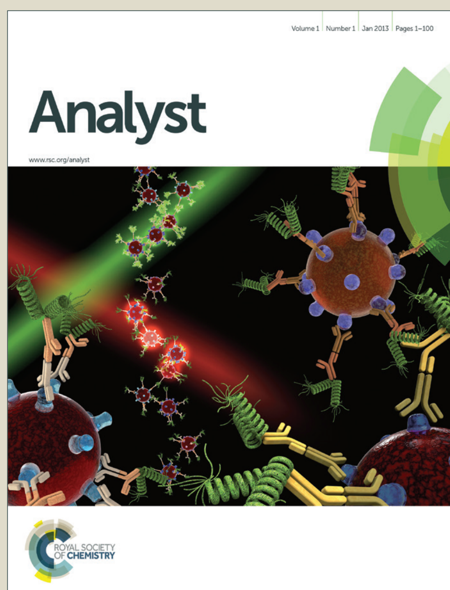
Analyst

Accepted Manuscript



This is an *Accepted Manuscript*, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this *Accepted Manuscript* with the edited and formatted *Advance Article* as soon as it is available.

You can find more information about *Accepted Manuscripts* in the Information for Authors.

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard Terms & Conditions and the Ethical guidelines still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this *Accepted Manuscript* or any consequences arising from the use of any information it contains.

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/analyst

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

Combined ^{1}H-NMR and ^{1}H-^{13}C HSQC-NMR to improve urinary screening in autism spectrum disorders †

Lydie Nadal-Desbarats,*‡ [abc] Nacima Aïdoud,‡[a] Patrick Emond,[abc] Hélène Blasco,[ab] Isabelle Filipiak,[f] Pierre Sarda,[e] Frédérique Bonnet-Brilhault,[dg] Sylvie Mavel[a] and Christian R Andres[ab]

[a] Equipe neurogénétique et neurométabolomique INSERM U930, Université François Rabelais, 37000 Tours, France

[b] Service de biochimie et de biologie moléculaire, CHRU Tours, 37000 Tours, France

[c] Département d'Analyses Chimique Biologique et Médicale. PPF "Analyses des Systèmes Biologiques", Tours, France

[d] Centre Universitaire de pédopsychiatrie, CHRU Tours, 37000 Tours, France.

[e] Service de Génétique Médicale, CHU Montpellier, 34000 Montpellier, France

[f] Equipe Imagerie et Ultrasons INSERM U930, Université François Rabelais, 37000 Tours, France

[g] Equipe Autisme INSERM U930, Université François Rabelais, 37000 Tours, France

‡These authors contributed equally to the work

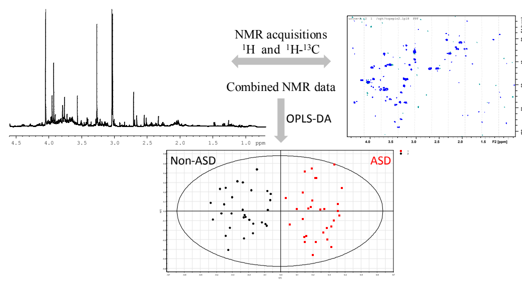
*Corresponding author:
Lydie Nadal-Desbarats,
INSERM U930
PPF "Analyses des Systèmes Biologiques"
UFR de Médecine
10 BlvdTonnellé, 37044 Tours Cedex 9
Tel : +33(0)2 47 36 61 64
Fax : +33(0)2 47 37 37 21
nadal@med.univ-tours.fr

† Electronic supplementary information (ESI) available. See DOI:

Analyst Accepted Manuscript

Table of contents

Improvement of urinary screening by combining ^{1}H and 2D HSQC NMR data in metabolomics: application in ASD.



Analyst Accepted Manuscript

Abstract

Autism spectrum disorders (ASD) are neurodevelopmental diseases with complex genetic and environmental etiological factors. Although genetic causes play a significant part in the etiology of ASD, metabolic disturbances may also play a causal role or modulate the clinical features of ASD. The number of ASD studies involving metabolomics is increasing, and sometime with conflicting findings. We assessed the metabolomics profiling of urine samples to determine a comprehensive biochemical signature of ASD. Furthermore, to date no study has combined metabolic profiles obtained from different analytical techniques to distinguish patient with ASD from healthy individuals. We obtained ^{1}H-NMR spectra and 2D ^{1}H-^{13}C HSQC NMR spectra from urine samples of patients with ASD or healthy controls. We analyzed these spectra by multivariate statistical data analysis. The OPLS-DA model obtained from ^{1}H NMR spectra showed a good discrimination between ASD samples and non-ASD samples (R^2Y(cum)=0.70 and Q^2=0.51). Combining the ^{1}H NMR spectra and the 2D ^{1}H-^{13}C HSQC NMR spectra increased the overall quality and predictive value of the OPLS-DA model (R^2Y(cum)=0.84 and Q^2=0.71), leading to a better sensitivity and specificity. Urinary excretion of succinate, glutamate and 3-methyl-histidine differed significantly between ASD and non-ASD samples. Urinary screening of children with neurodevelopmental disorders by combining NMR spectroscopies (1D and 2D) in multivariate analysis is a better sensitive and a straightforward method that could help the diagnosis ASD.

Analyst Accepted Manuscript

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

Introduction

Autistic spectrum disorder (ASD) refers to a group of complex neurodevelopmental disorders present from early childhood and that persist lifelong. The estimated prevalence worldwide is 1 in 150 children. There are five diagnostic subtypes including autism, pervasive developmental disorders-not otherwise specified (PDD-NOS), child disintegrative disorder, Rett syndrome, and Asperger syndrome.[1, 2] No diagnostic test is available (except for Rett syndrome), therefore diagnosis is based on a triad of criteria defined by the Diagnostic and Statistical Manual of Mental Disorders.[3] These criteria involve behavioral aspects that typically manifest before three years of age, including deficits in communication, impaired social interactions, and repetitive or restricted interests and behaviors.

The causes of ASD remain largely unclear, despite a considerable amount of research in the clinical, electrophysiological, and genetic aspects of ASD in recent years, ASD is a multifactorial disease that is associated with predisposing genetics factors and environmental influences. ASD is a multisystem disorder. Indeed, genetic, nutritional or environmental factors may affect a variety of cell types and would be expected to have consequences on multiple bodily systems. Chronic metabolic imbalances associated with complex diseases such as ASD may leave a metabolic fingerprint that can be followed analytically; thus, such analyses may provide new insights into the pathophysiology and pathogenesis of ASD[4] and may help diagnosis.

