



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:	<i>Analytical and Bioanalytical Chemistry</i>
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

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

1
2
3 ASD autism spectrum disorders

4
5 BSTFA bis(trimethylsilyl)trifluoroacetamide

6
7 GC-MS gas chromatography combined with mass spectroscopy

8
9 NMR nuclear magnetic resonance

10
11 OPLS-DA orthogonal partial least-squares discriminant analysis

12
13 Par pareto

14
15 PCA principal component analysis

16
17 PLS-DA partial least squares discriminant analysis

18
19 TMS trimethylsilylated derivative

20
21 TMSO trimethylsilylated and oximated derivative

22
23 UV unit variance

24
25 VIP variable importance on projection

26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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

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,

1
2
3 but many structures contain polar groups and therefore need to be chemically derivatized prior to
4 GC-MS analysis. The most commonly used derivatization methods involve silylation preceded
5 or not by oximation [8]. These derivatization strategies give an access to a large set of chemical
6 functions, including alcohol and carboxylic functions, amines, amides and aldehydes [8].
7
8
9

10
11
12 Several bio-fluids can be analyzed by GC-MS, of which urine has the advantages of
13 being easily and non-invasively accessible and containing hundreds of metabolites that represent
14 the endpoint of endogenous metabolism. GC-MS has been used for the description of the
15 metabolic status of patients with neuropsychiatric disorders such as autism ([14], for review, see
16 ref: [15]). In particular, gastrointestinal metabolites, abnormal neurotransmitter concentrations,
17 the creatine to creatinine ratio, and the guanidine acetate concentrations have been reported to be
18 discriminative markers. However, there is currently no evidence that any biomarker is useful in
19 routine practice, and consequently further studies are required to identify clinically useful
20 indicators [15].
21
22
23
24
25
26
27
28
29
30
31
32
33

34 We report an analysis of the urinary metabolic signatures of autism patients and healthy
35 subjects with the aim of identifying potential biomarkers of ASD. Two techniques were
36 compared, with and without oximation during the derivatization process before GC-MS.
37 Statistical analysis methods [partial least squares discriminant analysis (PLS-DA) and orthogonal
38 partial least squares discriminant analysis (OPLS-DA)] were used to identify metabolites that
39 discriminate between ASD and control populations.
40
41
42
43
44
45
46
47
48
49

50 **Materials and methods**

51 **Sample collection**

52
53
54
55
56
57
58
59
60

1
2
3 Between 2008 and 2010, urine samples were collected from children aged 6-14 years (mean 8
4 years) with ASD living in France (n = 26; male 85%, female 15%) and from control children (n =
5 24; male 67%, female 33%) aged 6-9 years (mean 7 years). Urine samples were collected into
6 untreated vials during routine medical consultations, principally in the morning and the exact
7 time of collection was recorded. Each urine sample was aliquoted into 1.5 mL Eppendorf tubes
8 and stored at -80°C immediately after collection until analysis. All study participants and their
9 parents or guardians provided informed consent.
10
11
12
13
14
15
16
17
18
19
20
21

22 Patients

23
24 The following data were collected for all autistic and control children: age, gender, treatment,
25 and clinical characteristics. The severity of autism was assessed according to the International
26 Classification of Diseases 10th Edition [16]. Autistic patients were recruited in three French
27 autism centers [Tours (n = 8), Montpellier (n = 13), and Orléans (n = 5)] and all twenty-four
28 urine samples were from healthy volunteers from Tours.
29
30
31
32
33
34
35
36
37
38

39 GC-MS study

40 *Sample preparation*

41
42 Samples were thawed at room temperature, centrifuged (at 3000 g) for 10 min and an aliquot was
43 used for creatinine analysis (Jaffé method, Olympus AU640, France). The urine volume used for
44 GC-MS was adjusted according to the urinary creatinine concentration as follows: for creatinine
45 concentrations lower than 1 mmol/L, 1 mL of urine was used; for concentrations higher than 5
46 mmol/L, 0.2 mL of urine was used; and for concentrations between 1-5 mmol/L, the urine
47 volume was calculated to obtain 1 μ mol of creatinine in the sample. To monitor the performance
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 of data acquisition, the samples were randomized and several samples were injected in duplicate
4
5 to ensure reproducibility. To estimate the accuracy of the analysis, QC samples were run after
6
7 every 10 patient samples.
8
9

10 11 12 *GC-MS experiments*

13
14
15 Organic metabolites isolated from urine were subjected to two preanalytical procedures
16
17 (derivatization with bis(trimethylsilyl)trifluoroacetamide (BSTFA) preceded, or not, by an
18
19 oximation step ; see below) so as to catch the widest urinary metabolome possible and then
20
21 injected into a GC-MS apparatus. Compounds were identified semi-quantitatively, and
22
23 concentrations are expressed relative to the amount of creatinine. 4-Phenylbutyric acid was used
24
25 as internal standard (3 μL of a 7 mM solution) because it is absent from urine, elutes in the
26
27 middle of the GC chromatogram and does not co-elute with any of the other metabolites.
28
29
30
31
32
33

34 *Procedure without oximation*

35
36 Urine samples were acidified with HCl (200 μL of a 2.4 N solution) and NaCl was added to
37
38 facilitate the extraction with an ethylacetate / diethylether mixture (1/1) ($3 \times 1 \text{ mL}$). After
39
40 centrifugation, the upper organic layers were pooled and dried under nitrogen at room
41
42 temperature. Each sample was derivatized by addition of 70 μL of a mixture of BSTFA and
43
44 trimethylchlorosilane (TMCS) (BSTFA/TMCS: 99/1) in 30 μL of acetonitrile for 40 min at 80°C
45
46 in a sand bath. The derivatized mixture was transferred to a silanized insert for GC-MS analysis.
47
48
49
50
51
52

