frontiers in Cellular Neuroscience

ORIGINAL RESEARCH
published: 30 April 2019
doi: 10.3389/fncel.2019.00150

Urine Organic Acids as Potential Biomarkers for Autism-Spectrum Disorder in Chinese Children

Qiao Chen[1†], You Qiao[1†], Xin-jie Xu[2], Xin You[1,3,4] and Ying Tao[5*]*

[1] Department of Rheumatology and Clinical Immunology, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing, China, [2] Central Research Laboratory, Department of Scientific Research, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China, [3] Key Laboratory of Rheumatology and Clinical Immunology, Ministry of Education, Beijing, China, [4] Autism Special Fund, Peking Union Medical Foundation, Beijing, China, [5] Institute of Artificial Intelligence, Ping An Technology (Shenzhen) Ltd., Beijing, China

OPEN ACCESS

Edited by:
Junyu Xu,
Zhejiang University, China

Reviewed by:
Zoltan Molnar,
University of Oxford, United Kingdom
Jing Zou,
Mayo Clinic, United States
Sylvie Mavel,
Université de Tours, France

***Correspondence:**
Xin You
youxin@pumch.cn
Ying Tao
taoying7001@139.com

[†] These authors have contributed equally to this work

Specialty section:
This article was submitted to Cellular Neuropathology, a section of the journal Frontiers in Cellular Neuroscience

Received: 10 December 2018
Accepted: 08 April 2019
Published: 30 April 2019

Citation:
Chen Q, Qiao Y, Xu X-j, You X and Tao Y (2019) Urine Organic Acids as Potential Biomarkers for Autism-Spectrum Disorder in Chinese Children. Front. Cell. Neurosci. 13:150. doi: 10.3389/fncel.2019.00150

Autism spectrum disorder (ASD) is a neurodevelopmental disorder that lacks clear biological biomarkers. Existing diagnostic methods focus on behavioral and performance characteristics, which complicates the diagnosis of patients younger than 3 years-old. The purpose of this study is to characterize metabolic features of ASD that could be used to identify potential biomarkers for diagnosis and exploration of ASD etiology. We used gas chromatography-mass spectrometry (GC/MS) to evaluate major metabolic fluctuations in 76 organic acids present in urine from 156 children with ASD and from 64 non-autistic children. Three algorithms, Partial Least Squares-Discriminant Analysis (PLS-DA), Support Vector Machine (SVM), and eXtreme Gradient Boosting (XGBoost), were used to develop models to distinguish ASD from typically developing (TD) children and to detect potential biomarkers. In an independent testing set, full model of XGBoost with all 76 acids achieved an AUR of 0.94, while reduced model with top 20 acids discovered by voting from these three algorithms achieved 0.93 and represent a good collection of potential ASD biomarkers. In summary, urine organic acids detection with GC/MS combined with XGBoost algorithm could represent a novel and accurate strategy for diagnosis of autism and the discovered potential biomarkers could be valuable for future research on the pathogenesis of autism and possible interventions.

Keywords: autism spectrum disorder, biomarker, urine organic acids, Chinese, metabolomics, diagnosis

INTRODUCTION

Autism spectrum disorder (ASD) is a developmental disorder characterized by impaired communication and social behavior, as well as displays of restricted and repetitive behavior (Keller and Persico, 2003). Although the pathogenesis of autism is uncertain, it is considered to involve an interaction between multiple genetic and environmental risk factors that are present in the few first years of life (Nair, 2000).

The diagnostic criteria for ASD require that symptoms become apparent in early childhood, typically before age three (Dieme et al., 2015). Autism diagnosis currently relies on scales and

professional surveyors using behavioral methods. For instance, the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-4) is the 1994 version of the American Psychiatric Association (APA) that provides a classification and diagnostic tool for ASD. Early identification and early intervention of autistic children are recognized as two of the most crucial factors for improving outcomes for individuals affected by ASD (Dawson et al., 2010, 2012; Warren et al., 2011; Zwaigenbaum et al., 2013; Klin et al., 2015). However, due to the challenges of early ASD diagnosis many older children miss the best intervention period.

Metabolic abnormalities associated with ASD include: phenylketonuria (PKU), disorders in purine metabolism, folate deficiency in brain development, succinic semialdehyde dehydrogenase deficiency, Smith-Lemli-Opitz syndrome (SLOS), organic acidurias (e.g., pyridoxine dependency, 3-methylcrotonyl-CoA carboxylase deficiency, and propionic acidemia), and mitochondrial disorders (Manzi et al., 2008; Zecavati and Spence, 2009; Ghaziuddin and Alowain, 2013). The presence of psychiatric, behavioral, and developmental regression together with metabolic disorders in autism (Wanders et al., 1999; Wang et al., 1999; Cox et al., 2001; Kompare and Rizzo, 2008) requires studies concerning the relationship between these pathological states and whether metabolic products of amino acid and lipid synthesis in urine or blood could be autism biomarkers (Schain and Freedman, 1961; Hanley et al., 1977; Bull et al., 2003; Kałużna-Czaplińska, 2011). After glomerular filtration and tubular condensation, the macromolecular proteins in the blood can be filtered and the urine becomes a concentrated organic acid. The natural physiological role of the kidney makes urine the best specimen for analyzing organic acid metabolism.

Several previous studies focused on organic acid biomarkers in autistic patients. Emond et al. (2013) found that levels of citrate, succinate, and glycolate were significantly increased in the urine sample of ASD children, whereas Mavel et al. (2013) found that β-alanine, glycine, taurine, and succinic acid levels were increased in the urine sample. Another study indicated that around 10 metabolites significantly differed between an autism group and the control group (Kałużna-Czaplińska, 2011). Some organic acids were highlighted by multiple studies, while others were seen only in a specific study. In general, microbial metabolites, niacin metabolism, mitochondria-related metabolites, and amino acid metabolites are the most common perturbations in autistic children. These results illustrates the complexity of metabolic disorders and etiology in autistic patients, leading to the exploration of building models for multivariate analysis. A metabolomics study of urine in 22 ASD children and 24 controls built an orthogonal partial least-squares discriminant analysis (OPLS-DA) model (AUROC = 0.91) (Hu, 2003), another one based on 14 ASD and 10 controls obtained a Principal Component Analysis (PCA) model (AUROC = 0.775) and identified a set of organic acids as potential biomarkers (Kałużna-Czaplińska, 2011). These studies do have some limitations, such as different races, limited regions, single algorithm used, and limited sample size. Similar researches with large sample size on Chinese children have rarely been reported. Moreover, some recently developed machine learning algorithms, such as XGBoost, have shown better performance over traditional algorithms on many tasks beyond biomedical domain. Therefore, we launched this representative study of a Chinese population with a larger sample size and a few more recent algorithms.

