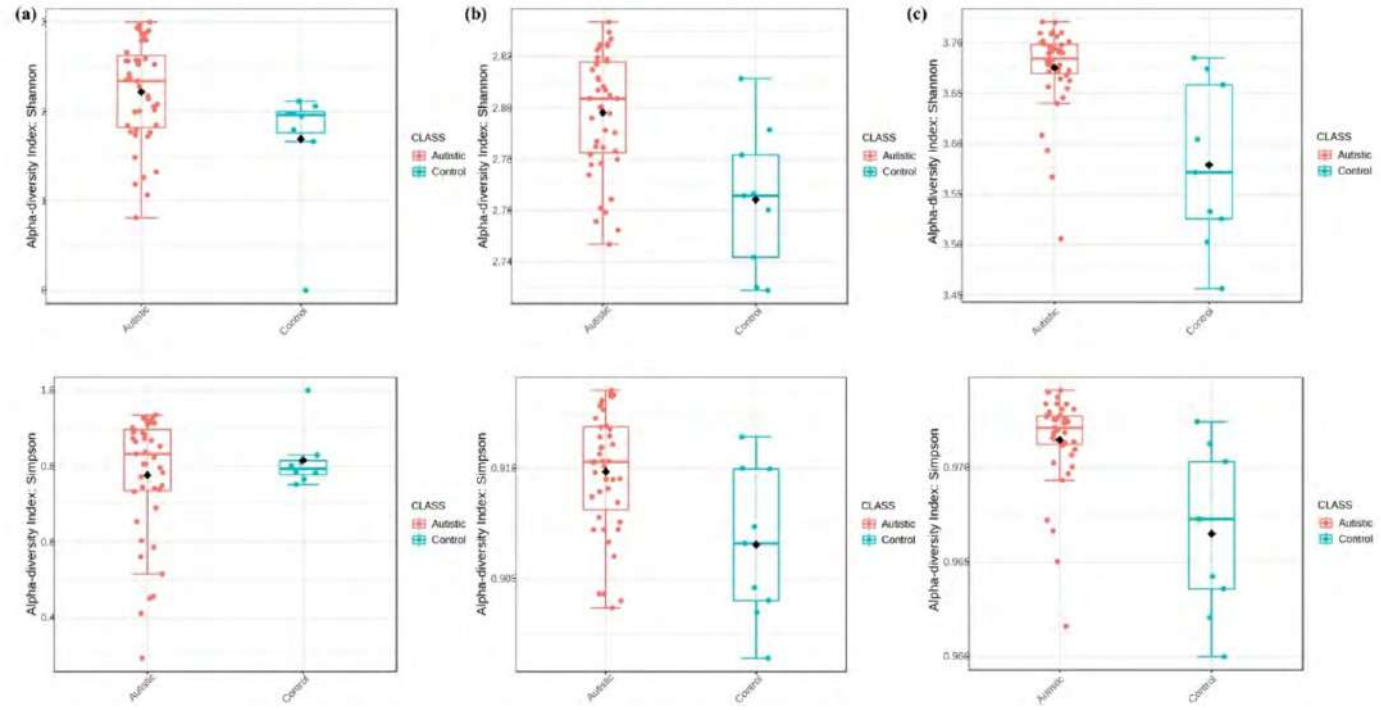

Figure 3. Alpha diversity of the gut microbiome samples using three independent metrics - (a) taxonomic classification profile, (b) expression pathway, and (c) functional profiling across different disease conditions

3.4. Beta diversity

The beta diversity analysis demonstrated by PCoA plots of gut sample taxonomic profiles (Figure 4(a)) showed considerable differences between control and autistic individual groupings. Likewise, considerable differences between control and autism disorder

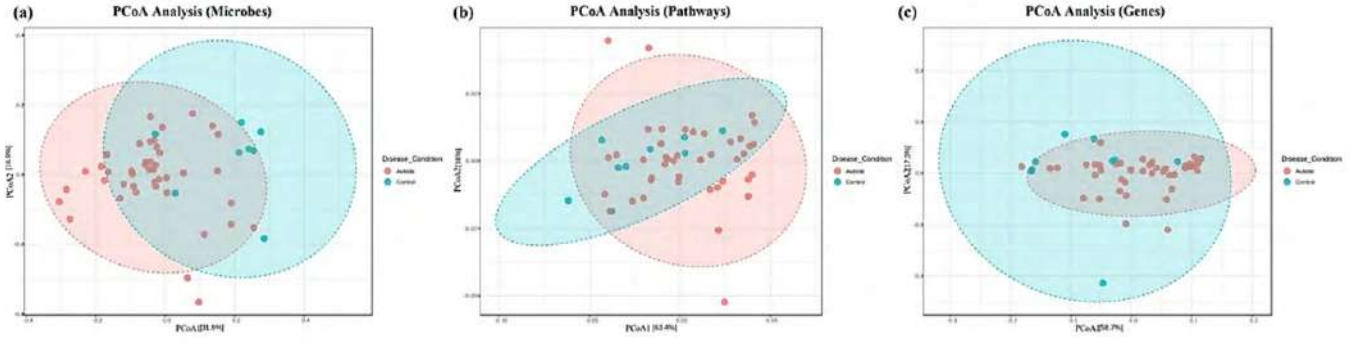

Figure 4. Beta-Diversity using (a) taxonomic classification of gut microbiome samples illustrating PCoA analysis for every microorganism and substantially differing microorganisms; and Beta-Diversity based on functional profiling of gut microbiome samples illustrating (b) PCoA analysis for every pathway and substantially differing pathways, and (c) PCoA analysis for every genes and substantially differing genes

substantial effects of disease condition on the functional profile of the samples. Figures 5 (b) and (c) show the impact of every pathway and gene respectively on the total gut microbiome of samples according to disorder state.