53 *Procedure with oximation*

1
2
3 Urine samples were basified with NaOH (100 μ L of a 6 N solution) and 200 μ L of
4 hydroxylamine (7 mg/mL) were added. The mixture was heated for 30 minutes at 60°C, cooled
5 to room temperature and acidified with HCl (400 μ L of a 6 N solution). NaCl was added until
6 saturation, and each sample was extracted with an ethylacetate / diethylether mixture (1/1) (3 \times 1
7 mL). After centrifugation, the upper organic layers were pooled and dried under nitrogen at room
8 temperature. Each sample was derivatized by addition of 70 μ L of a mixture of BSTFA /TMCS
9 (99/1) and 30 μ L of acetonitrile for 40 min at 80°C in a sand bath. The derivatized mixture was
10 transferred to a silanized insert for GC-MS analysis.
11
12
13
14
15
16
17
18
19
20
21
22
23

24 *GC-MS analysis*

25
26 A Shimadzu GC-MS system (Kyoto, Japan) was used. It is composed of an AOC-20S auto-
27 sampler, an AOC-20i auto-injector, a gas chromatograph 2010 and a QP-2010-Plus mass
28 spectrometer. The derivatized samples (3 μ L, split ratio = 10) were separated on a capillary CG
29 column (Phenomenex, Zebron ZB-5, 30 m \times 0.25 mm i.d., 0.25 μ m film thickness). The oven
30 temperature was set at 80°C for 6 min, ramped to 300°C at 5°C/min and then held for 10 min.
31 Helium was used as the carrier gas and set at 0.45 mL/min. The injection port, ion source and
32 interface temperature were 250°C, 250°C and 300°C, respectively. The mass spectra of all GC
33 peaks were generated by electronic impact (EI) at 70 eV and recorded in a positive total ion
34 monitoring mode scanning the 50-500 m/z range (event time = 0.1, scan speed = 5000).
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49

50 *Data preprocessing*

51
52 Each chromatogram obtained was processed for smoothing, library matching and area
53 calculation using an identical data processing method created using the GC-MS Solution Postrun
54
55
56
57
58
59
60

1
2
3 Analysis[®] software (Shimadzu, Japan) (Autoarea mode, maximum peak number = 300, width
4 time = 2 s, smoothing method = standard). Only peaks with minimum peak area = 50 000 were
5
6 selected for further analysis. The area of each peak was calculated using a unique quantifier ion
7
8 mass when its relative qualifier ion mass intensity was within 20% range ratio. To minimize
9
10 process errors, we manually checked each integrated peak for each sample.
11
12

13
14 The resulting data table was exported to Excel for normalization to the internal standard (4-
15 phenylbutyric acid) and then to the creatinine concentration. The normalization to the internal
16
17 standard was performed by dividing the integrated area for each analyte by the integrated area of
18
19 the internal standard. The results are expressed as ratios to the urinary creatinine concentration (
20
21 /mmole creatinine).
22
23
24
25
26
27
28

29 *Data processing - Peak Identification*

30
31 Compounds were identified from their electron impact mass spectra by comparison to the NIST
32
33 spectral mass library (NIST 05). Mass spectra of unknown compounds were labeled according to
34
35 the retention times as part of the identification, one quantifier peak and one or two fragmental
36
37 qualifier peaks. There is currently no commercially available reference standard for silyzed
38
39 metabolites to determine the derivatization recovery, so we checked the reproducibility of the
40
41 quantitative results without knowing the derivatization efficiency.
42
43
44
45
46
47

48 *Statistical Methods.*

49
50 The intensity of all peaks for all urine samples were studied by multivariate statistical methods,
51
52 following published protocols [17,18].
53
54
55
56
57
58
59
60

1
2
3 *Multivariate analysis* was performed using Simca-P+-12 software (version 12.0, Umetrics,
4 Umeå, Sweden). Unit variance (UV) scaling means that the variable is centered and scaled from
5 the standard deviation of the variable. With pareto scaling (Par), obtained by dividing each
6 variable by the square root of its standard deviation, the variance differs between variables, but
7 the range of variance across each spectrum is much smaller than that for the initial unscaled data
8 (small values being scaled up and large values being scaled down). The combination of scaling
9 and mean centering is termed autoscaling in MetaboAnalyst software [each descriptor (of high or
10 low intensity) is weighted equally] [19]. A logarithmic transformation (which is a nonlinear
11 conversion) such as an appropriate variance-stabilizing transformation, can be performed to
12 minimize the effects of noise or high variability of the variables [20]. Principal component
13 analysis (PCA) [21] was first performed as unsupervised clustering to identify the similarity or
14 the differences between sample profiles. Grouping, trends and outliers were revealed from the
15 scatter plot. To identify subsets (linear combinations) of metabolic features associated with a
16 specific sample class (ASD or control), partial least squares (PLS) analysis was used as
17 supervised clustering. PLS derives latent variables which describe the maximum proportion of
18 covariance between measured data (X matrix) and the response variable (Y matrix) [22].
19 Orthogonal partial least squares discriminant analysis (OPLS-DA), also used for discrimination,
20 is a refinement of this approach: variation in the data measured is partitioned into two blocks,
21 one containing variations that correlate with the class identifier (ASD or control) and the other
22 containing variations that are orthogonal to the first block and thus, do not contribute to
23 discrimination between the defined groups [23]. Discriminant metabolites were proposed by
24 OPLS-DA from one predictive and two or more orthogonal components. The quality of the
25 models was described by the cumulative modeled variation in the X matrix $R^2X(\text{cum})$, the
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 cumulative modeled variation in the Y matrix $R^2Y(\text{cum})$, where $R^2Y(\text{cum})$ is defined as the
4 proportion of variance in the data explained by the models and indicates goodness of fit, and the
5 cross-validated predictive ability $Q^2(\text{cum})$ values. Models were rejected if they presented
6 complete overlap of Q^2 distributions [$Q^2(\text{cum}) < 0$] or low classification rates [$Q^2(\text{cum}) < 0.05$
7 and eigenvalues should be > 2]. The features with variable importance on projection (VIP)
8 values and regression coefficients $|\text{CoeffCS}[1]|$ lower than 0.35 were deleted and evaluated
9 again. A number of variables was identified from PLS-DA and OPLS-DA as being responsible
10 for the difference between ASD and control urine samples with a VIP value > 1.0 .
11
12
13
14
15
16
17
18
19
20
21

