Received: 3 October 2014 Revised: 18 December 2014 Accepted article published: 7 January 2015 Published online in Wiley Online Library: 2 February 2015

(wileyonlinelibrary.com) DOI 10.1002/jsfa.7070

Metabolomic applications in nutritional research: a perspective

Aoife O'Gorman[a] and Lorraine Brennan[a,b*]

Abstract

Metabolomics focuses on the global study of metabolites in cells, tissues and biofluids. Analytical technologies such as nuclear magnetic resonance (NMR) spectroscopy and hyphenated mass spectrometry (MS) combined with advanced multivariate statistical methods allow us to study perturbations in metabolism. The close link between metabolism and nutrition has seen the application of metabolomics in nutritional research increase in recent times. Such applications can be divided into three main categories, namely (1) the area of dietary biomarker identification, (2) diet-related diseases and (3) nutritional interventions. The present perspective gives an overview of these applications and an outlook to the future.

Keywords: metabolomics; nutrition; diet-related disease; biomarkers; dietary intervention

INTRODUCTION

The field of metabolomics has grown considerably over the past decade and has proven to be a powerful tool in many disciplines in biology.[1] Through the identification of small molecules or metabolites present in biological samples, metabolomics allows us to study perturbations in metabolism. By comparing metabolomic profiles between study groups for example diseased *versus* healthy etc, one can identify alterations in metabolite levels and pathways.[2,3] Such profiles reveal useful biological information, as the metabolites represent biological end points which are now implicated in the development of a number of human diseases.[4]

Currently, the main technologies used are nuclear magnetic resonance (NMR) spectroscopy and mass spectrometry (MS)-based techniques. These techniques have their advantages and disadvantages, but, owing to the complex nature of the sample matrix and the vast number of metabolites present, no single analytical technique is capable of measuring and identifying all metabolites in a single sample simultaneously. As a result, a comprehensive view of the metabolome can only be obtained by merging data from different platforms.[5] NMR is highly selective, non-destructive and requires minimal sample preparation. However, a limitation associated with NMR-based metabolomics is the lower sensitivity compared with MS-based approaches.[6] There are a range of MS-based approaches, with the most recent advances coupled with chromatographic techniques such as liquid chromatography (LC), gas chromatography (GC) and capillary electrophoresis (CE).

Prior to statistical analysis, the generated metabolomic dataset is subjected to processing methods; examples include, but are not limited to, data conversion, data scaling and normalisation, peak alignment and identification of metabolites. At present, there are a number of multivariate statistical methods available for metabolomic data, with principal component analysis (PCA), partial least squares discriminant analysis (PLS-DA) and orthogonal PLS-DA (O-PLS-DA) being the most commonly used. PCA is usually applied first for data interrogation. It is an unsupervised technique that allows data to be visualised in order to identify similarities and/or differences between sample classes. It is also used for identifying extreme outliers, thereby obtaining a first impression of the quality of the data.[7,8]

Despite its widespread use, PCA has a number of shortcomings. The main one is that it does not have an associated probabilistic model, which makes assessing the fit of PCA to the data difficult,[9] thus limiting the scope of its application. PCA also fails to describe between-metabolite relationships, as its method covers all metabolic variation simultaneously.[10] In recent years, newer tools have been developed that may be more useful for future metabolomic analysis.[10,11] One such tool is probabilistic principal component and covariate analysis (PPCCA),[9] which incorporates covariates into the model and facilitates joint modelling of metabolomic data and covariates, ensuring that the principal components provide a clear picture of the underlying data. This is an important feature when studying human metabolism, as a number of phenotypic factors are known to impact on the metabolic profile.[4]

Supervised techniques refer to those that require prior knowledge of the class of a sample. Among the many methods available, PLS-DA and O-PLS-DA are at present the most popular in metabolomic research. PLS-DA is usually used for classification purposes. It can be used either to maximise the discrimination between predefined sample groups by identifying the metabolites that are different between them or to predict which class a sample should belong to using a set of known class distributions.[12] A