The aims of this study were to identify metabolic signatures of ASD and to find potential biomarkers for autism diagnosis and possible etiology. We used gas chromatography-mass spectrometry (GC/MS) to assess major metabolic perturbations in organic acid levels in urine from children with autism versus non-autistic subjects. Considering the complexity of ASD, the rise or fall of different organic acids is insufficient. A model using classification analyses of collected data for multiple organic acids that exhibit significant differences between healthy and autistic individuals should be feasible and may allow autistic patients to be distinguished.

MATERIALS AND METHODS

Participants

This prospective study involved children who had autism (AU) and typically developing children (TD) over the period from December 2014 through May 2018. Children in the autistic group were enrolled from the Beijing Herun Clinic. This study was approved by the Peking Union Medical College Hospital (study #ZS-824), written informed consent was obtained from the parents of the participants. All participants were examined by experienced pediatricians.

Inclusion criteria for Autistic Disorder (AU) were as defined by the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-4) (Hu, 2003). All autistic children were assessed by specialized clinicians.

Exclusion criteria were: (1) presence of other diseases such as diabetes or PKU; (2) presence of certain factors that would interfere with the detection of urine organic acids (e.g., renal failure, hepatic insufficiency, dietary intervention therapy); (3) diagnosis of other neuropsychiatric disorders; (4) parents who could not complete the assessment.

Typically developing children were enrolled from primary schools in Beijing.

Procedures

Several precautions were strictly followed both before and after sampling to ensure specimen quality. The precautions and sampling steps were:

Before sampling:

(1) The subjects could not have used antibiotics (oral or infusion) in the previous month. Since some indicators we detect are associated with the intestinal micro-environment, antibiotic usage could affect the results by altering the distribution of intestinal flora.
(2) Both groups were not allowed to take probiotics within 2 weeks of sample collection. Probiotics can also perturb the intestinal micro-environment and affect the accuracy of urine organic acid testing.

(3) Study participants were not allowed to consume fruits or tomatoes within 24 h of sample collection due to the phenol or acid contents of these foods. For example, apples contain polyphenols such as anthocyanins, flavanols, phenolic acids, and catechins. Grapes are also rich in polyphenols. Such compounds can affect various metabolic pathways, and could affect the consistency of our results.

Sampling Steps

Midstream urine from the first morning void was collected in sterile tubes. The samples were placed on dry ice or in a freezer as soon as possible to avoid bacterial growth.

Measures

Information concerning study population characteristics was obtained from the Peking Union Medical College Hospital electronic medical record system. Follow-up information was collected through regular clinic and telephone communication.

All assessments of the children's behavior and dietary habits were provided by the parents and professional third-party institutions. The forms were produced in strict accordance with relevant standards and were completed following delivery of a detailed introduction and description of the study. Samples were collected either in the home or outpatient environment to ensure that external factors did not affect the samples. The researcher confirmed by phone that study guidelines were being followed.

The urine samples were assayed at the Great Plains Laboratory, Inc. (Lenexa, KS, United States). The GC/MS was performed as described in a previous research (Shaw et al., 1995). Due to the limitation of available data, only concentrations of 76 organic acids were reported from the spectrum analysis. Before analyses, all sample concentrations were normalized with urine creatinine as a way of minimizing variability due to differences in urine concentration.

Data Processing and Modeling

The total processes of data processing and modeling are illustrated in **Figure 1**. Sample data from GC/MS were first standardized by creatinine to eliminate urine concentration variabilities. Then the data were further processed with scaling and centering. To avoid data contamination between model building and model testing processes, we set aside an independent testing set from the entire data set in advance. The independent testing set would be strictly excluded from any model building processes so that overfitting effect in testing stage could be minimized. The splitting between testing set and training set were through a random process, while the ratios of control and ASD samples were kept approximately equal in these two sets.

During the model building process, we first trained our models and adjusted algorithm parameters using the training set with all 76 organic acids by maximizing AURs from leave-one-out cross validations. The modeling algorithms included Partial Least Squares Discriminant Analysis (PLS-DA, R mixOmics package),

