HAL
open science

GC-MS-based urine metabolic profiling of autism spectrum disorders.

Patrick Emond, Sylvie Mavel, Nacima Aïdoud, Lydie Nadal-Desbarats, Frédéric Montigny, Frédérique Bonnet-Brilhault, Catherine Barthélémy, Marc Merten, Pierre Sarda, Frédéric Laumonnier, et al.

To cite this version:

Patrick Emond, Sylvie Mavel, Nacima Aïdoud, Lydie Nadal-Desbarats, Frédéric Montigny, et al.. GC-MS-based urine metabolic profiling of autism spectrum disorders.: GC-MS-based Urine Metabolic Profiling in ASD. Analytical and Bioanalytical Chemistry, 2013, 405 (15), pp.5291-300. 10.1007/s00216-013-6934-x . inserm-00908948

HAL Id: inserm-00908948

https://inserm.hal.science/inserm-00908948

Submitted on 23 Apr 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

GC-MS-based urine metabolic profiling of Autism Spectrum Disorders

Journal:	*Analytical and Bioanalytical Chemistry*
Manuscript ID:	ABC-00105-2013
Type of Paper:	Research Paper
Date Submitted by the Author:	18-Jan-2013

Complete List of Authors:	Emond, Patrick; Université François-Rabelais, INSERM U930 Mavel, Sylvie; University of Tours, INSERM U930; Université François-Rabelais, INSERM U930 Aïdoud, Nacima; Université François-Rabelais, INSERM U930 Nadal-Desbarats, Lydie; Université François-Rabelais, INSERM U930 Montigny, Frédéric; Université François-Rabelais, INSERM U930 Bonnet-Brilhault, Frédérique; Université François-Rabelais, INSERM U930 Barthélémy, Catherine; Université François-Rabelais, INSERM U930 Merten, Marc; Université Henri Poincaré, Laboratoire de Biochimie Sarda, Pierre; CHRU de Montpellier,

Keywords:	Mass spectrometry / ICP-MS, Chemometrics / Statistics, Bioanalytical methods, GC, Clinical / Biomedical analysis

SCHOLARONE™
Manuscripts

GC-MS-based urine metabolic profiling of Autism Spectrum Disorders

Patrick Emond[a,b*] . Sylvie Mavel[a] . Nacima Aïdoud[a] . Lydie Nadal-Desbarats[a,b] . Frédéric Montigny[b] . Frédérique Bonnet-Brilhault[c] . Catherine Barthélémy[c] . Marc Merten[d] . Pierre Sarda[e] . Frédéric Laumonnier[a] . Patrick Vourc'h[a,b] . Hélène Blasco[a] . Christian R Andres[a] .

[a] Université François-Rabelais, INSERM U930, Equipe neurogénétique et neurométabolomique, CHRU de Tours, Hôpital Bretonneau, 10 Bv Tonnellé, 37044 Tours, France

[b] Université François-Rabelais, PPF "Analyses des Systèmes Biologiques", UFR de Médecine, 10 Bv Tonnellé, 37044 Tours, France

[c] Université François-Rabelais, INSERM U930, Equipe Autisme, CHRU de Tours, 10 Bv Tonnellé, 37044 Tours, France

[d] Université Henri Poincaré, Faculté de Médecine, Laboratoire de Biochimie, 9 Av de la forêt de Haye, 54505 Vandoeuvre les Nancy, France

[e] CHRU de Montpellier, Hôpital Arnaud-de-Villeneuve, 34295 Montpellier cedex 5, France

Corresponding author: Université François Rabelais, INSERM U930, 10 Bv Tonnellé, 37044 Tours Tel: + 33 2 47 36 61.53 ; Fax: + 33 2 47 36 72 24; e-mail: patrick.emond@univ-tours.fr

Running Head: GC-MS-based Urine Metabolic Profiling in ASD

Abbreviations

For Peer Review

ASD autism spectrum disorders

BSTFA bis(trimethylsilyl)trifluoroacetamide

GC-MS gas chromatography combined with mass spectroscopy

NMR nuclear magnetic resonance

OPLS-DA orthogonal partial least-squares discriminant analysis

Par pareto

PCA principal component analysis

PLS-DA partial least squares discriminant analysis

TMS trimethylsilylated derivative

TMSO trimethylsilylated and oximated derivative

UV unit variance

VIP variable importance on projection

For Peer Review

Abstract

Autism Spectrum Disorders (ASD) are a group of neurodevelopmental disorders resulting from multiple factors. Diagnosis is based on behavioural and developmental signs detected before 3 years of age, and there is no reliable biological marker. The purpose of this study was to evaluate the value of gas chromatography combined with mass spectroscopy (GC-MS) associated with multivariate statistical modeling to capture the global biochemical signature of autistic individuals. CG-MS urinary metabolic profiles of 26 autistic and 24 healthy children were obtained by liq/liq extraction, and were or were not subjected to an oximation step, and then were subjected to a persilylation step. These metabolic profiles were then processed by multivariate analysis, in particular orthogonal partial least-squares discriminant analysis (OPLS-DA). Discriminating metabolites were identified. The relative concentrations of the succinate and glycolate were higher for autistic than healthy children; whereas those of hippurate, 3-hydroxyphenylacetate, vanillylhydracrylate, 3-hydroxyhippurate, 4-hydroxyphenyl-2-hydroxyacetate, 1*H*-indole-3-acetate, phosphate, palmitate, stearate, and 3-methyladipate relative concentrations were lower. Eight other metabolites, that were not identified but characterized by a retention time plus a quantifier and its qualifier ion masses, were found to differ between the two groups. Comparison of statistical models leads to the conclusion that the combination of data obtained from both derivatization techniques leads to the model best discriminating, between autistic and healthy groups of children.

Keywords: Mass spectrometry; Chemometrics / Statistics; Bioanalytical methods; GC; Clinical / Biomedical analysis

For Peer Review

Introduction

Autistic disorder (AD), Asperger syndrome (AS) and pervasive developmental disorder-not otherwise specified (PDD-NOS) are collectively termed autism spectrum disorders (ASD). The prevalence of ASD appears to be increasing (1 per 110 in 2009) [1] without any identified cause for this increase [2-5]. Autism is usually diagnosed in infancy between the second and the third years of life [6]. The disease is characterized by a behavioral triad as listed in the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV) [7]: impaired communication, impaired social interaction, and restricted and repetitive interests and activities. Diagnosis mostly involves clinical evaluation using subjective methods based on perceived behaviors in the patient. Thus, this diagnostic approach is dependent on the expertise of those administering the tests.

Metabolomics is the study of metabolites, including classification, identification and semi quantitative evaluation of metabolites levels [8]. Metabolomics has been successfully applied for disease diagnosis, therapeutics, and functional genomic and toxicology studies [9,10]. A metabolite is commonly defined as compound having a low molecular weight, from 50 to about 1000 Daltons. Biological fluids contain very large numbers of metabolites (more than 8000), so sensitive and robust analytical methods are required. The analytical techniques most commonly used to identify and quantify metabolites are gas chromatography or liquid chromatography combined with mass spectroscopy (GC-MS or LC-MS, respectively) and nuclear magnetic resonance spectroscopy (NMR) [11]. NMR is independent of ionization propensities but is less sensitive than MS: less than 60 different metabolites are commonly identified in biological samples using this technique [12]. GC-MS is a sensitive and reproducible analytical method and, combined with public databases, its power for compound identification makes it of great value for metabolomics [13]. Low-molecular weight metabolites may be analyzed directly by GC-MS,

For Peer Review

but many structures contain polar groups and therefore need to be chemically derivatized prior to GC-MS analysis. The most commonly used derivatization methods involve silylation preceded or not by oximation [8]. These derivatization strategies give an access to a large set of chemical functions, including alcohol and carboxylic functions, amines, amides and aldehydes [8].

Several bio-fluids can be analyzed by GC-MS, of which urine has the advantages of being easily and non-invasively accessible and containing hundreds of metabolites that represent the endpoint of endogenous metabolism. GC-MS has been used for the description of the metabolic status of patients with neuropsychiatric disorders such as autism ([14], for review, see ref: [15]). In particular, gastrointestinal metabolites, abnormal neurotransmitter concentrations, the creatine to creatinine ratio, and the guanidine acetate concentrations have been reported to be discriminative markers. However, there is currently no evidence that any biomarker is useful in routine practice, and consequently further studies are required to identify clinically useful indicators [15].