* *Correspondence to: Lorraine Brennan, UCD School of Agriculture and Food Science, Institute of Food and Health, UCD Conway Institute, Belfield, Dublin 4, Ireland. E-mail: lorraine.brennan@ucd.ie*

a *UCD School of Agriculture and Food Science, Institute of Food and Health, UCD Conway Institute, Belfield, Dublin 4, Ireland*

b *Institute of Health and Society, Newcastle University, Newcastle upon Tyne, NE1 7RU, UK*

general limitation of these supervised methods is the risk of overfitting, i.e. incorporation of noise into the model. To overcome such issues, a validation step is essential; validation can be performed using techniques including cross-validation[13] or bootstrapping.[14]

More recently, pathway-based analyses have become popular in metabolomic studies. Such tools aim to visualise all measurements on pathway maps to provide a more biological context to the data, moving away from the traditional analyses (e.g. PCA) towards a more systems approach, thus allowing better interpretability.[15] There are a number of commercial and freeware tools available such as Ingenuity Pathway Analysis (IPA; Ingenuity Systems, Redwood City, CA, USA) and Cytoscape[16] that have been developed to allow enriched views of metabolic pathways, usually with close integration of metabolic-based databases such as KEGG[17] or MetaCyc.[18]

APPLICATIONS OF METABOLOMICS IN NUTRITIONAL RESEARCH

In the past few years the application of metabolomics in nutritional research has increased significantly. Broadly speaking, its application can be divided into three categories, namely (1) the area of dietary biomarker identification, (2) diet-related diseases and (3) nutritional interventions, an overview of which can be seen in Fig. 1.

Dietary biomarker identification

Application of metabolomics in the identification of dietary biomarkers has risen significantly in the past five years.[19] Dietary biomarkers/markers can give information about intake of certain foods and can thus be used as an aid in dietary assessment. The application of metabolomics to identify novel dietary biomarkers has in general taken two approaches, (i) the identification of biomarkers of specific foods and (ii) the use of biomarkers reflective of dietary patterns.

Identification of biomarkers of specific foods

To date, a number of putative biomarkers of intake of certain foods including citrus fruit,[20] red meat,[21,22] fish,[23] coffee,[24] tea[25] and broccoli[23] have been identified. For a thorough review of the current putative biomarkers of food intake, the reader is referred to the review by Scalbert *et al.*[19] For the identification of dietary biomarkers, the design of these studies is usually consumption of foods followed by collection of biofluids either postprandially or following a short-term intervention (1–2 weeks). The biofluid of choice for these types of studies is urine; however, plasma and serum are also popular. The biofluids are then analysed using one or a combination of analytical technologies leading to the identification and quantification of dietary biomarkers. Validation of the identified biomarkers in an independent cohort is desirable following the discovery stage.[4] Although many biomarkers have been identified for a range of foods, the validation of these in a separate cohort is rarely performed. Moreover, for many of these biomarkers, confirmation of a quantitative relationship with food intake has not been demonstrated, which limits their potential use in nutritional research. Furthermore, for the field to progress, it is imperative that clear demonstration of use of such biomarkers emerges in coming years. There is danger that, if demonstrating the utility of such biomarkers does not emerge in coming years, the field will remain purely academic.

Dietary patterns of habitual diet

Work in our research group has pioneered the concept of using biomarkers to reflect dietary patterns. This approach has been used in a number of recent studies[26–28] where metabolic profiles have been linked to habitual dietary patterns. The identification of such dietary patterns may be important for studying relationships between diet and disease. Recent work in our research group applied a multivariate statistical approach to link dietary data with lipidomic data to identify dietary biomarkers. In this approach, PCA was used to reduce the lipid data into lipid patterns (LPs) and regressed against dietary data to identify biomarkers of dietary intake. Our study identified six novel LPs, with LP1 highly predictive of dietary fat intake with an area under the curve (AUC) of 0.82. LP4 was highly predictive of alcohol intake (AUC = 0.81), while LP6 had a reasonably good ability to predict dietary fish intake (AUC = 0.76).[28] The novelty of this work lies in the identification of the LPs, which can be referred to as biomarker panels, that have relationships with certain foods/nutrients. The identification of panels of biomarkers has emerged in recent times as a more accurate measure of dietary exposure in comparison with the more traditional single-biomarker approach, which has associated limitations. The identification of metabolite panels should greatly improve the assessment of exposure to classes of food bioactive compounds, groups or dietary patterns in future studies.[19]

