

Nutritional Metabolomics: What are the perspectives?

Jean-Louis SÉBÉDIO
Jean-François MARTIN
Estelle PUJOS

INRA UMR 1019, Plateforme d'exploration
du métabolisme: des gènes aux métabolites,
63122 Saint-Genès Champanelle
<jls@clermont.inra.fr>

Résumé : Une approche traditionnelle en nutrition a longtemps été d'étudier l'effet d'un régime ou bien d'un nutriment donné sur une fonction particulière ou un organe cible, ceci pour expliciter les mécanismes par lesquels les macro et micronutriments interviennent dans les voies métaboliques. Le développement de techniques analytiques très performantes et d'outils à haut débit comme par exemple la métabolomique ouvre maintenant un champ d'investigation beaucoup plus large permettant d'intégrer un ensemble de réponses biologiques résultant de la complexité de l'aliment et des régimes alimentaires. La métabolomique consiste en l'acquisition à partir de fluides biologiques (sang, urine, salive) de profils métaboliques complexes par l'analyse de centaines de métabolites le plus souvent par 1H-RMN, ou différentes techniques de couplage (HPLC-MS ou GC-MS) et leur comparaison par analyses statistiques multivariées. Elle s'est d'abord développée dans le champ de la toxicologie pour prédire les effets toxiques de médicaments dans les phases précoces de développement. Les études en nutrition sont encore récentes mais différents programmes de recherche concernent l'identification de marqueurs précoces de déséquilibres métaboliques associés à l'apparition de pathologies. Deux approches sont possibles: la première est une approche ciblée qui concerne l'étude d'une voie métabolique définie comme par exemple le métabolisme des glucides ou celui des lipides. La deuxième est une approche globale qui consiste à définir une empreinte métabolique en caractérisant le plus grand nombre possible de métabolites afin d'identifier les diverses voies métaboliques perturbées suite au stimuli. Toutefois, les bases de données « métabolites » permettant la saisie et la consultation de molécules identifiées lors d'explorations nutritionnelles sont encore insuffisantes pour permettre l'identification d'un grand nombre de métabolites. Chez l'homme, il existe également une variabilité interindividuelle importante ce qui représente un autre facteur limitant pour cette approche à haut débit. De plus, contrairement aux effets des toxiques, les effets liés à un changement de régime sont souvent de faible amplitude ce qui peut entraîner des difficultés de détection et d'identification des métabolites. La plateforme d'exploration du métabolisme de Clermont-Ferrand participe au développement de l'outil métabolomique dans le domaine de la nutrition et de la bioinformatique en partenariat avec le centre de bioinformatique de Bordeaux (CBiB) en particulier dans le cadre de l'ANR Metaprofile. Il s'agit de structurer une base de données en nutrition incluant non seulement des données de métabolomique mais également des données complémentaires de transcriptomique et de protéomique afin d'interroger et de croiser les différents jeux de données, et ainsi obtenir l'ensemble des informations des gènes aux métabolites.

Mots clés : nutrition, métabolomique, métabolites, spectrométrie de masse, bioinformatique, base de données

An important challenge of contemporary biology is to establish the relationship between phenotypes and disruptions in the underlying cellular functions and in the past, much effort has been devoted to a gene-based approach. However, such an approach, although successful, is far from sufficient, considering that most cellular components exert their function through intricate networks of regulation, protein interactions and metabolites.

The study of metabolism is a founding discipline of biochemistry but major parts of secondary metabolism, which determine important aspects of the biological function and phenotype of an organism, remain entirely unexplored. In the current postgenome era, metabolites are considered as the result of the in-

teraction of the system's genome with its environment and are not merely the end product of gene expression but also form part of the regulatory system in an integrated manner. Metabolomics, which seeks to systematically characterise metabolic perturbations in a dynamic and multivariate framework can provide a vehicle for defining phenotypes relating to health and disease.