We report an analysis of the urinary metabolic signatures of autism patients and healthy subjects with the aim of identifying potential biomarkers of ASD. Two techniques were compared, with and without oximation during the derivatization process before GC-MS. Statistical analysis methods [partial least squares discriminant analysis (PLS-DA) and orthogonal partial least squares discriminant analysis (OPLS-DA)] were used to identify metabolites that discriminate between ASD and control populations.

Materials and methods

Sample collection

For Peer Review

Between 2008 and 2010, urine samples were collected from children aged 6-14 years (mean 8 years) with ASD living in France (n = 26; male 85%, female 15%) and from control children (n = 24; male 67%, female 33%) aged 6-9 years (mean 7 years). Urine samples were collected into untreated vials during routine medical consultations, principally in the morning and the exact time of collection was recorded. Each urine sample was aliquoted into 1.5 mL Eppendorf tubes and stored at -80°C immediately after collection until analysis. All study participants and their parents or guardians provided informed consent.

Patients

The following data were collected for all autistic and control children: age, gender, treatment, and clinical characteristics. The severity of autism was assessed according to the International Classification of Diseases 10th Edition [16]. Autistic patients were recruited in three French autism centers [Tours (n = 8), Montpellier (n = 13), and Orléans (n = 5)] and all twenty-four urine samples were from healthy volunteers from Tours.

GC-MS study

Sample preparation

Samples were thawed at room temperature, centrifuged (at 3000 *g*) for 10 min and an aliquot was used for creatinine analysis (Jaffé method, Olympus AU640, France). The urine volume used for GC-MS was adjusted according to the urinary creatinine concentration as follows: for creatinine concentrations lower than 1 mmol/L, 1 mL of urine was used; for concentrations higher than 5 mmol/L, 0.2 mL of urine was used; and for concentrations between 1-5 mmol/L, the urine volume was calculated to obtain 1 µmole of creatinine in the sample. To monitor the performance

For Peer Review

of data acquisition, the samples were randomized and several samples were injected in duplicate to ensure reproducibility. To estimate the accuracy of the analysis, QC samples were run after every 10 patient samples.

GC-MS experiments

Organic metabolites isolated from urine were subjected to two preanalytical procedures (derivatization with bis(trimethylsilyl)trifluoroacetamide (BSTFA) preceded, or not, by an oximation step ; see below) so as to catch the widest urinary metabolome possible and then injected into a GC-MS apparatus. Compounds were identified semi-quantitatively, and concentrations are expressed relative to the amount of creatinine. 4-Phenylbutyric acid was used as internal standard (3 µL of a 7 mM solution) because it is absent from urine, elutes in the middle of the GC chromatogram and does not co-elute with any of the other metabolites.

Procedure without oximation

Urine samples were acidified with HCl (200 µL of a 2.4 *N* solution) and NaCl was added to facilitate the extraction with an ethylacetate / diethylether mixture (1/1) (3 × 1 mL). After centrifugation, the upper organic layers were pooled and dried under nitrogen at room temperature. Each sample was derivatized by addition of 70 µL of a mixture of BSTFA and trimethylchlorosilane (TMCS) (BSTFA/TMCS: 99/1) in 30 µL of acetonitrile for 40 min at 80°C in a sand bath. The derivatized mixture was transferred to a silanized insert for GC-MS analysis.

Procedure with oximation

For Peer Review

Urine samples were basified with NaOH (100 µL of a 6 *N* solution) and 200 µL of hydroxylamine (7 mg/mL) were added. The mixture was heated for 30 minutes at 60°C, cooled to room temperature and acidified with HCl (400 µL of a 6 *N* solution). NaCl was added until saturation, and each sample was extracted with an ethylacetate / diethylether mixture (1/1) (3 × 1 mL). After centrifugation, the upper organic layers were pooled and dried under nitrogen at room temperature. Each sample was derivatized by addition of 70 µL of a mixture of BSTFA /TMCS (99/1) and 30 µL of acetonitrile for 40 min at 80°C in a sand bath. The derivatized mixture was transferred to a silanized insert for GC-MS analysis.

GC-MS analysis

A Shimadzu GC-MS system (Kyoto, Japan) was used. It is composed of an AOC-20S auto-sampler, an AOC-20i auto-injector, a gas chromatograph 2010 and a QP-2010-Plus mass spectrometer. The derivatized samples (3 µL, split ratio = 10) were separated on a capillary CG column (Phenomenex, Zebron ZB-5, 30 m × 0.25 mm i.d., 0.25 µm film thickness). The oven temperature was set at 80°C for 6 min, ramped to 300°C at 5°C/min and then held for 10 min. Helium was used as the carrier gas and set at 0.45 mL/min. The injection port, ion source and interface temperature were 250°C, 250°C and 300°C, respectively. The mass spectra of all GC peaks were generated by electronic impact (EI) at 70 eV and recorded in a positive total ion monitoring mode scanning the 50-500 *m/z* range (event time = 0.1, scan speed = 5000).

Data preprocessing

Each chromatogram obtained was processed for smoothing, library matching and area calculation using an identical data processing method created using the GC-MS Solution Postrun

For Peer Review

Analysis® software (Shimadzu, Japan) (Autoarea mode, maximum peak number = 300, width time = 2 s, smoothing method = standard). Only peaks with minimum peak area = 50 000 were selected for further analysis. The area of each peak was calculated using a unique quantifier ion mass when its relative qualifier ion mass intensity was within 20% range ratio. To minimize process errors, we manually checked each integrated peak for each sample.

The resulting data table was exported to Excel for normalization to the internal standard (4-phenylbutyric acid) and then to the creatinine concentration. The normalization to the internal standard was performed by dividing the integrated area foreach analyte by the integrated area of the internal standard. The results are expressed as ratios to the urinary creatinine concentration (/mmole creatinine).

Data processing - Peak Identification

Compounds were identified from their electron impact mass spectra by comparison to the NIST spectral mass library (NIST 05). Mass spectra of unknown compounds were labeled according to the retention times as part of the identification, one quantifier peak and one or two fragmental qualifier peaks. There is currently no commercially available reference standard for silyzed metabolites to determine the derivatization recovery, so we checked the reproducibility of the quantitative results without knowing the derivatization efficiency.

Statistical Methods.

The intensity of all peaks for all urine samples were studied by multivariate statistical methods, following published protocols [17,18].

For Peer Review

Multivariate analysis was performed using Simca-P+-12 software (version 12.0, Umetrics, Umeå, Sweden). Unit variance (UV) scaling means that the variable is centered and scaled from the standard deviation of the variable. With pareto scaling (Par), obtained by dividing each variable by the square root of its standard deviation, the variance differs between variables, but the range of variance across each spectrum is much smaller than that for the initial unscaled data (small values being scaled up and large values being scaled down). The combination of scaling and mean centering is termed autoscaling in MetaboAnalyst software [each descriptor (of high or low intensity) is weighted equally] [19].A logarithmic transformation (which is a nonlinear conversion) such as an appropriate variance-stabilizing transformation, can be performed to minimize the effects of noise or high variability of the variables [20]. Principal component analysis (PCA) [21] was first performed as unsupervised clustering to identify the similarity or the differences between sample profiles. Grouping, trends and outliers were revealed from the scatter plot. To identify subsets (linear combinations) of metabolic features associated with a specific sample class (ASD or control), partial least squares (PLS) analysis was used as supervised clustering. PLS derives latent variables which describe the maximum proportion of covariance between measured data (X matrix) and the response variable (Y matrix) [22]. Orthogonal partial least squares discriminant analysis (OPLS-DA), also used for discrimination, is a refinement of this approach: variation in the data measured is partitioned into two blocks, one containing variations that correlate with the class identifier (ASD or control) and the other containing variations that are orthogonal to the first block and thus, do not contribute to discrimination between the defined groups [23]. Discriminant metabolites were proposed by OPLS-DA from one predictive and two or more orthogonal components. The quality of the models was described by the cumulative modeled variation in the X matrix R^2X(cum), the

For Peer Review

cumulative modeled variation in the Y matrix R^2Y(cum), where R^2Y(cum) is defined as the proportion of variance in the data explained by the models and indicates goodness of fit, and the cross-validated predictive ability Q^2(cum) values. Models were rejected if they presented complete overlap of Q^2 distributions [Q^2(cum) < 0] or low classification rates [Q^2(cum)< 0.05 and eigenvalues should be > 2]. The features with variable importance on projection (VIP) values and regression coefficients $|$CoeffCS[1]$|$ lower than 0.35 were deleted and evaluated again. A number of variables was identified from PLS-DA and OPLS-DA as being responsible for the difference between ASD and control urine samples with a VIP value > 1.0.