3.6. LEfSE and T-test analysis

Based on their LDA scores across various disease conditions, microbial, pathway, and gene biomarkers were identified by LEfSE analysis following data scaling and normalization. Five major genera, three major pathways, and seven major genes were found as biomarkers. Similar to this, the T-test identified significant biomarkers for various disease conditions based on their adjusted p-values and effect size values. Figure 6 shows the significant microbiological and functional biomarkers found using LEfSE and T-test analysis across control and autistic disease conditions.

3.7. Systems biology approach to construct network analysis

We analyzed interactions among microbes, pathways, and genes using Cytoscape 3.10.0 (Figure 7). Nodes represented microbes and pathways, while edges depicted their interactions. Color coding (red for microbes, green for pathways, and pink for genes) aided visualization. The node size reflected relative abundance, and the color gradient indicated interaction strength. Our findings revealed complex relationships, highlighting the interconnectedness and functional significance of biological components.

We integrated the above-mentioned co-occurrence abundance networks, resulting in a merged network (Figure 8) that provided a more comprehensive view of their interactions.

Subsequently, we subjected the merged network to analysis using CytoHubba to identify the top 25 nodes based on betweenness centrality. Interestingly, among these top nodes, only one represented a gene (K06147), while 18 nodes represented pathways and 6 represented

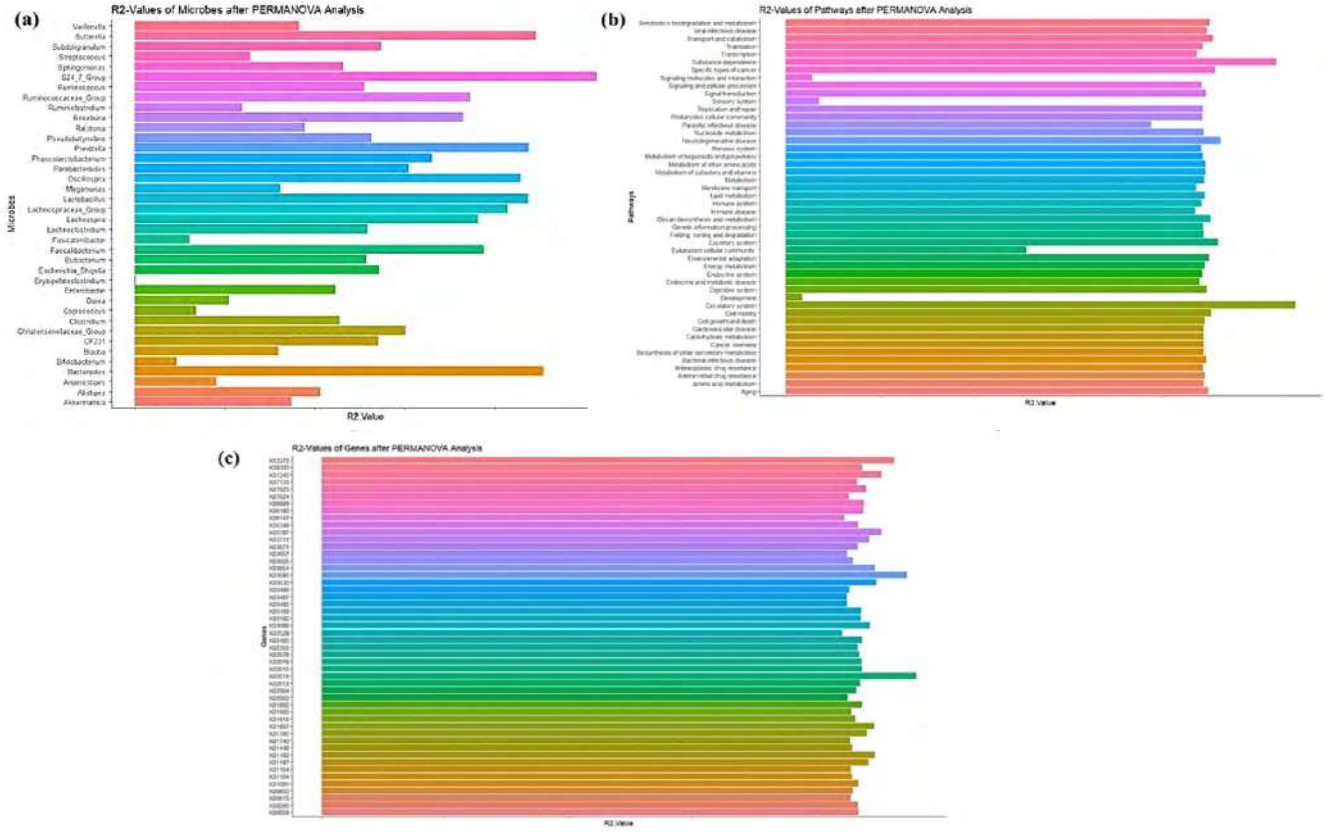

Figure 5. (a) Deviations identified of every microorganism comprising the gut microbiome samples based on their taxonomic classification across different disease conditions, (b) Deviations identified of every pathway comprising the gut microbiome samples; and (c) Deviations identified of every gene comprising the gut microbiome samples based on their functional profiling across different disease conditions