Diet-related diseases

It is well known that the diet is an important environmental exposure and its measurement is an essential component of much health-related research.[29] Epidemiological and clinical studies have concluded that many diseases with high rates of morbidity and mortality worldwide are associated with diet-related incidences, including cardiovascular disease, diabetes and cancer.[30] However, the link between specific diets and health outcomes is generally poorly understood. A primary goal of nutritional research is to optimise health and prevent or delay the disease. Therefore metabolomics is considered a powerful platform to study the interactions between diet, nutrients and human metabolism and how they together contribute to disease and health.[31] Furthermore, it could prove advantageous in terms of monitoring diseases and determining early markers of disease risk.

Recent metabolomic studies have identified biomarkers of diseases such as type 2 diabetes mellitus (T2DM) and cardiovascular disease (CVD). In a number of studies, elevated plasma concentrations of branched chain amino acids (BCAAs) have emerged as biomarkers for T2DM risk. Furthermore, disturbances of BCAA metabolism have also been described in obesity and in kidney and liver dysfunction.[32] Novel biomarkers for pre-diabetes were recently identified in a large-scale human study by Wang-Sattler *et al.*[33] Three candidate serum biomarkers (glycine, LPC18:2 and acetylcarnitine C2) of pre-diabetes which could potentially be used to develop novel strategies to prevent T2DM were identified and validated in an independent cohort. In relation to CVD, Wang *et al.*[34] identified choline, trimethylamine-*N*-oxide (TMAO) and betaine as predictors of CVD in mice. Furthermore, in a follow-up study the relationship between oral intake of phosphatidylcholine and the involvement of the intestinal microbiota in the formation of TMAO in humans was explored. The authors also examined the relationship between fasting plasma levels of TMAO and the long-term risk of incidence of major adverse cardiovascular events. The study confirmed that the production of TMAO from dietary phosphatidylcholine is dependent on metabolism by the gut flora

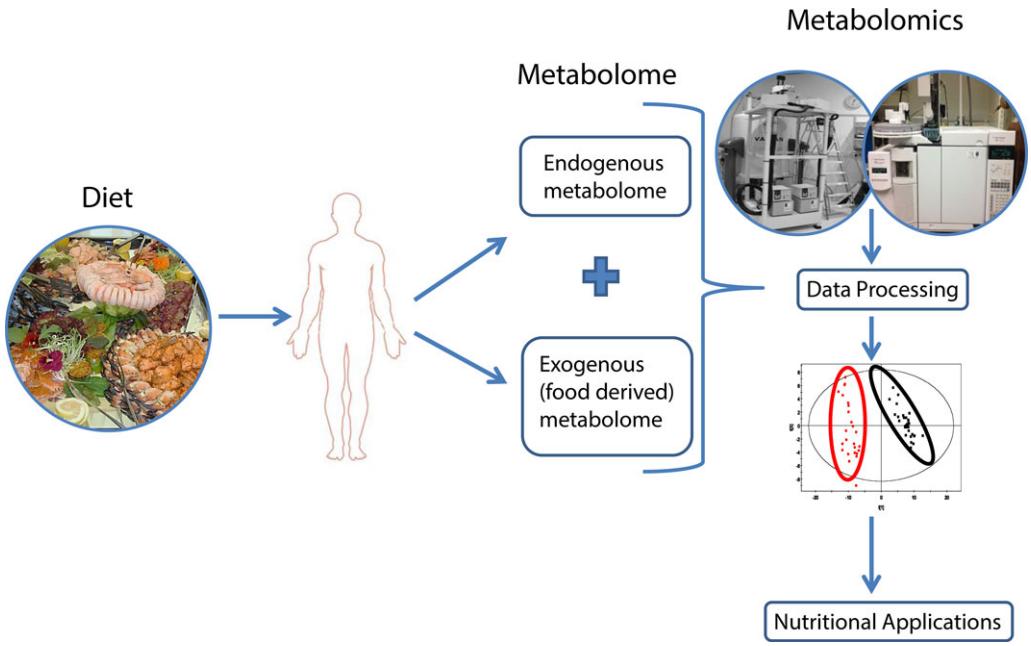

Figure 1. Metabolomic applications in nutritional research: an overview.