22 One of the main problems with PLS-DA is the data overfitting that can occur if the
23 algorithm picks up random noise as real signals. To validate the model, the data are divided into
24 seven parts: a model was built with 6/7th of data left in, and the left out data were predicted [24].
25 The predicted data are then compared with the original data and the sum of squared errors
26 calculated for the whole dataset. This was converted into Q^2 , which is an estimation of the
27 predictive ability of the model. The model was thus considered sufficiently well guarded against
28 overfitting and validated after 200 random permutation tests [25] as the Q^2 line intercepted the Y
29 axis at a negative value.
30
31
32
33
34
35
36
37
38
39
40

41 We also performed PLS-DA with autoscaling normalization which is very similar to
42 pareto scaling using the freely available web-based software called MetaboAnalyst [26].
43
44

45 Univariate analysis (Student's *t*-test) was performed using MetaboAnalyst [26] for all
46 metabolites with a VIP > 1.0 . The ratio of the peak areas of these metabolites to that of creatinine
47 was calculated and a non-parametric test was performed with the critical *p*-value set at 0.05
48
49
50
51
52
53 [27,18].
54
55
56
57
58
59
60

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.

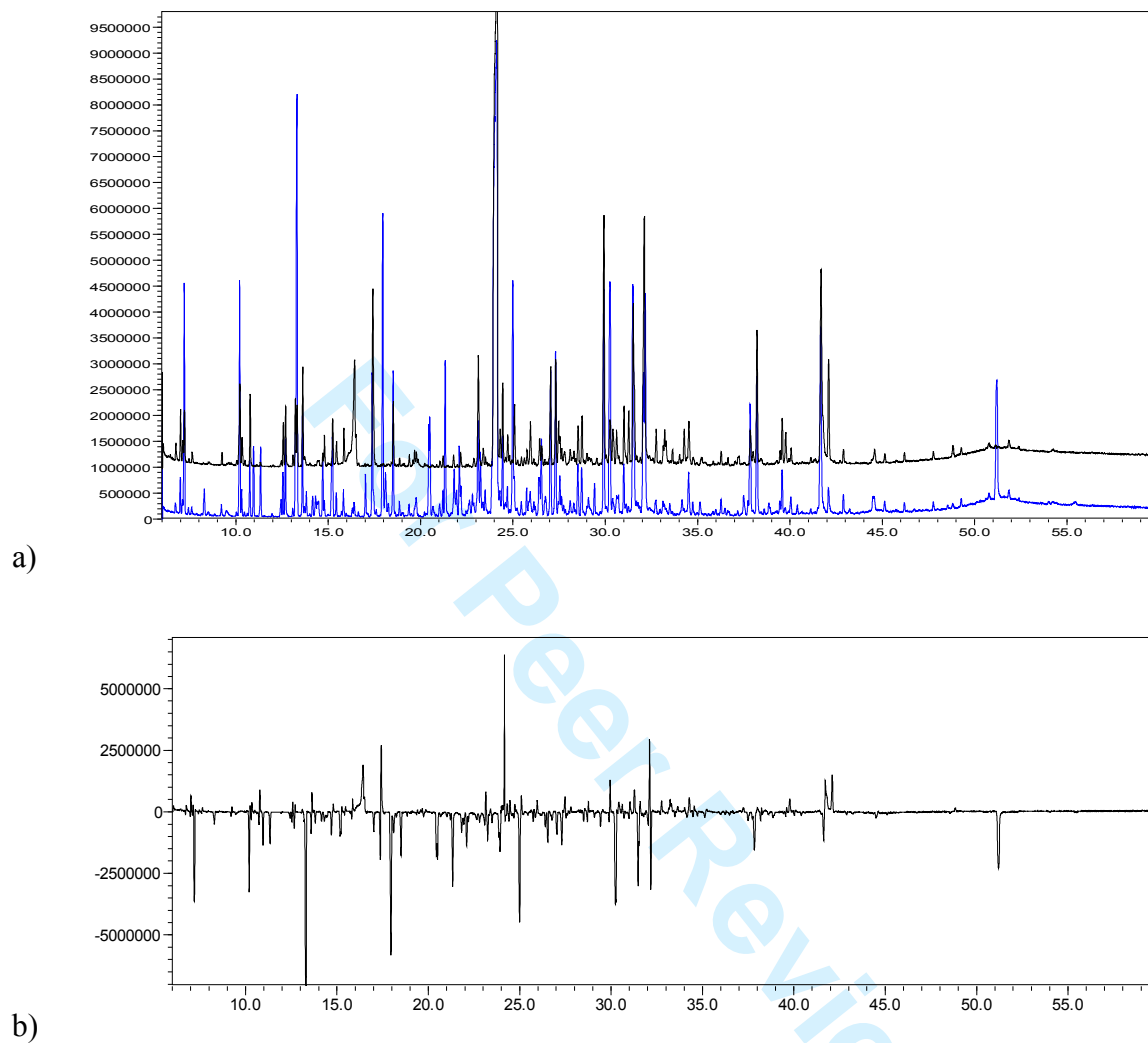


Fig. 1 a) Typical urine GC-MS spectra obtained for urine from ASD (blue) and healthy control (black) children; b) Subtraction of these two chromatograms revealing differential metabolite profiles.