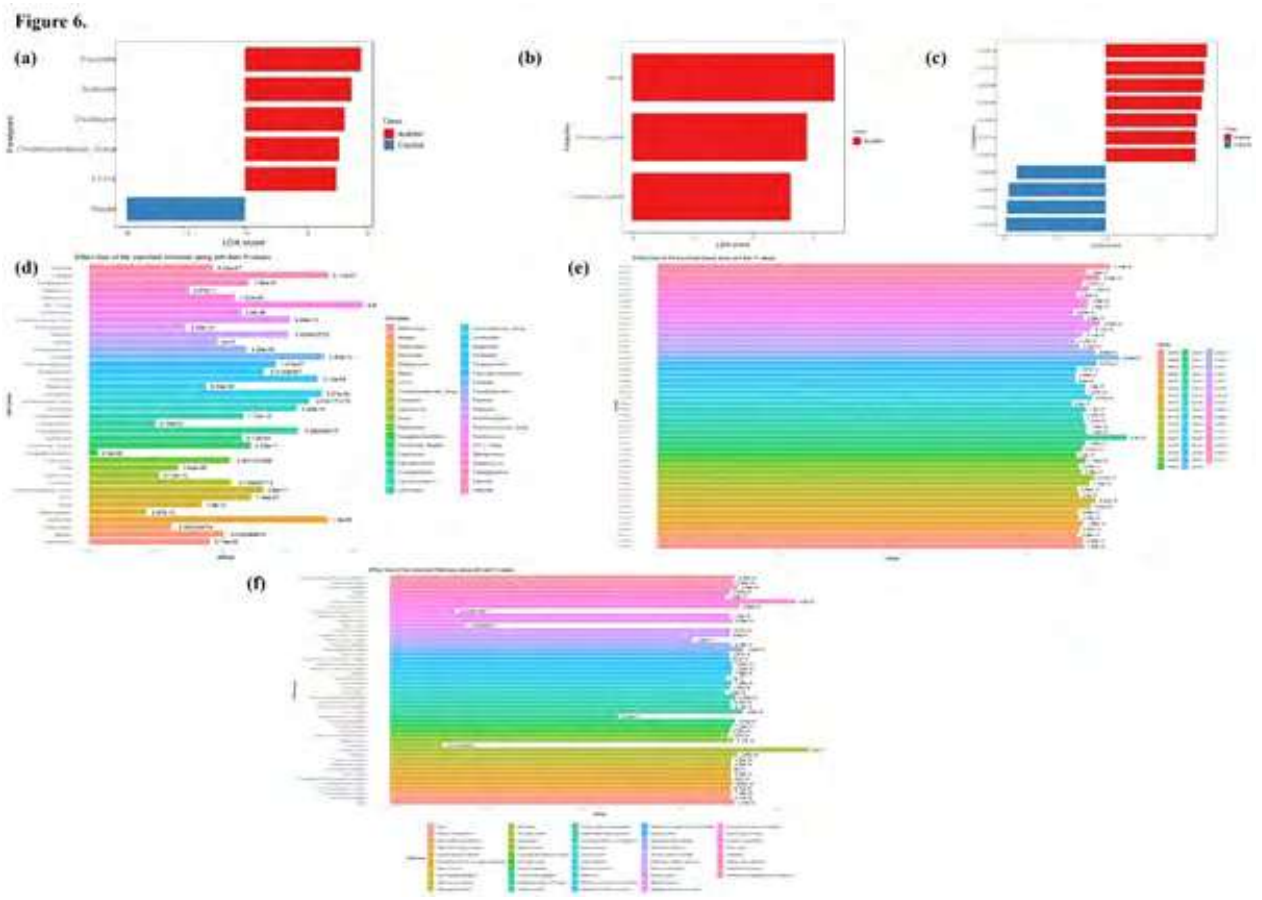

Figure 6. Important (a) microbial, (b) pathway, and (c) gene biomarkers discovered across different disease condition using LEfSE results based on LDA scores; and Important (d) microbial, (e) pathway, and (f) gene biomarkers discovered across different disease condition using T-test analysis based on p-values and effect size

