



HAL
open science

Evaluation and validation of an analytical approach for high-throughput metabolomic fingerprinting using direct introduction–high-resolution mass spectrometry: Applicability to classification of urine of scrapie-infected ewes

Estelle Rathahao-Paris, Sandra Alves, Nawel Boussaid, Nicole Picard-Hagen, Véronique V. Gayrard-Troy, Pierre-Louis Toutain, Jean-Claude Tabet, Douglas N. Rutledge, Alain Paris

► **To cite this version:**

Estelle Rathahao-Paris, Sandra Alves, Nawel Boussaid, Nicole Picard-Hagen, Véronique V. Gayrard-Troy, et al.. Evaluation and validation of an analytical approach for high-throughput metabolomic fingerprinting using direct introduction–high-resolution mass spectrometry: Applicability to classification of urine of scrapie-infected ewes. *European Journal of Mass Spectrometry*, 2019, 25 (2), pp.251-258. 10.1177/1469066718806450 . hal-02317795v2

HAL Id: hal-02317795

<https://agroparistech.hal.science/hal-02317795v2>

Submitted on 7 Nov 2019

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.



Evaluation and validation of an analytical approach for high-throughput metabolomic fingerprinting using direct introduction-high resolution mass spectrometry: applicability to classification of urine of scrapie-infected ewes

Journal:	<i>European Journal of Mass Spectrometry</i>
Manuscript ID	EJMS-18-0028.R2
Manuscript Type:	Jean-Claude Tabet honor issue
Date Submitted by the Author:	n/a
Complete List of Authors:	Rathahao-Paris, Estelle; INRA, UMR SPI ; Sorbonne Université - Faculté des Sciences et de l'Ingénierie, Institut Parisien de Chimie Moléculaire - UMR 8232 Alves, Sandra; Sorbonne Universités, Paris 06, IPCM-UMR 8232 Boussaid, Nawel; INRA-AgroParisTech Picard-Hagen, Nicole; Université de Toulouse, INRA, INP-ENVT, Toxalim Gayraud, Véronique; Université de Toulouse, INRA, INP-ENVT, Toxalim Toutain, Pierre-Louis ; Université de Toulouse, INRA, ENVT, INTHERES TABET , Jean-Claude; Sorbonne Université, Faculté des Sciences et de l'Ingénierie, Institut Parisien de Chimie Moléculaire (IPCM) Rutledge, Douglas; AgroParisTech, UMR Ingénierie Procédés Aliments Paris, Alain ; Muséum national d'Histoire naturelle, CNRS, Unité Molécules de Communication et Adaptation des Microorganismes (MCAM)
Keywords:	High-throughput metabolomics, Direct injection-mass spectrometry, High resolution mass spectrometry, Matrix effects, Data analysis
Abstract:	Direct injection-mass spectrometry (DIMS) can be used to perform high-throughput metabolomic fingerprinting. This work aims to evaluate a global analytical workflow in terms of sample preparation (urine sample dilution), high resolution detection (quality of generated data based on criteria such as mass measurement accuracy and detection sensitivity) and data analysis using dedicated bioinformatics tools. Investigation was performed on a large number of biological samples collected from sheep infected or not with scrapie. DIMS approach is usually affected by matrix effects, eventually hampering detection of some relevant biomarkers. Reference compounds were spiked in biological samples to help evaluate the quality of DIMS data produced by Fourier Transform mass spectrometry (FTMS). Despite the potential of high resolution detection, some drawbacks still remain. The most critical is the presence of matrix effects, which could be minimized by optimizing the sample dilution factor. The data quality in terms of mass measurement accuracy and reproducible intensity was evaluated. Good repeatability was obtained for the chosen dilution factor (i.e. 2000). More than 150 analyses were performed in less than 16 hours using the optimized DIMS approach. Discrimination of different status of sheep in relation to scrapie infection

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

	<p>(i.e. scrapie-affected, preclinical-scrapie or healthy) was obtained from the application of Shrinkage Discriminant Analysis (SDA) to the DIMS data. The most relevant variables related to this discrimination were selected and annotated.</p> <p>This study demonstrated that the choice of appropriated dilution faction is indispensable for producing quality and informative DIMS data. Successful application of DIMS approach for high throughput analysis of a large number of biological samples constitutes the proof of the concept.</p>

SCHOLARONE™
Manuscripts

For Peer Review

1
2
3 **Evaluation and validation of an analytical approach for high-throughput metabolomic**
4 **fingerprinting using direct introduction-high resolution mass spectrometry:**
5 **applicability to classification of urine of scrapie-infected ewes**
6
7

8
9 Estelle Rathahao-Paris^{1,2*}, Sandra Alves², Nawel Boussaid¹, Nicole Picard-Hagen³, Véronique
10 Gayrard³, Pierre-Louis Toutain⁴, Jean-Claude Tabet^{2,5}, Douglas N. Rutledge¹, Alain Paris⁶
11
12

13
14 ¹UMR Ingénierie Procédés Aliments, AgroParisTech, Inra, Université Paris-Saclay, 91300 Massy, France

15 ²Sorbonne Université, Faculté des Sciences et de l'Ingénierie, Institut Parisien de Chimie Moléculaire (IPCM),
16 75005 Paris, France

17
18 ³Toxalim, Université de Toulouse, INRA (Institut National de la Recherche Agronomique), INP (Institut
19 National Polytechnique de Toulouse) –ENVT (Ecole Nationale Vétérinaire de Toulouse), 31076 Toulouse,
20 France

21
22 ⁴INTHERES, Université de Toulouse, INRA, ENVT, 31076 Toulouse, France

23 ⁵CEA-INRA, Service de Pharmacologie et d'Immunoanalyse, Laboratoire d'Etude du Métabolisme des
24 Médicaments, MetaboHUB, Université Paris-Saclay, Gif-sur-Yvette cedex, France

25
26 ⁶Unité Molécules de Communication et Adaptation des Microorganismes (MCAM), Muséum national d'Histoire
27 naturelle, CNRS, CP54, 57 rue Cuvier, 75005 Paris, France
28
29

30
31
32
33 * Corresponding author: Dr. Estelle Rathahao-Paris

34 Present address: UMR CEA-INRA Service de Pharmacologie et d'Immunoanalyse, Laboratoire d'Immuno-
35 Allergie Alimentaire, CEA de Saclay - Bat 136, F-91191, Gif-sur-Yvette cedex, France.

36 E-mail: Estelle.Paris@inra.fr
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Abstract

Direct injection-mass spectrometry (DIMS) can be used to perform high-throughput metabolomic fingerprinting. This work aims to evaluate a global analytical workflow in terms of sample preparation (urine sample dilution), high resolution detection (quality of generated data based on criteria such as mass measurement accuracy and detection sensitivity) and data analysis using dedicated bioinformatics tools. Investigation was performed on a large number of biological samples collected from sheep infected or not with scrapie.

DIMS approach is usually affected by matrix effects, eventually hampering detection of some relevant biomarkers. Reference compounds were spiked in biological samples to help evaluate the quality of DIMS data produced by Fourier Transform mass spectrometry (FTMS). Despite the potential of high resolution detection, some drawbacks still remain. The most critical is the presence of matrix effects, which could be minimized by optimizing the sample dilution factor. The data quality in terms of mass measurement accuracy and reproducible intensity was evaluated. Good repeatability was obtained for the chosen dilution factor (*i.e.* 2000). More than 150 analyses were performed in less than 16 hours using the optimized DIMS approach. Discrimination of different status of sheeps in relation to scrapie infection (*i.e.* scrapie-affected, preclinical-scrapie or healthy) was obtained from the application of Shrinkage Discriminant Analysis (SDA) to the DIMS data. The most relevant variables related to this discrimination were selected and annotated.

This study demonstrated that the choice of appropriated dilution faction is indispensable for producing quality and informative DIMS data. Successful application of DIMS approach for high throughput analysis of a large number of biological samples constitutes the proof of the concept.