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 [$R^2X(\text{cum})$] and Y [$R^2Y(\text{cum})$] matrix on spectral datasets and predictability of the model [$Q^2(\text{cum})$] are given [observations (N)=50]

	Scaling/transformation	R^2X	$R^2X(\text{cum})$	$R^2Y(\text{cum})$	$Q^2(\text{cum})$
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 $R^2X(\text{cum})$, the models 1 and 2 with the same scaling (UV) interpreted around 40% of the total

1
2
3 variation in X (0.42 and 0.47, respectively, Table 1). The part of the variation that could not be
4 explained by the model might originate from the noise or high variability of the variables [20].
5
6 This variation was minimized by Pareto scaling (model 3) and is expressed by the formula [28]:
7
8
9
10 $1 - R^2X - R^2X(\text{cum}) = 1 - 0.173 - 0.83 = 0.003$, as the noise could account for less 1% in this
11 analysis. The quality of the models is expressed by $R^2Y(\text{cum})$ and $Q^2(\text{cum})$ values. UV scaling
12 (model 2, Table 1), explained 79% of the variations in the various peaks, whereas with Pareto
13 scaling $R^2Y(\text{cum}) = 0.59$. With non linear transformation (model 4, Table 1), by log
14 transformation minimizing the effects of noise or high variability of the variables [$R^2X(\text{cum}) =$
15 0.75], the model explained 65% of the variance in the data, but the value was lower than by UV
16 scaling where $R^2Y(\text{cum}) = 0.79$ (model 2, Table 1). The high $Q^2(\text{cum})$ value [$Q^2(\text{cum}) > 0.5$]
17 indicated good predictivity. Pareto scaling or log transformation led to lower predictability
18 [$Q^2(\text{cum}) = 0.33$ and 0.42 , respectively] (Table 1). As UV scaling seemed to be the best scaling
19 method in our study investigations [(confirmed by analysis of variance CV-ANOVA with the
20 lower value of p -value = 4.4×10^{-6})], it was used for the subsequent investigations. The OPLS-DA
21 cross-validated score scatter plots for model 2 [Fig. 2, (a)] showed good discrimination between
22 the two populations.
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42

43 With oximation

44
45 OPLS-DA models explained 83% (Table 2) of the variance of the data obtained after oximation-
46 silylation-based GC-MS compared to 79% [$R^2Y(\text{cum}) = 0.79$] of that obtained without oximation.
47
48 The cross-validated predictive ability $Q^2(\text{cum})$ values were also better after oximation (0.68
49 compared to 0.54, respectively).
50
51
52
53
54
55
56
57
58
59
60

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 [$R^2X(\text{cum})$] and Y [$R^2Y(\text{cum})$] matrix on spectral datasets and predictive power of the model [$Q^2(\text{cum})$] are given [observations (N)=50]

	R^2X	$R^2X(\text{cum})$	$R^2Y(\text{cum})$	$Q^2(\text{cum})$
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

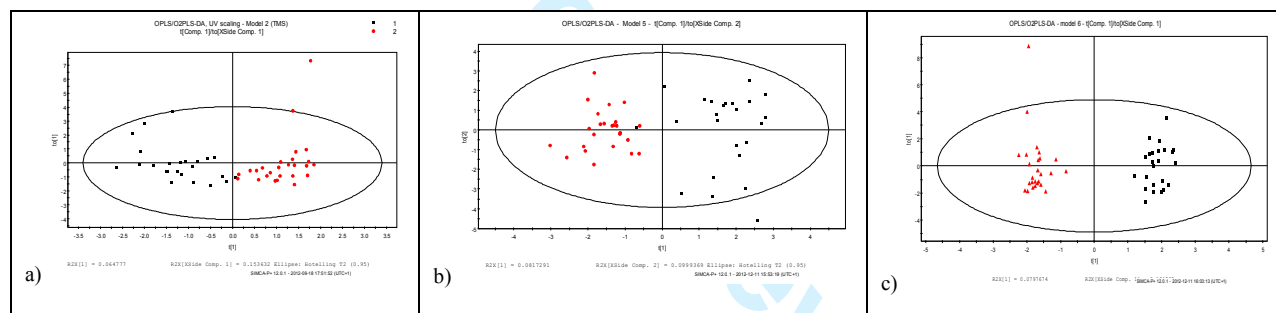


Fig. 2 OPLS-DA, UV-scaled, score scatter plots obtained from of GC-MS analysis of samples without oximation (a), with oximation (b), and with and without oximation (c) from urine from autistic (red dot) and control (black box) children, showing that the two populations are well separated with (a) a $R^2Y(\text{cum})=0.79$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

1
2
3 without an oximation process were included in the analysis (data obtained from samples
4
5 undergoing less manipulation were expected to be more repeatable). Consequently, 117 signals,
6
7
8 58 of them identified, were used.
9