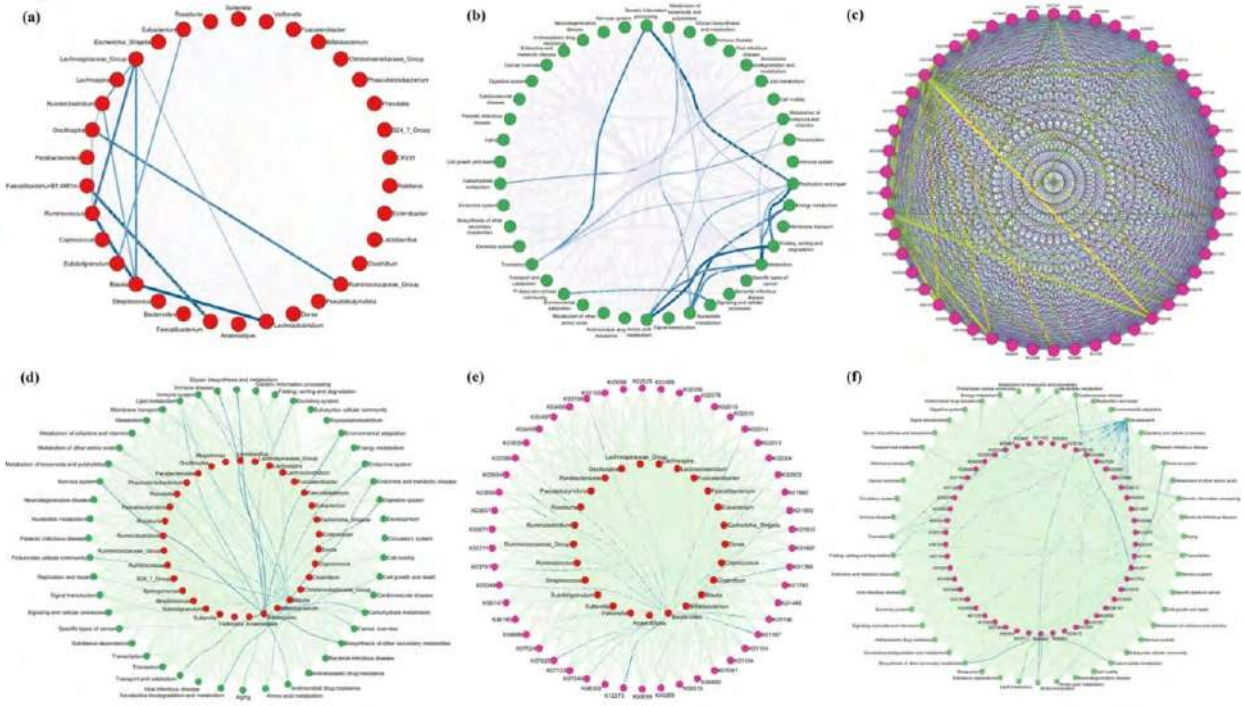

Figure 7. Co-occurrence abundance network representing (a) Microbe-Microbe interactions, (b) Pathway-Pathway interactions, (c) Genes-Genes interactions, (d) Microbe-Pathway interactions, (e) Microbes-Genes interactions, and (f) Pathways-Genes interactions

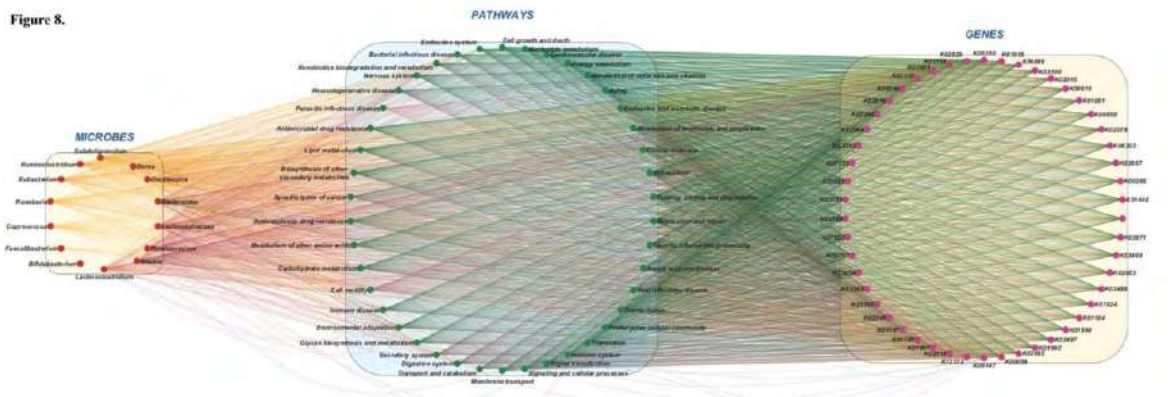

Figure 8. Merged network of co-occurrence abundance networks of microbes, pathways, and genes with reduced microbial diversity

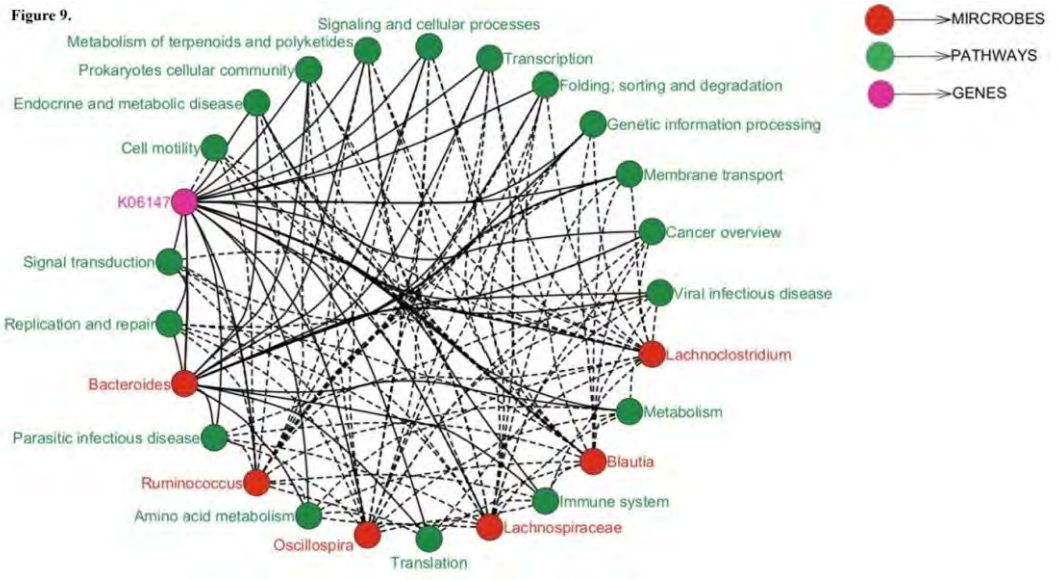

Figure 9. Topological analysis of merged biological network using CytoHubba demonstrating the top 25 features which include 6 microbes, 18 pathways, and 1 gene

microbes as can be visualized in Figure 9.

3.8. Important feature determination using Machine Learning (ML)

The combined dataset of top pathways and top genes yielded the greatest results with Gradient Boosting, Decision Tree, and Gaussian Naive Bayes achieving the highest accuracy of 0.91 each. The feature importance analysis of this dataset revealed that the top three influential pathways are the Substance dependence pathway, Circulatory system pathway, and Parasitic infectious disease pathway. Similarly, the top 3 gene features included the *K02014* (*TC.FEV.OM*) gene, the *K03585* (*acrA*) gene, and the *K06147* (*ABCB-BAC*) gene.

Similarly, the models trained using the dataset including the most significant microbes had favorable outcomes, with Gaussian Naive Bayes achieving the highest accuracy rate of 0.91. Feature importance for the top microbes dataset showed that the top 3 microbial features included *Sutterella*, *Prevotella*, and *Blautia*. The ROC curves, Correlation matrices, and the top important features identified have been depicted in Figure 10.