Keywords: High-throughput metabolomics, Direct injection-mass spectrometry, High resolution mass spectrometry, Matrix effects, Data analysis

Introduction

Metabolomics¹⁻² refers to the large-scale study of small-molecule metabolites present in a biological system (*e.g.*, cell, tissue or biofluid). It studies metabolic profile changes, which may reflect metabolic disruptions induced by environmental stress or diseases, with the objective of detecting metabolic biomarkers. Nuclear magnetic resonance (NMR)³ and mass spectrometry (MS)⁴⁻⁷ are the analytical technologies generally used to produce metabolomic data. These techniques enable simultaneous analysis of a large set of metabolites in a single experiment. Although NMR has demonstrated its high potential in measurement reproducibility, its sensitivity remains a weak point compared to that of MS. Thanks to its sensitivity, selectivity and specificity, MS has become an efficient analytical platform in metabolomics. Its coupling to a separation technique like gas chromatography (GC),⁸ liquid chromatography (LC)⁹ or capillary electrophoresis (CE)¹⁰ is usual practice, providing separation of isobaric and/or isomeric compounds and also reduction of matrix effects. However, such coupling techniques involve relatively long analysis times, limiting the throughput of analyses, especially technical drifts occur when performing large scale analyses such as in cohort studies.

Direct injection mass spectrometry (DIMS) is an attractive alternative to perform high-throughput metabolomic analyses.¹¹ For a long time direct analysis of biological samples has been successfully performed using pyrolysis mass spectrometry^{12,13} as well as electrospray ionization coupled to mass spectrometry (ESI-MS).¹⁴⁻¹⁹ ESI is a soft desorption/ionization method well adapted for polar molecule analysis, providing controlled and limited in-source fragmentations of charged molecular species.²⁰ It is one of the most suitable techniques for high throughput DIMS approach and the most often used in metabolomics because of the polar character of most metabolites.¹⁴⁻¹⁹

DIMS usually involves high resolution mass spectrometry (HRMS).^{11,18,21,22} HRMS technology offers a way to replace conventional chromatographic separation of molecules by compound distinction according to their accurate ion *m/z* values. Typically, Fourier transform mass spectrometer (FTMS) is used, providing **high mass resolving power** and distinction of isobaric species.²³ Its capacity to accurately measure *m/z* values with ppm error levels facilitates the assignment of elemental compositions of biomarker candidates. In addition, high dynamic mass range of FTMS instrumentation enables simultaneous detection of large number of metabolites present over a wide range of concentrations. DIMS approach using ESI source and FTMS detection has been successfully employed in several fields, *e.g.*, biological

1
2
3 fluid analysis,^{24,25} plant metabolomic studies²⁶ and microbial applications.²⁷ Although DIMS
4 approach does not allow exhaustive analysis of all metabolites present in a biological sample
5 even using the most powerful FTMS technology, detection of only metabolic profile change
6 (*i.e.* some discriminant metabolites or specific endogenous compound signal variations)
7 should be sufficiently indicative to qualify a global metabolic disruption or some coordinated
8 metabolic changes associated to the progression of a disease. ESI is highly susceptible to
9 matrix effects, *i.e.* signal suppression or enhancement phenomena due to matrix components,
10 which alter the ionization efficiency and can hamper detection of some (eventually
11 discriminant) compounds.^{28,29} Thus, the use of ESI technique for direct analysis of complex
12 mixtures is very challenging. The first step of analytical workflow implies the choice of the
13 best sample preparation conditions.³⁰ The separation of matrix components from target
14 analyte(s) could be done during the sample preparation or purification step to limit these
15 matrix effects. However, these processes are less suitable for metabolomic fingerprinting,
16 which aims, as much as possible, for an exhaustive detection of metabolites. Nevertheless, for
17 urine samples the dilution constitutes an easy way of minimizing, although not totally
18 eliminating the matrix effects.

19 The objective of this work is to demonstrate the efficiency of DIMS approach using ESI
20 technique and an LTQ-Orbitrap instrument³¹ for acquiring high-throughput metabolomic
21 fingerprints from a large set of biological samples. The evaluation of different critical aspects
22 such as selectivity, specificity, linearity, sensitivity and repeatability of analytical method was
23 first performed. Finally, a large number of biological samples (*i.e.* more than 150 samples)
24 collected in sheep infected or not with scrapie were analyzed using our direct approach. The
25 generated spectral data were subsequently analyzed using multivariate statistical analysis,
26 providing the discrimination of animals according to their health status (healthy control,
27 preclinical and clinical scrapie-infected sheeps).

28 **Materials and Methods**

29 ***Animals***

30 Urines were collected from healthy, preclinical and scrapie-infected ewes with a diagnosis
31 confirmation as published previously.³² Samples were stored at -20°C until analysis.

32 ***Material***

33 HPLC-grade methanol was purchased from VWR International S.A.S. (Fontenay-sous-Bois,
34 France) and ultrapure water was provided by a Milli-Q system (Millipore, Molsheim, France).

1
2
3 Three reference compounds were used to evaluate the quality of data produced by DIMS:
4 isoniazid ($C_6H_7N_3O$, Mw 137.0589 u) and oxytetracycline ($C_{22}H_{24}N_2O_9$, Mw 460.1482 u)
5 were purchased from Sigma–Aldrich Chimie (Saint-Quentin Fallavier, France) and 2-amino-1-
6 methyl-6-phenylimidazo [4,5-b] pyridine (PhIP, $C_{13}H_{12}N_4$, Mw 224.1062 u) was obtained
7 from Toronto Research Chemicals (Ontario, Canada).
8
9

10 All urine samples were centrifuged and the supernatants were collected. First two urine
11 samples (one from a healthy sheep and another from a scrapie-affected sheep) were diluted in
12 CH_3OH/H_2O (50:50, v/v) solvent mixture at different dilution factors (*i.e.* 100, 500, 1000,
13 2000, 3000 and 5000). Based on the preliminary results from dilution study, all urine samples
14 were prepared in the same conditions at a fixed dilution factor (*i.e.* 1:2000). All samples were
15 spiked with the three reference compounds at final concentrations of 5 ng/ μ L for isoniazid
16 and oxytetracycline, and 0.05 ng/ μ L for PhIP.
17
18
19
20
21
22
23

24 Analytical methods

25 Mass spectrometry experiments were performed using an LTQ-Orbitrap-XL hybrid
26 instrument (Thermo Fisher Scientific, Villebon sur Yvette, France) equipped with an ESI
27 source operated in positive ion mode.³³
28
29

30 All mass spectra were recorded in the 100-1000 m/z range. The Ion Trap and FTMS Full
31 AGC (automatic gain control) target values were set at 1×10^4 and 2×10^5 , respectively and
32 the ion trap and FTMS Full Max ion times at 50 ms and 200 ms, respectively. Mass
33 measurements were performed with a theoretical mass resolving power of 100,000 FWHM
34 (full width at half maximum) at m/z 400. An internal calibration mode (“lock mass”) was
35 applied based on impurity ions at m/z 236.07157 corresponding to $[M+Na]^+$ ions of N-butyl
36 benzene-sulfonamide.³⁴
37
38
39
40

41 Positive ion source conditions were as follows: electrospray needle, -4 kV; heated transfer
42 capillary temperature, 250°C; heated transfer capillary voltage, 30 V, tube lens offset, 100 V;
43 sheath gas at a flow rate, 40 % (arbitrary units).
44
45

46 Flow injection analysis (FIA) mode was used to introduce samples into the mass spectrometer
47 through a HP 1100 HPLC system constituting of a quaternary pump and an auto-sampler
48 (Agilent, Les Ulis, France). 10 μ L of sample were injected in the flow of methanol/water
49 isocratic gradient (50:50, v/v) at 30 μ L.min⁻¹. Mass spectra were recorded for 5 min,
50 including a 2.5 min automatic washing step, consisting of an increase of the flow rate to
51 200 μ L/min after a 2.5 min run.
52
53
54
55
56
57
58
59
60

Data processing methods

A program written in the R 2.15.0 (<http://www.R-project.org>) was used to process DI-HRMS data. From the raw mass spectra, a single large data matrix was generated.

Data filtering steps were performed to eliminate non-relevant variables and to limit the presence of missing values from the resulting matrix. Only ions present in most of the major part of samples were selected, *i.e.* the tail part of the distribution curve when a power law distribution is used.