and that increased levels of TMAO are associated with an increased risk of CVD. It also supports earlier evidence that TMAO may have potential as a biomarker of CVD.[35]

While metabolomics has identified biomarkers for certain diet-related diseases, few have been translated into clinical diagnostics. A major reason is that the clinical community is largely unfamiliar with the field of metabolomics, its methodologies, technologies and, more importantly its clinical uses. Therefore, in order for biomarker development and translation into diagnostics to succeed, close collaboration between academic institutions, industry and clinicians is essential.[36]

Nutritional interventions

A third area where metabolomics has played a role and is set to play a further role is the application in nutritional intervention studies.[37–39] Application of metabolomics pre- and post-intervention can reveal information about the mechanism of the intervention. In one such study the authors investigated the anti-inflammatory effects induced by nutritional intervention in overweight men with mildly increased C-reactive protein (CRP) concentrations. The five-week study used a supplement or 'anti-inflammatory dietary mix' containing fish oil, green tea extract, resveratrol, vitamin E, vitamin C and tomato extracts. Through use of a combination of omics techniques, changes in protein and metabolite levels that reflected improved endothelial function, oxidative stress and fatty acid oxidation were determined. This gave the authors an insight into the mechanisms by which the anti-inflammatory mix was working.[40]

The majority of nutritional intervention studies have utilised metabolomics to characterise the response of foods such as fruits, vegetables, tea and nuts compared with a control in healthy individuals. Although many of these studies yielded a handful of endogenous metabolites that differ in abundance between the interventions, these compounds are generally reported as differences in metabolite profiles owing to a lack of adequate analysis tools. Therefore it is often unclear if such differences are indicative of perturbations in specific pathways or molecular targets in response to the dietary intervention, or are unrelated compounds identified by chance. However, although such studies have provided some hypothesis on mechanisms through which foods may promote health, the interpretation of metabolite changes is often difficult and speculative. Progress in mapping metabolites to annotated biological pathways is essential to help move this field forward.[41]

OUTLOOK

Metabolomics has been increasingly utilised in nutritional studies with significant impact. However, for it to reach its full potential, a number of challenges need to be addressed. One such challenge relates to the area of validation. In order for identified metabolites to be properly used as robust biomarkers, they require full validation, which should involve confirmation in a separate cohort(s) using authentic standards and/or previously published and validated methods. This essential step in metabolomics is very often omitted but is clearly required to allow clear biological interpretation.[42]

Another major challenge facing groups employing metabolomics is the identification of metabolites; there is a lack of resources to support metabolite identification. Furthermore, chemical standards for many of the metabolites of interest from a nutrition viewpoint are not available. To progress the field, there is a need for (1) freely available annotated databases and (2) chemical standard libraries. Although advances have been seen with the development of a number of metabolite reference databases such as Metlin, MassBank, FiehnLib and Chenomx, much work is still needed. Novel algorithms that score metabolic signals and help in their annotation in order to increase the quality of metabolomics are the next step. At present, there is no software available to perform this much needed task.[43]

Shared chemical standard libraries have the potential to aid metabolomics to be a success in nutritional research. The availability of reference standards for the whole community could allow the creation of in-house databases which would further

facilitate identification and quantification. Notwithstanding these challenges, metabolomics has proven to be a valuable technology for significantly advancing nutritional research and has a bright, exciting future ahead.

ACKNOWLEDGEMENTS

We kindly acknowledge financial support from Nutritech project number 289511 and from the Food for Health Research initiative (NDP 2007-2013; 07FHRIUCD1) of the Irish Department of Agriculture, Fisheries and Food, the Health Research Board and the Department of Health and Children.