One of the main problems with PLS-DA is the data overfitting that can occur if the algorithm picks up random noise as real signals. To validate the model, the data are divided into seven parts: a model was built with 6/7th of data left in, and the left out data were predicted [24]. The predicted data are then compared with the original data and the sum of squared errors calculated for the whole dataset. This was converted into Q^2, which is an estimation of the predictive ability of the model. The model was thus considered sufficiently well guarded against overfitting and validated after 200 random permutation tests [25] as the Q^2 line intercepted the Y axis at a negative value.

We also performed PLS-DA with autoscaling normalization which is very similar to pareto scaling using the freely available web-based software called MetaboAnalyst [26].

Univariate analysis (Student's *t*-test) was performed using MetaboAnalyst [26] for all metabolites with a VIP >1.0. The ratio of the peak areas of these metabolites to that of creatinine was calculated and a non-parametric test was performed with the critical *p*-value set at 0.05 [27,18].

For Peer Review

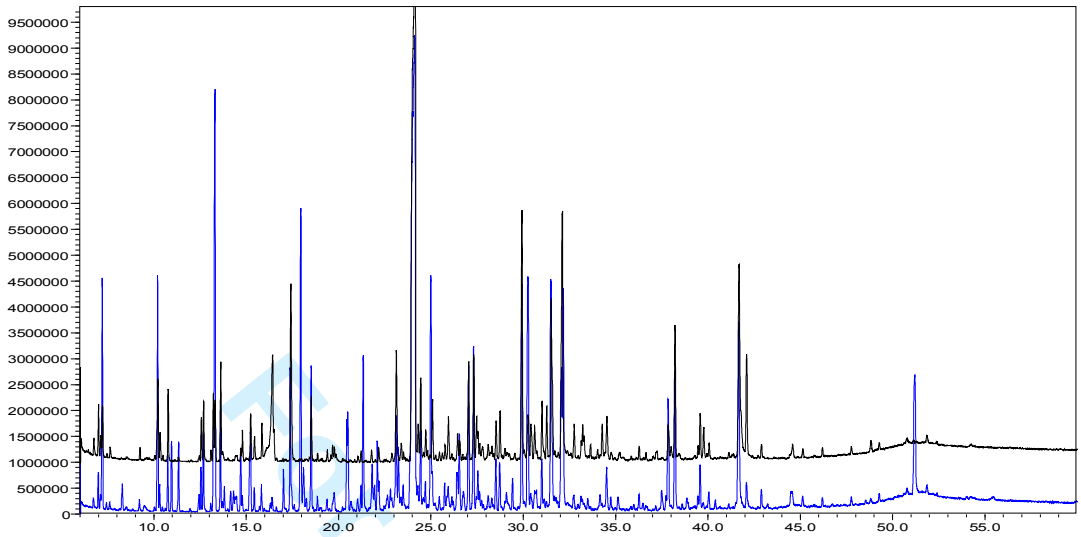
Results

We used two preanalytical preparations methods: a simple liq./liq. separation, followed by trimethylsilylation process (TMS); and oximation (with hydroxylamine) before trimethylsilylation derivatization (TMSO). All samples were subjected to a single GC-MS run, and explored on the basis of the total ion current (TIC) and one quantifier fragment ion as responses (see Table 3).

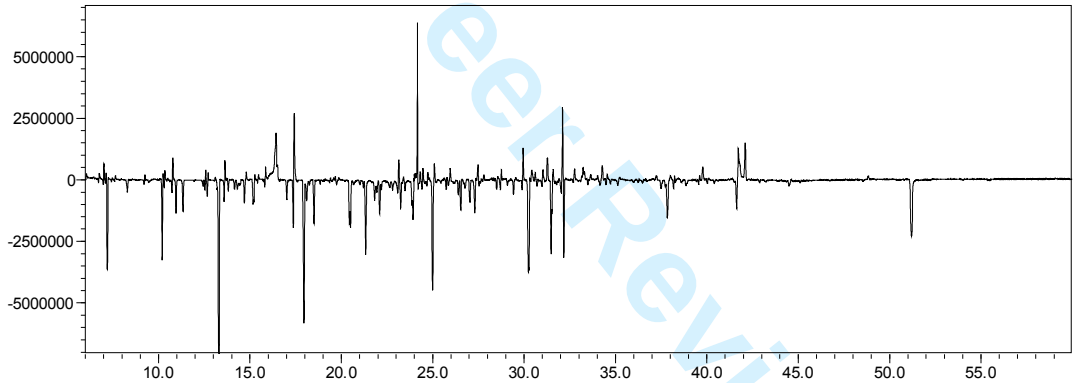
GC-MS data

Representative GC-MS TIC chromatograms of urine samples is displayed in Fig 1. We focused on 56 signals obtained using the method without oximation: 37 of them could be identified by comparison of retentions times and fragmentation patterns with the GC-MS software library. For samples subjected to oximation, we obtained 76 signals and 36 of them could be identified.

For Peer Review



a)



b)

Fig. 1 a) Typical urine GC-MS spectra obtained for urine from ASD (blue) and healthy control (black) children; b) Subtraction of these two chromatograms revealing differential metabolite profiles.

For Peer Review

Statistical studies – choice of pre-treatment technique

Both unsupervised and supervised statistical analyses were done for the results from samples subjected to both pre analytical techniques to assess the possibility of discriminating between control and ASD children using urinary metabolite profiling

Without oximation

PCA analysis did not identify any particular similarity or large differences between sample profiles, such that there were no identified outliers. Potentially discriminant metabolites were identified by PLS-DA and OPLS-DA. The impact of data pretreatment (scaling, transformation) before multivariate analysis was assessed (see statistical values shown in Table 1).

Table 1 Summary of statistical values of PLS-DA and OPLS-DA with different scaling methods for data obtained from GC-MS analyses, without oximation. The different cumulated modeled variations in X [R^2X(cum)] and Y [R^2Y(cum)] matrix on spectral datasets and predictability of the model [Q^2(cum)] are given [observations (N)=50]

	Scaling/transformation	R^2**X**	R^2**X (cum)**	R^2**Y (cum)**	Q^2 **(cum)**
Model 1 [a]	UV	0.21	0.427	0.608	0.443
Model 2 [b]	UV	0.0648	0.469	0.792	0.541
Model 3 [c]	Pareto	0.173	0.83	0.591	0.333
Model 4 [d]	log transformed, UV scaling	0.080	0.753	0.654	0.422

[a] PLS-DA, from 2 components, Variables X=12
[b] OPLS-DA, Variables X=28, 2 orthogonal projections
[c] OPLS-DA, Variables X=11, 2 orthogonal projections
[d] OPLS-DA, Variables X=34, 2 orthogonal projections

From the predictive variation between X (metabolites) and Y (urine samples) given by R^2X(cum), the models 1 and 2 with the same scaling (UV) interpreted around 40% of the total

For Peer Review

variation in X (0.42 and 0.47, respectively, Table 1). The part of the variation that could not be explained by the model might originate from the noise or high variability of the variables [20]. This variation was minimized by Pareto scaling (model 3) and is expressed by the formula [28]: 1 - R^2X - R^2X(cum) = 1 - 0.173 - 0.83 = 0.003, as the noise could account for less 1% in this analysis. The quality of the models is expressed by R^2Y(cum) and Q^2(cum) values. UV scaling (model 2, Table 1), explained 79% of the variations in the various peaks, whereas with Pareto scaling R^2Y(cum) = 0.59. With non linear transformation (model 4, Table 1), by log transformation minimizing the effects of noise or high variability of the variables [R^2X(cum) = 0.75)], the model explained 65% of the variance in the data, but the value was lower than by UV scaling where R^2Y(cum) = 0.79 (model 2, Table 1). The high Q^2(cum) value [Q^2(cum) > 0.5] indicated good predictivity. Pareto scaling or log transformation led to lower predictability [Q^2(cum) = 0.33 and 0.42, respectively] (Table 1). As UV scaling seemed to be the best scaling method in our study investigations [(confirmed by analysis of variance CV-ANOVA with the lower value of p-value = 4.4 e^{-6})], it was used for the subsequent investigations. The OPLS-DA cross-validated score scatter plots for model 2 [Fig. 2, (a)] showed good discrimination between the two populations.