The resulting data matrix was then processed using Shrinkage Discriminant Analysis³⁵ in order to distinguish different status of sheeps according to scrapie infection. Canonical correlation analysis was also applied to detect the most discriminant variables. Putative ion annotation was done using measured accurate m/z values to make queries through Human Metabolome Database (HMDB)^{36,37} or METLIN.³⁸

Results and discussion

1. Sample dilution as an efficient way to minimize matrix effects

The determination of the best conditions for sample preparation was performed in order to minimize matrix effects and to obtain a sensitive MS response from analysis of a complex mixture containing a wide concentration range of metabolites. Two urine samples constituted the quality control (QC) samples used for evaluation and validation of the analytical conditions (see experimental section). They were prepared at different dilution levels and spiked with the same volume of the solution containing the three reference compounds. Mass spectra generated from DIMS approach and FTMS detection are shown in **Figure 1**. Only one spiked reference compound, *i.e.* isoniazid (STD1) was detected at m/z 138.0658 with low abundance (7%) for sheep urine diluted one hundred times (**Figure 1a**), reflecting strong matrix effects through the ion suppression phenomenon. Ionization competition probably occurred between matrix components present at high levels and the reference compounds to the detriment of the ionization of the latter. The same phenomenon could also impact the ionization of some endogenous compounds present in the matrix, including potential metabolic biomarkers.^{39,40} The 1:2000 dilution appears to be the best, providing MS response better than other dilution samples as shown in **Figure 1b**. In addition to the three reference compound ions (*i.e.* isoniazid (or STD1), PhIP (or STD2) and oxytetracycline (or STD3) detected as $[M+H]^+$ species at m/z 138.0658, m/z 225.1133 and m/z 461.1548, respectively), a larger number of other intense peaks were detected. At larger dilution factors (*i.e.* 3000 and 5000) the ions formed from the two reference STD1 and STD3 compounds (m/z 138.0657

1
2
3 and m/z 461.1548) became the most abundant ions, reflecting too high dilution conditions
4 **(Figure 1c)**.

5
6 **(Figure 1 near here)**

7 Note that the matrix effects could not be easily detected without addition of the reference
8 compounds. However, the use of these reference compounds to optimize the analytical
9 conditions does not totally guarantee the detection of all endogenous species as some of them
10 could still undergo the matrix effects.
11

12
13
14 An interesting way to assess the global analytical response is to examine the relative peak
15 intensity variations for most ions (including both reference and endogenous compounds)
16 detected in the mass spectra recorded for a given urine sample at different concentrations
17 **(Figure 1d)**. The relative intensity values should be less affected by analytical drifts
18 compared to absolute abundances and representative of the mass spectrum profile. The
19 relative abundances of the most abundant ions increased with increasing dilution factor until
20 2,000, indicating the reduction of the matrix effects. Indeed, the largest number of ions with
21 relative intensities higher than 10 % was obtained for the 1:2000 dilution. Higher sample
22 dilution resulted in a decrease in relative intensities for most ions, reflecting the dilution
23 effects.
24

25
26
27 Alternatively, relative intensities of some ions were plotted as a function of sample
28 concentration to examine ion behavior through sample dilution. Different trends are observed
29 in the **Figure 2**. This suggests that the matrix effects depend on the ionic species considered
30 and thus on their structure. For example, some ions having relatively high abundance at the
31 1:100 dilution underwent strong ion suppression phenomena at the 1:500 dilution.
32 Nevertheless, most species displayed a bell-shaped curve with the maximum intensity at a
33 dilution factor of 2,000 as confirmed in **Figure 2b**, which shows the total ion current at
34 different dilution factors.
35

36
37 Investigation was also done for another urine sample and similar results were obtained (see
38 **Figure S1 of the supplementary information**), confirming the choice of the dilution factor at
39 2,000 and the representativeness of the sample set.
40

41
42 Note that sample preparation conditions depend on the sample type (e.g., biofluids or tissues)
43 and the origin of species (e.g., humans or animals). Thus, each type of matrix should be
44 systematically investigated.
45

46
47 Finally, a 1:2000 dilution was applied for performing metabolomic fingerprinting of sheep
48 urines.
49

1
2
3 (Figure 2 near here)
4
5

6 2. Evaluation of the quality of data produced by DI-HRMS

7 The use of HRMS instruments provides high selectivity and specificity of detection. In
8 addition, a large number of species, among them most isobaric ions can be detected thanks to
9 the high mass resolving power and high dynamic mass range of FTMS instruments. Although
10 sample dilution minimized matrix effects, the quality of data produced from sample diluted at
11 the chosen dilution factor should be examined based on criteria of linearity and repeatability
12 of analytical response.
13
14
15
16
17

18 a) Linearity

19 Metabolomics aims to detect metabolic profile changes. The measurements of metabolite
20 levels in biological samples should provide information about metabolic function and
21 evidence eventual metabolic disruptions. Hence, analytical response of every metabolite
22 should be representative of its concentration.
23
24
25
26

27 The relative ion abundance variations were examined for randomly selected ions according to
28 sample concentration (Figure 3). Good linearity was observed only under high dilution
29 conditions (*i.e.* dilution higher than 1:2000) with satisfactory correlation coefficient values for
30 most of the selected ions whatever their abundances (as shown in Table S1 of supplementary
31 information). Under low dilution conditions, non-linearity was observed, probably due to the
32 presence of matrix effects which impact quantitative response. Such observations confirmed
33 that the best dilution factor was 2,000.
34
35
36
37
38
39

40 (Figure 3 near here)
41
42

43 b) Repeatability of the method

44 Repeatability of the method used was also evaluated by repeating 20 injections of a same
45 urine sample (diluted at 1:2000 and spiked with the three reference compounds). To do this,
46 mass spectrum data quality was assessed according to spectral data features, *i.e.* peak intensity
47 and accurate m/z value.
48
49
50

51 The first criterion was examined by plotting the relative standard deviations (RSD) of
52 absolute intensity of fifty most abundant ions from the twenty repeated analysis of the same
53 sample (see Figure S2 of the supplementary information). Most RSD values were in the
54 range of about 20 %. For the twenty five most abundant ions (*i.e.* with relative intensity
55
56
57
58
59
60

1
2
3 greater than 10%), the RSD values were below 10 % (except for one peak). Of course, greater
4 variations of peak intensities are unavoidable for lower abundance ions.

5
6 Similar RSD values have been reported with the use of DIMS approach on a 7-T FTICR
7 instrument.^{41,42} Kirwan and co-workers⁴¹ obtained a value of 15.9 % as the median RSD of
8 peak intensity after application of computational tools for correcting intra and inter-batch
9 variations based on the QC spectra. Similar median RSD values were also obtained without
10 any prior computational correction using an instrument equipped with a dynamically
11 harmonized Fourier transform ion cyclotron resonance cell.⁴² Here, our results using LTQ-
12 Orbitrap detection are in the same range.
13
14
15
16
17
18

19 **3. Application of high throughput metabolomics to biological samples**

20 The global metabolomic workflow includes the optimization of sample preparation
21 conditions, the choice of an analytical platform but also bioinformatic tools for pre-processing
22 metabolomic data and for selecting biologically-relevant metabolites related to the group
23 discrimination.
24
25

26 The quality of accurate mass measurements was examined for all data based on measured
27 accurate m/z values of the spiked reference compound species. The use of internal calibration
28 (lock mass) was also tested, providing improvement in the mass measurement accuracy, e.g.,
29 for m/z 461.1548 protonated oxytetracycline 1.5 ppm mass error was obtained in external
30 calibration and error below 0.5 ppm in internal calibration, data not shown). The 'lock mass'
31 mode was then applied for analysing all biological samples (**Figure S3 of the supplementary
32 information**). A good stability of mass measurement accuracy was also obtained all along the
33 analysis time of 159 biological samples during approximately 16 hours. Mass error was
34 around 3 ppm for protonated isoniazid at m/z 138.0658 and below 1 ppm for the two other
35 reference compound ions at m/z 225.1133 and m/z 461.1548 from oxytetracycline and PhIP,
36 respectively.
37
38
39

40 After generating sheep urine fingerprints, the first step of data analysis is the pre-processing,
41 which consists in transforming the raw data into a matrix suitable for subsequent statistical
42 and/or chemometric analyses.⁴³ Generated data are often affected by measurement errors due
43 to technical variability in terms of mass accuracy drifts and peak intensity variations from
44 scan-to-scan and between injections. The presence of missing values and the detection of non-
45 relevant variables are not favourable for the subsequent statistical analysis. Non-relevant
46 signals could arise from artefact signals (e.g., due to Fourier Transform processing of the
47 primary signal) or contaminant compounds (e.g., sample preparation contaminants).
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 Otherwise, there are also redundant signals corresponding to different species (*i.e.* adduct or
4 in-source fragment ions) coming from the same compound and usually formed when using
5 atmospheric pressure ionization techniques. To eliminate missing values and non-relevant
6 variables, a data filtering step was applied to retain only m/z values detected in most of the
7 biological samples at a given percentage (*i.e.* 85 %). This approach is similar that used by
8 Taylor et al.⁴⁴

9
10
11 Here, the distribution of ions was investigated from the data set obtained from urinary
12 samples collected on 159 different animals. Surprisingly, from an initial population of 9684
13 ions only 122 ions were detected in almost all samples, *i.e.* with 15% or less of the "non-
14 availability" in the resulting matrix (**Figure S4 of the supplementary information**). The global
15 distribution of this ion population follows a power law distribution. The 122 variables were
16 selected in the heavy tail of such a distribution and used for subsequent statistical analysis
17 using a supervised method for discriminant analysis, *i.e.* Shrinkage Discriminant Analysis
18 (SDA)³⁵. SDA was applied to a prior corrected dataset using PLS-based orthogonal signal
19 correction performed with Simca-P v10.0 (Umetrics, Umea, Sweden) with one component
20 considering the health status of animals (3 groups).