4. Discussion

In our investigation based on 16s rRNA analysis, 38 microbial taxa with variable abundance in autistic individuals were found; many of which have been linked to ASD pathogenicity in the past. Additionally, we

Figure 10.

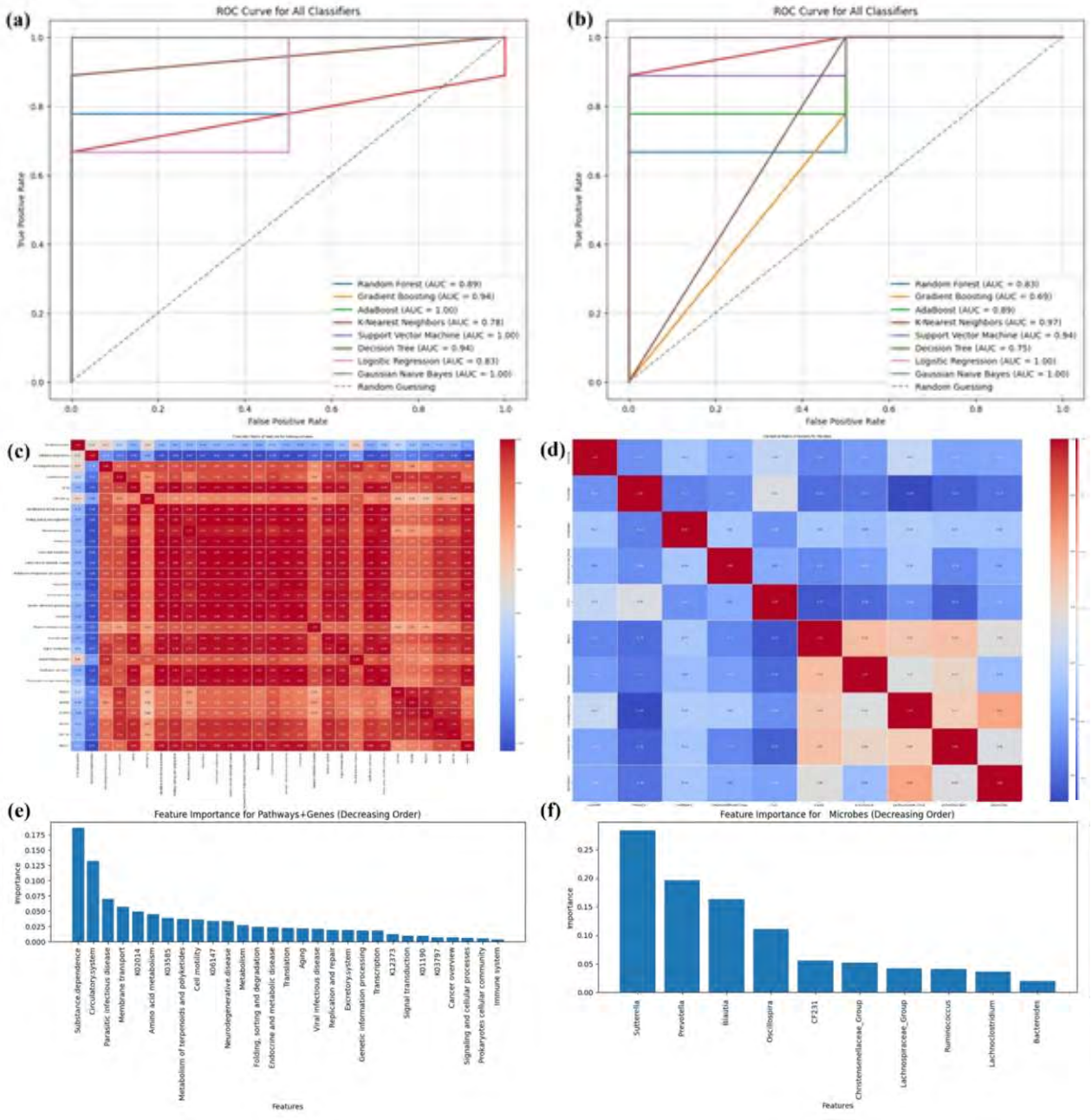

Figure 10. (a) ROC-curve for the model trained on functional profiling data (Pathway and genes combined) (b)ROC-curve for the model trained on taxonomic classification data (Microbes only) (c) Correlation matrix for top pathways and top genes, and (d) Correlation matrix for top microbial features; Feature importance for (e) pathways and genes, and (f) microbes depicting the importance of each in ASD in the decreasing order

identified 48 pathways and 7695 genes linked to ASD. Alpha diversity, assessed using the Shannon and Simpson indices, indicated higher diversity within the autistic group, demonstrating higher within-group diversity in the autistic condition compared to the control, with a higher number of outliers in the autistic group. Beta diversity of samples visualized using PCoA plots depicted data points in the form of 2D representation based on distance matrix wherein the data points more similar to each other tend to cluster together. The beta diversity analysis revealed significant differences between the two subcategories, with the control community showing more variation compared to autistic group. Additionally, the PERMANOVA