The rapidly expanding field of metabolomics has been driven by major advances in analytical tools such as NMR (Nuclear Magnetic Resonance) [1] and Mass Spectrometry and the corresponding hyphenated methods such as high performance liquid chromatography coupled with mass spectrometry (LC-MS) in particular [2-4] but also in chemometrics

and bioinformatics technologies [5-8]. Metabolomics [9, 10] can be described as a global analysis of small molecules of a biofluid (blood, urine, culture broth, cell extract, etc), which are produced or modified as a result of a stimuli (nutritional intervention, environmental stressors, drug etc). The production and utilisation of metabolites is more directly connected to the phenotype exhibited by an organism than the presence of mRNAs or proteins.

The purpose of this short review is not to be exhaustive but to give a few examples showing how metabolomics can be applied to nutrition research and to evaluate what should be developed in that field (analytical instruments, sample preparation, statistics and bioinformatics) if

metabolomics is to realise its potential in the field of nutrition.

Nutritional metabolomics

What are the perspectives?

The metabolomics field has been mainly developed in pharmaceutical, medical and plant sciences. However, its application to nutrition is still relatively new although this topic constitutes a growing domain of interest for our society [11-13]. Many challenges exist in nutritional metabolomics. For example the amplitude of the effect of a nutrient and especially of a micro-nutrient on the metabolic pathways is expected to be low compared to that of a drug. Furthermore, the metabolome will be the result of the food metabolome and endogenous metabolism. Nutritional metabolomics will then have two main objectives, i) to determine biomarkers of exposure and ii) to identify biomarkers of metabolic function in order to correct a dysfunction by an appropriate dietary intervention. Factors influencing the metabolic profile of healthy individuals must be fully understood.

While pharmacology may apply when disease is established, nutrition has an important role to play at the stage when it is possible to detect very early biomarkers of homeostatic dysregulation where minor metabolic effects will have to be identified against a background of genetic and environmental variation requiring the development of specific methodologies and sophisticated data processing tools.

A typical metabolomic experiment (figure 1) includes i) preparation of the sample, ii) analysis of bio fluids or tissues via HPLC-MS, NMR or GC-MS, iii) treatment of data set, and iii) the identification of key or discriminatory metabolites. The data acquisition and retrieval may be achieved by a targeted analysis where specific pathways are analysed or an untargeted analysis where no prior selection of pathways is performed. Both have advantages and disadvantages and the approach taken depends ultimately on the study in question. To maximize metabolite coverage the current trend is to use a combination of NMR and MS techniques. However, combining the data from these different platforms is not trivial and much work needs to be done to develop this area.

Metabolomics is a very promising tool to evaluate the exposure of animals or humans to micronutrients such as for example polyphenols which may participate to the prevention of cardiovascular diseases and cancer [14]. In one study, Sang *et al.*, 2008 identified 25 metabolites of tea polyphenols from urine sample collected at different time periods after consump-

tion of green tea using LC/MS/MS. These metabolites resulted from the glucuronidation, sulfation, methylation and ring fission of tea polyphenols (epicatechin, epigallocatechin). Ring fission metabolites were the results of the action of the intestinal flora. As epicatechin (EC) is also present in cocoa and in many fruits, metabolites of EC and the ring fission products cannot be used to evaluate tea consumption. On the contrary, conjugated metabolites of epigallocatechin and its ring fission products may be used as the exposure markers to reflect tea consumption.