REFERENCES

1 Putri SP, Nakayama Y, Matsuda F, Uchikata T, Kobayashi S, Matsubara A, *et al*, Current metabolomics: practical applications. *J Biosci Bioeng* **115**:579–589 (2013).
2 Nicholson JK, Holmes E, Kinross JM, Darzi AW, Takats Z and Lindon JC, Metabolic phenotyping in clinical and surgical environments. *Nature* **491**:384–392 (2012).
3 Fiehn O, Metabolomics – the link between genotypes and phenotypes. *Plant Mol Biol* **48**:155–171 (2002).
4 Brennan L, Metabolomics in nutrition research: current status and perspectives. *Biochem Soc Trans* **41**:670–673 (2013).
5 German JB, Hammock BD and Watkins SM, Metabolomics: building on a century of biochemistry to guide human health. *Metabolomics* **1**:3–9 (2005).
6 Lei Z, Huhman DV and Sumner LW, Mass spectrometry strategies in metabolomics. *J Biol Chem* **286**:25435–25442 (2011).
7 Rezzi S, Ramadan Z, Fay LB and Kochhar S, Nutritional metabonomics: applications and perspectives. *J Proteome Res* **6**:513–525 (2007).
8 Koulman A and Volmer DA, Perspectives for metabolomics in human nutrition: an overview. *Nutr Bull* **33**:324–330 (2008).
9 Nyamundanda G, Brennan L and Gormley IC, Probabilistic principal component analysis for metabolomic data. *BMC Bioinformatics* **11**:571 (2010).
10 Jansen JJ, Szymanska E, Hoefsloot HC, Jacobs DM, Strassburg K and Smilde AK, Between metabolite relationships: an essential aspect of metabolic change. *Metabolomics* **8**:422–432 (2012).
11 Sugimoto M, Kawakami M, Robert M, Soga T and Tomita M, Bioinformatics tools for mass spectroscopy-based metabolomic data processing and analysis. *Curr Bioinform* **7**:96–108 (2012).
12 Bartel J, Krumsiek J and Theis FJ, Statistical methods for the analysis of high-throughput metabolomics data. *CSBJ* **4**:e201301009 (2013).
13 Westerhuis JA, Hoefsloot HCJ, Smit S, Vis DJ, Smilde AK, van Velzen EJJ, *et al*, Assessment of PLSDA cross validation. *Metabolomics* **4**:81–89 (2008).
14 Wehrens R, Putter H and Buydens LMC, The bootstrap: a tutorial. *Chemometrics Intell Lab Syst* **54**:35–52 (2000).
15 Kotze HL, Armitage EG, Sharkey KJ, Allwood JW, Dunn WB, Williams KJ, *et al*, A novel untargeted metabolomics correlation-based network analysis incorporating human metabolic reconstructions. *BMC Syst Biol* **7**:107 (2013).
16 Smoot ME, Ono K, Ruscheinski J, Wang PL and Ideker T, Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics* **27**:431–432 (2011).
17 Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M and Tanabe M, Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Res* **42**:D199–D205 (2014).
18 Caspi R, Altman T, Dreher K, Fulcher CA, Subhraveti P, Keseler IM, *et al*, The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res* **40**:D742–D753 (2012).
19 Scalbert A, Brennan L, Manach C, Andres-Lacueva C, Dragsted LO, Draper J, *et al*, The food metabolome: a window over dietary exposure. *Am J Clin Nutr* **99**:1286–1308 (2014).
20 Heinzmann SS, Brown IJ, Chan Q, Bictash M, Dumas ME, Kochhar S, *et al*, Metabolic profiling strategy for discovery of nutritional biomarkers: proline betaine as a marker of citrus consumption. *Am J Clin Nutr* **92**:436–443 (2010).
21 Stella C, Beckwith-Hall B, Cloarec O, Holmes E, Lindon JC, Powell J, *et al*, Susceptibility of human metabolic phenotypes to dietary modulation. *J Proteome Res* **5**:2780–2788 (2006).
22 Cross AJ, Major JM and Sinha R, Urinary biomarkers of meat consumption. *Cancer Epidemiol Biomarkers Prev* **20**:1107–1111 (2011).
23 Lloyd AJ, Fave G, Beckmann M, Lin W, Tailliart K, Xie L, *et al*, Use of mass spectrometry fingerprinting to identify urinary metabolites after consumption of specific foods. *Am J Clin Nutr* **94**:981–991 (2011).