variance analysis revealed that the *S24_7 group* of microbes had the highest variance, followed by *Sutterella* and *Prevotella*. Similarly, among pathways, the circulatory system exhibited the highest variance, followed by substance dependence pathways. Regarding genes, the K02014 and K03585 genes showed the highest variance. Research has revealed that critical microbes such as those belonging to the *S24_7 family*, can cause significant differences between the gut microbial profiles between ASD and Control cases by associating with chemokine disorders (serum levels of MCP-3, MIP-1α, eotaxin, and RANTES) (10). These changes might impact the gut-brain axis, which could affect behavior and neurodevelopment. Moreover, the *S24-7 group* may affect metabolite production that interacts with the immunological and neurological systems of the host, causing systemic inflammation and gastrointestinal symptoms associated with autism spectrum disorders (11-12). Additionally, ASD patients exhibit elevated levels of *Sutterella*, which showed high variance across the two groups in our study, consistent with previous reports (13), potentially inducing pro-inflammatory responses and intestinal permeability, leading to systemic inflammation and neuroinflammation (14). This overgrowth may affect immunological responses, neurodevelopment, and neurotransmitter synthesis, influencing the microbiome-gut-brain axis and contributing to the pathophysiology of the disorder (15). Interestingly, *Prevotella* dysbiosis in ASD has shown conflicting findings, with some studies reporting decreased abundances (16) and others, including our study, reporting increased levels (17). *Prevotella*'s involvement in intestinal fermentation and carbohydrate metabolism suggests its potential role in ASD-related neurobehavioral phenotypes, affecting metabolic processes, nutrition intake, and microbial metabolite synthesis (18). *Prevotella* overgrowth may also contribute to immunological dysregulation, systemic inflammation, changes in immune function (19), and gastrointestinal disorders associated with ASD (18).

Additionally, the dysfunction of the Circulatory system pathway (which showed highest variance) observed in ASD may lead to cerebral hypoperfusion and oxygenation deficits, affecting brain regions crucial for language processing and social cognition (22). Similarly, the substance dependence pathway involves neurotransmitter systems such as glutamate, serotonin, and dopamine, which play crucial roles in mood, behavior, and cognition - all significant features of ASD (20). Environmental factors like maternal substance use during pregnancy have been associated with increased ASD risk, possibly influencing fetal neurodevelopment or epigenetic pathways (21). Additionally, comorbid mental health conditions like anxiety and attention-deficit hyperactivity disorder (ADHD) often co-occur with ASD, potentially sharing neurobiological mechanisms leading to substance abuse.

Furthermore, the *K02014* (*TC.FEV.OM*) gene (highest variance), encoding an outer membrane receptor protein for iron complexes, may play a role in iron uptake, transport, and utilization. Although the precise relationship between this gene and ASD is unknown, dysregulated iron homeostasis may affect neurodevelopmental processes and add to the pathogenesis of the disorder (23). Similarly, the *K03585* (*acrA*) gene encodes for the membrane fusion protein in a multidrug efflux system. Compounds such as immune mediators, neurotransmitters, and signaling molecules are regulated by multidrug efflux systems as they traverse cell membranes. Multidrug efflux system dysfunction may impact immunological responses, which in turn may contribute to the inflammatory processes seen in ASD pathology, as well as neurotransmitter homeostasis, which has been linked to ASD (24). Additionally, the blood-brain barrier's (BBB) integrity is preserved in part by multidrug efflux mechanisms. The integrity of the blood-brain barrier may be jeopardized by impaired multidrug efflux system activity, which could result in increased permeability and possibly neuroinflammation linked to ASD (24).

Lastly, Machine Learning (ML) was employed to

find the important features related to ASD. The results indicated nine significant features namely *Sutterella*, *Prevotella*, *Blautia*, Substance dependence pathway, Circulatory system pathway, Parasitic infectious disease, *K02014* (*TC.FEV.OM)* gene, *K03585* (*acrA)* gene, and *K06147* (*ABCB-BAC)* gene.

Interestingly, GM of children with ASD was found to include lower levels of *Blautia* in certain studies (25). Contrasting to this, our findings showed an increased abundance of *Blautia* in ASD individuals compared to control. It is known that *Blautia* ferments dietary fiber to create short-chain fatty acids (SCFAs) (26). SCFAs – particularly butyrate – have been linked to immune system regulation, gut barrier integrity maintenance, and altered neural development and function. Modulation of SCFA synthesis by *Blautia* may be a factor in immunological dysregulation, gastrointestinal dysregulation, and neurodevelopmental problems linked to ASD (26). Similar to *Prevotella*, *Blautia* is also involved in carbohydrate metabolism and energy production in the gut, affiliating it with ASD (18). Parasitic infections, especially during pregnancy or early childhood have been implicated in immune-mediated pathways disrupting neurodevelopment, possibly exacerbating ASD pathology (27). Furthermore, the *K06147* (*ABCB-BAC*) gene encodes for ATP-binding cassette (ABC) transporter, and functions similarly to multidrug efflux systems, except that it needs ATP (adenosine triphosphate) as energy to carry out the transport. The nervous system has ABC transporters, which are involved in the movement of chemicals necessary for the growth and operation of neurons. ASD may be exacerbated by disruptions in these mechanisms (28). The relation of *Sutterella*, *Prevotella*, Substance dependence pathway, Circulatory system pathway, *K02014* (*TC.FEV.OM)* gene and *K03585* (*acrA)* gene with autism has already been discussed.

With this, it can be clearly demonstrated that the features namely – *Sutterella*, *Prevotella*, Substance dependence pathway, Circulatory system pathway, *K02014* (*TC.FEV.OM)* gene and *K03585* (*acrA)* gene have the highest statistical proof with the combined approach of PERMANOVA analysis, network analysis and predictive modeling analysis.

5. Conclusion

Our research contributes to the expanding corpus of research linking the gut microbiome to the pathophysiology of ASD and offers insightful information about possible biomarkers and treatment targets for the condition. Our results can potentially guide the development of tailored therapies targeting the gut microbiome in the management of ASD by clarifying the intricate relationship between microbial dysbiosis, metabolic pathways, and ASD pathogenesis.