In another study, Fardet *et al.* [15] compared the urinary metabololms of rats fed two lignins isolated either from wheat bran (LEWB) or poplar wood (PL) and compared them to urinary profiles of rats fed either ferulic acid (FA) or syringic acid (SA). Comparison of the urine LC-QToF metabolic profiles (figure 2) indicated similarity between both lignin supplemented groups (PL and LEWB) and the control groups which confirms that lignins are inert and not absorbed in the body. On the contrary, metabolic fingerprints of the FA and SA supplemented groups were distinct from the control

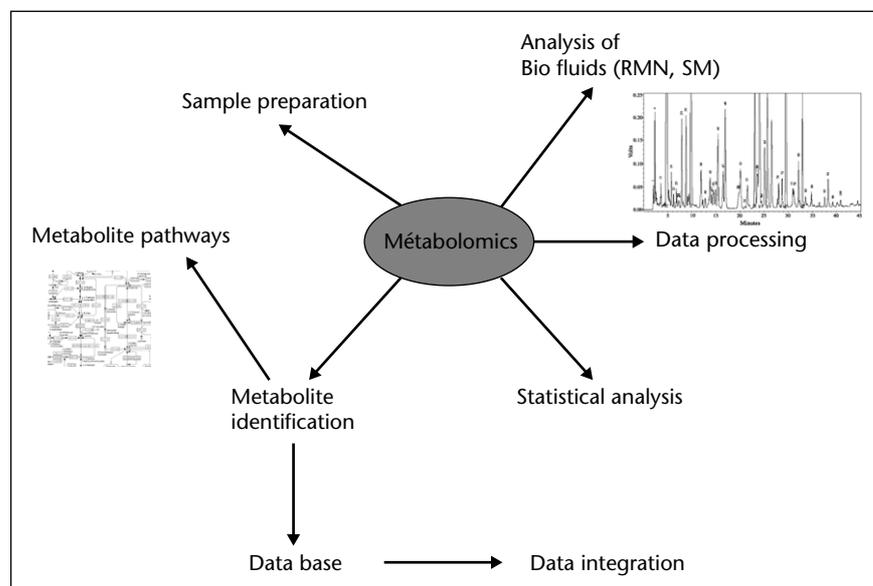


Figure 1. The different steps of a nutritional study using a metabolomics approach.

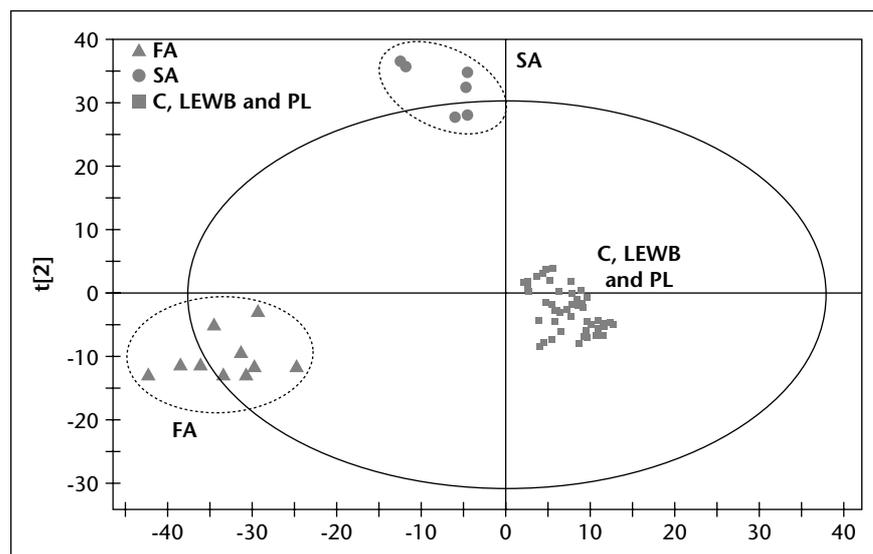


Figure 2. PLS-DA score plot of the LC-QToF metabolic profiles of rats fed C, LEWB, PL, SA, or FA diets. Three classes were predefined: FA group at day 1 and 2; SA group at day 1 and 2; SA group at day 1 and 2; C, LEWB, and PL groups at day 0, 1 and 2 and FA and SA groups at day 0. Reproduced by permission from [15].

group. Differences came from non metabolized FA and SA and from metabolites excreted in urine. Some of the identified metabolites were the sulfate esters, glucuronide and glycine conjugates of the phenolic acids, but also of dihydrosinaptic, vanillic and benzoic acids. Again, this study shows the power of metabolomics to reveal biomarkers of exposition to food phytochemicals. In the same field of research, Walsh *et al.* [16] showed that utilization of two complementary techniques such as NMR and LC-MS and a cross platform comparison of two resulting data sets using co-inertia analysis will be a promising approach for metabolite and exposure marker identifications.