24 Mennen LI, Sapinho D, Ito H, Bertrais S, Galan P, Hercberg S, *et al*, Urinary flavonoids and phenolic acids as biomarkers of intake for polyphenol-rich foods. *Br J Nutr* **96**:191–198 (2006).
25 Hodgson JM, Chan SY, Puddey IB, Devine A, Wattanapenpaiboon N, Wahlqvist ML, *et al*, Phenolic acid metabolites as biomarkers for tea- and coffee-derived polyphenol exposure in human subjects. *Br J Nutr* **91**:301–306 (2004).
26 Menni C, Zhai G, Macgregor A, Prehn C, Romisch-Margl W, Suhre K, *et al*, Targeted metabolomics profiles are strongly correlated with nutritional patterns in women. *Metabolomics* **9**:506–514 (2013).
27 Pere-Trepat E, Ross AB, Martin FP, Rezzi S, Kochhar S, Hasselbalch AL, *et al*, Chemometric strategies to assess metabonomic imprinting of food habits in epidemiological studies. *Chemometrics Intell Lab Syst* **104**:95–100 (2010).
28 O'Gorman A, Morris C, Ryan M, O'Grada CM, Roche HM, Gibney ER, *et al*, Habitual dietary intake impacts on the lipidomic profile. *J Chromatogr B* **966**:140–146 (2014).
29 Fave G, Beckmann ME, Draper JH and Mathers JC, Measurement of dietary exposure: a challenging problem which may be overcome thanks to metabolomics? *Genes Nutr* **4**:135–141 (2009).
30 Armitage EG, Ruperez FJ and Barbas C, Metabolomics of diet-related diseases using mass spectrometry. *Trends Anal Chem* **52**:61–73 (2013).
31 Hyotylainen T, Bondia-Pons I and Oresic M, Lipidomics in nutrition and food research. *Mol Nutr Food Res* **57**:1306–1318 (2013).
32 Adeva MM, Calvino J, Souto G and Donapetry C, Insulin resistance and the metabolism of branched-chain amino acids in humans. *Amino Acids* **43**:171–181 (2012).
33 Wang-Sattler R, Yu Z, Herder C, Messias AC, Floegel A, He Y, *et al*, Novel biomarkers for pre-diabetes identified by metabolomics. *Mol Syst Biol* **8**:615 (2012).
34 Wang Z, Klipfell E, Bennett BJ, Koeth R, Levison BS, Dugar B, *et al*, Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. *Nature* **472**:57–63 (2011).
35 Tang WH, Wang Z, Levison BS, Koeth RA, Britt EB, Fu X, *et al*, Intestinal microbial metabolism of phosphatidylcholine and cardiovascular risk. *New Engl J Med* **368**:1575–1584 (2013).
36 Collino S, Martin FP and Rezzi S, Clinical metabolomics paves the way towards future healthcare strategies. *Br J Clin Pharmacol* **75**:619–629 (2012).
37 Perez-Cornago A, Brennan L, Ibero-Baraibar I, Hermsdorff HH, O'Gorman A, Zulet MA, *et al*, Metabolomics identifies changes in fatty acid and amino acid profiles in serum of overweight older adults following a weight loss intervention. *J Physiol Biochem* **70**:593–602 (2014).
38 Rezzi S, Ramadan Z, Martin FP, Fay LB, van Bladeren P, Lindon JC, *et al*, Human metabolic phenotypes link directly to specific dietary preferences in healthy individuals. *J Proteome Res* **6**:4469–4477 (2007).
39 Navarro SL, Chen Y, Li L, Li SS, Chang JL, Schwarz Y, *et al*, UGT1A6 and UGT2B15 polymorphisms and acetaminophen conjugation in response to a randomized, controlled diet of select fruits and vegetables. *Drug Metabol Dispos* **39**:1650–1657 (2011).
40 Bakker GC, van Erk MJ, Pellis L, Wopereis S, Rubingh CM, Cnubben NH, *et al*, An antiinflammatory dietary mix modulates inflammation and oxidative and metabolic stress in overweight men: a nutrigenomics approach. *Am J Clin Nutr* **91**:1044–1059 (2010).
41 Lampe JW, Navarro SL, Hullar MA and Shojaie A, Inter-individual differences in response to dietary intervention: integrating omics platforms towards personalised dietary recommendations. *Proc Nutr Soc* **72**:207–218 (2013).
42 Villas-Boas SG, Koulman A and Lane GA, Analytical methods from the perspective of method standardization, in *Metabolomics*, ed. by Nielsen J and Jewett MC. Springer, Berlin, pp. 11–52 (2007).
43 Fiehn O, *Quality control in metabolomics*. Restek Corporation, Bellefonte (2014).