21
22 The score plot reported in the **Figure 4** shows a visualization of all samples in a two
23 dimensional space. A discrimination between the three animal classes (*i.e.* healthy,
24 preclinical-scrapie and scrapie-affected status) was obtained. For example, on the second
25 component the healthy animal group is separated from the scrapie-infected one. A leave-one-
26 out procedure was used to estimate the misclassification ratio at 4.5%, indicating an overall
27 metabolic response able to distinguish the animal status. Besides the plot of the probabilities
28 coming from the transformation of the Mahalanobis distances⁴⁵ between every statistical
29 individual and the barycenter of the 3 phenotype groups (see **Figure S5 of the supplementary
30 information**) can detect some eventual outliers from the data set. The classification property
31 of SDA comes from the efficient shrinkage procedure applied on the correlation matrix, the
32 covariance matrix and the prior probabilities. Here, few misclassified individuals, *i.e.* only 7,
33 were detected among all individuals (see **Figure S5 of the supplementary information**).

34
35 Additionally, the canonical correlation analysis was performed to detect more precisely
36 variables related to every health status (**Figure S6 of the supplementary information**). The
37 experimental accurate m/z values of these discriminating variables were used to query freely
38 available databases (DBs) for putative annotation (**Table S2 of the supplementary
39 information**). Several variables (*i.e.* different m/z values) could correspond to a same
40 metabolite, which reinforce the compound identity assignment. For example, three
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 discriminating variables could be putatively annotated as different forms of creatinine ions: i)
4 the protonated ions at m/z 114.0656 annotated as $C_4H_8ON_3$ with 5.0 ppm error, ii) their
5 corresponding ^{13}C isotope detected at m/z 115.068 and annotated as $^{13}C_1C_3H_8ON_3$ with 4.3
6 ppm error and iii) the m/z 152.0217 $[M+K]^+$ adducts annotated as $C_4H_7ON_3K$ with 2.5 ppm
7 error. From the fifteen most discriminating variables, nine candidates could be proposed with a
8 mass measurement accuracy below 5.0 ppm, leading to the putative assignment of the identity
9 of six following biomarkers: creatine, creatinine, dideoxyhexitol, glycyl-lysine, methyl-
10 hippuric acid and octanetriol. Note that according to the canonical correlation protonated
11 creatinine ions (including monoisotopic peak at m/z 114.066 and its (M+1) isotope at m/z
12 115.068) display a positive correlation with scrapie-infected status whereas negative values
13 were observed with samples from healthy animals or those at preclinical stage.
14
15
16
17
18
19
20
21
22
23

24 **(Figure 4 near here)**

25 26 27 28 **Conclusions**

29 This study demonstrates that the investigation for optimizing sample dilution factor is
30 essential to minimize matrix effects in order to produce quality and informative DIMS data.
31 High throughput analyses can be efficiently performed only after this optimization step. In our
32 study, discrimination of different health status (clinical or preclinical-scrapie and healthy) **was**
33 **obtained** from the generated urinary fingerprints using Shrinkage Discriminant Analysis. This
34 study constitutes a proof of concept of the feasibility of DI-HRMS approach for high
35 throughput metabolomics at large scale.
36
37
38
39

40 Our approach enabled acquisition of more than 150 metabolomic fingerprints in less than 16
41 hours. Such a short analysis time cannot be obtained when using LC/MS approach, even if
42 recent ultra-high-performance liquid chromatography (UHPLC) systems allow the decrease of
43 LC/MS experiments duration. Rapid production of metabolomic fingerprintings is essential
44 when analyzing a large number of samples to limit technical drifts which can impact the
45 subsequent statistical analysis.
46
47
48
49

50 Of course, such DIMS approach using low resolution instrument will provide less informative
51 data (no distinction of isobaric ions). Instead, the use of higher performance instrumentation
52 in terms of mass measurement accuracy and mass resolving power enables separation of very
53 close isobaric ions and improves data quality. More, the determination of isotopic fine
54
55
56
57

1
2
3 structures consisting in the identification of the different isotopologues of M+1 and M+2
4 isotopic clusters is very useful for metabolomic investigations, particularly for the metabolite
5 identity assignment.
6
7

8
9 **Compliance with ethical standards**

10 **Conflict of interest:** The authors declare that they have no conflict of interest.

11
12 **Ethical approval:** All procedures performed in studies involving animals were in accordance
13 with the ethical standards of the institution and the French legal requirements regarding the
14 protection of laboratory animals.
15
16

17
18 **Acknowledgement:** The “Région Midi-Pyrénées” is acknowledged for the financial support
19 of this exploratory study.
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

References

- ¹ Oliver SG, Winson MK, Kell DB, et al. Systematic functional analysis of the yeast genome. *Trends Biotechnol* 1998; 16: 373-378.
- ² Goodacre R, Vaidyanathan S, Dunn WB, et al. Metabolomics by numbers: acquiring and understanding global metabolite data. *Trends Biotechnol* 2004; 22: 245-252.
- ³ Nicholson JK, Lindon JC and Holmes E. 'Metabonomics': understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. *Xenobiotica* 1999; 29: 1181-1189.
- ⁴ Bedair M and Sumner LW. Current and emerging mass-spectrometry technologies for metabolomics. *Trends Anal Chem* 2008; 27: 238-250.
- ⁵ Dettmer K, Aronov PA and Hammock BD. Mass spectrometry-based metabolomics. *Mass Spectrom Rev* 2007; 26: 51-78.
- ⁶ Dunn WB, Bailey NJ and Johnson HE. Measuring the metabolome: current analytical technologies. *Analyst* 2005; 130: 606-625.
- ⁷ Dunn WB. Current trends and future requirements for the mass spectrometric investigation of microbial, mammalian and plant metabolomes. *Phys Biol* 2008; 5: 011001.
- ⁸ Kanani H, Chrysanthopoulos PK, Klapa MI. Standardizing GC-MS metabolomics. *J Chromatogr B* 2008; 871: 191-201.
- ⁹ Zhou B, Xiao JF, Tuli L, Resson HW. LC-MS-based metabonomics. *Mol Biosyst* 2012;8: 470-81.
- ¹⁰ Ramautar R, Somsen GW and de Jong GJ. CE-MS for metabolomics: developments and applications in the period 2010-2012. *Electrophoresis* 2013; 34: 86-98.
- ¹¹ Habchi B, Alves S, Paris A, et al. How to really perform high throughput metabolomics analyses efficiently? *Trends Anal Chem*. 2016; 85: 128-139.
- ¹² Goodacre R and Kell DB. Pyrolysis mass spectrometry and its applications in biotechnology. *Cur Opin Biotechnol* 1996; 7: 20-28.
- ¹³ Dumas ME, Debrauwer L, Beyet L, et al. Analyzing the Physiological Signature of Anabolic Steroids in Cattle Urine Using Pyrolysis/Metastable Atom Bombardment Mass Spectrometry and Pattern Recognition. *Anal Chem* 2002; 74: 5393-5404.
- ¹⁴ Goodacre R, York EV, Heald JK, et al. Chemometric discrimination of unfractionated plant extracts analyzed by electrospray mass spectrometry. *Phytochemistry* 2003; 62: 859-863.
- ¹⁵ Kaderbhai NN, Broadhurst DI, Ellis DI, et al. Functional genomics via metabolic footprinting: monitoring metabolite secretion by *Escherichia coli* tryptophan metabolism mutants using FT-IR and direct injection electrospray mass spectrometry. *Comp Funct Genomics* 2003; 4: 376-391.
- ¹⁶ Castrillo JI, Hayes A, Mohammed S, et al. An optimized protocol for metabolome analysis in yeast using direct infusion electrospray mass spectrometry. *Phytochemistry*. 2003; 62: 929-937.
- ¹⁷ Koulman A, Tapper BA, Fraser K, et al. High-throughput direct-infusion ion trap mass spectrometry: a new method for metabolomics. *Rapid Commun Mass Spectrom* 2007; 21: 421-428.
- ¹⁸ Madalinski G, Godat E, Alves S, et al. Direct introduction of biological samples into a LTQ-Orbitrap hybrid mass spectrometer as a tool for fast metabolome analysis. *Anal Chem* 2008; 80: 3291-3303.
- ¹⁹ Draper J, Lloyd AJ, Goodacre R. et al. Flow infusion electrospray ionisation mass spectrometry for high throughput, non-targeted metabolite fingerprinting: a review. *Metabolomics* 2013; 9: S4-S29.
- ²⁰ Yamashita M and Fenn JB. Electrospray ion source. Another variation on the free-jet theme. *J Phys Chem* 1984; 88: 4451-4459.
- ²¹ Junot C, Madalinski G, Tabet JC, et al. Fourier transform mass spectrometry for metabolome analysis. *Analyst* 2010; 135: 2203-2219.