Metabolomics is the study of the metabolome.[5] The metabolome consists of a repertoire of low-molecular weight compounds that are intermediates or endpoints of metabolism and are present in biological fluids, cells, or tissues.[6] The metabolites are the final product of interactions between the regulation of gene expression, protein abundance, and the cellular environment. Therefore metabolites may serve as reporters of intermediary or disease phenotypes.[7] This promising approach may help to define new candidate biomarkers and physiological pathways involved in disease pathology. Recently, the analysis of biological fluids to identify biomarkers has become an area of active investigation. This approach has been widely used to characterize metabolic signatures of several neurological disorders including depression,[8] motor neuron disease,[9] neurodegenerative disease,[10] addiction to drugs[11] and schizophrenia.[12, 13] An integrative analysis of the metabolome from biological fluids may reveal biological disruptions common to ASD patients. This would allow the defining of a metabolic profile (metabotype) made up of composite biomarkers of ASD.[14] For

Analyst Accepted Manuscript

1
2
3 ethical and methodological reasons, urine samples are suitable for the analysis of metabolic perturbations, and have already been used for investigations of metabolic abnormalities in ASD.[15, 16]

The number of publications involving metabolomics and biomarkers is increasing.[17] Studies have investigated metabolic profiles in numerous biologicals fluids, giving rise to many clinical applications.[18, 19] The most commonly used analytical tools are chromatographic methods (GC-MS, LC-MS)[20] and nuclear magnetic spectroscopy (NMR).[21] ^{1}H-NMR spectroscopy is a rapid, robust and reliable analytical tool with high reproducibility. Lately, there have been many conflicting findings in studies involving metabolomics in autism spectrum disorders, depending on biological fluid used, mostly urine[16, 21-24] but also plasma.[25-27] The large number of biological fluids and analytical techniques used means that the list of metabolites studies is long. From these, there may be one, or a few, that are relevant to autism. Yap *et al.* published the only metabolomics study to date involving ^{1}H-NMR analysis of urine.[21] They examined metabolic profiles in three groups: ASD children, their unaffected siblings, and unrelated controls. ASD children showed a distinct profile of gut microbial metabolism, amino-acid metabolism and nicotinic acid metabolism. However, assignment of 1D ^{1}H-NMR spectra is challenging because of significant peak overlap and the presence of uncharacterized metabolites. Our contribution in the field was a study using ^{1}H-^{13}C heteronuclear single quantum coherence (HSQC) spectra to improve the assessment of the metabolite content of biological fluids such as urine.[28, 29] 2D ^{1}H-^{13}C HSQC NMR was used to compare urinary profiles from autistic patients and non-autistic controls. We described, urinary metabolic imbalance in autistic individuals similar to that reported by Yap *et al.*[30] For the discovery of metabolomics biomarkers, it is important to identify, among the many potential compounds analyzed, the combination of metabolites (variables) that best discriminates diseased from healthy individuals. Therefore, we used multivariate analyses, following well established protocols now,[31] to reduce, summarize, and transform all the data to a few key components that corresponded to the most discriminating biomarkers.

In this study, we investigate whether the combination of ^{1}H-NMR and ^{1}H-^{13}C HSQC NMR metabolic profiling of urine samples may facilitate the identification of biochemical signatures of ASD. Using this approach, we also attempted to replicate biomarkers of ASD that have already been described. To our knowledge, this is the first study combining data from ^{1}H-NMR and ^{1} H-^{13}C HSQC spectra of urine. The use of these two NMR modalities associated with multivariate statistical data analysis was expected to increase the accuracy of the dis-

Analyst Accepted Manuscript

crimination between ASD patients and controls by optimizing the model performance. We used a receiver operating curve (ROC) to assess the diagnosis accuracy of our combined biomarkers.[17]

Experimental

Sample collection

Urine samples were collected in vials without preservative. Samples were collected during medical consultation from thirty children with ASD. All children were living in France [24 boys and six girls, median age of 8 (6-14)]. Urine samples were also obtained from 28 healthy individuals from Tours, France [17 boys and 11 girls median, age of 8 (6-9)]. Diagnosis of autism was made according to the International Classification of Diseases (ICD) Edition 10th [2] and the DSM-IV-TR Edition 4th.[32] Each individual and their family gave informed consent for the study. Each urine sample was centrifuged, aliquoted in 1.5 mL Eppendorf tubes and stored at -80°C immediately after collection until analysis.

Sample preparation

Urine samples were thawed at room temperature, and centrifuged at 3000*g* for 5 min. Urine samples were prepared by mixing 500 µL of urine supernatant, 100 µL of D_2O solution (deuterium oxide) and 100 µL of phosphate buffer to obtain a pH = 7.4 ± 0.5. The samples were then transferred to 5-mm NMR tubes for ^{1}H -NMR analysis.

Magnetic Resonance Spectroscopy experiments

^{1}H NMR experiments:

The ^{1}H NMR spectra were obtained by a Bruker DRX-500 spectrometer (Bruker SADIS, Wissembourg, France), operating at 11.7 T, with a Broad Band Inverse (BBI) probehead equipped with Z gradient coil. NMR measurements were done at 298 K. Conventional ^{1}H NMR spectra were recorded with a 90° pulse (p_1=10 µs, pl=0 dB) using a pulse-and-acquire sequence with residual water presaturation (single-frequency irradiation during the relaxation delay). ^{1}H spectra were collected with 64 transients (and 8 dummy scans) in 32K data points with a spectral width of 7500 Hz, and a recycling time of 30 s. CPMG spin echo spectra were obtained with 80 ms total echo times and 32K data points. This spin echo sequence avoided broad short T_2 resonance (provided by macromolecules). Sample shimming was performed automatically on the water signal.

Analyst Accepted Manuscript

Spectra were processed using XWinNMR version 3.5 software (Bruker Daltonik, Karlsruhe, Germany). Prior to Fourier transformation (FT), the FIDs were zero-filled to 64K data points which provided sufficient data points for each resonance, and a line broadening factor of 0.3 Hz was applied. All spectra were corrected for phase distortion and baseline was manually corrected for each spectrum.

The ^{1}H NMR spectra were referenced to the creatinine methylene resonance at δ=4.05 ppm and automatically reduced to ASCII files using the AMIX software package (Analysis of MIXture, version 3.1.5, Bruker Biospin, Karslruhe, Germany). The regions containing the water (δ 4.70 – 5.51 ppm) and urea (δ 5.58 – 6 ppm) signals were removed from each spectrum to eliminate baseline effects of imperfect water saturation. Spectral intensities were scaled to the total intensity and reduced to equidistant integrated regions of 0.005 ppm (buckets) over the chemical shift range of 0.7-9.5 ppm. Before the multivariate analysis, the NMR spectral datasets were preprocessed using the peak alignment algorithm *i*coshift[33] (http://www.models.life.ku.dk) to minimize spectral peak shift due to residual pH differences amongst samples. The corresponding realigned bucket tables were then exported to SIMCA-P$^+$ software (version 12.0, Umetrics, Umeå, Sweden)[34] for statistical analysis. In a second dataset, spectral intensities were scaled to the creatinine area peak (δ=4.05 ppm) and reduced to equidistant integrated regions of 0.005 ppm (buckets) over the chemical shift range of 0.7-9.5 ppm. A realignment using *i*coshift before multivariate analysis was also used.

^{1}H-^{13}C NMR experiments:

HSQC-NMR experiments were performed and processed as previously described.[30]

Data analysis and statistics

An unsupervised method, principal component analysis (PCA), was performed with SIMCA-P$^+$ software. Data were scaled using pareto unit (Par) (for 1D NMR) scaling prior to PCA. A plot of the first two principal components (score plot) provided the most effective 2D representation of the information contained in the data set. The overall quality of the models was judged by the cumulative R^2, and the predictive ability by cumulative Q^2 extracted according to the internal cross-validation default method of the SIMCA-P$^+$ software.