With oximation

OPLS-DA models explained 83% (Table 2) of the variance of the data obtained after oximation-silylation-based GC-MS compared to 79% [R^2Y(cum)= 0.79] of that obtained without oximation. The cross-validated predictive ability Q^2(cum) values were also better after oximation (0.68 compared to 0.54, respectively).

For Peer Review

Table 2 Summary of statistical values of OPLS-DA UV scaling obtained for data from GC-MS analyses, from TMSO procedure (model 5) and, without and with oximation (TMS + TMSO, model 6). The different cumulated modeled variations in X [R^2X(cum)] and Y [R^2Y(cum)] matrix on spectral datasets and predictive power of the model [Q^2(cum)] are given [observations (N)=50]

	R^2X	R^2X (cum)	R^2Y (cum)	Q^2 (cum)
Model 5 [a]	0.0817	0.385	0.826	0.679
Model 6 [b]	0.0798	0.446	0.973	0.878

[a] OPLS-DA, UV scaling, variables X=39, 2 orthogonal projections
[b] OPLS-DA, UV scaling, variables X=42, 4 orthogonal projections

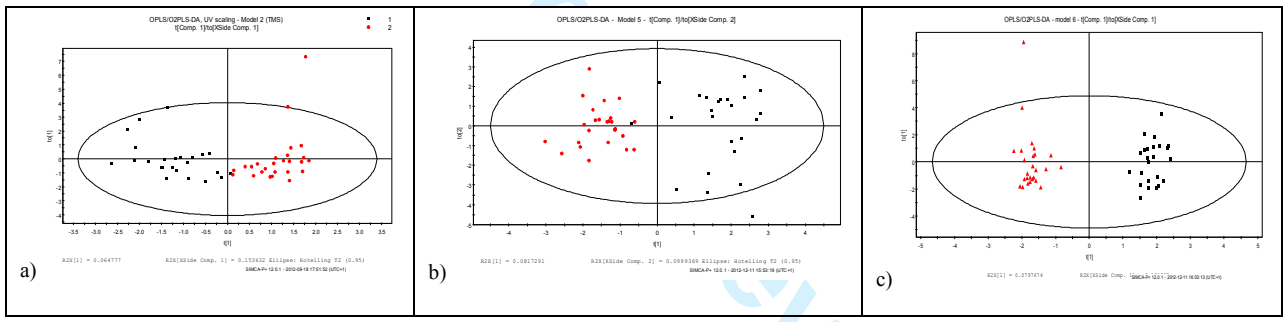


Fig. 2 OPLS-DA, UV-scaled, score scatter plots obtained from of GC-MS analysis of samples without oximation (a), with oximation (b), and with and without oximation (c) from urine from autistic (red dot) and control (black box) children, showing that the two populations are well separated with (a) a R^2Y(cum)=0.79 and a Q^2(cum)=0.54 (model 2, Table 1), with (b) a R^2Y(cum)=0.83 and a Q^2(cum) =0.68 (model 5, Table 2) and (c) a R^2Y(cum)=0.97 and a Q^2(cum)=0.88 (model 6, Table 2).

Combination of data obtained by both TMS and TMSO derivatizations

We also tested the hypothesis that a better statistical model could be obtained by combining data for samples subjected to the TMS process (56 metabolites) with those obtained from the oximation+silylation process (TMSO, 76 metabolites) into a single matrix. About 20% of the metabolites identified were detected in both sample sets and in these cases, only data obtained

For Peer Review

without an oximation process were included in the analysis (data obtained from samples undergoing less manipulation were expected to be more repeatable). Consequently, 117 signals, 58 of them identified, were used.

Multivariate analysis of the two sets of data combined showed that OPLS-DA (Table 2, model 6) led to a very good discrimination between the two populations [0.45 for R^2X(cum) and 0.97 for R^2Y(cum)]; this was confirmed by analysis of variance CV-ANOVA (p-value = 9.6 e^{-15}, see supporting information), and good predictive ability value with Q^2(cum) = 0.88. Using the cumulated data (that for samples with or without an oximation process), the multivariate analysis clearly gave more discriminatory results than obtained using data from samples subjected to just one pre-analytical chemical process. The OPLS-DA cross-validated score plots for model 6 [Fig. 2, (c)] showed excellent discrimination between the two populations. The loading scatter plot (Figure 3) shows which variables expressed similarity between ASD and control children in model 6.

For Peer Review

Table 3 Study of both analyses without and with an oximation process

Putative Assigment [a]	Quantifier ion mass	Retention time (minutes)	Qualifier ion mass		Differentiation for autistic samples	OPLS-DA-VIP values [b] Model 6
Z18	219.1	23.91	189.1		↓	2.00 *(2.10^{-5})*
Hippurate-Ox	236.1	32.06	206		↓	1.65 *(0.0008)*
Succinate	147	18.60	246.9		↑	1.53 *(0.002)*
Z26	267.15	30.60	341.15		↓	1.48 *(0.003)*
3-Hydroxyphenylacetate	295.9	26.63	164	280.8	↓	1.47 *(0.003)*
Vanillylhydracrylate-Ox	297.15	32.75	371.15		↓	1.42 *(0.004)*
Z21	170	25.93	122.1		↓	1.35 *(0.007)*
3-Hydroxy-hippurate-Ox	294.15	39.77	193.1		↓	1.32 *(0.008)*
***p*-Hydroxy mandelate**	266.9	30.72	267.9	341.8	↓	1.30 *(0.010)*
1*H*-Indole-3-acetate	201.9	34.65	73	319	↓	1.24 *(0.014)*
Phosphate	298.8	17.42	299.8	313.8	↓	1.18 *(0.019)*
X16	288.9	24.81	125	147	↓	1.17 *(0.021)*
Palmitate-Ox	313	36.26	145		↓	1.16 *(0.021)*
Stearate-Ox	341	40.06	117		↓	1.15 *(0.023)*
3-Methyladipate-Ox	199.1	24.70	186.15		↓	1.15 *(0.005)*
Z41	338	33.65	323		↓	1.11 *(0.029)*
Glycolate	218	12.55	190.10		↑	1.07 *(0.046)*
Z5	238	14.68	208.05		↑	1.06 *(0.016)*
Z37	324	27.54	309	324	↑	1.02 *(0.046)*

[a] Metabolites characterized after oximation are marked with the suffix –Ox. "X" compounds are TMS derivatives, "Z" compounds are TMSO derivatives.

For Peer Review

[b] Magnitude of variation of variable importance in the projection (VIP) with a threshold of 1.0 obtained using Simca-P^{+} software.
[c] *p*-values for the *t*-test were calculated with MetaboAnalyst software for these 19 metabolites. (↑) denotes higher and (↓) a lower concentration for the ASD population than contols.