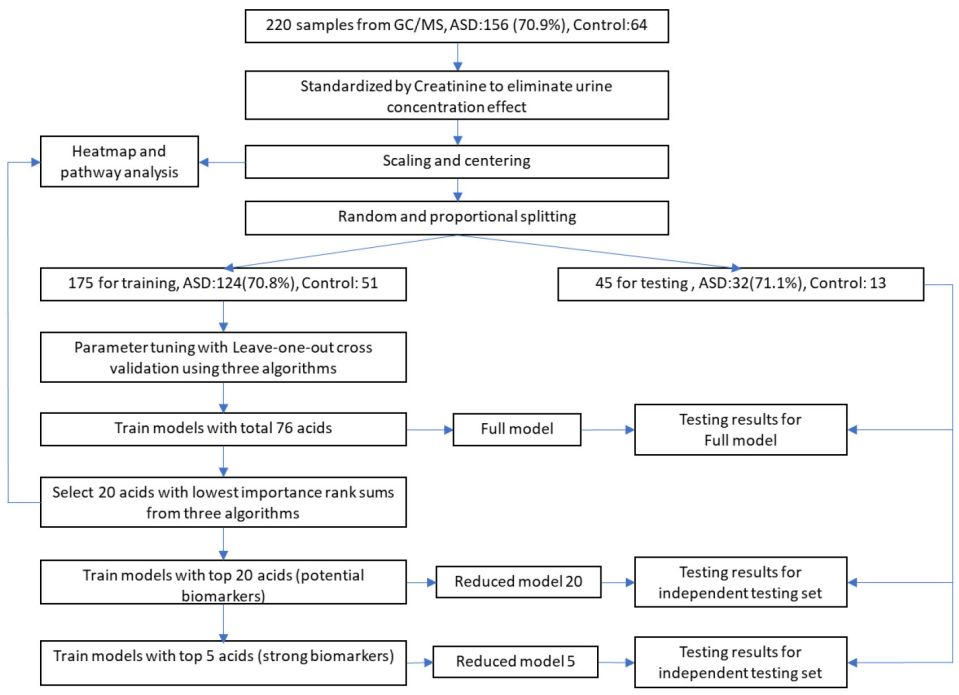

FIGURE 1 | Workflow of data processing and modeling. After standardization, scaling, and centering, sample data were split into training and testing sets, while testing set would be strictly excluded from any model building processes to minimize overfitting. Potential biomarkers were discovered by selecting top N acids from an importance rank sum list generated by three different classification algorithms, while N was determined from testing results of different *N*-values. The discovered biomarkers' potential mechanisms were investigated through heatmap along with associated metabolic pathways.

Support Vector Machine (SVM, R e1071 package), and XGBoost (eXtreme Gradient Boosting, R XGBoost package). The generated models with total 76 acids were designated as full models. Then, full models' classification performance was evaluated using the independent testing set.

To identify potential biomarkers for ASD, we exploited a voting mechanism from all three algorithms to avoid possible biases. First, importance scores of all acids were determined by all the three algorithms using R caret package (Gevrey et al., 2003). Next, each algorithm provided a rank order of all acids according to their importance scores. Then, a list of acids with each acid's sum of importance rank from the three algorithms in descending order was generated. Last, we trained models with only top N acids on the list and tested their classification performance on the testing set to identify the possible biomarkers. The models with top a few acids are referred as reduced models. Biomarkers were determined by observing the testing results of these reduced models on the testing set.

After the detection of biomarkers, to investigate possible mechanism behind them, we produced a heatmap with hierarchical clusters of all sample data and mark these biomarkers on the map after standardization processes. The heatmap was split aligning two dimensions, sample groups, and related metabolic pathways.

Classification algorithms were evaluated using AURs and their confidence intervals were estimated using bootstrapping methods with 2,000 bootstrap steps. Mann–Whitney *U*-test was used to compare the values for important biomarker acids. Multiple comparisons were adjusted using the false discovery rate (FDR) method (Benjamini and Yekutieli, 2001). Part of evaluations of PLS-DA algorithm was conducted using SIMCA-P Version 11.5 (Umetrics, Umeå, Sweden).

RESULTS

A total of 220 participants were enrolled including 156 autism patients (ASD group) and 64 typically developing children (TD group). The population characteristics of the ASD and TD groups have been summarized in **Table 1**. Among the ASD group, 80.13% were male. In TD group, 73.44% were male. The ASD and TD groups showed no significant differences in gender ($P = 0.285$).

TABLE 1 | Characteristics of ASD group and TD group.

	ASD	TD
	***n* = 156**	***n* = 64**
Age, years	6 (4, 9.75)*	5 (4, 7)*
No. of males (%)	125 (80.13%)	47 (73.44%)
No. of females (%)	31 (19.87%)	17 (26.56%)

*All values are expressed as numbers (percentages) or median value (P25, P75). A total of 220 participants (156 autism patients and 64 typically developing children) were enrolled and divided into the ASD group and TD group. Among the ASD group, 80.13% were male. In TD group, 73.44% were male. The ASD and TD groups showed no significant differences in gender (P = 0.285). *median values (P25, P75).*

TABLE 2 | Potential marker metabolites found in GC/MS of urine samples.

No	Metabolite	Differentiation for autistic samples	*p*-value* after FDR adjustment
1	Phenylactic	↑	0.000
2	3-Hydroxy-3-methylglutaric	↑	0.004
3	Phosphoric	↓	0.001
4	Fumaric	↓	0.003
5	3-Oxoglutaric	↓	0.001
6	Aconitic	↓	0.000
7	*N*-Acetylcysteine (NAC)	↓	0.056
8	Malonic	↓	0.031
9	Tricarballylic	↓	0.052
10	Glycolic	↓	0.140
11	Creatinine	↑	0.010
12	Malic	↓	0.055
13	Oxalic	↑	0.025
14	Tartaric	↓	0.046
15	Pyruvic	↑	0.013
16	4-Cresol	↑	0.030
17	Carboxycitric	↓	0.001
18	3-Hydroxyglutaric	↓	0.071
19	2-Hydroxybutyric	↑	0.330
20	2-Oxoglutaric	↓	0.408

*A voting mechanism from all three algorithms was applied to generate the most important 20 organic acids as potential biomarkers for ASD. We performed U-Mann–Whitney test to determine whether there were significant differences in the levels of these 20 organic acids between ASD and TD children. *p-values were calculated using Mann–Whitney test; ↑, increased level compared with TD; ↓, decreased level compared with TD.*

Data Sets

Two sets, a training set (80%) and a testing set (20%), were randomly separated from the total data set, and each had a similar proportion of ASD children. The training set contained 124 (70.9%) ASD children and 51 TD children, whereas the testing set had 32 (71.1%) ASD and 13 TD children.

Model Building Using Training Set

Using training set, we fine-tuned parameters of the three algorithms. For PLS-DA, the best major parameter, Ncomp, is 2. For SVM, we obtained the best results using linear kernel. For XGBoost, we optimized three major parameters, max_depth, eta, and nrounds with optimal values of 2, 0.15, and 200, respectively.

The model building process employed leave-one-out cross validation as guidance for parameter tuning. The final results of the full models in this process were shown in **Table 2**. The AURs for these three algorithms were 0.864 (PLS-DA), 0.833 (SVM), and 0.931 (XGBoost) in training set with leave-one-out cross validation.