A partial least-squares discriminant analysis (PLS-DA) was performed, as a supervised model of classes, with SIMCA-P$^+$ software. Data were scaled using unit pareto (Par) scaling. PLS-DA is a prediction and regression method that finds information in the X data (variables) that is related to known information, the Y data (classes). PLS-DA exploits the class information

Analyst Accepted Manuscript

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

to attempt to maximize the separation between groups of observations. To check the validity and the degree of overfit for the PLS-DA model, a validating model, after 200 random permutations, was plotted (Fig. S2†). This plot displays the correlation coefficient between the original y-variable and the permuted y variable versus the cumulative R^2 and Q^2, and the regression line. Q^2 is the estimation of the predictive ability of the model and should intercept the Y axis at a negative value,[35] and R^2Y intercept should be <0.3. An extension of PLS model is an orthogonal PLS (OPLS-DA). OPLS-DA can rotate the projection so that the model focuses on the effect of interest. The overall quality of the models was judged by the cumulative R^2, and the predictive ability by cumulative Q^2. R^2 is defined as the proportion of variance in the data explained by the model and indicates goodness of fit. Q^2 is defined as the proportion of variance in the data that can be predicted by the model and thus indicates predictability. To evaluate further the significance of the findings, cross-validation analysis of variance (CV-ANOVA) was applied.[36] The contribution plot provides information about the variables that influence any observed clustering of samples. According to these criteria, metabolites with greater contribution in the separation of the groups were identified and quantified in the NMR spectra. The features with variables importance on projection (VIP) values>1.0, obtained from OPLS-DA, were responsible for the differences between ASD and control urine samples.

To improve the screening, the minimum number of features (spectral buckets or cross-correlation intensities) needed for optimal classification of the two previous models (OPLS-DA obtained with ^{1}H-NMR spectral data and OPLS-DA obtained with ^{1}H-^{13}C HSQC) was determined. An alternative model was then used to combine the two datasets: the X matrix was composed of the minimum number of features of the combined ^{1}H-NMR variables and the ^{1}H-^{13}C HSQC cross-correlation variables. To avoid the domination of one type of measurement over the other one, the variables from the same type of spectrum were block-scaled (1/sqrt) prior to multivariate analysis using SIMCA-P$^+$. OPLS-DA model was fitted using the above Y and X matrices. Results from cross-model validation were compared to the results from models using one dataset only. The OPLS-DA models were summarized in terms of sensitivity (Sn, proportion of diseased subjects that are correctly classified) and specificity (Sp, proportion of healthy subjects that are correctly classified).

To evaluate the prediction performance of the obtained OPLS-DA models, the receiver operating characteristics (ROC) (sensitivity values on the Y-axis and 1-specificity values on the X-axis) curve was used. The area under the ROC curve (AUC) and 95% CI (confidence intervals) were calculated for each model with ROCCET a freely available web-based tool.[17] The

Analyst Accepted Manuscript

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

linear Support Vector Machine (SVM) algorithm (the default), without scaling, was used for classification and feature selection. ROCCET uses repeated random sub-sampling cross-validation to test the performance of a model created with different numbers of features. AUC is a measure of how well a parameter can distinguish between ASD patients and controls, and accuracy can be determined from sensitivity (proportion of "ASD" that are correctly classified as "ASD") and specificity (proportion of "control" that are correctly classified as "control") [accuracy = (TrueNeg + TPos)/(TN+TP+FalseN+FP) = (Number of correct assessments)/Number of all assessments)].

Concentration ratios of urinary metabolites selected from the Multivariate Statistical Analysis

^{1}H NMR experiments:

To calculate the relative mean concentrations of the selected urinary metabolites, the peak areas of the selected NMR signals of the chosen metabolites were integrated using XWinNMR version 3.5 software (Bruker Daltonik, Karlsruhe, Germany). The ratios of the peak areas of these selected metabolites to the methylene creatinine peak (δ 4.05 ppm) were then calculated. SigmaStat 3.1 software (Systat Software, Inc., California, USA) was used for univariate statistical analysis of these ratios. Mann-Whitney rank sum test was performed to compare metabolite concentrations between groups, and $p < 0.05$ was considered as significant.

Results and discussion

Urinary ^{1}H NMR spectroscopic profiles

The analysis of biological fluids by NMR-based metabolomics may identify potential biomarkers associated with disease. Indeed, the differences in metabolite content between pathological and normal samples may be biologically relevant.[37] However, working with a biological fluid such as urine is challenging, and requires appropriate standardization of the procedures for sample preparation to avoid bias from sample handling.[38] Also, numerous factors, including age,[39] gender,[40, 41] ethnicity,[42] nutrition[43-45] and medical treatment may affect the metabolome and complicate the identification of relevant biomarkers. It is also important for all individuals included in the study to be exposed to a common environment,[46] and to maintain a similar level of physical activity,[47, 48] because these parameters can affect

Analyst Accepted Manuscript

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

the urinary metabolome. In contrast to animal studies, the standardization of such factors that influence the urinary metabolome is difficult in clinical studies.

Typical NMR spectra of urine samples from control and ASD individuals are shown in Fig. 1. Spectral ^{1}H assignments were made according to the values for chemical shifts reported in the literature, and in the human metabolome database (HMDB).[49] Urine spectra contained signals for low-molecular-weight metabolites, including amino-acids, organic acids, and carbohydrates derived from the diet and from microbial and human metabolism. In particular, the NMR spectra revealed the presence of endogenous metabolites in urine (creatine, creatinine; lactate, citrate, succinate, and formate as organic acids), methylamine compounds such as microbial-derived metabolites dimethylamine (DMA), trimethylamine (TMA) and trimethylamine-*N*-oxide (TMAO), aromatic metabolites (hippurate and phenylacetylglutamine (PAG)], and amino-acids [alanine, glycine, phenylalanine, tyrosine, *N*-methyl nicotinic acid (NMNA), and *N*-methyl nicotinamide (NMND)].

Please, insert Figure 1

Multivariate Statistical Analysis of the ^{1}H-NMR spectral data

To create a normal or Gaussian distribution of metabolites levels, the choice of scaling parameter is important, because it defines the relationships between variables. We chose Pareto scaling because it gives a greater weight to variables with large values than variables with small values. This contrasts with ‘unit variance’ scaling that forces all x values to have equal weight, irrespective of the starting intensity, and thus tends to enhance distortion from poor baseline and other spectral artefacts. We carried out Principal Component Analysis (PCA), which is an unsupervised classification technique (Fig. S1†), (R^2X(cum) = 0.61). This lack of discrimination between the two groups could indicate that the major source of variation in the data was not related to ASD. Instead, variation may be due to inter individual differences arising from a lack of standardization that could occur in clinical studies. Yap *et al.* reported similar findings for NMR data normalized to the total NMR spectral intensity.[21] PCA identifies the largest variations in the NMR data, but the latent variables (fundamental relationships) that allow the discrimination between ASD patients and controls did not necessarily show the largest variation.[37]

OPLS-DA is a regression model that reflects the correlation between multivariate data and dependent variables with class information[37], thereby minimizing any effects of non-relevant metabolite variability. Using ^{1}H-NMR data and OPLS-DA, we demonstrated differences in

Analyst Accepted Manuscript

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

the urine metabolite content ASD children and controls. The scatter plot scores for total area normalization were R^2X(cum)=0.19, R^2Y(cum)=0.70, Q^2=0.51 (Fig. 2). To assess the reliability of the OPLS-DA model, we applied CV-ANOVA which gave a p value of 8.37 10^{-8} (Table 1). The corresponding contribution score plot obtained from NMR data (Fig. S3†) showed differences in the urinary metabolic profiles of ASD patients and controls.