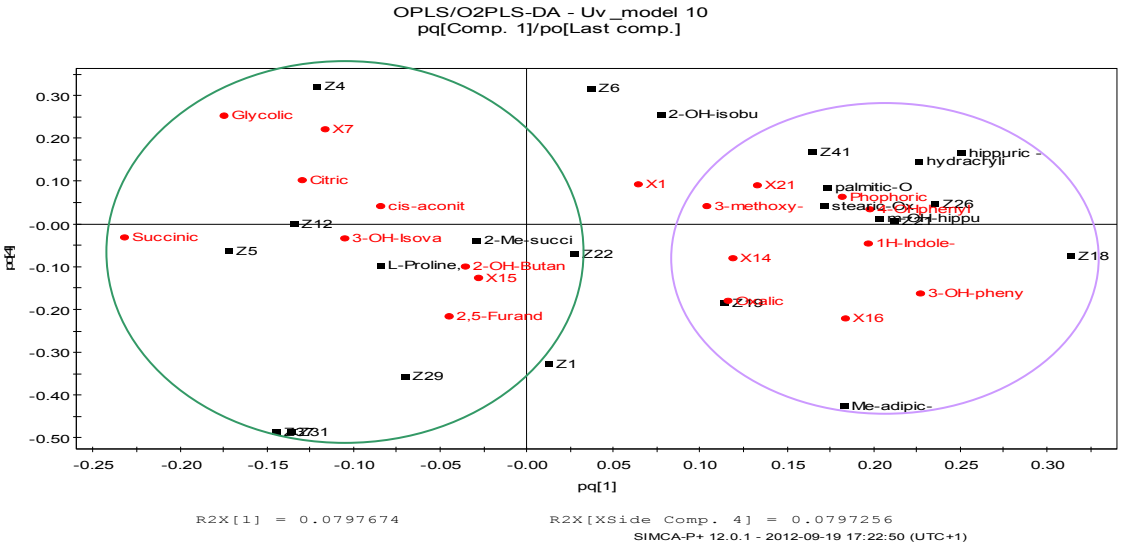


Fig. 3 pq loading plot of the OPLS-DA model 6 (Table 2): scatter plot of the X- and Y-loadings (p and q). This plot shows how the responses (Y's) varied in relation to each other, i.e. which provided similar information and their relationship to the terms of the model. Two tendencies can be seen. Peaks framed in pink were lower in ASD urine, and peaks framed in green were higher. Peaks in red were obtained without (TMS), and peaks in black with (TMSO) the oximation process.

The metabolites making the greatest contribution to the discrimination between the two populations were screened according to the variable importance on projection (VIP) values >1.0. The concentrations of all the nineteen metabolites (Table 3) were found to be significantly different ($p < 0.05$) between the two groups.

Discussion

Statistical studies – choice of pre-treatment technique

OPLS-DA is recommended as it allows a clearer and more straightforward interpretation than other statistical methods [19], so we focused on this approach. Variation between samples can generally be classified into "technical" or "biological" variance; the impact of technical variability should be minimized and useful biological data needs to be discrimination from noise. Data processing methods are highly dependent on the pre-treatment technique used for the data. Several types of scaling are commonly used: UV, Par, and autoscaling. The choice of pre-treatment methods depends on several factors (numbers of samples, magnitude of concentration, similarities, etc), and different methods emphasize different aspect of the data. As each method has its own merits and drawback, and as, to our knowledge, there is non consensus in the pre-treatment for urines analyzed by GC-MS, we studied the effect of all these scaling methods on identification of biomarkers.

We found that noise (and/or high variability of the variables) was minimized by Pareto scaling: (R^2X(cum) was highest in model 3, but it led to lower predictability [Q^2(cum) = 0.23] (Table 1). UV scaling seemed to be the best scaling method for our data set (confirmed by analysis of variance CV-ANOVA, see supporting information). We therefore focused on results obtained from OPLS-DA using UV scaling (model 6, Tables 2 and 3).

Statistical analysis of TMS and TMSO results – Analysis of models and identification of important features

Factors such as disease, drugs and diet modify the concentrations of individual metabolites [29,30]. It is known that autistic patients display dysfuntions in the levels of hormones, peptides, metabolites associated with neurological, gastrointestinal, immunological and toxicological effects (for review see [31]). Although the number of samples included in this study was too

For Peer Review

small for powerful statistical analysis, we applied a metabolomics strategy to screen for metabolites potentially associated with a pathological context, ASD. Few targeted or untargeted metabolomics studies by MS have been performed on urine samples from ASD patients ([32], and for review see ref [31,15]). For example, the targeted studies include evaluation of tartaric acid concentrations [33,34], homocysteine levels [35], and aminoacid excretions [36,37] in the urine of autistic and healthy children. Fewer than one hundred metabolites, including amino acids,have been screened in untargeted studies. To expand the number of metabolites detected by GC-MS, we tested two types of pre-analytical chemical treatment. As described in previously study (without an oximation step before derivatization) [32], we found higher citrate concentrations in the ASD than control group (Table 3). However, isocitric acid, 2-oxoglutaric acid, adipic acid, suberic acid (all rejected in the different models), 4-hydroxyphenylacetic acid (0.6<VIP<1) and 4-hydroxybenzoic acid (VIP<0.5) were not found to provide any discrimination between our two groups (see supporting info). Furthermore, we found significantly lower concentrations of hippurate (p=0.002, Table 3), contrasting with a previous report [32], the same for *m*-hydroxyhippurate. The succinate concentration was found to be higher, consistent with a previous ^{1}H-NMR study [14]. Another targeted GC-MS study [38] reported higher than control levels of homovanillic acid (HVA) and vanillylmandelic acid (VMA) in the urine of autistic children. In our study, no significant differences in the level of HVA (observed as a TMSO-derivative, VIP values in OPLS-DA were lower than 1, and the p-value=0.1) or VMA (observed as a TMS-derivative as the previous study, p-value=0.8) were detected.

The discriminant urinary metabolites were vanillylhydracrylate, 3-methyladipate, *p*-hydroxymandelate**,** glycolate, palmitate, stearate, succinate, phosphate, hippurate, 3-hydroxyphenylacetate, 3-hydroxyhippurate, 1*H*-indole-3-acetate (these last metabolites could be

For Peer Review

implicated in microbial pathways associated with gut bacterial), and also eight metabolites identified only according to their retention time and fragmentation patterns (Table 3).

We compared metabolic fingerprinting obtained by GC-MS after a silylation step (which is the most versatile and universally applicable derivatization method) with that after oximation of keto-derivatives followed by a BSTFA silylation step. This study clearly shows that the two derivatization procedures are complementary, and allow analysis of a wider range of metabolites for metabolomics studies.

Conclusion

Evaluation of urinary metabolite profiles using a GC-MS method showed promising results. First, this study clearly suggest that sample pre-treatment techniques are relevant, and we highlighted complementarities between the two derivatization procedures assessed: TMS derivatization (economic, minimal sample handling suitable for routine analysis), and a TMSO procedure (higher sensitivity). Combined, the two procedures allowed analysis of 132 metabolites as their TMS/TMSO ether/ester derivatives, 73 of which could be identified as known compounds. Secondly, relevant metabolites, with different concentrations in children with or without a diagnosis of ASD were successfully extracted by multivariate analysis. Further work including targeted studies, is needed to identify chemically those features that are potentially discriminant and to validate their clinical value.

Acknowledgements

This work was supported by the “Institut National de la Santé et de la Recherche” INSERM, and the University François-Rabelais. We thank the center “Sésame Autisme Loiret” for participation

For Peer Review

in this study. We thank the "Département d'Analyses Chimiques et S.R.M. Biologique et Médicale" (PPF, Tours, France) for GC-MS analyses.

References

1. Weintraub K (2011) The prevalence puzzle: Autism counts. Nature 479 (7371):22-24

2. Falck-Ytter T, von Hofsten C (2011) How special is social looking in ASD: a review. Prog Brain Res 189:209-222

3. Franke B, Faraone SV, Asherson P, Buitelaar J, Bau CH, Ramos-Quiroga JA, Mick E, Grevet EH, Johansson S, Haavik J, Lesch KP, Cormand B, Reif A (2011) The genetics of attention deficit/hyperactivity disorder in adults, a review. Mol Psychiatry 17:960-987

4. McPartland JC, Coffman M, Pelphrey KA (2011) Recent advances in understanding the neural bases of autism spectrum disorder. Curr Opin Pediatr 23 (6):628-632

5. Schaefer GB, Lutz RE (2006) Diagnostic yield in the clinical genetic evaluation of autism spectrum disorders. Genet Med 8 (9):549-556

6. Kolvin I (1971) Studies in the childhood psychoses. I. Diagnostic criteria and classification. Br J Psychiatry 118 (545):381-384

7. Association AP (2000) Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition: DSM-IV-TR®. Washington

8. Koek M, Jellema R, van der Greef J, Tas A, Hankemeier T (2011) Quantitative metabolomics based on gas chromatography mass spectrometry: status and perspectives. Metabolomics 7 (3):307-328

For Peer Review

9. Madsen R, Lundstedt T, Trygg J (2010) Chemometrics in metabolomics-A review in human disease diagnosis. Analytica Chimica Acta 659 (1-2):23-33