10 Multivariate analysis of the two sets of data combined showed that OPLS-DA (Table 2,
11
12 model 6) led to a very good discrimination between the two populations [0.45 for $R^2X(\text{cum})$ and
13
14 0.97 for $R^2Y(\text{cum})$]; this was confirmed by analysis of variance CV-ANOVA ($p\text{-value} = 9.6 \text{ e}^{-15}$,
15
16 see supporting information), and good predictive ability value with $Q^2(\text{cum}) = 0.88$. Using the
17
18 cumulated data (that for samples with or without an oximation process), the multivariate analysis
19
20 clearly gave more discriminatory results than obtained using data from samples subjected to just
21
22 one pre-analytical chemical process. The OPLS-DA cross-validated score plots for model 6 [Fig.
23
24 2, (c)] showed excellent discrimination between the two populations. The loading scatter plot
25
26 (Figure 3) shows which variables expressed similarity between ASD and control children in
27
28
29
30
31
32 model 6.
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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

Putative Assignment ^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)
1H-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.

^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.

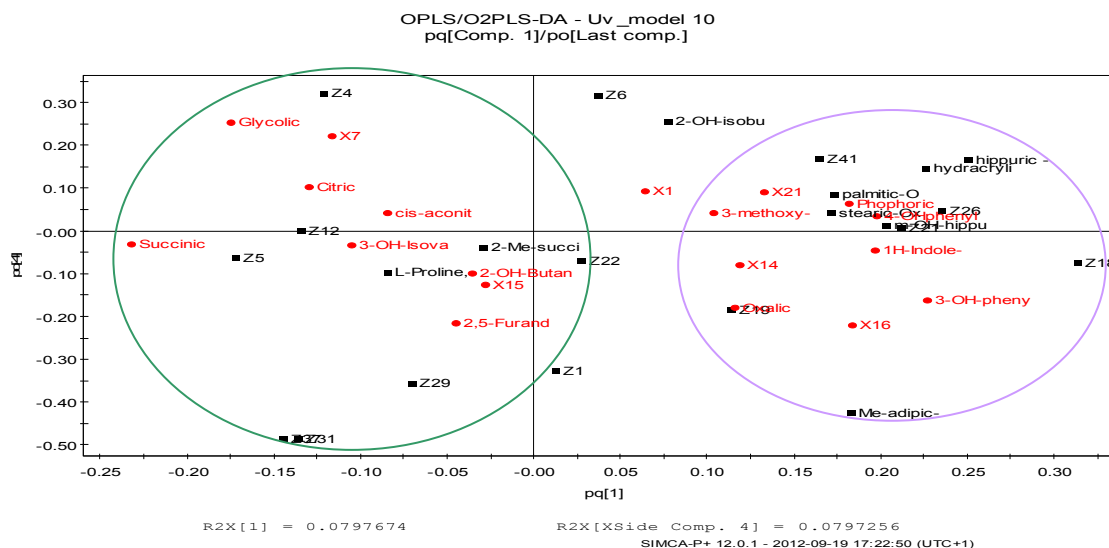


Fig. 3 pq loading plot of the OPLS-DA model 6 (Table 2): scatter plot of the X- and Y-loadings (p and q). This plot shows how the responses (Y's) varied in relation to each other, i.e. which provided similar information and their relationship to the terms of the model. Two tendencies can be seen. Peaks 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

1
2
3 OPLS-DA is recommended as it allows a clearer and more straightforward interpretation than
4 other statistical methods [19], so we focused on this approach. Variation between samples can
5 generally be classified into “technical” or “biological” variance; the impact of technical
6 variability should be minimized and useful biological data needs to be discrimination from noise.
7
8 Data processing methods are highly dependent on the pre-treatment technique used for the data.
9
10 Several types of scaling are commonly used: UV, Par, and autoscaling. The choice of pre-
11 treatment methods depends on several factors (numbers of samples, magnitude of concentration,
12 similarities, etc), and different methods emphasize different aspect of the data. As each method
13 has its own merits and drawback, and as, to our knowledge, there is non consensus in the pre-
14 treatment for urines analyzed by GC-MS, we studied the effect of all these scaling methods on
15 identification of biomarkers.
16
17
18
19
20
21
22
23
24
25
26
27
28

29 We found that noise (and/or high variability of the variables) was minimized by Pareto
30 scaling: ($R^2X(\text{cum})$) was highest in model 3, but it led to lower predictability [$Q^2(\text{cum}) = 0.23$]
31 (Table 1). UV scaling seemed to be the best scaling method for our data set (confirmed by
32 analysis of variance CV-ANOVA, see supporting information). We therefore focused on results
33 obtained from OPLS-DA using UV scaling (model 6, Tables 2 and 3).
34
35
36
37
38
39
40
41
42

43 Statistical analysis of TMS and TMSO results – Analysis of models and identification of
44 important features
45

46 Factors such as disease, drugs and diet modify the concentrations of individual metabolites
47 [29,30]. It is known that autistic patients display dysfunctions in the levels of hormones, peptides,
48 metabolites associated with neurological, gastrointestinal, immunological and toxicological
49 effects (for review see [31]). Although the number of samples included in this study was too
50
51
52
53
54
55
56
57
58
59
60

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

23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50 The discriminant urinary metabolites were vanillylhydracrylate, 3-methyladipate, *p*-
51 hydroxymandelate, glycolate, palmitate, stearate, succinate, phosphate, hippurate, 3-
52 hydroxyphenylacetate, 3-hydroxyhippurate, 1*H*-indole-3-acetate (these last metabolites could be
53
54
55
56
57
58
59
60