Unlike urine analysis for routine medical practice which is referenced to the creatinine concentration of the samples, metabolomics data are commonly normalized to the total molecule signals of the sample as detected by ^{1}H NMR analysis. In this study, we also compared the two spectral normalization methods: normalization to the creatinine peak or to total spectral intensity. The scatter plot scores for creatinine normalization were R^2X(cum)=0.36, R^2Y(cum)=0.69, Q^2=0.36, and CV-ANOVA p=2.1 10^{-4} (Figure S4†). These results suggest that there are metabolic differences between the two groups that can be distinguished irrespective of the spectral normalization used.

ASD patients had higher urinary levels of citrate, glycine, succinate, phenylacetylglutamine (PAG), formate and an unidentified compound “Und” (d= 0.88ppm), than controls. The urinary levels of creatine (Cr), 4-cresol sulfate (4-CS), hippurate (Hip), glutamate (Glu), 3-methyl-histidine (3-MH), trimethylamine-*N*-oxyde (TMAO) and dimethylamine (DMA), were lower in ASD patients than in controls. The 1D NMR analysis replicated some of the findings of 2D analysis, that was carried out with the same cohort,[30] including findings with respect to citrate, succinate, glycine and 3-methyl-histidine.

Please, insert Figure 2

Using the OPLS-DA model, we identified the most relevant variables, from the ^{1}H-NMR spectral data and tested their significance. Table 2 shows the medians and p values for the urinary metabolites. In particular, the urinary concentration of 3-methyl-histidine, succinate and glutamate differed significantly between the ASD patients and controls.

Please, insert Table 1

Please, insert Table 2

Multivariate Statistical Analysis of the combined ^{1}H-NMR and ^{1}H-^{13}C-HSQC-NMR

Analyst Accepted Manuscript

To improve the discrimination between ASD patients and controls, we fitted an alternative model combining the minimum number of variables from ^{1}H-NMR OPLS-DA model and ^{1}H-^{13}C HSQC model.[30] We obtained a good discrimination between ASD patients and controls in the OPLS-DA analysis. The scatter plot scores were R^2X(cum) = 0.14, R^2Y(cum) = 0.84, Q^2 = 0.71 (Fig. 3). Combining the spectra reduce the number of misclassified samples and resulted in a high sensitivity and specificity (Table 1). We used CV-ANOVA to assess the reliability of the OPLS-DA model, which gave a p value of $9.22\ 10^{-13}$.

Please, insert Figure 3

We analyzed classification performance by evaluating receiver operating characteristic (ROC) plots obtained by the ROCCET web-based tool.[17] The ROC curve for one model is constructed by plotting the true positive rate against the false positive rate. ROC results of the three OPLS-DA based models (^{1}H-NMR, ^{1}H-^{13}C-HSQC, and combined ^{1}H-NMR and ^{1}H-^{13}C-HSQC-NMR) are shown in Table 1. The area under the ROC curve (AUC) is an indicator of how well a given model can predict ASD (Fig. S5†). The combined ^{1}H-NMR and ^{1}H-^{13}C-HSQC model had the largest AUC (0.92) of the three models. This demonstrates the ability of this model to effectively discriminate between control and ASD samples. Indeed, this AUC value correspond to a prediction accuracy of 83.2%, which validates the proposed model structure.[17] These two complementary evaluations (predictive abilities and ROC curves) of the OPLS-DA models show that the model based combining spectra better discriminates ASD patients and controls than the model based on single spectral data.

Metabolites analysis

We also compared our findings to those of other studies involving metabolic profiles in ASD. Conflicting findings have been reported amongst these studies (Table 3).[16, 20-23, 25-27, 50-53]

Please, insert Table 3

We show that the urinary metabolite content of ASD patients differs from that of children without ASD. These finding agree with a previous study.[21] The multivariate models show that urinary levels of creatine, TMAO, hippurate and formate, were lower in ASD patients than in controls. Urinary levels of citrate, glycine, and PAG were higher in the autistic group than the control group. However, when these metabolites were tested individually, the difference in

Analyst Accepted Manuscript

1
2
3 their abundance between ASD and control urine samples was not statistically significant. The urinary concentration of succinate was significantly higher in ASD patients than in controls, whereas urinary concentrations of glutamate and 3-methyl-histidine were significantly lower in ASD patients than in control. The creatinine normalization can face some biological challenges because of changes in creatinine concentration caused by metabolic responses but in our population, results were similar when normalization was based on creatinine or total spectral intensity. To be in agreement with the clinical values in the literature, the concentrations are voluntarily expressed as relative concentration to creatinine (mM/μM of creatinine).

Kaluzna-Czaplinska *et al.*, by GC-MS analysis, reported high urinary concentrations of organic acids, such as citrate, in children with ASD.[24] These authors also reported an increase in the urinary concentrations of succinic acid in ASD children.[50] We found that the urinary concentrations of these two organic acids were higher in ASD patients than in controls, a finding that was also reported in study involving GC-MS.[16] High urinary concentrations of succinate is a marker of perturbation to the citric acid cycle, resulting in a deficiency in the production of cellular energy.[51] Consequently, the higher than normal urinary concentrations of citrate and succinate that we report for ASD patients suggests that ASD is associated with a disturbance to energy metabolism. Several studies have described an association between ASD and mitochondrial dysfunction.[54-56] Indeed ASD patients display peripheral markers of mitochondrial energy metabolism dysfunction, including elevated levels of lactate,[57] pyruvate and alanine in blood, urine, and/or cerebrospinal fluid.[58-60] A study involving phosphorus (^{31}P) nuclear magnetic resonance (NMR) spectroscopy demonstrated low abundance of phosphocreatine (PCr) and ATP levels in the frontal lobe of ASD patients.[61] This suggests that mitochondrial dysfunction in the central nervous system (CNS) is a feature of ASD. More recently Kubas B. *et al.*[62] used ^{1}H-MRS *in vivo* and revealed a lower ratio of Glx (glutamine+glutamate) to Cr in the frontal lobe region of autistic children than in the frontal lobe of healthy controls. Therefore perturbation to some metabolite ratios may contribute to the pathogenesis of autism. At 1.5-Tesla MRI, it is not possible to examine which compound (i.e., glutamate or glutamine) contributes most to the decrease in Glx. Joshi G. *et al.*[63] used 4-Tesla MRI and showed that in adolescent male with autism, that there was high abundance of glutamate in the anterior cingulate cortex but a low abundance of glutamate in the right medial temporal lobe. These observations support the glutamatergic dysregulation hypothesis in autism. In our study, the urinary glutamate concentration was lower in the ASD group than in controls. Several studies have shown that children with ASD show perturbations of amino-

Analyst Accepted Manuscript

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

acid metabolism.[52, 58] For example, concentrations of alanine, valine, leucine, asparate, glutamine and glutamate levels[52] are lower in autistic children than in controls.[64]

Clayton *et al.*[65] found that gut microbial metabolism of phenylalanine and tyrosine is associated with autism and suggested that this is involved in disease pathogenesis. These findings are in agreement with those of Yap *et al*.[21], and Kaluzna-Czaplinska J.[24] Similarly to Yap *et al.*, we observed perturbation to urinary concentrations of glycine and glycine-conjugated compounds, such as hippurate, in patients with ASD. This is consistent with the hypothesis of the involvement of gut microbial amino-acid metabolism in ASD. However, unlike Yap *et al.*, urinary concentrations of PAG were not remarkably low in patients with ASD.