10. Suhre K, Shin SY, Petersen AK, Mohney RP, Meredith D, Wagele B, Altmaier E, Deloukas P, Erdmann J, Grundberg E, Hammond CJ, de Angelis MH, Kastenmuller G, Kottgen A, Kronenberg F, Mangino M, Meisinger C, Meitinger T, Mewes HW, Milburn MV, Prehn C, Raffler J, Ried JS, Romisch-Margl W, Samani NJ, Small KS, Wichmann HE, Zhai G, Illig T, Spector TD, Adamski J, Soranzo N, Gieger C (2011) Human metabolic individuality in biomedical and pharmaceutical research. Nature 477 (7362):54-60

11. Gebregiworgis T, Powers R (2012) Application of NMR metabolomics to search for human disease biomarkers. Comb Chem High Throughput Screen 15 (8):595-610

12. Mercier P, Lewis MJ, Chang D, Baker D, Wishart DS (2011) Towards automatic metabolomic profiling of high-resolution one-dimensional proton NMR spectra. J Biomol NMR 49 (3-4):307-323

13. Garcia A, Barbas C (2011) Gas chromatography-mass spectrometry (GC-MS)-based metabolomics. Methods Mol Biol 708:191-204

14. Yap IK, Angley M, Veselkov KA, Holmes E, Lindon JC, Nicholson JK (2010) Urinary metabolic phenotyping differentiates children with autism from their unaffected siblings and age-matched controls. J Proteome Res 9 (6):2996-3004

15. Wang L, Angley MT, Gerber JP, Sorich MJ (2011) A review of candidate urinary biomarkers for autism spectrum disorder. Biomarkers 16 (7):537-552

16. Organization WH (1991) International Classification of Diseases (ICD-10). World Health Organization, Geneva

For Peer Review

17. Chan EC, Pasikanti KK, Nicholson JK (2011) Global urinary metabolic profiling procedures using gas chromatography-mass spectrometry. Nat Protoc 6 (10):1483-1499

18. Xia J, Wishart DS (2011) Web-based inference of biological patterns, functions and pathways from metabolomic data using MetaboAnalyst. Nat Protoc 6 (6):743-760

19. Trygg J, Holmes E, Lundstedt T (2007) Chemometrics in metabonomics. J Proteome Res 6 (2):469-479

20. van den Berg RA, Hoefsloot HC, Westerhuis JA, Smilde AK, van der Wer MJ (2006) Centering, scaling, and transformations: improving the biological information content of metabolomics data. BMC Genomics 7:142-157

21. Kemsley EK, Le Gall G, Dainty JR, Watson AD, Harvey LJ, Tapp HS, Colquhoun IJ (2007) Multivariate techniques and their application in nutrition: a metabolomics case study. Br J Nutr 98 (1):1-14

22. Wold S, Sjöström M, Eriksson L (2001) PLS-regression: a basic tool of chemometrics. Chemometr Intell Lab 58 (2):109-130

23. Trygg J (2002) O2-PLS for qualitative and quantitative analysis in multivariate calibration. J Chemometr 16 (6):283-293

24. Cloarec O, Dumas ME, Trygg J, Craig A, Barton RH, Lindon JC, Nicholson JK, Holmes E (2004) Evaluation of the Orthogonal Projection on Latent Structure Model Limitations Caused by Chemical Shift Variability and Improved Visualization of Biomarker Changes in ^{1}H NMR Spectroscopic Metabonomic Studies. Anal Chem 77 (2):517-526

25. Westerhuis J, Hoefsloot H, Smit S, Vis D, Smilde A, van Velzen E, van Duijnhoven J, van Dorsten F (2008) Assessment of PLSDA cross validation. Metabolomics 4 (1):81-89

For Peer Review

26. Xia J, Bjorndahl TC, Tang P, Wishart DS (2008) MetaboMiner--semi-automated identification of metabolites from 2D NMR spectra of complex biofluids. BMC Bioinformatics 9:507

27. Xia J, Psychogios N, Young N, Wishart DS (2009) MetaboAnalyst: a web server for metabolomic data analysis and interpretation. Nucleic Acids Res 37 (Web Server issue):W652-660

28. SIMCA-P+ 12 User Guide http://wwwumetricscom/

29. Krug S, Kastenmüller G, Stückler F, Rist MJ, Skurk T, Sailer M, Raffler J, Römisch-Margl W, Adamski J, Prehn C, Frank T, Engel K-H, Hofmann T, Luy B, Zimmermann R, Moritz F, Schmitt-Kopplin P, Krumsiek J, Kremer W, Huber F, Oeh U, Theis FJ, Szymczak W, Hauner H, Suhre K, Daniel H (2012) The dynamic range of the human metabolome revealed by challenges. FASEB J 26 (6):2607-2619

30. Rosenling T, Stoop MP, Attali A, Aken H, Suidgeest E, Christin C, Stingl C, Suits F, Horvatovich P, Hintzen RQ, Tuinstra T, Bischoff R, Luider TM (2012) Profiling and Identification of Cerebrospinal Fluid Proteins in a Rat EAE Model of Multiple Sclerosis. J Proteome Res 11 (4):2048-2060

31. Ratajczak HV (2011) Theoretical aspects of autism: biomarkers--a review. J Immunotoxicol 8 (1):80-94

32. Kaluzna-Czaplinska J (2011) Noninvasive urinary organic acids test to assess biochemical and nutritional individuality in autistic children. Clin Biochem 44 (8-9):686-691

33. Lord RS, Burdette CK, Bralley JA (2005) Significance of urinary tartaric acid. Clin Chem 51 (3):672-673

For Peer Review

34. Shaw W, Kassen E, Chaves E (1995) Increased urinary excretion of analogs of Krebs cycle metabolites and arabinose in two brothers with autistic features. Clin Chem 41 (8):1094-1104

35. Kaluzna-Czaplinska J, Michalska M, Rynkowski J (2011) Homocysteine level in urine of autistic and healthy children. Acta Biochim Pol 58 (1):31-34

36. Adams JB, Audhya T, McDonough-Means S, Rubin RA, Quig D, Geis E, Gehn E, Loresto M, Mitchell J, Atwood S, Barnhouse S, Lee W (2011) Nutritional and metabolic status of children with autism vs. neurotypical children, and the association with autism severity. Nutr Metab 8 (1):34

37. Evans C, Dunstan RH, Rothkirch T, Roberts TK, Reichelt KL, Cosford R, Deed G, Ellis LB, Sparkes DL (2008) Altered amino acid excretion in children with autism. Nutr Neurosci 11 (1):9-17

38. Kaluzna-Czaplinska J, Socha E, Rynkowski J (2010) Determination of homovanillic acid and vanillylmandelic acid in urine of autistic children by gas chromatography/mass spectrometry. Med Sci Monit 16 (9):CR445-450

For Peer Review

List of figures

Fig. 1 a) Typical GC-MS spectra obtained for urine from ASD (blue) and healthy control (black) children; b) Subtraction of these two chromatograms revealing differential metabolite profiles

Fig. 2 OPLS-DA, UV-scaled, score scatter plots obtained from of GC-MS analysis of samples without oximation (a), with oximation (b), and with and without oximation (c) from urine from autistic (red dots) and control (black box) children, showing that the two populations are well separated with (a) a R^2Y(cum)=0.79 and a Q^2(cum)=0.54 (model 2, Table 1), with (b) a R^2Y(cum)=0.86 and a Q^2(cum) =0.73 (model 5, Table 2) and (c) a R^2Y(cum)=0.97 and a Q^2(cum)=0.88 (model 6, Table 2).

Fig. 3 pq loading plot of the OPLS-DA model 6 (Table 2): scatter plot of the X- and Y-loadings (p and q). This plot shows how the responses (Y's) varied in relation to each other, i.e. which provided similar information, and their relationship to the terms of the model. Two tendencies can be seen. Peaks framed in pink were lower in ASD urine, and peaks framed in green were higher. Peaks in red were obtained without, and peaks in black with, the oximation process

For Peer Review

Supporting Information

GC-MS-based urine metabolic profiling of Autism Spectrum Disorders

Patrick Emond[a,b,*] . Sylvie Mavel[a] . Nacima Aïdoud[a] . Lydie Nadal-Desbarats[a,b] . Fréderic Montigny[b] . Frédérique Bonnet-Brilhault[c] . Catherine Barthélémy[c] . Marc Merten[d] . Pierre Sarda[e] . Frédéric Laumonnier[a] . Patrick Vourc'h[a,b] . Hélène Blasco[a] . Christian R Andres[a] .