Acknowledgment

We would like to acknowledge Dr. Ashok K. Chauhan, Founder, and President, of Amity University, Uttar Pradesh for providing us with the opportunity to conduct research. We would also like to thank the Centre for Computational Biology and Bioinformatics, Amity Institute of Biotechnology, Amity University for providing us with the necessary resources.

Funding Information

No funding was received for this work.

Conflict of Interest

The authors declare no conflict of interest.

Declaration of Generative AI and AI-assisted Technologies in the Writing Process

During the preparation of this work, the authors used Quillbot to improve the readability and understandability of the article. After using this tool/service, the authors reviewed and edited the content as needed and took full responsibility for the content of the publication.

Contributions

PN, SK, AS, and SSP carried out Conceptualization, and BN, PS, SK, RG, and MS carried out Data curation, Investigation, Methodology, Visualization, and Writing (original draft). PN, SK, SSP, and AS carried out the

formal analysis, Supervision, Validation, and Writing (Review and Editing). All authors read and approved the final manuscript.

Figure Legends

Figure 1. Gut microbiome sample distribution based on disease condition into Control (n = 47) and Autistic (n = 51) samples respectively.

Figure 2. Bar plots representing normalized abundances of the different **(a)** microorganisms and (b) pathways identified in gut microbiome samples across control and autistic disease conditions.

Figure 3. Alpha diversity of the gut microbiome samples using three independent metrics - (a) taxonomic classification profile, (b) expression pathway, and (c) functional profiling across different disease conditions

Figure 4. Beta-Diversity using (a) taxonomic classification of gut microbiome samples illustrating PCoA analysis for every microorganism and substantially differing microorganisms; and Beta-Diversity based on functional profiling of gut microbiome samples illustrating (b) PCoA analysis for every pathway and substantially differing pathways, and (c) PCoA analysis for every genes and substantially differing genes

Figure 5. (a) Deviations identified of every microorganism comprising the gut microbiome samples based on their taxonomic classification across different disease conditions, (b) Deviations identified of every pathway comprising the gut microbiome samples; and (c) Deviations identified of every gene comprising the gut microbiome samples based on their functional profiling across different disease conditions

Figure 6. Important (a) microbial, (b) pathway, and (c) gene biomarkers discovered across different disease conditions using LEfSE results based on LDA scores; and Important (d) microbial, (e) pathway, and (f) gene biomarkers discovered across different disease condition using T-test analysis based on p-values and effect size

Figure 7. Co-occurrence abundance network representing (a) Microbe-Microbe interactions, (b) Pathway-Pathway interactions, (c) Genes-Genes interactions, (d) Microbe-Pathway interactions, (e) Microbes-Genes interactions, and (f) Pathways-Genes interactions

Figure 8. Merged network of co-occurrence abundance networks of microbes, pathways, and genes with reduced microbial diversity

Figure 9. Topological analysis of merged biological network using CytoHubba demonstrating the top 25 features which include 6 microbes, 18 pathways, and 1 gene

Figure 10. (a) ROC-curve for the model trained on functional profiling data (Pathway and genes combined) (b)ROC-curve for the model trained on taxonomic classification data (Microbes only) (c) Correlation matrix for top pathways and top genes, and (d) Correlation matrix for top microbial features; Feature importance for (e) pathways and genes, and (f) microbes depicting the importance of each in ASD in the decreasing order

Table Legends

Table 1. Details of the data utilized in our study.

References

1. Vuong HE, Hsiao EY. Emerging Roles for the Gut Microbiome in Autism Spectrum Disorder. Biological Psychiatry 2017;81:411–23. https://doi.org/10.1016/j.biopsych.2016.08.024.

2. Rylaarsdam L, Guemez-Gamboa A. Genetic Causes and Modifiers of Autism Spectrum Disorder. Front Cell Neurosci 2019;13:385. https://doi.org/10.3389/fncel.2019.00385.

3. Hossain MD, Ahmed HU, Jalal Uddin MM, Chowdhury WA, Iqbal MS, Kabir RI, et al. Autism Spectrum disorders (ASD) in South Asia: a systematic review. BMC Psychiatry 2017;17:281. https://doi.org/10.1186/s12888-017-1440-x.

4. Lasheras I, Seral P, Latorre E, Barroso E, Gracia-García P, Santabárbara J. Microbiota and gut-brain axis dysfunction in autism spectrum disorder: Evidence for functional gastrointestinal disorders. Asian Journal of

Psychiatry 2020;47:101874.
https://doi.org/10.1016/j.ajp.2019.101874.

5. Wang LW, Tancredi DJ, Thomas DW. The Prevalence of Gastrointestinal Problems in Children Across the United States With Autism Spectrum Disorders From Families With Multiple Affected Members. Journal of Developmental & Behavioral Pediatrics 2011;32:351–60.
https://doi.org/10.1097/DBP.0b013e31821bd06a.

6. Coury DL, Ashwood P, Fasano A, Fuchs G, Geraghty M, Kaul A, et al. Gastrointestinal Conditions in Children With Autism Spectrum Disorder: Developing a Research Agenda. Pediatrics 2012;130:S160–8. https://doi.org/10.1542/peds.2012-0900N.

7. Chaidez V, Hansen RL, Hertz-Picciotto I. Gastrointestinal Problems in Children with Autism, Developmental Delays or Typical Development. J Autism Dev Disord 2014;44:1117–27.
https://doi.org/10.1007/s10803-013-1973-x.

8. Kheirouri S, Kalejahi P, Noorazar SG. Plasma levels of serotonin, gastrointestinal symptoms,and sleep problems in children with autism. Turk J Med Sci 2016;46:1765–72. https://doi.org/10.3906/sag-1507-68.