Conclusion

The selection of metabolomics biomarkers that may be helpful for diagnosis of ASD has been complicated by conflicting findings amongst metabolomics studies. This is probably due to the large variety of biological fluids and analytical techniques used in metabolomics studies. From the list of metabolites implicated in autism, there may be one, or a few, that are relevant to the disease. In this NMR study, we used multivariate data analysis to reveal differences in the urinary concentrations of various metabolites between children with ASD and controls.

We used a combination of ^{1}H-NMR and ^{1}H-^{13}C HSQC NMR to analyse the metabolite content of urine. ^{1}H-NMR is quantitative and reproducible, and ^{1}H-^{13}C HSQC NMR can identify compounds with high accuracy. We show that combined use of these complementary spectroscopies improved classification. Furthermore using combined ^{1}H-NMR and ^{1}H-^{13}C-HSQC NMR and multivariate statistical techniques, we identified that an urinary metabolic profile of ASD was distinct from that of healthy controls and demonstrated strong predictive power for this disease.

Analyst Accepted Manuscript

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

Acknowledgments

This work was supported by the "Institut National de la Santé et de la Recherche" INSERM and the University François-Rabelais. We would like to thank the children and their parents/guardians who volunteered to participate in this study. We thank the center "Sésame Autisme Loiret" for their participation in this study.

References:

1. M. Wingate, B. Mulvihill, R. Kirby, S. Pettygrove, C. Cunniff, F. Meaney, E. Schulz, L. Miller, C. Robinson, G. Quintana, M. Kaiser, L. Lee, R. Landa, C. Newschaffer, J. Constantino, R. Fitzgerald, W. Zahorodny, J. Daniels, E. Giarelli, J. Pinto-Martin, S. Levy, J. Nicholas, J. Charles, J. Zimmerman, M. Maenner, M. Durkin, C. Rice, J. Baio, K. Van Naarden Braun, K. Phillips, N. Doernberg and M. Yeargin-Allsopp, *MMWR Surveill Summ*, 2012, **61**, 1-19.
2. W. H. Organization, *World Health Organization*, 2010.
3. A. P. Association, *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition, DSM-5*, Arlington, VA, 2013.
4. A. W. Zimmerman, *Autism - Current Theories and Evidence*, Humana Press, Berlin, 2008.
5. J. C. Lindon, J. K. Nicholson, E. Holmes, H. Antti, M. E. Bollard, H. Keun, O. Beckonert, T. M. Ebbels, M. D. Reily, D. Robertson, G. J. Stevens, P. Luke, A. P. Breau, G. H. Cantor, R. H. Bible, U. Niederhauser, H. Senn, G. Schlotterbeck, U. G. Sidelmann, S. M. Laursen, A. Tymiak, B. D. Car, L. Lehman-McKeeman, J. M. Colet, A. Loukaci and C. Thomas, *Toxicol Appl Pharmacol*, 2003, **187**, 137-146.
6. J. K. Nicholson, J. C. Lindon and E. Holmes, *Xenobiotica*, 1999, **29**, 1181-1189.
7. M. E. Bollard, E. G. Stanley, J. C. Lindon, J. K. Nicholson and E. Holmes, *NMR Biomed*, 2005, **18**, 143-162.
,8. L. Xiao-Jie, L. Zhen-Yu, L. Zhong-Feng, G. Xiao-Xia, Z. Yu-Zhi, S. Hai-Feng, Z. Li-Zeng, G. Xiao-Qing, D. Guan-Hua and Q. Xue-Mei, *Magnetic Resonance Chemistry*, 2012, **50**, 187-192.
9. H. Blasco, P. Corcia, C. Moreau, S. Veau, C. Fournier, P. Vourc'h, P. Emond, P. Gordon, P. F. Pradat, J. Praline, D. Devos, L. Nadal-Desbarats and C. R. Andres, *PloS one*, 2010, **5**, e13223.
10. W. T. Hu, A. Chen-Plotkin, S. E. Arnold, M. Grossman, C. M. Clark, L. M. Shaw, L. McCluskey, L. Elman, J. Karlawish, H. I. Hurtig, A. Siderowf, V. M. Lee, H. Soares and J. Q. Trojanowski, *Acta Neuropathol*, 2010, **120**, 385-399.
11. C. J. Sands, M. Coen, A. D. Maher, T. M. Ebbels, E. Holmes, J. C. Lindon and J. K. Nicholson, *Analytical chemistry*, 2009, **81**, 6458-6466.
12. Y. He, Z. Yu, I. Giegling, L. Xie, A. M. Hartmann, C. Prehn, J. Adamski, R. Kahn, Y. Li, T. Illig, R. Wang-Sattler and D. Rujescu, *Transl Psychiatry*, 2012, **2**, e149.
13. H. L. Cai, H. D. Li, X. Z. Yan, B. Sun, Q. Zhang, M. Yan, W. Y. Zhang, P. Jiang, R. H. Zhu, Y. P. Liu, P. F. Fang, P. Xu, H. Y. Yuan, X. H. Zhang, L. Hu, W. Yang and H. S. Ye, *J Proteome Res*, 2012, **11**, 4338-4350.
14. H. V. Ratajczak, *J Immunotoxicol*, 2011, **8**, 80-94.
15. L. Wang, M. T. Angley, J. P. Gerber and M. J. Sorich, *Biomarkers*, 2011, **16**, 537-552.