[a] *Université François-Rabelais, INSERM U930, Equipe neurogénétique et neurométabolomique, CHRU de Tours, 10 Bv Tonnellé, 37044 Tours, France*
[b] *Université François-Rabelais, PPF "Analyses des Systèmes Biologiques", UFR de Médecine, 10 Bv Tonnellé, 37044 Tours, France*
[c] *Université François-Rabelais, INSERM U930, Equipe Autisme, CHRU de Tours, 10 Bv Tonnellé, 37044 Tours, France*
[d] *Université Henri Poincaré, Faculté de Médecine, Laboratoire de Biochimie, 9 Av de la forêt de Haye, 54505 Vandoeuvre les Nancy, France*
[e] *CHRU de Montpellier, Hôpital Arnaud-de-Villeneuve, 34295 Montpellier cedex 5, France*

Fig. S1 Principal component analysis (PCA) score plot of GC-MS	**S2**
Fig. S2 Screenshots of Validation Plots of PLS-DA, UV scaling, model 6 obtained after 200 permutations tests of valid	**S2**
Tables S1, S2, S3 and S4 CV-ANOVA analyses of models 6, 7, 8 and 9	**S3**
Fig. S3 Correlation plot of model 6, OPLS-DA, UV scaling.	**S4**
Table S5 Analysis of urinary metabolites: ions monitored for each metabolites and retention time (RT), variable importance (VIP values) for OPLS-DA UV scaling (model 6), and *p*-value (*t*-test)	**S5-6**
Table S6 Analysis of urinary metabolites in children: ions monitored for each metabolites and retention time (RT), variable importance (VIP values > 1) for OPLS-DA, UV scaling, models 2 and 5	**S7**

For Peer Review

Fig. S1
Principal component analysis (PCA) score plot of GC-MS data obtained without and with oximation process before derivatization

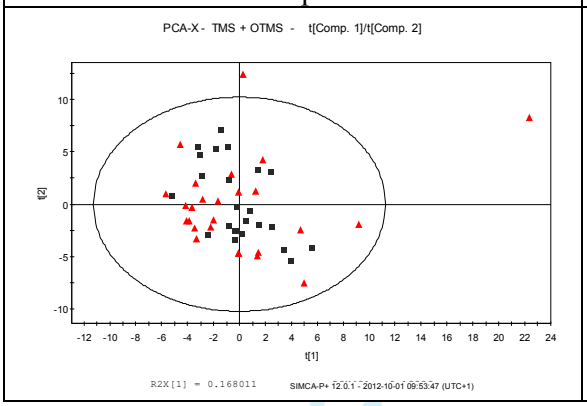
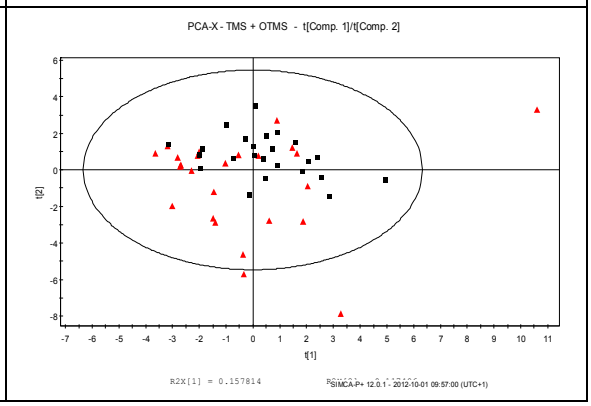
PCA from all the cross peaks 116 variables	PCA from the 42 variables of model 6
	

Fig. S2
Screenshots of Validation Plots of PLS-DA, UV scaling, model 6 obtained after 200 permutations tests of valid

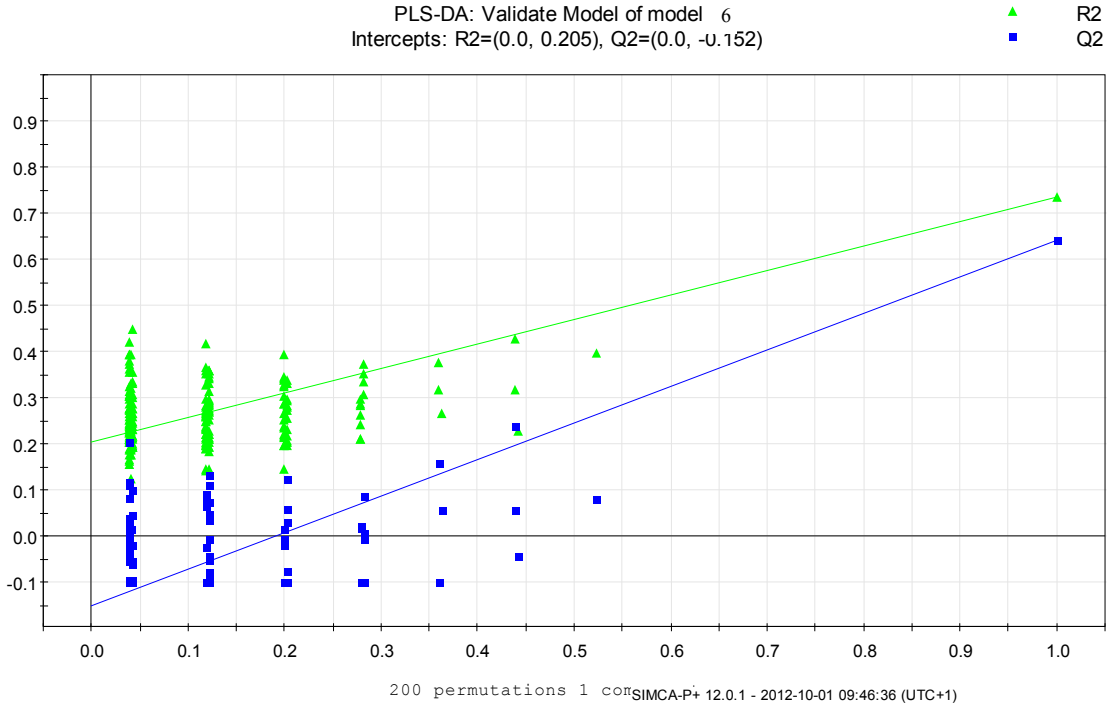


Table S1. TMS and TMSO data

CV-ANOVA analysis of PLS-DA, UV scaling, model 7

Model 7	SS	DF	MS	F	p	SD
Total corr.	49	49	1			1
Regression	31.3777	2	15.6888	41.8432	3.65492e-011	3.96091
Residual	17.6223	47	0.374943			0.612326

Table S2.
CV-ANOVA analysis of model 6, OPLS-DA, UV scaling

Model 6	SS	DF	MS	F	p	SD
Total corr.	49	49	1			1
Regression	43.0028	10	4.30028	27.9646	9.62845e-015	2.07371
Residual	5.99725	39	0.153776			0.392142

Table S3.
CV-ANOVA analysis OPLS-DA, Par scaling, model 8

Model 8	SS	DF	MS	F	p	SD
Total corr.	49	49	1			1
Regression	21.2664	6	3.5444	5.49546	0.000270877	1.88266
Residual	27.7336	43	0.644968			0.803099

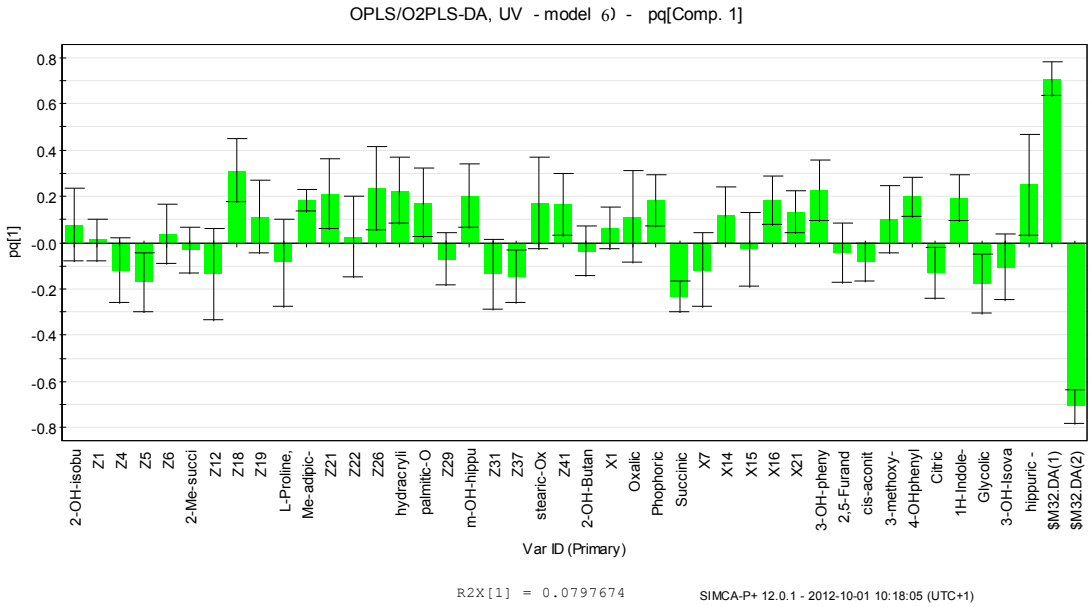
Table S4.
CV-ANOVA analysis of OPLS-DA, log. transformation, UV scaling, model 9