The PCA result on training set with all acids is on **Figure 2A**. From the figure, we see that PCA could not distinguish between ASD and TD groups, since the new components variables generated with maximal variances might not be aligned with the outcome groups. However, it does identify some outliers. To make models more robust, we did not remove these outliers in the following analyses.

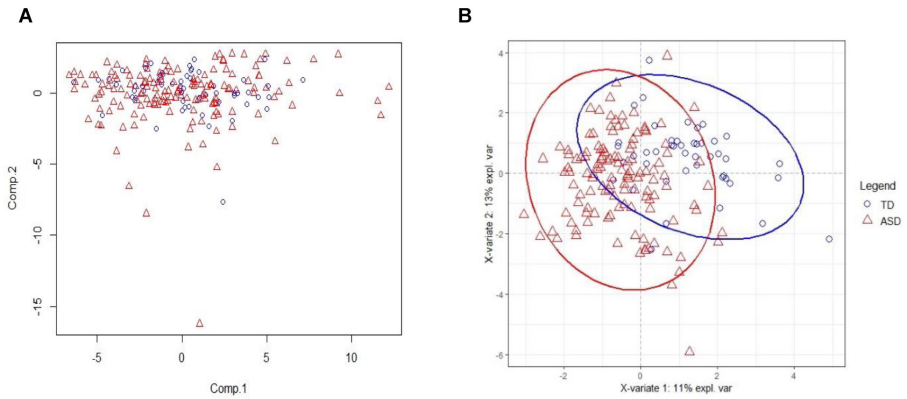

FIGURE 2 | PCA and PLSDA score plots. **(A)** The Principal Component Analysis (PCA) score plot on training set with all 75 acids. **(B)** The PLS-DA score plot on training set with selected 20 biomarker acids. With first two components, R2X (cum) = 0.26, R2Y (cum) = 0.535, Q2 (cum) = 0.386.

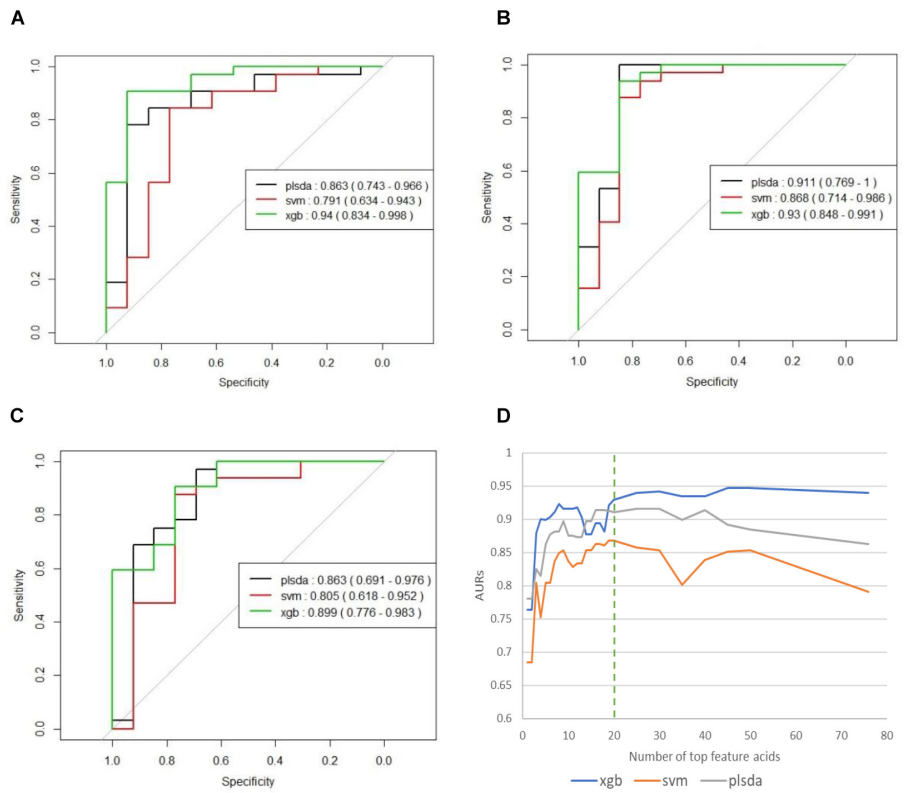

FIGURE 3 | (A) ROCs of final models on independent testing set based on all 76 organic acids. **(B)** ROCs of final models on independent testing set based on top 20 organic acids. **(C)** ROCs of final models on independent testing set based on top five organic acids. **(D)** Curves of AURs against selected top acids. Top 20 acids represent the best collection of possible ASD biomarkers, while adding more will make AURs for SVM and PLS-DA decrease and make AUR for XGB appear platformed.

Model Testing Using Independent Testing Set

To avoid any possible overfitting, we tested the full model on the independent testing set and obtained AURs of 0.863 (PLS-DA), 0.791 (SVM), and 0.94 (XGBoost). These results had shown similar values with those in training stage showing that training stage has generated little overfitting to the training set.

Potential Marker Metabolites

We used the testing resulting of reduced models to identify potential markers. The results of these reduced models are shown

in **Supplementary Table S1** and **Figure 3**. **Figure 3D** the curves of AURs against different N selected top acids on testing set. Clearly, top 20 acids represent the best collection of possible ASD biomarkers, while adding more acids to the model will make AURs for SVM and PLS-DA decrease and make AUR for XGB appear platformed (The ensemble mechanism of XGBoost might make it more robust to irrelevant features). Actually, XGBoost achieved an AUR of 0.93 which was very close to the value of 0.94 in full model, and this suggest that these top 20 acids could capture most of the features of ASD. Even top 5 acids could get an AUR of 0.899.

The 20 identified potential marker metabolites are listed in **Table 2**. Their levels compared with TD group are also shown in **Table 2**. Using these 20 identified marker metabolites, we draw the score plot of PLA-DA of training set on **Figure 2B**. There is a separation between TD and ASD groups with R2X (cum) = 0.26, R2Y (cum) = 0.535, Q2 (cum) = 0.386, *p*-value of CV-anova = 1.26183e-006.