Analyst Accepted Manuscript

16. P. Emond, S. Mavel, N. Aidoud, L. Nadal-Desbarats, F. Montigny, F. Bonnet-Brilhault, C. Barthelemy, M. Merten, P. Sarda, F. Laumonnier, P. Vourc'h, H. Blasco and C. R. Andres, *Anal Bioanal Chem*, 2013, **405**, 5291-5300.
17. J. Xia, D. I. Broadhurst, M. Wilson and D. S. Wishart, *Metabolomics*, 2013, **9**, 280-299.
18. S. Bouatra, F. Aziat, R. Mandal, A. C. Guo, M. R. Wilson, C. Knox, T. C. Bjorndahl, R. Krishnamurthy, F. Saleem, P. Liu, Z. T. Dame, J. Poelzer, J. Huynh, F. S. Yallou, N. Psychogios, E. Dong, R. Bogumil, C. Roehring and D. S. Wishart, *PloS one*, 2013, **8**, e73076.
19. R. Powers, *Magnetic Resonance Chemistry*, 2013, **47**, S2-S11.
20. E. Zurawicz, J. Kaluzna-Czaplinska and J. Rynkowski, *Biomed Chromatogr*, 2013, **27**, 1273-1279.
21. I. K. Yap, M. Angley, K. A. Veselkov, E. Holmes, J. C. Lindon and J. K. Nicholson, *J Proteome Res*, 2010, **9**, 2996-3004.
22. J. Kaluzna-Czaplinska, M. Michalska and J. Rynkowski, *Acta Biochim Pol*, 2011, **58**, 31-34.
23. X. Ming, T. P. Stein, V. Barnes, N. Rhodes and L. Guo, *J Proteome Res*, 2012, **11**, 5856-5862.
24. J. Kaluzna-Czaplinska, *Clin Biochem*, 2011, **44**, 686-691.
25. R. Tirouvanziam, T. V. Obukhanych, J. Laval, P. A. Aronov, R. Libove, A. G. Banerjee, K. J. Parker, R. O'Hara, L. A. Herzenberg, L. A. Herzenberg and A. Y. Hardan, *J Autism Dev Disord*, 2012, **42**, 827-836.
26. W. J. Tu, H. Chen and J. He, *J Clin Biochem Nutr*, 2012, **51**, 248-249.
27. H. Kuwabara, H. Yamasue, S. Koike, H. Inoue, Y. Kawakubo, M. Kuroda, Y. Takano, N. Iwashiro, T. Natsubori, Y. Aoki, Y. Kano and K. Kasai, *PloS one*, 2013, **8**, e73814.
28. J. Xia, T. C. Bjorndahl, P. Tang and D. S. Wishart, *BMC Bioinformatics*, 2008, **9**, 507.
29. W. Gronwald, M. S. Klein, H. Kaspar, S. R. Fagerer, N. Nurnberger, K. Dettmer, T. Bertsch and P. J. Oefner, *Analytical chemistry*, 2008, **80**, 9288-9297.
30. S. Mavel, L. Nadal-Desbarats, H. Blasco, F. Bonnet-Brilhault, C. Barthelemy, F. Montigny, P. Sarda, F. Laumonnier, P. Vourc'h, C. R. Andres and P. Emond, *Talanta*, 2013, **114**, 95-102.
31. J. Xia and D. S. Wishart, *Nature protocols*, 2011, **6**, 743-760.
32. A. P. Association, *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision (DSM-IV-TR)*, Arlington, VA, 2000.
33. F. Savorani, G. Tomasi and S. B. Engelsen, *J Magn Reson*, 2010, **202**, 190-202.
34. Umetrics, 2012.
35. E. C. Chan, K. K. Pasikanti and J. K. Nicholson, *Nature protocols*, 2011, **6**, 1483-1499.
36. L. Eriksson, J. Trygg and S. Wold, *J Chemometr*, 2008, **22**, 594-600.
37. T. Gebregiworgis and R. Powers, *Comb. Chem. High Throughput Screen*, 2012, **15**, 595-610.
38. O. Beckonert, H. C. Keun, T. M. Ebbels, J. Bundy, E. Holmes, J. C. Lindon and J. K. Nicholson, *Nature protocols*, 2007, **2**, 2692-2703.
39. H. Gu, Z. Pan, B. Xi, B. E. Hainline, N. Shanaiah, V. Asiago, G. A. Gowda and D. Raftery, *NMR Biomed*, 2009, **22**, 826-833.
40. C. M. Slupsky, K. N. Rankin, J. Wagner, H. Fu, D. Chang, A. M. Weljie, E. J. Saude, B. Lix, D. J. Adamko, S. Shah, R. Greiner, B. D. Sykes and T. J. Marrie, *Analytical chemistry*, 2007, **79**, 6995-7004.
41. H. C. Bertram, J. O. Duus, B. O. Petersen, C. Hoppe, A. Larnkjaer, L. Schack-Nielsen, C.

Analyst Accepted Manuscript

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

Molgaard and K. F. Michaelsen, *Metabolism*, 2009, **58**, 1039-1045.
42. I. K. Yap, I. J. Brown, Q. Chan, A. Wijeyesekera, I. Garcia-Perez, M. Bictash, R. L. Loo, M. Chadeau-Hyam, T. Ebbels, M. De Iorio, E. Maibaum, L. Zhao, H. Kesteloot, M. L. Daviglus, J. Stamler, J. K. Nicholson, P. Elliott and E. Holmes, *J Proteome Res*, 2010, **9**, 6647-6654.
43. L. G. Rasmussen, H. Winning, F. Savorani, H. Toft, T. M. Larsen, L. O. Dragsted, A. Astrup and S. B. Engelsen, *Nutrients*, 2012, **4**, 112-131.
44. H. Winning, E. Roldan-Marin, L. O. Dragsted, N. Viereck, M. Poulsen, C. Sanchez-Moreno, M. P. Cano and S. B. Engelsen, *Analyst*, 2009, **134**, 2344-2351.
45. S. S. Heinzmann, C. A. Merrifield, S. Rezzi, S. Kochhar, J. C. Lindon, E. Holmes and J. K. Nicholson, *J Proteome Res*, 2012, **11**, 643-655.
46. F. Dieterle, A. Ross, G. Schlotterbeck and H. Senn, *Analytical chemistry*, 2006, **78**, 4281-4290.
47. C. Enea, F. Seguin, J. Petitpas-Mulliez, N. Boildieu, N. Boisseau, N. Delpech, V. Diaz, M. Eugene and B. Dugue, *Anal Bioanal Chem*, 2010, **396**, 1167-1176.
48. L. Le Moyec, L. Mille-Hamard, M. N. Triba, C. Breuneval, H. Petot and V. L. Billat, *Anal Bioanal Chem*, 2012, **404**, 593-602.
49. D. S. Wishart, D. Tzur, C. Knox, R. Eisner, A. C. Guo, N. Young, D. Cheng, K. Jewell, D. Arndt, S. Sawhney, C. Fung, L. Nikolai, M. Lewis, M. A. Coutouly, I. Forsythe, P. Tang, S. Shrivastava, K. Jeroncic, P. Stothard, G. Amegbey, D. Block, D. D. Hau, J. Wagner, J. Miniaci, M. Clements, M. Gebremedhin, N. Guo, Y. Zhang, G. E. Duggan, G. D. Macinnis, A. M. Weljie, R. Dowlatabadi, F. Bamforth, D. Clive, R. Greiner, L. Li, T. Marrie, B. D. Sykes, H. J. Vogel and L. Querengesser, *Nucleic Acids Res*, 2007, **35**, D521-526.
50. J. Kaluzna-Czaplinska, M. E. Zurawicz, W. Strucb and M. Markuszewski, *J. Chrom. B*, 2014, in press.
51. J. Kaluzna-Czaplinska, E. Socha and J. Rynkowski, *Nutr Res*, 2011, **31**, 497-502.
52. C. Evans, R. H. Dunstan, T. Rothkirch, T. K. Roberts, K. L. Reichelt, R. Cosford, G. Deed, L. B. Ellis and D. L. Sparkes, *Nutr Neurosci*, 2008, **11**, 9-17.
53. J. B. Adams, T. Audhya, S. McDonough-Means, R. A. Rubin, D. Quig, E. Geis, E. Gehn, M. Loresto, J. Mitchell, S. Atwood, S. Barnhouse and W. Lee, *Nutrition & metabolism*, 2011, **8**, 34.
54. J. J. Fillano, M. J. Goldenthal, C. H. Rhodes and J. Marin-Garcia, *J Child Neurol*, 2002, **17**, 435-439.
55. G. Oliveira, L. Diogo, M. Grazina, P. Garcia, A. Ataide, C. Marques, T. Miguel, L. Borges, A. M. Vicente and C. R. Oliveira, *Dev Med Child Neurol*, 2005, **47**, 185-189.
56. D. A. Rossignol and R. E. Frye, *Mol Psychiatry*, 2012, **17**, 290-314.
57. M. Coleman and J. P. Blass, *J Autism Dev Disord*, 1985, **15**, 1-8.
58. S. Aldred, K. M. Moore, M. Fitzgerald and R. H. Waring, *J Autism Dev Disord*, 2003, **33**, 93-97.
59. A. Laszlo, E. Horvath, E. Eck and M. Fekete, *Clin Chim Acta*, 1994, **229**, 205-207.
60. L. Palmieri and A. M. Persico, *Biochim Biophys Acta*, 2010, **1797**, 1130-1137.
61. N. J. Minshew, G. Goldstein, S. M. Dombrowski, K. Panchalingam and J. W. Pettegrew, *Biol Psychiatry*, 1993, **33**, 762-773.
62. B. Kubas, W. Kulak, W. Sobaniec, E. Tarasow, U. Lebkowska and J. Walecki, *Adv Med Sci*, 2012, **57**, 152-156.
63. G. Joshi, J. Biederman, J. Wozniak, R. L. Goldin, D. Crowley, S. Furtak, S. E. Lukas and A.