Model 9	SS	DF	MS	F	p	SD
Total corr.	49	49	1			1
Regression	42.2228	8	5.27785	31.9294	3.02587e-015	2.29736
Residual	6.77719	41	0.165297			0.406568

For Peer Review

Fig S3

Correlation plot, cross peaks with negative correlations were in higher concentration in ASD population, with positive correlation in higher concentration in control urines (model 6)



/----------TMSO derivatives----------/ /--------TMS derivatives ----------/

Peer Review

Table S5
Analysis of urinary metabolites in children: ions monitored for each metabolites TMS + OTMS and retention time (RT), variable importance (VIP values) for PLS and OPLS-DA with different scaling/transformation, model 6, and *p*-value (*t*-test)

Putative Assigment [a]	Quantifier ion mass	Retention time minutes	Qualifier ion mass		Differentiation for autistic samples	OPLS-DA-UV-VIP values [b] Model 6	*p* value [c]
Z18	219.1	23.91	189.1		↓	2.00	*1.10^{-4}*
Hippurate-Ox	236.1	32.06	206		↓	1.65	*0.002*
Succinate	147	18.60	246.9		↑	1.53	*0.002*
Z26	267.15	30.60	341.15		↓	1.48	*0.030*
3-Hydroxyphenylacetate	295.9	26.63	164	280.8	↓	1.47	*0.003*
Vanillylhydracrylate-Ox	297.15	32.75	371.15		↓	1.42	*0.017*
Z21	170	25.93	122.1		↓	1.35	
3-Hydroxyhippurate-Ox	294.15	39.77	193.1		↓	1.32	*0.012*
***p*-Hydroxy mandelate**	266.9	30.72	267.9	341.8	↓	1.30	*0.010*
1*H*-Indole-3-acetate	201.9	34.65	73	319	↓	1.24	*0.014*
Phosphate	298.8	17.42	299.8	313.8	↓	1.18	*0.019*
X16	288.9	24.81	125	147	↓	1.17	*0.021*
Palmitate-Ox	313	36.26	145		↓	1.16	
Stearate-Ox	341	40.06	117		↓	1.15	*0.032*
3-Methyladipate-Ox	199.1	24.70	186.15		↓	1.15	*0.005*
Z41	338	33.65	323		↓	1.11	
Glycolic	218	12.55	190.10		↑	1.07	*0.046*
Z5	238	14.68	208.05		↑	1.06	*0.016*
Z37	324	27.54	309	324	↑	1.02	*0.036*
Z31	234	19.77	219		↑	0.98	
Citrate	272.9	31.63	374.8	464.7	↑	0.82	
β-Lactate	147	10.29	117	190.9	↑	-	
4-Hydroxyphenylacetate	295.9	27.41	252	280.9	↓	-	
X21	157	26.03	147	75	↓	0.87	
X14	221.1	24	146.9	103	↓	0.82	
Z12	252.05	22.17	191.10		↑	0.82	
Z4	221.10	14.46	187.10		↑	0.77	
Z19	131.15	24.09	117.10		↓	0.77	
Oxalate	73	12.75	147	190	↓	0.72	
X7	219	21.50	103	72.90	↑	0.71	
3-Methoxy-4-hydroxy phenyl acetate	325.90	30.51	208.90	266.9	↓	0.64	

For Peer Review

Z29	118.10	37.87	91.10		↑	0.52	
Cis-aconitate	374.80	30.02	228.9	284.8	↑	0.51	
2-Hydroxy-isobutyrate-Ox	205.15	10.31	131.15		↓	0.50	
L-Proline, 5-oxo-Ox	258.1	24.28	230		↑	0.48	
X1	244.7	12.52	142.9	146.9	↓	0.42	
Z6	103.1	17.14	147.1		↓	0.32	
2,5-Furandicarboxylate	284.8	27.55	147	73	↑	0.32	
2-Hydroxybutanoate	131.1	10.41	142.9	146.9	↑	0.28	
X15	317.4	24.19	217		↑	0.22	
2-Me-succinate-Ox	261.10	27.74	202.1		↑	0.16	
Z22	226.1	27.74	202.1		↓	0.14	
Z1	221.1	10.45	133.1		↓	0.03	

[a] Metabolites characterized after oximation were marked by the suffix –Ox. "X" compounds were TMS derivatives, "Z" compounds were TMSO derivatives.
[b] Magnitude of variation of variable importance in the projection (VIP) with a threshold of 1.0 obtained using Simca-P+ software.
[c] *p*-values from *t*-test were obtained from MetaboAnalyst software.
(↑) denotes an increased concentration for ASD population, (↓): decreased
(-) denotes not present in the model

For Peer Review

Table S6
Analysis of urinary metabolites in children: ions monitored for each metabolites and retention time (RT), variable importance (VIP values > 1) for OPLS-DA, UV scaling, models 2 and 5

Without oximation: TMS process							**With oximation: TMSO process**					
Putative Assigment	Quantifier ion mass	Retention time minutes	Qualifier ion mass		Differentiation for autistic samples	VIP coeff	**Putative Assigment**	Quantifier ion mass	Retention time minutes	Qualifier ion mass	Differentiation for autistic samples	VIP coeff
Succinic	147	18.60	246.9		↑	1.90	**Z18**	219.1	23.91	189.1	↓	2.17
3-OH-phenylacetic	295.9	26.63	164	280.8	↓	1.83	**Hippuric**	236.1	32.06	206	↓	1.78
4-OHphenyl-2-OH-acetic	266.9	30.72	267.9	341.8	↓	1.61	**Z26**	267.15	30.60	341.15	↓	1.60
1H-Indole-3-acetic acid	201.9	34.65	73	319	↓	1.57	**Vanillylhydracrylate**	297.15	32.75	371.15	↓	1.54
Phophoric	298.8	17.42	299.8	313.8	↓	1.47	**Z21**	170	25.93	122.1	↓	1.46
X16	288.9	24.81	125	147	↓	1.44	***m*-OH-hippurate**	294.15	39.77	193.1	↓	1.43
Glycolic	218	12.55	190.10		↑	1.33	**Palmitic**	313.25	36.26	145.15	↓	1.26
X21	157	26.03	147	75	↓	1.09	**Stearic**	341	40.06	117	↓	1.25
X20	348.8	25.88	129.1	246.9	↑	1.09	**Z41**	338	33.65	323	↓	1.20
X14	189	24.01	146.9	103	↓	1.04	**Z5**	238.05	14.68	208.05	↑	1.15
Citric	272.9	31.63	374.8	464.7	↑	1.02	**Z37**	324	27.54	309	↑	1.10
X7	219	21.50	103	72.9	↑	0.90	**Z31**	234	19.77	219	↑	1.06
Oxalic	73	12.73	147	190	↓	0.90	**Z34**	220	22.89	205	↑	1.03
3-OH-isovaleric						0.87	**Z36**	293	23.91	251	↑	1.03
3-methoxy-4-OHphenyl-acetic	325.9	30.51	208.9	266.9	↓	0.80	**Z9**	73.05	20.90	131.15	↑	0.97
…							**Z12**	252.05	22.17	191.1	↑	0.89
***p*-cresol**						0.48	**Succinic**	247.1	18.50	172.05	↑	0.85

For Peer Review

For Peer Review