Heatmap Analysis of Metabolic Pathway

We tried to use heatmap with hierarchical clustering to discovery possible related metabolic pathways (**Supplementary Figure S1**). Rows of the heatmap represent different samples from TD and ASD groups, while columns represent different metabolites grouped in different metabolic pathways. The pathway names are listed in the figure legend. The heatmap shows that the identified biomarker metabolite acids are distributed across a wide variety of pathways: Intestinal Microbial Overgrowth, Amino Acid Metabolism, nutritional, Krebs Cycle, Oxalate Metabolism, Glycolytic Cycle, and Mineral Metabolism. This diverse distribution suggests that these organic acids may act on a variety of metabolic pathways and reflects the complexity of metabolic abnormalities in autism.

DISCUSSION

To identify metabolic signatures of ASD and find organic acids in urine that could act as potential biomarkers for diagnosis and disease treatment, three algorithms (PLS-DA, SVM, and XGBoost) were used to analyze GC/MS data for urine samples. The results showed the effectiveness of this method in distinguishing ASD children from TD children. XGBoost model produced the best results (AUROC = 0.94) among the three algorithms. The modeling was performed on the basis of all 76 organic acids, among which the top 20 acids were identified as potential biomarkers with a voting mechanism from all three algorithms. To go a step further, we selected top 5 acids as strong biomarkers. The amount of phenylactic acid was significantly higher in the ASD group, whereas the amounts of aconitic acid, phosphoric acid, 3-oxoglutaric acid, and carboxycitric acid were significantly lower in the ASD group. These organic acids are involved in a variety of metabolic pathways including amino acid metabolism, intestinal flora, energy metabolism (Krebs Cycle), and bone salt metabolism. Although a total of 76 organic acids contributed to modeling, we just involved the top 5 ones in discussion part since they made significant contributions in modeling, which may indicate the major metabolic abnormality of autism.

Complex Relationships Among Urinary Organic Acids and ASD Pathogenesis

The heatmap generated from GC/MS analysis of urinary organic acids showed the complex relationship among these compounds in ASD. Several organic acids were in the same pathway, whereas others are involved in multiple pathways. To date, the metabolites that have been explored as possible ASD biomarkers include: nutritional markers, microbiome metabolites, amino acid metabolites, Krebs cycle metabolites, pyrimidine metabolites, neurotransmitter metabolites, products of ketone, and fatty acid oxidation and mineral metabolism, as well as indicators of detoxification and fluid intake (e.g., creatinine) (Kałużna-Czaplińska, 2011).

These organic acids may affect the function of intestinal flora. In our study, we also collected stool specimens from the study participants. Analysis of stool samples and the intestinal flora is underway, and the abundance of intestinal flora combined with findings for urinary organic acid metabolism should strengthen the diagnostic potential of these compounds.

The organic acids we identified may affect nervous system development and thus we included assessments of neurological symptoms (e.g., unexplained excitability or mania) on the study scales. We will examine whether the severity of these symptoms in ASD and the CARS and ABC scores are relevant in a future study.

The Diagnostic Potential of the Established Model

Calibration and optimization of parameters is a critical step in model building. Three algorithms (PLS-DA, SVM, and XGBoost) were examined to achieve this task. Among them, the first two algorithms, PLS-DA and SVM, were previously described (West et al., 2014; Dieme et al., 2015). To our knowledge, application of the XGBoost algorithm in a model of urine organic acids to distinguish children with ASD from TD children has not been previously reported.

Among the three algorithms, XGBoost had an AOC of 0.94. Use of the XGBoost algorithm is an innovation in autism-related research (Chen and Guestrin, 2016), and the efficiency of this model differs from that described in earlier studies. XGBoost has been proved to have better performance than other more traditional models in many machine learning tasks outside biomedical domains. This is largely due to its built-in ensemble mechanism and its ability to capture non-linear features. In contrast, traditional linear algorithm for metabolite analysis, PLS-DA, is limited in capturing non-linear relations. This has also been observed in this study. In addition, XGBoost also shows more robustness to adding more irrelevant features than SVM and PLA-DA. Conclusively, the establishment of this model increases the possibility of early diagnosis of autism. The examination of organic acids in urine is non-invasive and relatively inexpensive, the

requirements for sample collection are not strict and the operability is very high.

Notable Changes in Urinary Organic Acid Levels in ASD Patients

The PLS-DA score plot shows a clear distinction among the distribution of metabolite profiles between TD and ASD children. Our analyses showed that 5 urinary organic acids had significant differences between ASD and TD children and thus could have diagnostic potential as ASD biomarkers.

The ASD group had higher levels of phenylactic acid but decreased amounts of aconitic acid, phosphoric acid, 3-oxoglutaric acid, and carboxycitric acid compared to TD children. These metabolites are associated with multiple biochemical processes (Koulman et al., 2009). Phenylactic acid is a byproduct of amino acid metabolism, and the higher levels seen for ASD children could indicate abnormalities in the function of enzymes involved in amino acid metabolism. Moreover, phenylactic acid can inhibit the growth of Gram-negative and Gram-positive bacteria, as well as some fungi. Thus, elevated phenylactic acid levels could inhibit the normal function of the intestinal microflora and exacerbate metabolic disorders. Intestinal microbes can affect neurotransmitter production in the central nervous system and in turn affect the induction of endogenous sensations, production of bacterial metabolites, and mucosal immune-related activity (Carabotti et al., 2015). Moreover, phenylactic acid is a metabolite of phenethylamine, which acts as a monoaminergic neuromodulator and as a neurotransmitter in the human central nervous system to promote neuron excitation (Sabelli et al., 1976).