Analyst Accepted Manuscript

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

Gonenc, *Eur Arch Psychiatry Clin Neurosci*, 2013, **263**, 379-384.
64. L. H. Rolf, F. Y. Haarmann, K. H. Grotemeyer and H. Kehrer, *Acta Psychiatr Scand*, 1993, **87**, 312-316.
65. T. A. Clayton, *FEBS Lett*, 2012, **586**, 956-961.

Analyst Accepted Manuscript

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

ASSOCIATED CONTENT

Figures:

Figure 1: Typical ^{1}H-NMR spectra for urine from controls or ASD patients. Principal metabolites giving peaks in the spectrum: creatine, creatinine, dimethylamine (DMA), *N*-acetyl glycoprotein, trimethylamine-*N*-oxyde (TMAO), hippurate, 4-cresol sulfate (4-CS) phenylacetylglutamine (PAG), lactate, succinate, citrate, formate, alanine, glycine, glutamate (Glu).

Figure 2: Scatter plot of OPLS-DA scores obtained from ^{1}H-NMR spectra of urine samples from control children (grey dots) or autistic children (black squares). (R^2Y(cum)=0.70, Q^2(cum)=0.51, CV-ANOVA p=8.37 . 10^{-8})

Figure 3: Scatter plot of OPLS-DA scores obtained from combined ^{1}H-NMR and ^{1}H-^{13}C HSQC spectra of urine samples from control children (grey dots) or autistic children (black squares). (R^2Y(cum)=0.84, Q^2(cum)=0.71, CV-ANOVA p=9.22 . 10^{-13})

Tables

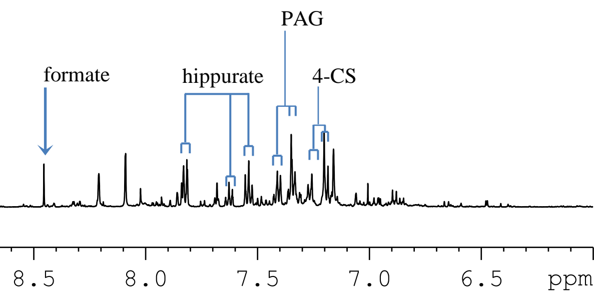
Table 1: Predictive abilities of the models constructed and classification results.

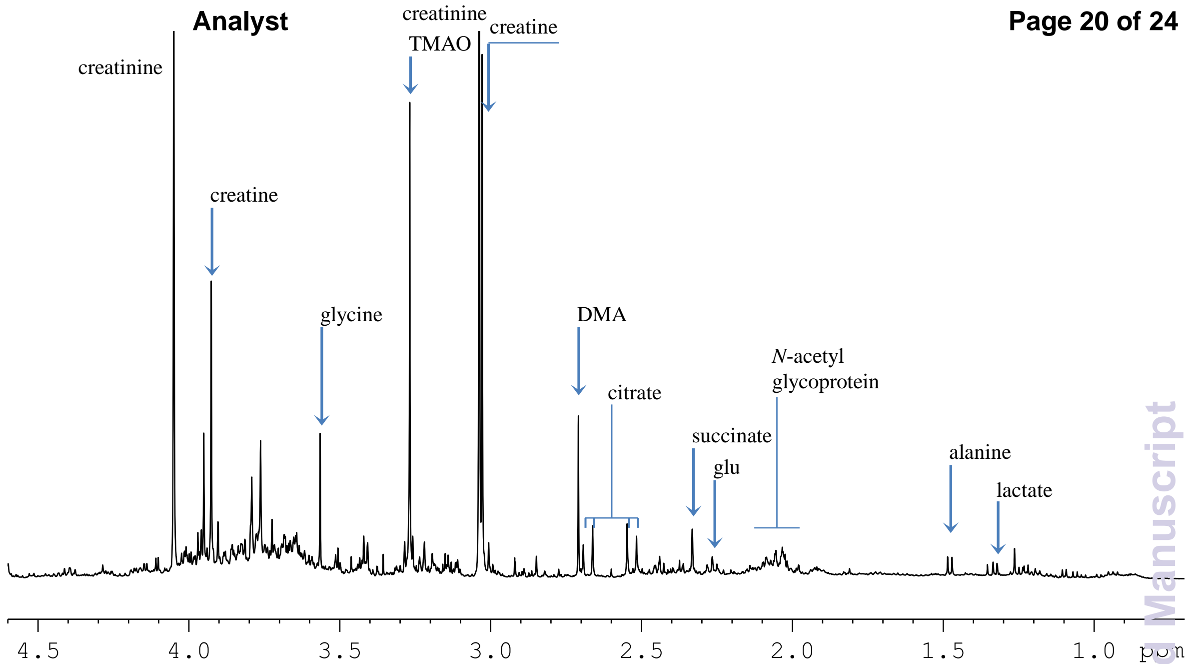
Table 2: Ratio of the concentration of a relevant metabolite to the methylene peak of creatinine. *P* values were calculated with the Mann-Whitney rank sum test. Relevant metabolites were determined as those most capable of distinguishing ASD patient from controls in the ^{1}H-NMR OPLS-DA model.

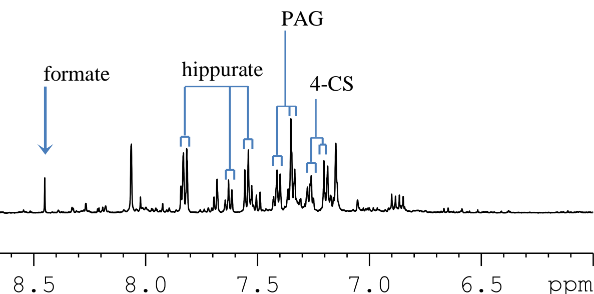
Table 3: Comparison of findings from different metabolomics studies. Listed studies that have used NMR or other analytical techniques to investigate differences in the urinary concentrations of metabolites in ASD patients and healthy subjects.