9. Li H, Liu C, Huang S, Wang X, Cao M, Gu T, et al. Multi-omics analyses demonstrate the modulating role of gut microbiota on the associations of unbalanced dietary intake with gastrointestinal symptoms in children with autism spectrum disorder. Gut Microbes 2023;15:2281350.
https://doi.org/10.1080/19490976.2023.2281350.

10. Liu J, Gao Z, Liu C, Liu T, Gao J, Cai Y, et al. Alteration of Gut Microbiota: New Strategy for Treating Autism Spectrum Disorder. Front Cell Dev Biol 2022;10:792490.
https://doi.org/10.3389/fcell.2022.792490.

11. Alamoudi MU, Hosie S, Shindler AE, Wood JL, Franks AE, Hill-Yardin EL. Comparing the Gut Microbiome in Autism and Preclinical Models: A Systematic Review. Front Cell Infect Microbiol 2022;12:905841.
https://doi.org/10.3389/fcimb.2022.905841.

12. Liu F, Horton-Sparks K, Hull V, Li RW, Martínez-Cerdeño V. The valproic acid rat model of autism presents with gut bacterial dysbiosis similar to that in human autism. Molecular Autism 2018;9:61. https://doi.org/10.1186/s13229-018-0251-3.

13. Wang L, Christophersen CT, Sorich MJ, Gerber JP, Angley MT, Conlon MA. Increased abundance of Sutterella spp. and Ruminococcus torques in feces of children with autism spectrum disorder. Molecular Autism 2013;4:42. https://doi.org/10.1186/2040-2392-4-42.

14. Hsiao EY, McBride SW, Hsien S, Sharon G, Hyde ER, McCue T, et al. Microbiota Modulate Behavioral and Physiological Abnormalities Associated with Neurodevelopmental Disorders. Cell 2013;155:1451–63.
https://doi.org/10.1016/j.cell.2013.11.024.

15. Mayer EA, Knight R, Mazmanian SK, Cryan JF, Tillisch K. Gut Microbes and the Brain: Paradigm Shift in Neuroscience. J Neurosci 2014;34:15490–6. https://doi.org/10.1523/JNEUROSCI.3299-14.2014.

16. Kang D-W, Park JG, Ilhan ZE, Wallstrom G, LaBaer J, Adams JB, et al. Reduced Incidence of Prevotella and Other Fermenters in Intestinal Microflora of Autistic Children. PLoS ONE 2013;8:e68322. https://doi.org/10.1371/journal.pone.0068322.

17. Strati F, Cavalieri D, Albanese D, De Felice C, Donati C, Hayek J, et al. New evidences on the altered gut microbiota in autism spectrum disorders. Microbiome 2017;5:24. https://doi.org/101186/s40168-017-0242-1.

18. Kang D-W, Ilhan ZE, Isern NG, Hoyt DW, Howsmon DP, Shaffer M, et al. Differences in fecal microbial metabolites and microbiota of children with autism spectrum disorders. Anaerobe 2018;49:121–31. https://doi.org/10.1016/j.anaerobe.2017.12.007.

19. Finegold SM, Molitoris D, Song Y, Liu C, Vaisanen M, Bolte E, et al. Gastrointestinal Microflora Studies in

Late-Onset Autism. CLIN INFECT DIS 2002;35:S6–16. https://doi.org/10.1086/341914.

20. Rubenstein E, Wiggins LD, Lee L-C. A Review of the Differences in Developmental, Psychiatric, and Medical Endophenotypes Between Males and Females with Autism Spectrum Disorder. J Dev Phys Disabil 2015;27:119–39. https://doi.org/10.1007/s10882-014-9397-x.

21. Roberts AL, Koenen KC, Lyall K, Robinson EB, Weisskopf MG. Association of autistic traits in adulthood with childhood abuse, interpersonal victimization, and posttraumatic stress. Child Abuse & Neglect 2015;45:135–42.
https://doi.org/10.1016/j.chiabu.2015.04.010.

22. Krakowiak P, Goines PE, Tancredi DJ, Ashwood P, Hansen RL, Hertz-Picciotto I, et al. Neonatal Cytokine Profiles Associated With Autism Spectrum Disorder. Biological Psychiatry 2017;81:442–51.
https://doi.org/10.1016/j.biopsych.2015.08.007.

23. Adams JB, Audhya T, McDonough-Means S, Rubin RA, Quig D, Geis E, et al. Nutritional and metabolic status of children with autism vs. neurotypical children, and the association with autism severity. Nutr Metab (Lond) 2011;8:34. https://doi.org/10.1186/1743-7075-8-34.

24. Frazier TW, Hardan AY. A Meta-Analysis of the Corpus Callosum in Autism. Biological Psychiatry 2009;66:935–41.
https://doi.org/10.1016/j.biopsych.2009.07.022.

25. Finegold SM, Dowd SE, Gontcharova V, Liu C, Henley KE, Wolcott RD, et al. Pyrosequencing study of fecal microflora of autistic and control children. Anaerobe 2010;16:444–53.
https://doi.org/10.1016/j.anaerobe.2010.06.008.

26. Sharon G, Cruz NJ, Kang D-W, Gandal MJ, Wang B, Kim Y-M, et al. Human Gut Microbiota from Autism Spectrum Disorder Promote Behavioral Symptoms in Mice. Cell 2019;177:1600-1618.e17.
https://doi.org/10.1016/j.cell.2019.05.004.

27. Estes ML, McAllister AK. Immune mediators in the brain and peripheral tissues in autism spectrum disorder. Nat Rev Neurosci 2015;16:469–86. https://doi.org/10.1038/nrn3978.

28. Jha NK, Kar R, Niranjan R. ABC Transporters in Neurological Disorders: An Important Gateway for Botanical Compounds Mediated Neuro-Therapeutics. CTMC 2019;19:795–811.
https://doi.org/10.2174/1568026619666190412121811.