Also in the context of intestinal flora, levels of carboxycitric acid and 3-oxoglutarate acid were significantly decreased in the ASD group relative to the TD group. Carboxycitric acid can be a marker of intestinal microbial overgrowth, particularly yeast and fungi. Certain strains of the mold *Aspergillus niger* have efficient citric acid production and can be used for industrial-scale citric acid production (Lotfy et al., 2007). Although to our knowledge, this study is the first to report a significant decrease of carboxycitric acid in urine samples from ASD children, other studies identified intestinal microbe metabolites as potential agents that can affect nervous system function. Meanwhile, carboxycitric acid, a product of the Krebs Cycle, showed decreased levels in our assays and may be indicative of energy metabolism disorders in children with autism. We also found that 3-oxoglutarate, a common metabolite of yeast and fungi (Thomas et al., 2010; MacFabe et al., 2011; Kocovska et al., 2012), was significantly lower in children with autism. The low concentrations of both carboxycitric acid and 3-oxoglutarate that we observed in urine from autistic patients could be due to increased uptake of these compounds across the blood-brain barrier of the brain. Our results are consistent with previous studies that showed anti-fungal treatments for children with autism can effectively reduce the amounts of corresponding organic acid indicators (Cobb and Cobb, 2010), and suggests that gastrointestinal yeast could provide a basis for dietary adjustments such as gluten/casein-free diets that are important for children's nervous system development and could mitigate autism symptoms. 3-oxoglutarate in urine is associated with the presence of harmful gut flora such as *Candida albicans* (Schmidt, 1994). These results support the reliability of the gut-brain axis and suggest new avenues of study for autism.

Aconitic acid is produced from citric acid dehydration that occurs during the Krebs Cycle and is a marker of mitochondrial activity. Mitochondrial disease, either through maternal inheritance or other causes, is present in up to 5% of autistic children (Rossignol and Frye, 2012; Frye et al., 2013). Previous studies reported that *cis*-acotinic acid levels are increased in children with autism (Noto et al., 2014; Mussap et al., 2016). Here, we found that acotinic acid levels were decreased in the ASD group relative to the TD group, which is indicative of energy metabolism deficiencies in energy metabolism of ASD. In the Krebs Cycle, citrate undergoes stereospecific isomerization to isocitrate by the enzyme aconitase hydratase and the intermediate *cis*-aconitate (Mussap et al., 2016). Meanwhile, *trans*-aconitic acid (TAA) acts as an anti-inflammatory agent in plant-based treatments for rheumatoid arthritis used in Brazil, and could be one explanation for the decreased levels of aconitic acid in the ASD group. Similarly, it has been reported that inflammatory mediators may play crucial role in some neuropsychiatric diseases. Dan et al. (2015) found that homocysteine (Hcy) and uric acid (UA) may contribute to the pathogenesis of multiple system atrophy (MSA) and serum Hcy together with UA levels could be a diagnostic tool of MSA (AUROC = 0.736). In addition, another cross-sectional study supported that low serum UA levels may indicate a higher risk of Parkinson's disease (PD) and serum UA level could serve as an indirect biomarker of prediction in PD (Mengqiu et al., 2013).

Phosphoric acid is important for bone metabolism. In our study population we observed decreased amounts of phosphoric acid in ASD children relative to TD children, which could suggest that ASD pathology involves abnormal bone metabolism, although this possibility requires further investigation. Vitamin D regulates bone formation and density by promoting absorption of key intestinal compounds such as calcium and phosphate. Imbalances in phosphoric acid could be related to an imbalance of several other substances. In pregnant women, vitamin D deficiencies can affect regulatory T cell function and in turn immune responses. Such vitamin D deficiencies can impact the developing fetus and could increase the risk for autism. Vitamin D is also critical during development of the fetal nervous system through regulation of the expression of several nerve growth factors as well as transforming factor beta 2 (TGF-b2) and neurotrophin 3 and 4. Previous studies showed that some children with autism have vitamin D deficiency (Pioggia et al., 2014; Uğur and Gürkan, 2014). The amount of serum 25 (OH) D_3 is significantly lower in children with ASD, indicating that lower 25 (OH) D levels could be an independent risk factor for autism, and may be independently associated with disease severity (Gong et al., 2014). Our findings support observations of disorders in bone salt metabolism in children with autism, and are also consistent with clinical symptoms

indicating that reduced bone mineral density is common in children with autism.

CONCLUSION

In this study, we used GC/MS to evaluate major metabolic fluctuations in 76 organic acids present in urine from 156 children with ASD and from 64 non-autistic children. Three algorithms, Partial Least Squares-Discriminant Analysis (PLS-DA), Support Vector Machine (SVM), and eXtreme Gradient Boosting (XGBoost), were used to develop models to distinguish ASD from TD children and to detect potential biomarkers. By a voting mechanism, 20 acids have been successfully identified as potential ASD biomarkers and reduced model with top 20 acids achieved 0.93 and represent a good collection of potential ASD biomarkers. These biomarkers were distributed across a wide variety of metabolic pathways, indicating the complicated mechanism behind ASD. XGBoost algorithm has shown better classification performance and more robustness than other traditional algorithms.

In summary, urine organic acids detection with GC/MS combined with XGBoost algorithm could represent a novel, non-invasive and accurate strategy for diagnosis of autism and the discovered potential biomarkers could be valuable for future research on the pathogenesis of autism and possible interventions, and have a range of clinical applications.

ETHICS STATEMENT

This study was carried out in accordance with the recommendations of Chinese Academy of Medical Sciences Peking Union Medical College Hospital Ethics Review Committee The protocol was approved by the Chinese Academy of Medical Sciences Peking Union Medical College Hospital Ethics Review Committee.

AUTHOR CONTRIBUTIONS

QC, YQ, XY, and X-jX contributed to the conception and design of the study. XY and X-jX enrolled the patients and controls, and collected all the clinical data, and urine samples. QC and YQ transferred the data to a database, analyzed the data, and wrote the first draft of the manuscript. YT analyzed the data, set up diagnosis model, and revised the figures. All authors contributed to the manuscript revision, and read, and approved the submitted version of the manuscript.

FUNDING

This study was supported by Peking Union Medical Foundation.