Metabolomics may also be used to reveal individuals at risk for metabolic deregulation or pathologies. One of the first metabolomics approach to study biochemical modifications following a dietary intervention (soy proteins) in humans was published by Solanki *et al.* [17]. Briefly, 5 women of 21-29 years old were fed for one month a diet of non-soy-containing food. During a second month period, the diet was modified to include 60 g/day of soy protein corresponding to 45 mg isoflavones. Blood samples were collected during the whole study and analyzed by NMR in order to identify metabolite change during the feeding period. Statistical analyses (figure 3) show that even if the extent of the metabolic response was subject specific, the nature of this response was consistent across subjects and a two component PLS-DA model could predict the dietary intervention period accurately as reported in. Identification of metabolites revealed that a soy intervention resulted in an increase of 3-

hydroxybutyrate, N-acetyl glycoproteins, a decrease in sugars and an increase in lactate which indicates an increase in anaerobic metabolism and suggests a possible inhibition of gluconeogenesis.

Some studies have already shown the possibility of identifying markers of pathologies such as cancer, diabetes, cardiac diseases [18-22] using a metabolomic approach. For example, Salek *et al.* [19] used NMR and multivariate statistics to examine the disease urinary profiles changes of mice, rat and human suffering of type 2 diabetes mellitus. Metabolic similarities between the 3 species were observed. Furthermore, in addition to the expected changes in energy metabolism (fatty acid oxidation and gluconeogenesis), major effects were detected in nucleotide metabolism, with a decrease in N-methylnicotinate and an increase in N-methylnicotinamide (NMN amide) and in N-methyl-2-pyridone-5-carboxamide (2PY). The authors concluded that the "effects in NMN amide and 2PY metabolism may provide novel biomarkers for following the progression of type 2 diabetes mellitus".

Even if NMR and HPLC-MS have been used extensively in the field of metabolomics, gas-liquid chromatography coupled to mass spectrometry (GC-MS) may also be a powerful tool to pinpoint the metabolic pathways which have been deregulated as a consequence of diagnosed pathologies. In a recent study using GC-ToF based metabolomics, Denkert *et al.* [20] showed that this method could be used for biochemical phenotyping of normal and colon cancer tissues. The analysis of 45 samples (18 from normal colon mucosa and 27 from

colon carcinomas) found 82 metabolites to be discriminant. These metabolites were connected to abnormalities in metabolic pathways using the KEGG REACTION database. Urea cycle metabolites, purines, pyrimidines, and amino acids were found at higher levels in cancer tissues compared to normal mucosa while the intermediates of TCA cycle and lipids were decreased. The authors consequently suggested that this type of metabolic profiling approach may be useful to monitor changes in tumor metabolism and therefore of response to therapy.

What needs to be improved and developed?

Sample collection and preparation

In nutritional metabolomics, there is often a large variability in the samples compared to the changes of interest induced by the nutritional intervention. So it is very important to quantify and minimize the analytical variability. Sources of variation during a metabolomic study on humans have been described by Maher *et al.* [23]: it includes sample collection, storage and pre treatment, instrument stability, and intra/inter-individual variations. The development of sample preparation methods is challenging in metabolomics, as biological media contain a high number of metabolites, with a wide chemical variety and at very different concentrations. Bruce *et al.* [24] showed a method to evaluate a protocol from sample extraction to data analysis for a metabolomic profiling approach. Many studies give recommendations for sample treatment but there are still no standard operating procedures available. Moreover, requirements for the validation of metabolomic methods have not been described yet.