- 1
2
3
4 ²² Alves S, Rathahao-Paris E and Tabet JC. Potential of Fourier transform mass spectrometry for high-throughput metabolomics analysis. In: Rolin D, (eds) *Metabolomics coming of age with its technological diversity*. Amsterdam: Adv Bot Res, 2013, pp. 219-302.
- 5
6
7 ²³ Xian F, Hendrickson CL and Marshall AG. High resolution mass spectrometry. *Anal Chem*. 2012; 84:708–719.
- 8
9
10 ²⁴ Habchi B, Alves S, Jouan-Rimbaud Bouveresse D, et al. An innovative chemometric method for processing Direct Introduction High Resolution Mass Spectrometry metabolomic data: Independent Component - Discriminant Analysis (IC-DA). *Metabolomics* 2017; 13: 45.
- 11
12
13 ²⁵ Chekmeneva E, dos Santos Correia G, Chan Q. et al. Optimization and Application of Direct Infusion Nanoelectrospray HRMS Method for Large-Scale Urinary Metabolic Phenotyping in Molecular Epidemiology. *J Proteome Res*. 2017; 16:1646–1658.
- 14
15
16
17 ²⁶ Sommella E, Pagano F, Salviati E et al. Chemical profiling of bioactive constituents in hop cones and pellets extracts by online comprehensive two-dimensional liquid chromatography with tandem mass spectrometry and direct infusion Fourier transform ion cyclotron resonance mass spectrometry. *J Sep Sci*. 2018; 41:1548- 1557.
- 18
19
20 ²⁷ Chamoun R, Aliferis KA, and Jabaji S. Identification of signatory secondary metabolites during mycoparasitism of *Rhizoctonia solani* by *Stachybotrys elegans*. *Front. Microbiol*. 2015; 6: 353.
- 21
22 ²⁸ Taylor PJ. Matrix effects: the Achilles heel of quantitative high-performance liquid chromatography–electrospray–tandem mass spectrometry. *Clin Biochem* 2005; 38: 328-334.
- 23
24 ²⁹ Cappiello A, Famigliani G, Palma P, et al. Overcoming Matrix Effects in Liquid Chromatography–Mass Spectrometry. *Anal Chem* 2008; 80: 9343-9348.
- 25
26
27 ³⁰ González-Domínguez R, Castilla-Quintero R and García-Barrera T. Development of a metabolomic approach based on urine samples and direct infusion mass spectrometry. *Anal Biochem* 2014; 465:20-27.
- 28
29 ³¹ Makarov A. Electrostatic axially harmonic orbital trapping: a high-performance technique of mass analysis. *Anal Chem* 2000; 72: 1156-1162.
- 30
31
32 ³² Picard-Hagen N, Gayraud V, Viguié C, et al. Prion protein in the cerebrospinal fluid of healthy and naturally scrapie-affected sheep. *J Gen Virol* 2006; 87: 3723-3727.
- 33
34 ³³ Hu Q, Noll RJ, Li H, et al. The Orbitrap: a new mass spectrometer. *J. Mass Spectrom*. 2005, 40, 430-443.
- 35
36
37 ³⁴ Olsen JV, de Godoy LMF, Li G. et al. Parts per Million Mass Accuracy on an Orbitrap Mass Spectrometer via Lock Mass Injection into a C-trap. *Molecular & Cellular Proteomics* 2005, 4:2010-2021.
- 38
39 ³⁵ Ahdesmäki M, Strimmer K. Feature selection in omics prediction problems using cat scores and false non-discovery rate control. *Ann Appl Stat* 2010; 4: 503-519.
- 40
41
42 ³⁶ Wishart DS, Knox C, Guo AC, et al. HMDB: a knowledgebase for the human metabolome. *Nucleic Acids Res* 2009; 37: D603-D610.
- 43
44 ³⁷ Wishart DS, Tzur D, Knox C, et al. HMDB: the human metabolome database. *Nucleic Acids Res* 2007; 35: D521-D526.
- 45
46
47 ³⁸ Smith CA, O'Maille G, Want EJ, et al. Custodio DE, Abagyan R, Siuzdak G. METLIN: a metabolite mass spectral database. *Ther Drug Monit* 2005; 27: 747-51.
- 48
49
50 ³⁹ Ikonomidou MG, Blades AT and Kebarle P. Investigations of the electrospray interface for liquid chromatography/mass spectrometry. *Anal Chem* 1990; 62: 957-967.
- 51
52
53 ⁴⁰ Constantopoulos TL, Jackson GS and Enke CG. Effects of salt concentration on analyte response using electrospray ionization mass spectrometry. *J Am Soc Mass Spectrom* 1999; 10: 625-634.
- 54
55
56 ⁴¹ Kirwan JA, Broadhurst DI, Davidson RL, et al. Characterising and correcting batch variation in an automated direct infusion mass spectrometry (DIMS) metabolomics workflow. *Anal Bioanal Chem* 2013; 15: 5147-5157.
- 57
58
59 ⁴² Habchi B, Alves S, Jouan-Rimbaud Bouveresse D, et al. Potential of dynamically harmonized Fourier transform ion cyclotron resonance cell for high throughput metabolomics fingerprinting: control of data quality. *Anal Bioanal Chem* 2018; 410: 483-490.
- 60 ⁴³ van den Berg RA, Hoefsloot HC, Westerhuis JA, et al. Centering, scaling, and transformations: improving the biological information content of metabolomics data. *BMC Genomics* 2006; 7: 142.

1
2
3
4 ⁴⁴ Taylor NS, Weber RJM, Southam AD, et al. A new approach to toxicity testing in *Daphnia magna*:
5 Application of high throughput FT-ICR mass spectrometry metabolomics. *Metabolomics* 2009; 5:44-58.

6 ⁴⁵ Schulz-Trieglaff O, Machtejevas E, Reinert K, et al. Statistical quality assessment and outlier detection for
7 liquid chromatography-mass spectrometry experiments. *BioData Mining* 2009; 2:4.
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

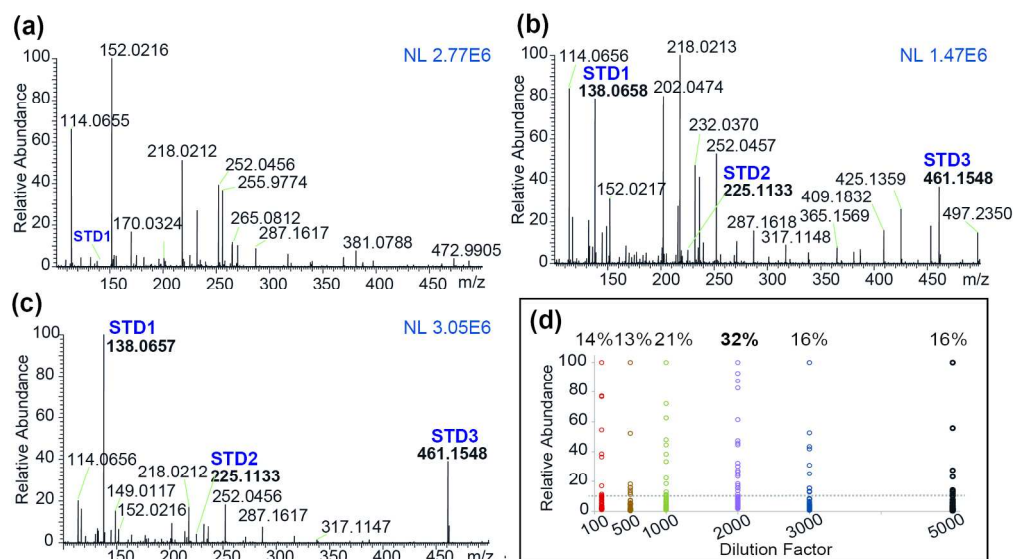
Figure Captions

Figure 1 Mass spectra of sheep urine samples diluted at a) 1:100, b) 1:2000 and c) 1:5000 and spiked with three reference compounds (STD1, STD2 and STD3 corresponding to protonated isoniazid, oxytetracycline and 2-amino-1-methyl-6-phenylimidazo [4, 5-b] pyridine (PhIP), respectively) (The NL (normalized level) value corresponds to the base peak intensity in arbitrary units), and d) plot of relative intensities of the most abundant ions present in the DI-HRMS mass spectra as a function of dilution factors. For each dilution factor, percentage of peaks displaying a relative intensity greater than 10% was reported, suggesting the dilution factor of 2,000 as the best