ACKNOWLEDGMENTS

We would like to thank the reviewers for their comments. We were particularly indebted to the families and children involved in this study; the study would not have been possible without their collaboration.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fncel.2019.00150/full#supplementary-material

REFERENCES

Benjamini, Y., and Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Ann. Stat.* 29, 1165–1188. doi: 10.1186/1471-2105-9-114

Bull, G., Shattock, P., Whiteley, P., Anderson, R., Groundwater, P. W., Lough, J. W., et al. (2003). Indolyl-3-acryloylglycine (IAG) is a putative diagnostic urinary marker for autism spectrum disorders. *Med. Sci. Monit.* 9:CR422.

Carabotti, M., Scirocco, A., Maselli, M. A., and Severi, C. (2015). The gut-brain axis: interactions between enteric microbiota, central and enteric nervous systems. *Ann. Gastroenterol. Quart. Publ. Hellen. Soc. Gastroenterol.* 28, 203–209.

Chen, T., and Guestrin, C. (2016). "XGBoost: a scalable tree boosting system," in *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, (New York, NY: ACM).

Cobb, M. L., and Cobb, A. (2010). *Treatment of Autism Using Probiotic Composition*. New Zealand: Cobb And Company.

Cox, K. B., Hamm, D. A., Millington, D. S., Matern, D., Vockley, J., Rinaldo, P., et al. (2001). Gestational, pathologic and biochemical differences between very long-chain acyl-CoA dehydrogenase deficiency and long-chain acyl-CoA dehydrogenase deficiency in the mouse. *Hum. Mol. Genet.* 10, 2069–2077. doi: 10.1093/hmg/10.19.2069

Dan, C., Xiaobo, W., Jing, Z., Rui, W., Xu, L., Xiaofeng, X., et al. (2015). Contra-directional expression of serum homocysteine and uric acid as important biomarkers of multiple system atrophy severity: a cross-sectional study. *Front. Cell. Neurosci.* 9:247. doi: 10.3389/fncel.2015.00247

Dawson, G., Jones, E. J., Merkle, K., Venema, K., Lowy, R., Faja, S., et al. (2012). Early behavioral intervention is associated with normalized brain activity in young children with autism. *J. Am. Acad. Child Adolesc. Psychiatry* 51, 1150–1159. doi: 10.1016/j.jaac.2012.08.018

Dawson, G., Rogers, S., Munson, J., Smith, M., Winter, J., Greenson, J., et al. (2010). Randomized, controlled trial of an intervention for toddlers with autism: the early start denver model. *Pediatrics* 125:e17. doi: 10.1542/peds.2009-0958

Dieme, B., Mavel, S., Blasco, H., Tripi, G., Bonnetbrilhault, F., Malvy, J., et al. (2015). Metabolomics study of urine in autism spectrum disorders using a multiplatform analytical methodology. *J. Proteome Res.* 14, 5273–5282. doi: 10.1021/acs.jproteome.5b00699

Emond, P., Mavel, S., Aidoud, N., Nadal-Desbarats, L., Montigny, F., Bonnet-Brilhault, F., et al. (2013). GC-MS-based urine metabolic profiling of autism spectrum disorders. *Anal. Bioanal. Chem.* 405, 5291–5300. doi: 10.1007/s00216-013-6934-x

Frye, R. E., Melnyk, S., and Macfabe, D. F. (2013). Unique acyl-carnitine profiles are potential biomarkers for acquired mitochondrial disease in autism spectrum disorder. *Transl. Psychiatry* 3:e220. doi: 10.1038/tp.2012.143

Gevrey, M., Dimopoulos, I., and Lek, S. (2003). Review and comparison of methods to study the contribution of variables in artificial neural network models. *Ecol. Modell.* 160, 249–264. doi: 10.1016/s0304-3800(02)00257-0

Ghaziuddin, M., and Alowain, M. (2013). Autism spectrum disorders and inborn errors of metabolism: an update. *Pediatr. Neurol.* 49, 232–236. doi: 10.1016/j.pediatrneurol.2013.05.013

Gong, Z. L., Luo, C. M., Wang, L., Shen, L., Wei, F., Tong, R. J., et al. (2014). Serum 25-hydroxyvitamin D levels in Chinese children with autism spectrum disorders. *Neuroreport* 25, 23–27. doi: 10.1097/WNR.0000000000000034

Hanley, H. G., Stahl, S. M., and Freedman, D. X. (1977). Hyperserotonemia and amine metabolites in autistic and retarded children. *Arch. Gen. Psychiatry* 34, 521–531.

Hu, R. J. (2003). Diagnostic and statistical manual of mental disorders (DSM-IV). *Encycl. Neurol. Sci.* 25, 4–8.

Kałużna-Czaplińska, J. (2011). Noninvasive urinary organic acids test to assess biochemical and nutritional individuality in autistic children. *Clin. Biochem.* 44, 686–691. doi: 10.1016/j.clinbiochem.2011.01.015

Keller, F., and Persico, A. M. (2003). The neurobiological context of autism. *Mol. Neurobiol.* 28, 1–22. doi: 10.1385/mn%3A28%3A1%3A1

Klin, A., Klaiman, C., and Jones, W. (2015). Reducing age of autism diagnosis: developmental social neuroscience meets public health challenge. *Rev. Neurol.* 60(Suppl. 1), S3–S11.

Kocovska, E., Fernell, E., Billstedt, E., Minnis, H., and Gillberg, C. (2012). Vitamin D and autism: clinical review. *Res. Dev. Disabil.* 33, 1541–1550. doi: 10.1016/j.ridd.2012.02.015

Kompare, M., and Rizzo, W. B. (2008). Mitochondrial fatty-acid oxidation disorders. *Semin. Pediatr. Neurol.* 15, 140–149. doi: 10.1016/j.spen.2008.05.008

Koulman, A., Lane, G. A., Harrison, S. J., and Volmer, D. A. (2009). From differentiating metabolites to biomarkers. *Anal. Bioanal. Chem.* 394, 663–670. doi: 10.1007/s00216-009-2690-3

Lotfy, W. A., Ghanem, K. M., and El-Helow, E. R. (2007). Citric acid production by a novel *Aspergillus niger* isolate: II. Optimization of process parameters through statistical experimental designs. *Bioresourc. Technol.* 98, 3470–3477. doi: 10.1016/j.biortech.2006.11.032

MacFabe, D. F., Cain, N. E., Boon, F., Ossenkopp, K. P., and Cain, D. P. (2011). Effects of the enteric bacterial metabolic product propionic acid on object-directed behavior, social behavior, cognition, and neuroinflammation in adolescent rats: relevance to autism spectrum disorder. *Behav. Brain Res.* 217, 47–54. doi: 10.1016/j.bbr.2010.10.005

Manzi, B., Loizzo, A. L., Giana, G., and Curatolo, P. (2008). Autism and metabolic diseases. *J. Child Neurol.* 11, 90–90.