Data processing tools and bioinformatics

After acquisition of LC-MS data, many difficulties and pitfalls appear at each step of the data analysis workflow:

Data extraction

The metabolomic data analysis workflow begins with extraction of information from the different chromatograms and spectra. It is a critical step because all the further statistical analyses depend on the accuracy of this extraction. A lot of automated extraction softwares exist. Some are developed by the different instrument manufacturer and others are being developed by research laboratories and are free software. They all use different algorithms and have a lot of parameters to be optimized. As a consequence, they don't give the same dataset results and it is really difficult to define which software is the best.

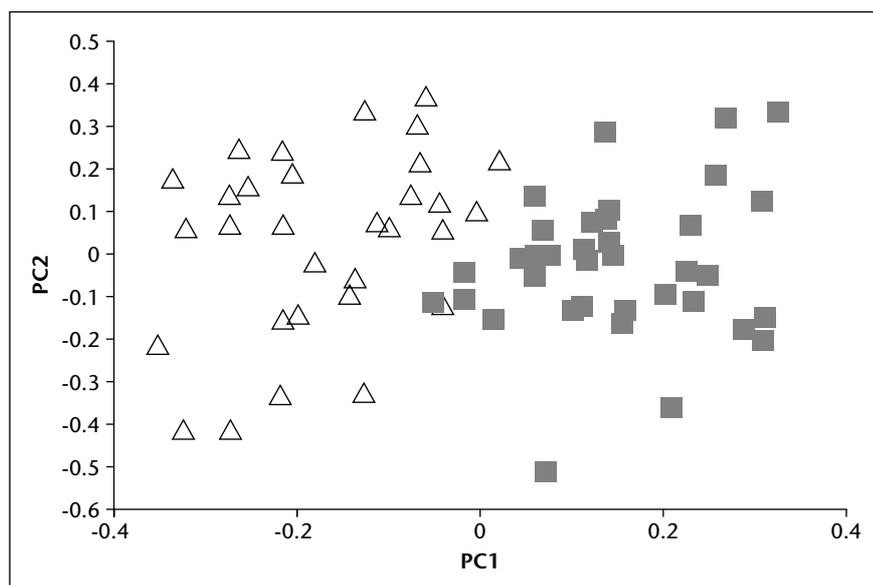


Figure 3. PLS-DA of the OSC filtered data based on spectra collected pre soy (Δ) and post soy (\blacksquare) intervention period. Adapted from [16].

Statistical analysis

Nutritional intervention studies generally induce subtle changes on metabolites intensities which may be lower than inter-individual variability. As a consequence, it is sometimes difficult to observe a significant difference between treatments if data are not obtained under well designed experimental conditions on a reasonable number of subjects. Another difficulty is the choice of statistical tools. Supervised techniques like PLS-DA (Partial least square discriminant analysis) if used with only few observations may lead to overfitted models which are good for explanation but weak for predicting. These techniques are very powerful but must be validated before biological interpretation [25]. Moreover, the variability induced by variation of the retention time during the experiment can lead to misalignment and generate duplicate ions with a problem of co linearity in the statistical analysis. Another typical LC-MS problem is dirtying of the ion source which affects the intensity of ions and consequently the statistical analysis results and which must be corrected by an appropriate normalisation.