Figure 2 a) Plot of relative intensities of some ions (with low and high abundances) from DI-HRMS mass spectra as a function of dilution factor, and b) Total ion current (TIC) at different dilution factors (note that the TIC values are calculated from the sum of the relative intensities of the 35 most abundant ions)

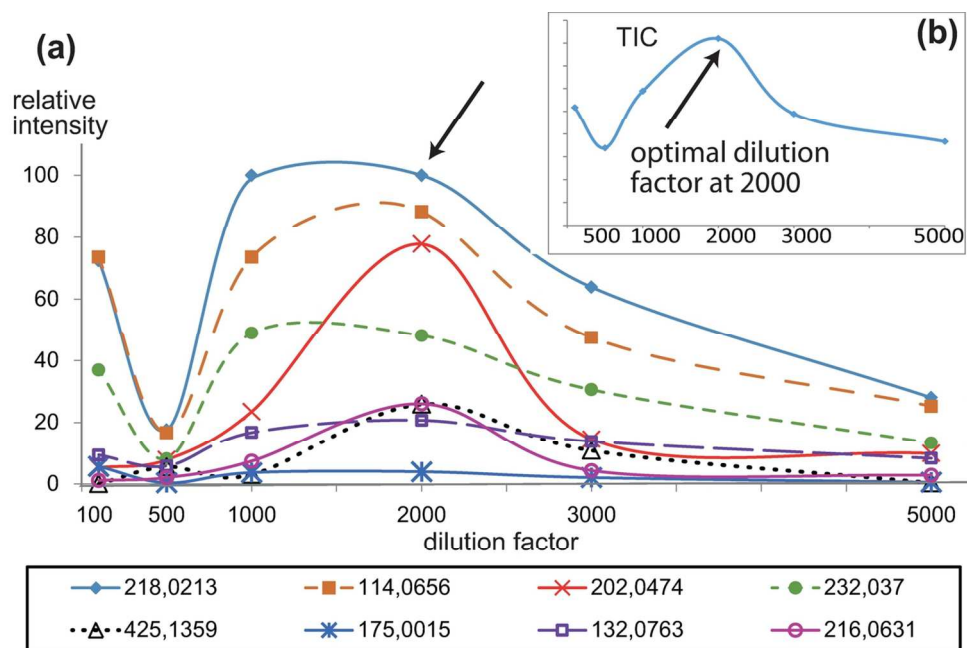
Figure 3 Plot of ion relative intensities as a function of sample concentration (*i.e.* 1:dilution factor) from DI-HRMS mass spectra. A linear response was obtained at high dilution conditions (*i.e.* above 1:2000) for most ions with correlation coefficient > 0.988 (see **Table S1** in supplementary information). Were reported in this plot low and high abundance ions

Figure 4 Score plot from Shrinkage Discriminant Analysis (SDA) performed on the 122 selected variables after data pre-processing



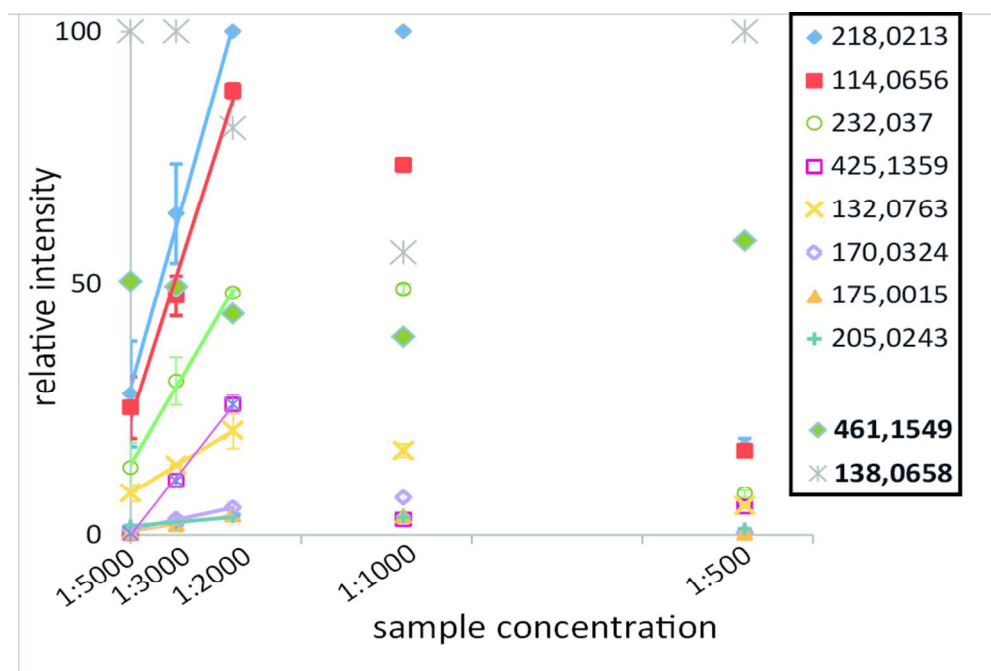
Mass spectra of sheep urine samples diluted at a) 1:100, b) 1:2000 and c) 1:5000 and spiked with three reference compounds (STD1, STD2 and STD3 corresponding to protonated isoniazid, oxytetracycline and 2-amino-1-methyl-6-phenylimidazo [4, 5-b] pyridine (PhIP), respectively) (The NL (normalized level) value corresponds to the base peak intensity in arbitrary units), and d) plot of relative intensities of the most abundant ions present in the DI-HRMS mass spectra as a function of dilution factors. For each dilution factor, percentage of peaks displaying a relative intensity greater than 10% was reported, suggesting the dilution factor of 2,000 as the best

189x105mm (300 x 300 DPI)



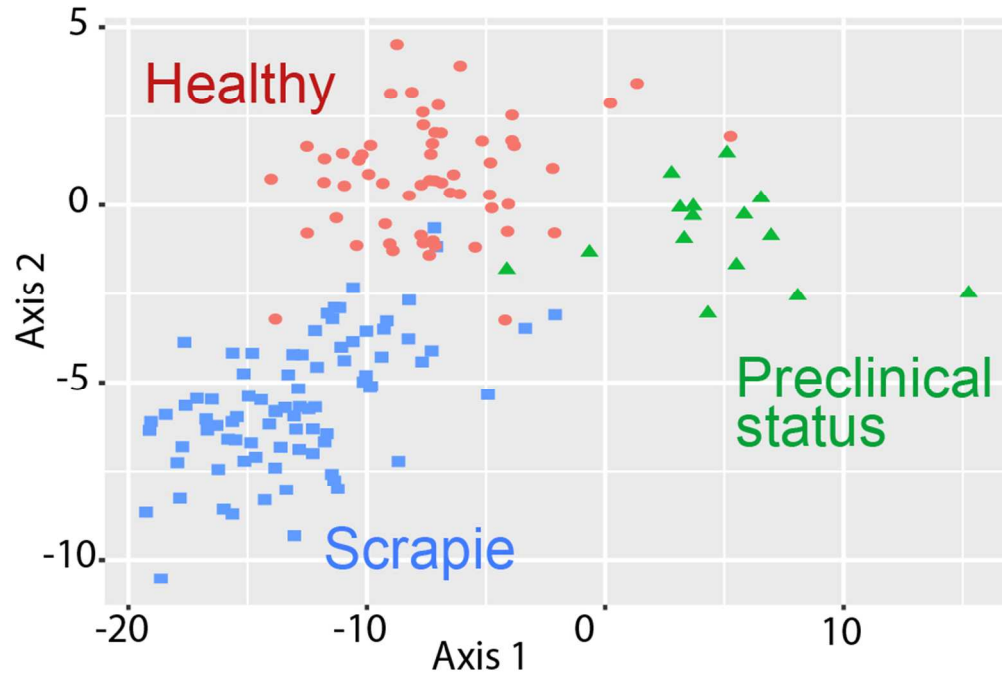
a) Plot of relative intensities of some ions (with low and high abundances) from DI-HRMS mass spectra as a function of dilution factor, and b) Total ion current (TIC) at different dilution factors (note that the TIC values are calculated from the sum of the relative intensities of the 35 most abundant ions)