1
2
3 implicated in microbial pathways associated with gut bacterial), and also eight metabolites
4 identified only according to their retention time and fragmentation patterns (Table 3).
5
6

7
8 We compared metabolic fingerprinting obtained by GC-MS after a silylation step (which
9 is the most versatile and universally applicable derivatization method) with that after oximation
10 of keto-derivatives followed by a BSTFA silylation step. This study clearly shows that the two
11 derivatization procedures are complementary, and allow analysis of a wider range of metabolites
12 for metabolomics studies.
13
14
15
16
17
18

19 20 21 22 **Conclusion**

23 Evaluation of urinary metabolite profiles using a GC-MS method showed promising results.
24 First, this study clearly suggest that sample pre-treatment techniques are relevant, and we
25 highlighted complementarities between the two derivatization procedures assessed: TMS
26 derivatization (economic, minimal sample handling suitable for routine analysis), and a TMSO
27 procedure (higher sensitivity). Combined, the two procedures allowed analysis of 132
28 metabolites as their TMS/TMSO ether/ester derivatives, 73 of which could be identified as
29 known compounds. Secondly, relevant metabolites, with different concentrations in children with
30 or without a diagnosis of ASD were successfully extracted by multivariate analysis. Further work
31 including targeted studies, is needed to identify chemically those features that are potentially
32 discriminant and to validate their clinical value.
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49

50 51 **Acknowledgements**

52 This work was supported by the “Institut National de la Santé et de la Recherche” INSERM, and
53 the University François-Rabelais. We thank the center “Sésame Autisme Loiret” for participation
54
55
56
57
58
59
60

1
2
3 in this study. We thank the “Département d’Analyses Chimiques et S.R.M. Biologique et
4
5 Médicale” (PPF, Tours, France) for GC-MS analyses.
6
7
8
9

10 **References**

- 11
12
13
14
15 1. Weintraub K (2011) The prevalence puzzle: Autism counts. *Nature* 479 (7371):22-24
- 16
17
18 2. Falck-Ytter T, von Hofsten C (2011) How special is social looking in ASD: a review. *Prog*
19
20 *Brain Res* 189:209-222
- 21
22
23 3. Franke B, Faraone SV, Asherson P, Buitelaar J, Bau CH, Ramos-Quiroga JA, Mick E, Grevet
24
25 EH, Johansson S, Haavik J, Lesch KP, Cormand B, Reif A (2011) The genetics of attention
26
27 deficit/hyperactivity disorder in adults, a review. *Mol Psychiatry* 17:960-987
- 28
29
30 4. McPartland JC, Coffman M, Pelphrey KA (2011) Recent advances in understanding the neural
31
32 bases of autism spectrum disorder. *Curr Opin Pediatr* 23 (6):628-632
- 33
34
35 5. Schaefer GB, Lutz RE (2006) Diagnostic yield in the clinical genetic evaluation of autism
36
37 spectrum disorders. *Genet Med* 8 (9):549-556
- 38
39
40 6. Kolvin I (1971) Studies in the childhood psychoses. I. Diagnostic criteria and classification. *Br*
41
42 *J Psychiatry* 118 (545):381-384
- 43
44
45 7. Association AP (2000) Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition:
46
47 DSM-IV-TR®. Washington
- 48
49
50 8. Koek M, Jellema R, van der Greef J, Tas A, Hankemeier T (2011) Quantitative metabolomics
51
52 based on gas chromatography mass spectrometry: status and perspectives. *Metabolomics* 7
53
54 (3):307-328
55
56
57
58
59
60

- 1
2
3 9. Madsen R, Lundstedt T, Trygg J (2010) Chemometrics in metabolomics-A review in human
4 disease diagnosis. *Analytica Chimica Acta* 659 (1-2):23-33
5
6
- 7
8 10. Suhre K, Shin SY, Petersen AK, Mohny RP, Meredith D, Wagele B, Altmaier E, Deloukas P,
9 Erdmann J, Grundberg E, Hammond CJ, de Angelis MH, Kastenmuller G, Kottgen A,
10 Kronenberg F, Mangino M, Meisinger C, Meitinger T, Mewes HW, Milburn MV, Prehn C,
11 Raffler J, Ried JS, Romisch-Margl W, Samani NJ, Small KS, Wichmann HE, Zhai G, Illig T,
12 Spector TD, Adamski J, Soranzo N, Gieger C (2011) Human metabolic individuality in
13 biomedical and pharmaceutical research. *Nature* 477 (7362):54-60
14
15
- 16
17 11. Gebregiworgis T, Powers R (2012) Application of NMR metabolomics to search for human
18 disease biomarkers. *Comb Chem High Throughput Screen* 15 (8):595-610
19
20
- 21
22 12. Mercier P, Lewis MJ, Chang D, Baker D, Wishart DS (2011) Towards automatic
23 metabolomic profiling of high-resolution one-dimensional proton NMR spectra. *J Biomol NMR*
24 49 (3-4):307-323
25
26
- 27
28 13. Garcia A, Barbas C (2011) Gas chromatography-mass spectrometry (GC-MS)-based
29 metabolomics. *Methods Mol Biol* 708:191-204
30
31
- 32
33 14. Yap IK, Angley M, Veselkov KA, Holmes E, Lindon JC, Nicholson JK (2010) Urinary
34 metabolic phenotyping differentiates children with autism from their unaffected siblings and age-
35 matched controls. *J Proteome Res* 9 (6):2996-3004
36
37
- 38
39 15. Wang L, Angley MT, Gerber JP, Sorich MJ (2011) A review of candidate urinary biomarkers
40 for autism spectrum disorder. *Biomarkers* 16 (7):537-552
41
42
- 43
44 16. Organization WH (1991) International Classification of Diseases (ICD-10). World Health
45 Organization, Geneva
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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 ^1H 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 PLS-DA cross validation. *Metabolomics* 4 (1):81-89

- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
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://www.umetrics.com/>
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

- 1
2
3 34. Shaw W, Kassen E, Chaves E (1995) Increased urinary excretion of analogs of Krebs cycle
4 metabolites and arabinose in two brothers with autistic features. *Clin Chem* 41 (8):1094-1104
5
6
7
8 35. Kaluzna-Czaplinska J, Michalska M, Rynkowski J (2011) Homocysteine level in urine of
9 autistic and healthy children. *Acta Biochim Pol* 58 (1):31-34
10
11
12 36. Adams JB, Audhya T, McDonough-Means S, Rubin RA, Quig D, Geis E, Gehn E, Loresto
13 M, Mitchell J, Atwood S, Barnhouse S, Lee W (2011) Nutritional and metabolic status of
14 children with autism vs. neurotypical children, and the association with autism severity. *Nutr*
15 *Metab* 8 (1):34
16
17
18
19
20
21
22 37. Evans C, Dunstan RH, Rothkirch T, Roberts TK, Reichelt KL, Cosford R, Deed G, Ellis LB,
23 Sparkes DL (2008) Altered amino acid excretion in children with autism. *Nutr Neurosci* 11 (1):9-
24
25
26
27 17
28
29 38. Kaluzna-Czaplinska J, Socha E, Rynkowski J (2010) Determination of homovanillic acid and
30 vanillylmandelic acid in urine of autistic children by gas chromatography/mass spectrometry.
31
32
33
34 *Med Sci Monit* 16 (9):CR445-450
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 List of figures
4
5

6 **Fig. 1** a) Typical GC-MS spectra obtained for urine from ASD (blue) and healthy control (black)
7 children; b) Subtraction of these two chromatograms revealing differential metabolite profiles
8
9
10

11
12
13
14
15 **Fig. 2** OPLS-DA, UV-scaled, score scatter plots obtained from of GC-MS analysis of samples
16 without oximation (a), with oximation (b), and with and without oximation (c) from urine from
17 autistic (red dots) and control (black box) children, showing that the two populations are well
18 separated with (a) a $R^2Y(\text{cum})=0.79$ and a $Q^2(\text{cum})=0.54$ (model 2, Table 1), with (b) a
19 $R^2Y(\text{cum})=0.86$ and a $Q^2(\text{cum})=0.73$ (model 5, Table 2) and (c) a $R^2Y(\text{cum})=0.97$ and a
20 $Q^2(\text{cum})=0.88$ (model 6, Table 2).
21
22
23
24
25
26
27
28
29
30
31
32
33

34 **Fig. 3** pq loading plot of the OPLS-DA model 6 (Table 2): scatter plot of the X- and Y-loadings
35 (p and q). This plot shows how the responses (Y's) varied in relation to each other, i.e. which
36 provided similar information, and their relationship to the terms of the model. Two tendencies
37 can be seen. Peaks framed in pink were lower in ASD urine, and peaks framed in green were
38 higher. Peaks in red were obtained without, and peaks in black with, the oximation process
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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} . Frédéric Montigny^b .
Frédérique Bonnet-Brilhault^c . Catherine Barthélémy^c . Marc Merten^d . Pierre Sarda^c . Frédéric
Laumonier^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

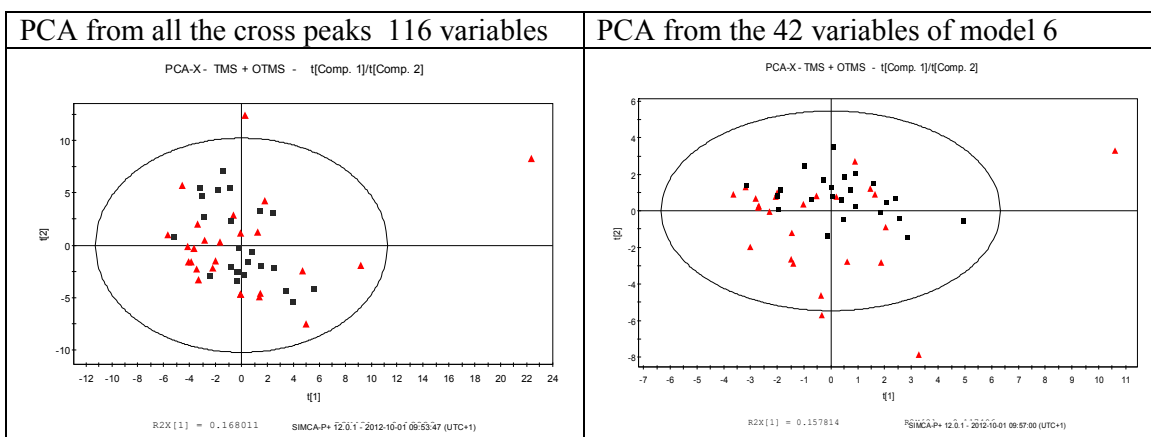
^d Université Henri Poincaré, Faculté de Médecine, Laboratoire de Biochimie, 9 Av de la forêt de Haye, 54505
Vandoeuvre les Nancy, France

^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 <i>p</i> -value (<i>t</i> -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