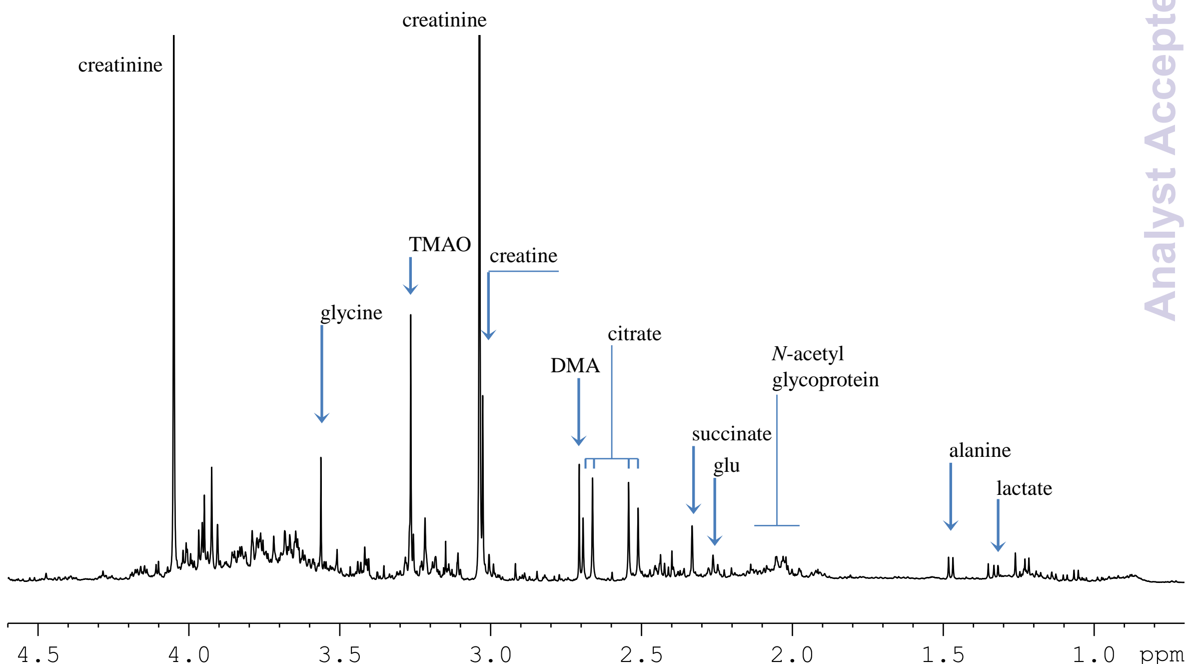
Analyst Accepted Manuscript

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









Analyst Accepted Manuscript

Fig 2 : Scatter plot of OPLS-DA scores obtained from ^{1}H-NMR spectra of urine samples from control children (grey dots) or autistic children (black squares). (R^2Y(cum)=0.70, Q^2(cum)=0.51, CV-ANOVA p=8.37 . 10^{-8})

R2X[1] = 0,0780601 R2X[XSide Comp. 1] = 0,111596
Ellipse: Hotelling T2 (0,95)

SIMCA-P+ 12.0.1 - 2013-11-07 11:55:39 (UTC+1)

Fig 3 : Scatter plot of OPLS-DA scores obtained from combined ^{1}H-NMR and ^{1}H-^{13}C HSQC spectra of urine samples from control children (grey dots) or autistic children (black squares). (R^2Y(cum)=0.84, Q^2(cum)=0.7[illegible], CV-ANOVA p=9.22 . 10^{-13})

R2X[1] = 0,0587383 R2X[XSide Comp. 1] = 0,0824334 Ellipse: Hotelling T2 (0,95)

SIMCA-P+ 12.0.1 - 2013-11-07 11:58:12 (UTC+1)

Analyst Accepted Manuscript

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

Table 1: Predictive abilities of the models constructed and classification results.

OPLS-DA Models	**Predictive Abilities**			**Misclassified**		**ROC curves**
	R^2Y(cum)	Q^2(cum)	CV-ANOVA	ASD samples *(Sn)*	Control samples *(Sp)*	AUC [95% CI] *(Average Accuracy)*
^{1}H-NMR	0.70	0.51	$8.37.10^{-8}$	4/30 *(86.6%)*	0/28 *(100%)*	0.91 [0.761-1] *(79%)*
^{1}H-^{13}C-HSQC	0.78	0.60	$7.77.\ 10^{-9}$	1/30 *(96.6%)*	2/28 *(92.8%)*	0.84 [0.707-0.965] *(74.8%)*
Combined ^{1}H-NMR and HSQC	0.84	0.71	$9.22.\ 10^{-13}$	0/30 *(100%)*	0/28 *(100%)*	0.92 [0.803-1] *(83.2%)*

Sn = Sensitivity (The number of diseased subjects that are correctly identified as diseased)
Sp = Specificity (The number of healthy subjects that are correctly identified as healthy)

Analyst Accepted Manuscript

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

Table 2 Ratio of the concentration of a relevant metabolite to the methylene peak of creatinine. *P* values were calculated with the Mann-Whitney rank sum test. Relevant metabolites were determined as those most capable of distinguishing ASD patient from controls in the ^{1}H-NMR OPLS-DA model.

Metabolites	Urine level in *μM/mM Creatinine* Median [interquartile range]		*p* score
	ASD	Control	
Und	51 [41-66]	54 [47-68]	0.534
Succinate	**28 [17-52]**	**17 [13-21]**	**<0.001**
Citrate	340 [266-476]	328 [261-432]	0.453
DMA	47 [41-57]	50 [41-82]	0.172
Creatine	158 [62-286]	205 [83-456]	0.164
TMAO	76 [52-135]	71 [53-143]	0.617
Glycine	138 [111-241]	157 [112-196]	0.612
3-MH	**45 [36-51]**	**53 [44-78]**	**0.014**
Hippurate	348 [223-626]	348 [270-701]	0.817
4-CS	96 [59-153]	90 [66-128]	0.705
Formate	36 [25-76]	52 [36-63]	0.336
Glutamate	**284 [212-325]**	**331 [287-376]**	**0.012**

Metabolite levels in urine (μM/mM Creatinine) are indicated as median values (with interquartile range i.e 25th and 75th percentiles in brackets). Non-parametric statistics were used due to lack of normal distribution for most of the metabolites. For metabolites indicated in bold, *p* values are below 0.05.

Analyst Accepted Manuscript

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

Table 3 : Comparison of findings from different metabolomic studies. Listed studies that have used NMR or other analytical techniques to investigate differences in the urinary concentrations of metabolites in ASD patients and healthy subjects.

Metabolites	Our study 1D NMR	Our study 1D + 2D[a] NMR	Yap IKS *et al.* 1D NMR[21]	Others analytical platforms
Succinate	↑	↑	↑	↑[16]; ↑.[50]
Citrate	(↑)	(↑)		↑[16];↑[22];↑.[50]
Glutamate	↓	↓	(↓)	↑[25]; ↓[26]; ↑[52]
Alanine	(↑)	(↑)	(↑)	↓[25]; ↓[23]; ↓[52]
Hippurate	(↓)	(↓)	(↓)	↓[16]; ↑[22]
Glycine	(↑)	(↑)	(↑)	(↓)[25]; ↓[23]; (↓)[52]
3-MH[b]	↓	↓		
Taurine		↑[a]	(↑)	↓[26];↓[23]; ↑[27]
Creatine	(↓)	↓	(↓)	
Histidine		↓[a]		↓[52]; ↓[23]

(Upward (downward) arrows indicate significantly higher (lower) urine metabolite concentrations in ASD patients than in controls. Upward (downward) arrows in brackets indicate metabolites with a trend toward higher (lower) concentrations in ASD patients than in controls).

[a] only metabolites identified in the HSQC analysis[30]

[b] 3-MH: 3-methyl-histidine

Analyst Accepted Manuscript