Mavel, S., Nadal-Desbarats, L., Blasco, H., Bonnet-Brilhault, F., Barthelemy, C., Montigny, F., et al. (2013). 1H-13C NMR-based urine metabolic profiling in autism spectrum disorders. *Talanta* 114, 95–102. doi: 10.1016/j.talanta.2013.03.064

Mengqiu, P., Huimin, G., Ling, L., Yunqi, X., Mei, L., Jing, Z., et al. (2013). Serum uric acid in patients with Parkinson's disease and vascular parkinsonism: a cross-sectional study. *Neuroimmunomodulation* 20, 19–28. doi: 10.1159/000342483

Mussap, M., Noto, A., and Fanos, V. (2016). Metabolomics of autism spectrum disorders: early insights regarding mammalian-microbial cometabolites. *Expert Rev. Mol. Diagnost.* 16:869. doi: 10.1080/14737159.2016.1202765

Nair, M. K. (2000). Autism spectrum disorders. *Neuron.* 28, 355–363.

Noto, A., Fanos, V., Barberini, L., Grapov, D., Fattuoni, C., Zaffanello, M., et al. (2014). The urinary metabolomics profile of an Italian autistic children population and their unaffected siblings. *J. Matern. Fetal Neonatal Med.* 27((Suppl. 2), 46–52. doi: 10.3109/14767058.2014.954784

Pioggia, G., Tonacci, A., Tartarisco, G., Billeci, L., Muratori, F., Ruta, L., et al. (2014). Autism and lack of D3 vitamin: a systematic review. *Res. Autism Spectr. Disord.* 8, 1685–1698. doi: 10.1016/j.rasd.2014.09.003

Rossignol, D. A., and Frye, R. E. (2012). Mitochondrial dysfunction in autism spectrum disorders: a systematic review and meta-analysis. *Mol. Psychiatry* 17, 290–314. doi: 10.1038/mp.2010.136

Sabelli, H. C., Mosnaim, A. D., Vazquez, A. J., Giardina, W. J., Borison, R. L., and Pedemonte, W. A. (1976). Biochemical plasticity of synaptic transmission: a critical review of Dale's Principle. *Biol. Psychiatry* 11, 481–524.

Schain, R. J., and Freedman, D. X. (1961). Studies on 5-hydroxyindole metabolism in autistic and other mentally retarded children. *J. Pediatr.* 58, 315–320. doi: 10.1016/s0022-3476(61)80261-8

Schmidt, M. A. (1994). *Tired of Being Tired: Overcoming Chronic Fatigue and Low Vitality*. Berkeley, CA: Frog, Ltd.

Shaw, W., Kassen, E., and Chaves, E. (1995). Increased urinary excretion of analogs of Krebs cycle metabolites and arabinose in two brothers with autistic features. *Clin. Chem.* 41, 1094–1104.

Thomas, R. H., Foley, K. A., Mepham, J. R., Tichenoff, L. J., Possmayer, F., and MacFabe, D. F. (2010). Altered brain phospholipid and acylcarnitine profiles in propionic acid infused rodents: further development of a potential model of autism spectrum disorders. *J. Neurochem.* 113, 515–529. doi: 10.1111/j.1471-4159.2010.06614.x

Uğur, Ç, and Gürkan, C. K. (2014). Serum vitamin D and folate levels in children with autism spectrum disorders. *Res. Autism Spectr. Disord.* 8, 1641–1647. doi: 10.1016/j.rasd.2014.09.002

Wanders, R. J., Vreken, P., den, Boer ME, Wijburg, F. A., van, Gennip AH, and IJlst, L. (1999). Disorders of mitochondrial fatty acyl-CoA β-oxidation. *J. Inherit. Metab. Dis.* 22, 442–487.

Wang, S. S., Fernhoff, P. M., Hannon, W. H., and Khoury, M. J. (1999). Medium chain acyl-CoA dehydrogenase deficiency human genome epidemiology review. *Genet. Med.* 1, 332–339. doi: 10.1097/00125817-199911000-00004

Warren, Z., Mcpheeters, M. L., Sathe, N., Foss-Feig, J. H., Glasser, A., and Veenstra-Vanderweele, J. (2011). A systematic review of early intensive intervention for autism spectrum disorders. *Pediatrics* 127:e1303. doi: 10.1542/peds.2011-0426

West, P. R., Amaral, D. G., Bais, P., Smith, A. M., Egnash, L. A., Ross, M. E., et al. (2014). Metabolomics as a tool for discovery of biomarkers of autism spectrum disorder in the blood plasma of children. *PLoS One* 9:e112445. doi: 10.1371/journal.pone.0112445

Zecavati, N., and Spence, S. J. (2009). Neurometabolic disorders and dysfunction in autism spectrum disorders. *Curr. Neurol. Neurosci. Rep.* 9, 129–136. doi: 10.1093/brain/awx054

Zwaigenbaum, L., Bryson, S., and Garon, N. (2013). Early identification of autism spectrum disorders. *Behav. Brain Res.* 251, 133–146. doi: 10.1016/j.bbr.2013.04.004

Conflict of Interest Statement: YT was employed by Ping An Technology (Shenzhen) Ltd., Institute of Artificial Intelligence, Beijing, China.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Chen, Qiao, Xu, You and Tao. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.