Data base and metabolite identifications

The identification of metabolites is one of the main limiting factors of metabolomics when using mass spectrometry. The main reason is the huge chemical diversity that is present in the biological media. Exact mass measurements, especially done on high resolution mass spectrometers, enable the elemental composition of detected peaks but there are still a lot of compounds that can match a given atomic composition. MS/MS strategies can be applied but there are still some issues, like lack of signal intensity or nonspecific losses. Stoll et al. [26] have developed an identification strategy based on the isotope pattern evaluation to reduce the elemental compositions determined with a FTICR mass spectrometer. King et al. [27] have also described a series of rules for chemical formula extraction and compound ranking from high resolution mass spectra. Another complementary approach is databases searching but there is today a lack of exhaustive databases dealing with the nature and the levels of the metabolites present in the biological media of interest. The most useful databases for compound identification are the MS-based metabolite databases, which provide reference MS spectra of pure compounds, like the Metlin database [28], the HMDB library [29]. However, the metabolomics community still has to establish standards for these databases, especially concerning the analytical experimental details and queries, and to share their databases to cover the huge diversity of metabolites of interest.

Conclusion

The metabolomics field has been mainly developed in pharmaceutical, medical and plant sciences. However, its application to nutrition is still relatively new although this topic constitutes a growing domain of interest for our society. Metabolomics is a very promising tool to evaluate the exposure of animals or humans to nutrients and consequently to determine biomarkers of exposure. Metabolomics may also be used to reveal individuals at risk for metabolic deregulation or pathologies, and determination of early biomarkers of metabolism perturbation before the detection of clinical signs is of major importance. However, in nutritional metabolomics, there is often a large variability in the samples compared to the changes of interest induced by the nutritional intervention. So it is very important to quantify and minimize the analytical variability. The identification of metabolites is one of the main limiting factors of nutritional metabolomics when using mass spectrometry. The main reason is the huge chemical diversity of metabolites which are present in the biological fluids. In order to progress in the metabolite identification, the metabolomics community is developing some databases but it still has to establish standards for these databases, especially for analytical experimental details and queries, and to share these databases to cover the huge diversity of metabolites of interest.

REFERENCES

1. Eisenreich W, Bacher A. Advances of high resolution NMR techniques in the structural and metabolic analysis of plant biochemistry. *Phytochem* 2007; 68: 2799-815.
2. Dettmer K, Aronov PA, Hammock BD. Mass spectrometry-based metabolomics. *Mass Spectrom Rev* 2007; 26: 51-78.
3. Wilson ID, Plumb R, Granger J, Major H, Williams R, Lenz EM. HPLC-MS based methods of the study of metabonomics. *J Chromatogr B* 2005; 817: 67-76.
4. Pan Z, Raftery D. Comparing and combining NMR spectroscopy and mass spectrometry in metabolomics. *Anal Bioanal Chem* 2007; 387: 525-7.
5. Idle JR, Gonzalez FJ. Metabolomics. *Cell Metabolism* 2007; 6: 348-51.
6. Jonsson P, Gullberg J, Nordstrom A, Kusano M, Kowalczyk M, Sjoström M, Moritz T. A strategy for identifying differences in large series of metabolomic samples analyzed by GC/MS. *Anal Chem* 2004; 76: 1738-45.
7. Moco S, Vervoort J, Moco S, Bino RJ, De Vos RCH, BINO R. Metabolomics technologies and

metabolite identification. *Trends in Analytical Chemistry* 2007; 26: 855-66.