121x78mm (300 x 300 DPI)



Plot of ion relative intensities as a function of sample concentration (*i.e.* 1:dilution factor) from DI-HRMS mass spectra. A linear response was obtained at high dilution conditions (*i.e.* above 1:2000) for most ions with correlation coefficient > 0.988 (see Table S1 in supplementary material). Were reported in this plot low and high abundance ions

87x58mm (300 x 300 DPI)



Score plot from Shrinkage Discriminant Analysis (SDA) performed on the 122 selected variables after data pre-processing

87x59mm (300 x 300 DPI)

Review

Supplementary information

Evaluation and validation of analytical approach for high-throughput metabolomic fingerprinting using direct introduction-high resolution mass spectrometry: applicability to urine of scrapie-infected ewes

Estelle Rathahao-Paris^{1,2*}, Sandra Alves², Nawel Boussaid¹, Nicole Picard-Hagen³, Véronique Gayrard³, Pierre-Louis Toutain⁴, Jean-Claude Tabet^{2,5}, Douglas N. Rutledge¹, Alain Paris⁶

¹ UMR Ingénierie Procédés Aliments, AgroParisTech, Inra, Université Paris-Saclay, 91300 Massy, France

² Sorbonne Université, Faculté des Sciences et de l'Ingénierie, Institut Parisien de Chimie Moléculaire (IPCM), 75005 Paris, France

³ Toxalim, Université de Toulouse, INRA (Institut National de la Recherche Agronomique), INP (Institut National Polytechnique de Toulouse) –ENVT (Ecole Nationale Vétérinaire de Toulouse), 31076 Toulouse, France

⁴ INTHERES, Université de Toulouse, INRA, ENVT, 31076 Toulouse, France

⁵ CEA-INRA, Service de Pharmacologie et d'Immunoanalyse, Laboratoire d'Etude du Métabolisme des Médicaments, MetaboHUB, Université Paris-Saclay, Gif-sur-Yvette cedex, France

⁶ Sorbonne Universités, Muséum national d'Histoire naturelle, CNRS, UMR7245 MCAM, 75005 Paris, France

* Corresponding author:

Dr. Estelle Rathahao-Paris

Present address: UMR CEA-INRA Service de Pharmacologie et d'Immunoanalyse, Laboratoire d'Immuno-Allergie Alimentaire, CEA de Saclay - Bat 136, F-91191, Gif-sur-Yvette cedex, France.

E-mail: Estelle.Paris@inra.fr

Table S1. List of some ions detected in the mass spectra acquired at different dilution factors. Their linearity response appears in the correlation coefficient value calculated for the highest dilution conditions, *i.e.* above 1:2000). Note that some ions which are not detected in both urine samples. This could be due to the difference in the sample groups: the sample 1 is collected from a healthy animal whereas the sample 2 is from a scrapie-affected one.

m/z	Putative ion annotation ^(a)	Ion elemental formulae	Error ^(b) (ppm)	Sample 1		Sample 2	
				Rel. Int ^(b)	R ²	Rel. Int ^(b)	R ²
180.0654	Hippuric acid + H	C ₉ H ₁₀ O ₃ N	0.8	3	0,581	nd	
202.0474	Hippuric acid + Na	C ₉ H ₉ O ₃ NNa	0.2	78	0,850	23	0,977
218.0213	Hippuric acid + K	C ₉ H ₉ O ₃ NK	0.7	100	0,996	43	0,994
219.0246	Hippuric acid (¹³ C) + K	¹³ C ₁ C ₈ H ₉ O ₃ NK	0	6	0,891	3	1.000
220.0194	Hippuric acid + ⁴¹ K	C ₉ H ₉ O ₃ N ⁴¹ K	4.5	4	0,843	nd	
114.0656	Creatinine + H	C ₄ H ₈ ON ₃	5.2	88	0,989	47	0,986
136.0477	Creatinine + Na	C ₄ H ₇ ON ₃ Na	3.7	8	0,829	2	0,960
152.0217	Creatinine + K	C ₄ H ₇ ON ₃ K	2.3	32	0,986	17	0,990
132.0763	Creatine + H	C ₄ H ₁₀ O ₂ N ₃	4.0	21	1,000	10	0,981
170.0324	Creatine + K	C ₄ H ₉ O ₂ N ₃ K	1.3	5	0,998	3	0,989
216.0631	Methyl hippuric acid + Na	C ₁₀ H ₁₁ O ₃ N ₁ Na	0.4	26	0,848	3	0,983
232.0370	Methyl-hippuric acid + K	C ₁₀ H ₁₁ O ₃ N ₁ K	0.3	48	0,996	9	0,995
425.1359	Porson + K	C ₂₂ H ₂₆ O ₆ K	0.5	26	0,998	nd	
175.0015	Hypoxanthine+K	C ₅ H ₄ ON ₄ K	1.1	4	1,000	nd	
205.0243	Ipomeanine + K	C ₉ H ₁₀ O ₃ K	9	4	0,988	nd	

(a) Only one candidate is proposed for each m/z value from DB search (putative annotation)

(b) The mass error and relative intensity are obtained from a mass spectrum acquired at 1:2000 dilution

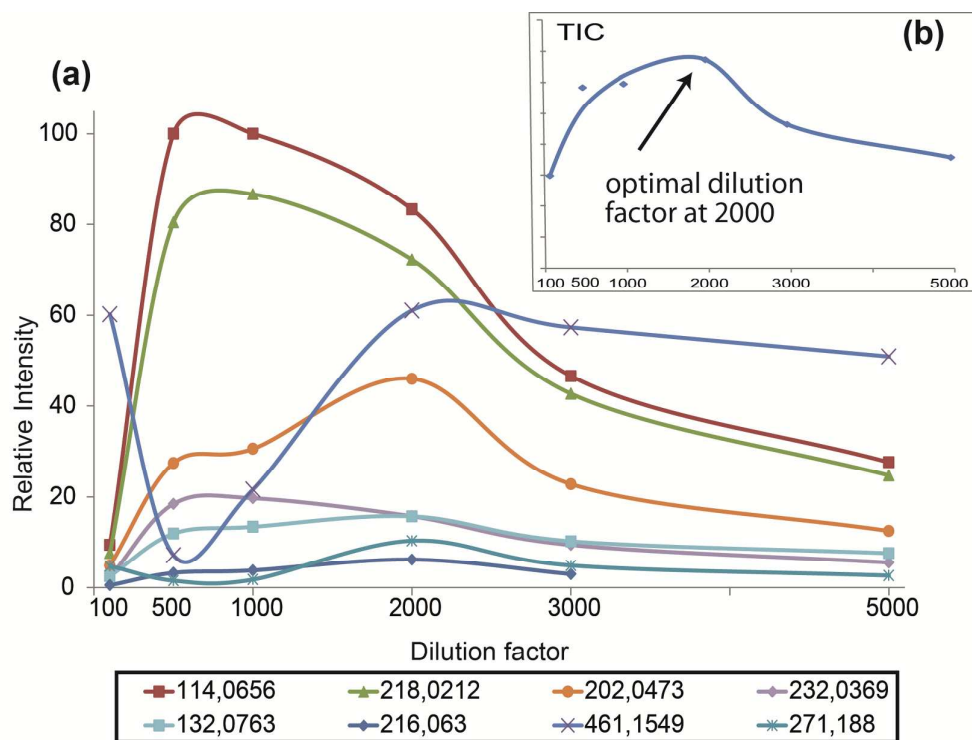


Figure S1. Plot of a) relative intensities of some ions (with low and high abundances) and b) total ion current (TIC) at different dilution factors (note that the TIC values are calculated from the sum of the relative intensities of the 30 most abundant ions). These values were reported from DI-HRMS mass spectra acquired from a second urine sample coming from a scrapie-affected animal. Different trends are observed for different ions, related to the matrix. The optimal dilution factor is identical to that obtained from the first urine sample coming from a healthy individual.