Fig. S1

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

**Fig. S2**

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

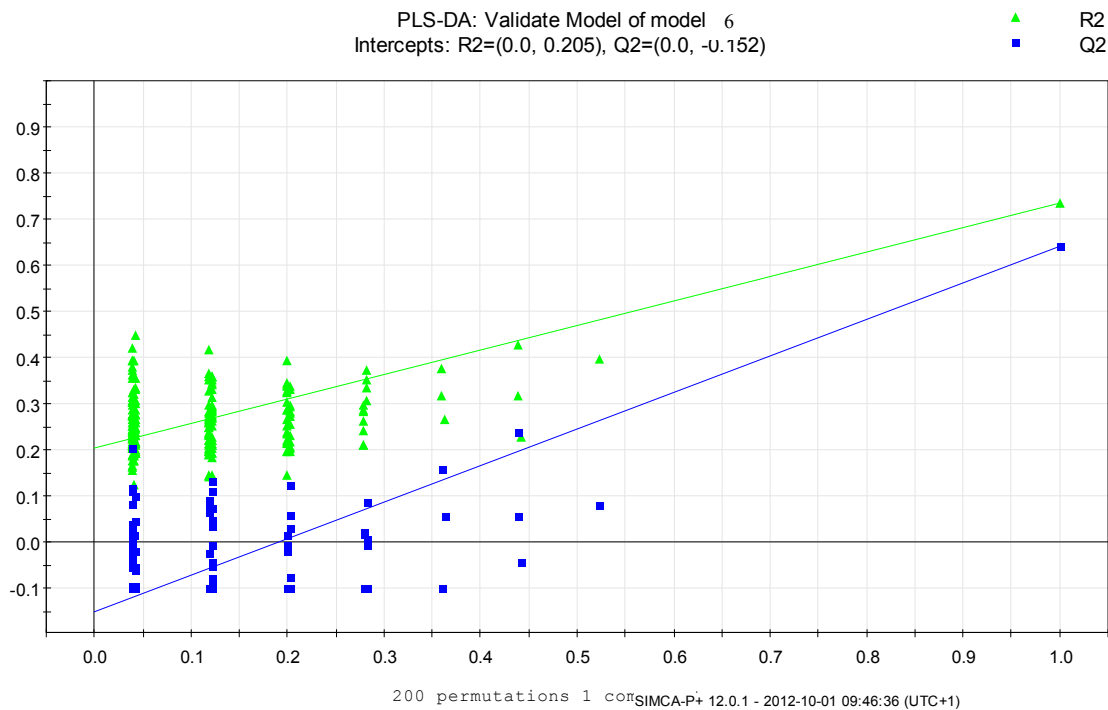


Table S1. TMS and TMSO data

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

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)

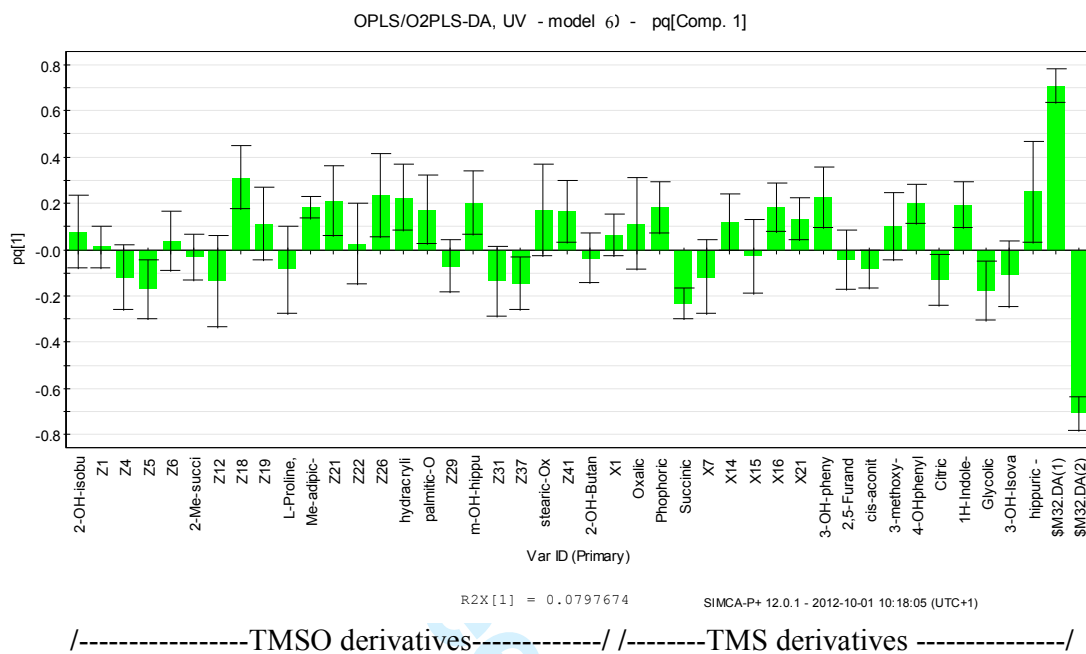


Table S5

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 Assignment ^a	Quantifier ion mass	Retention time minutes	Qualifier ion mass		Differentiation for autistic samples	OPLS-DA-UV-VIP values ^b Model 6	<i>p</i> 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
<i>p</i>-Hydroxy mandelate	266.9	30.72	267.9	341.8	↓	1.30	0.010
1<i>H</i>-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	

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

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 Assignment	Quantifier ion mass	Retention time minutes	Qualifier ion mass		Differentiation for autistic samples	VIP coeff	Putative Assignment	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
Phosphoric	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

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For Peer Review