8. Wishart DS. Current Progress in computational metabolomics. *Brief Bioinform* 2007; 8: 279-93.
9. Nicholson JK, Lindon JC, 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: 1181-9.
10. Fiehn O, Kopka J, Dormann P, Altmann T, Trethewey N, Willmitzer L. Metabolite profiling for plant functional genomics. *Nature Biotech* 2000; 18: 1157-61.
11. Watkins SM, Hammock BD, Newman JW, German JB. Individual metabolism should guide agriculture toward foods for improved health and nutrition. *Am J Clin Nutr* 2001; 74: 283-6.
12. German JB, Roberts MA, Fay L, Watkins SM. Metabolics and individual metabolic assessment: the next great challenge for nutrition. *J Nutr* 2002; 132: 2486-7.
13. Gibney MJ, Walsh M, Brennan L, Roche HM, German B, Van Ommen B. Metabolomics in human nutrition: opportunities and challenges. *Am J Clin Nutr* 2005; 82: 497-503.
14. Scalbert A, Manach C, Morand C, Remesy C, Jimenez L. Dietary polyphenols and the prevention of diseases. *Crit Rev Food Sci Nutr* 2005; 45: 287-306.
15. Fardet A, Llorach R, Orsoni A, Martin JF, Pujos-Guillot E, Lapiere C, Scalbert A. Metabolomics provide new insight on the metabolism of dietary phytochemicals in rats. *J Nutr* 2008; 138: 1262-87.
16. Solanky KS, Bailey NJC, Beckwith-Hall B, Davis A, Bingham S, Holmes E, Nicholson JK, Cassidy A. Application of ¹H nuclear magnetic resonance-based metabolomic techniques for the analysis of the biochemical effects of dietary isoflavones on human plasma profile. *Anal Biochem* 2003; 323: 197-204.
17. Walsh MC, Brennan L, Pujos-Guillot E, Sebedio JL, Scalbert A, Fagan A, Higgins DG, Gibney MJ. Influence of acute phytochemical intake on human urinary metabolomic profile. *Am J Clin Nutr* 2007; 86: 1687-93.
18. Chen J, Zhao X, Fritsche J, et al. Practical approach for the identification and isomer elucidation of biomarkers detected in a metabolomic study for diabetes by integrating the chromatographic and mass spectrometric information. *Anal Chem* 2008; 80: 1280-9.
19. Salek, RM, Maguire M, Bentley E, et al. A metabolomic comparison of urinary changes in type 2 diabetes in mouse, rat and human. *Physiol Genomics* 2007; 29: 99-108.
20. Denkert C, Budczies J, Weichert W, et al. Metabolite profiling of human colon carcinoma-deregulation of TCA cycle and

- amino acid turnover. *Molecular Cancer* 2008; 7: 72 (doi: 10.1186/1476-4598-7-72).
21. Yang J, Xu G, Zheng Y, Kong H, Pang T, Lv S, Yang Q. Diagnosis of liver cancer using HPLC-based metabolomics avoiding false positive result from hepatitis and hepatocirrhosis diseases. *J Chromatogr B* 2004; 813: 59-65.
 22. Ellis DI, Dunn WB, Griffin JL, Allwood JW, Goodacre R. Metabolic fingerprinting as a diagnostic tool. *Pharmacogenomics* 2007; 8: 1243-66.
 23. Maher AD, Severine F, Zirah M, Holmes E, Nicholson JK. Experimental and Analytical Variation in Human Urine in ¹H NMR Spectroscopy-Based Metabolic Phenotyping Studies. *Anal Chem* 2007; 79: 5204-11.
 24. Bruce SJ, Jonsson P, Antti H, *et al.* Evaluation of a protocol for metabolic profiling studies on human blood plasma by combined ultra-performance liquid chromatography/mass spectrometry: From extraction to data analysis. *Anal Biochem* 2008; 372: 237-49.
 25. Rubingh CM, Bijlsma S, Derks Eppa, *et al.* Assessing the performance of statistical validation tools for megavariable metabolomics data. *Metabolomics* 2006; 2: 53-61.
 26. Stoll N, Schmidt E, Thurow K. Isotope pattern evaluation for the reduction of elemental compositions assigned to high-resolution mass spectral data from electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry. *J Am Soc Mass Spectr* 2006; 17: 1692-9.
 27. Kind T, Fiehn O. Seven Golden Rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry. *Bmc Bioinformatics* 2007; 8: 105.
 28. Smith CA, O'maille G, Want EJ, *et al.* METLIN - A metabolite mass spectral database. *Therapeutic Drug Monitoring* 2005; 27: 747-51.
 29. Wishart DS, Tzur D, Knox C, Eisner R, Guo AC, Young N, Cheng D, *et al.* HMDB: the human metabolome database. *Nucleic Acids Research* 2007; 35: D521-6.