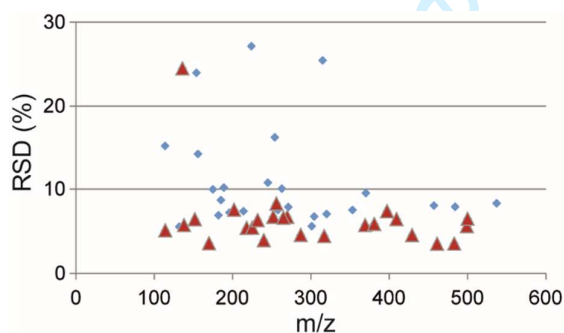


Figure S2. Plot of RSD values calculated from absolute ion abundances of fifty most abundant peaks detected from the DI-HRMS mass spectra of a urine sample at a 1:2000 dilution for 20 repeated sample injections. The most abundant ions (>10% compared to the base peak) were marked by red triangles

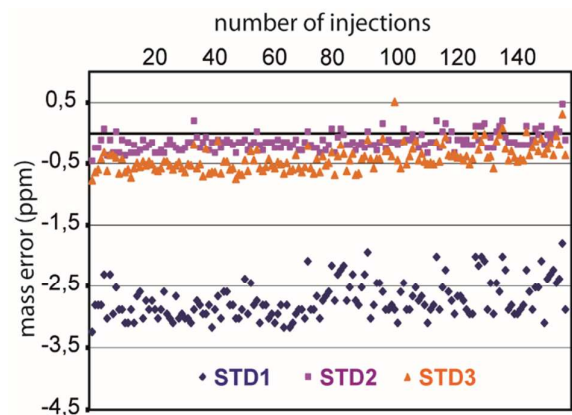


Figure S3. Plot of mass measurement accuracy of the three reference compound ions using internal calibration from injections of 159 samples (STD1: isoniazid; STD2: PhIP and STD3: oxytetracycline)

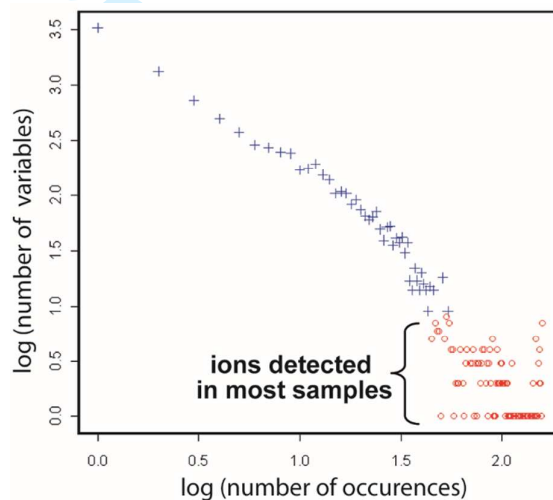


Figure S4. Plot of the seemingly power law distribution for 9684 variables detected from all urine samples. 122 variables shared in most samples were selected for subsequent statistical analysis.

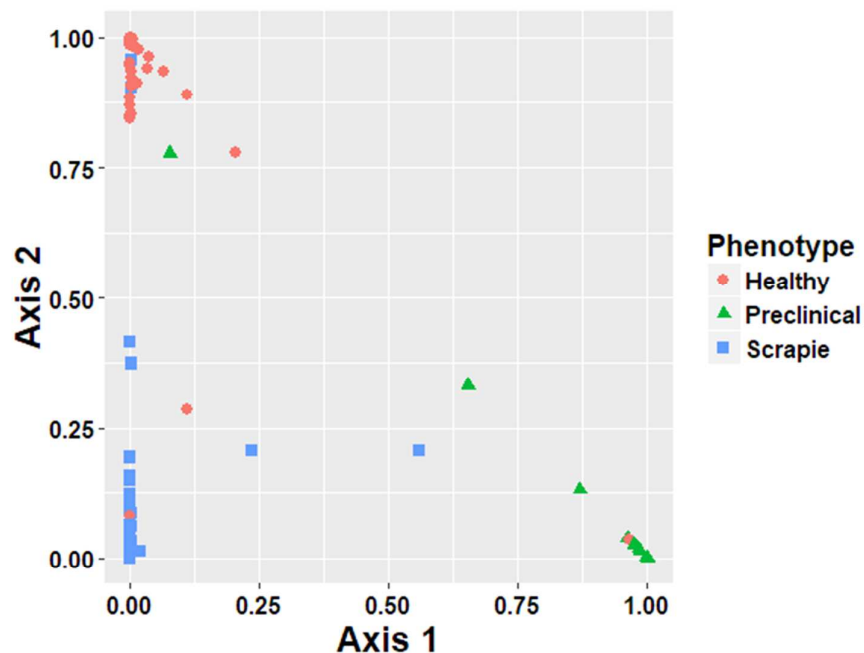


Figure S5. Plot of probabilities coming from the Mahalanobis distances between every statistical individual and the barycenter of the 3 groups, i.e. healthy, preclinical and scrapie-infected animals. The SDA-mediated discrimination is typically derived from this first classification where the minimal distance is retained for deciding the final assignment of a given individual to its final belonging group. With such a procedure based on the use of a shrinkage applied to the calculation of the correlation, the covariance matrices and the prior probabilities vector, all these three members entering in the calculation of the matrix of Mahalanobis distances, the misclassification ratio was estimated to nearly 4.5% using a leave-one-out procedure.

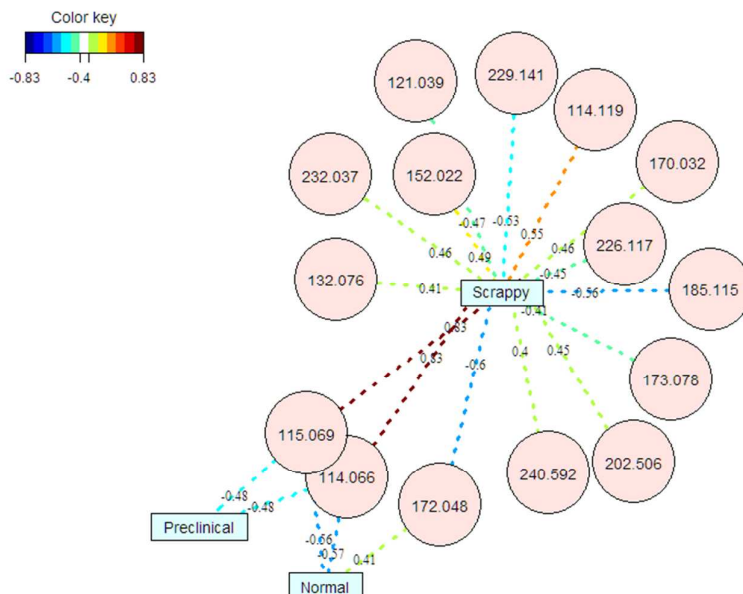


Figure S6. Canonical correlations calculated according to the second component ($\rho > 0.40$). Note that protonated creatinine ions (m/z 114.066 annotated as $C_4H_8ON_3$ and 115.068 annotated as $^{13}C_1C_3H_8ON_3$) display a positive correlation with scrapie-infected status whereas negative values were observed with samples from healthy animals or those at preclinical stage.

Table S2. Putative annotation of the most discriminants biomarkers detected from the canonical correlation analysis reported in the Figure S5

m/z	Ion elemental formulae	Error (ppm)	Adduct type	Putative annotation
114.0656	$C_4H_8ON_3$	5.0	M+H	Creatinine
115.0688	$^{13}C_1C_4H_8ON_3$	4.3	$(^{13}C)M+H$	Creatinine
152.0217	$C_4H_7ON_3K$	2.3	M+K	Creatinine
132.0762	$C_4H_{10}O_2N_3$	4.0	M+H	Creatine
170.0324	$C_4H_9O_2N_3K$	1.3	M+K	Creatine
173.0782	$C_6H_{14}O_4Na$	1.6	M+Na	dideoxyhexitol
185.1148	$C_8H_{18}O_3Na$	0.3	M+Na	Octanetriol
226.1167	$C_8H_{17}N_3O_3 Na$	2.3	M+Na	Glycyl-Lysine
232.0370	$C_{10}H_{11}O_3N_1K$	0.3	M+K	Methyl-hippuric acid