Proceedings of the Nutrition Society (2017), **76**, 295–302 doi:10.1017/S0029665116002974
 First published online 16 January 2017

Nutrition Society Summer Meeting 2016 held at the University College Dublin on 11–14 July 2016

Conference on 'New technology in nutrition research and practice' Symposium 2: Use of biomarkers in dietary assessment and dietary exposure

The role of metabolomics in determination of new dietary biomarkers

A. O'Gorman* and L. Brennan

Institute of Food & Health, UCD School of Agriculture and Food Science, University College Dublin, Belfield Road, Dublin 4, Ireland

Traditional methods for the assessment of dietary intake are prone to error; in order to improve and enhance these methods increasing interest in the identification of dietary biomarkers has materialised. Metabolomics has emerged as a key tool in the area of dietary biomarker discovery and to date the use of metabolomics has identified a number of putative biomarkers. Applications to identify novel biomarkers of intake have in general taken three approaches: (1) specific acute intervention studies to identify specific biomarkers of intake; (2) searching for biomarkers in cohort studies by correlating to self-reported intake of a specific food/food group(s); (3) analysing dietary patterns in conjunction with metabolomic profiles to identify biomarkers and nutritypes. A number of analytical technologies are employed in metabolomics as currently there is no single technique capable of measuring the entire metabolome. These approaches each have their own advantages and disadvantages. The present review will provide an overview of current technologies and applications of metabolomics in the determination of new dietary biomarkers. In addition, it will address some of the current challenges in the field and future outlooks.

Metabolomics: Dietary intake: Biomarkers: Nutrition

Metabolomics is the youngest member of the 'omics' family, joining genomics, proteomics and transcriptomics as tools in global systems biology[1]. Metabolomics studies the small molecular weight molecules or metabolites that are present in biological samples with an aim to identify perturbations in metabolism under different conditions[2]. It complements other 'omic' technologies such as trancriptomics and proteomics and is considered to best reflect activities at a functional level[3]. The metabolome responds to nutrients, stress or disease long before the transcriptome or proteome making it an attractive approach for multiple fields, with metabolite alterations now implicated in the development of a number of human diseases[4–6].

The metabolomics pipeline is composed of a number of steps. In general, these steps involve: (i) experimental design; (ii) sample preparation; (iii) data acquisition; (iv) data processing; (v) statistical analysis[2,7], an overview of which is illustrated in Fig. 1. All stages should be carefully designed and executed in order to provide valid datasets and ultimately valid experimental conclusions and hypotheses[7]. Numerous comprehensive reviews on the experimental strategies in metabolomics are available elsewhere[8–11]. The focus of this review will provide an overview of current technologies and applications of metabolomics in the determination of new dietary biomarkers. In addition, it will also address challenges in the field and future outlooks.

Metabolomic technologies

A number of analytical technologies[12] are employed in metabolomics with the ultimate goal of analysing a large fraction or all of the metabolites present. Due to metabolite diversity and the range of concentrations in which they are present (pM–mM), a range of these technologies is often used as at present no single technique has the capability to measure the whole metabolome[8,10,11]. An overview of metabolites captured using different

Abbreviations: 2D, two-dimensional; PC, principal components.

***Corresponding author**: Dr A. O'Gorman, email aoife.ogorman@ucd.ie

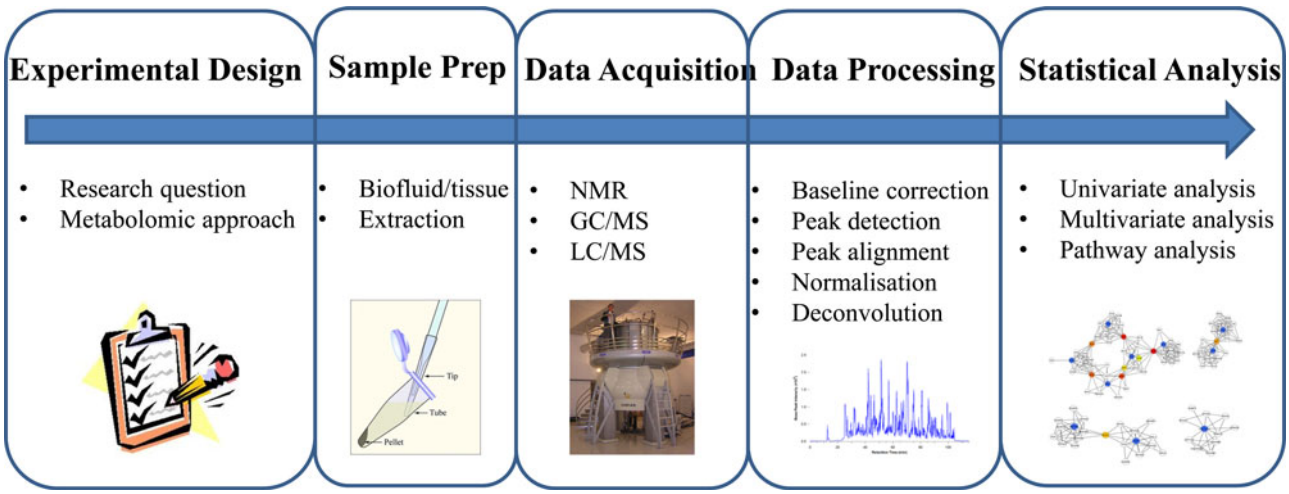

Fig. 1. (Colour online) An overview of the metabolomics pipeline. LC/MS, liquid chromatography–MS.

technologies is depicted in Table 1. In general, metabolomic analyses have been classified into targeted or non-targeted approaches and the type of approach used for a particular study will be dependent on the research question/design. The two major metabolomic platforms used in these approaches are NMR spectroscopy and MS. Non-targeted metabolomics involves measuring as many metabolites as possible in a biological sample simultaneously, therefore providing a broader coverage of metabolites[13]. This approach offers the opportunity for novel target discovery; however, the challenges lie in the time required to process the extensive amounts of raw data produced, difficulties in relation to characterising unknowns and bias towards high-abundance molecules[14,15]. In contrast, targeted approaches are taken when specific classes of metabolites are to be measured. Through the use of internal standards, analysis can be carried out in a quantitative manner[14]. The major limitation of this approach is that it requires the compounds of interest to be known *a priori*, which need to be available commercially in purified form as standards in order to be quantified[16].

NMR spectroscopy

NMR spectroscopy has been utilised extensively in the field of metabolomics research and has played a key role in our understanding of metabolism for many decades[17,18]. It benefits from being perhaps the most selective analytical technique, with its ability to provide unambiguous information about a molecule, an important aspect in terms of characterising components of complex mixtures[7,17]. NMR does not require extensive sample preparation time, has high reproducibility and little inter-laboratory variability[18]. In addition, the analysis is non-destructive and does not require pre-selection of the analysis conditions, such as ion source conditions for MS or chromatographic operating conditions[7]. However, a major drawback of NMR in comparison with MS-based methods is its limited sensitivity[19]. Another barrier delaying more prevalent use of NMR as a metabolomic tool is the need for manual spectral profiling[20].

The majority of metabolomic applications employ (one-dimensional) ^{1}H (proton) NMR as the majority of known metabolites contain hydrogen atoms. In comparison with other methods, NMR is non-biased to particular metabolites, i.e. all metabolites will be detected once they are present in concentrations above the limit of detection. Two-dimensional (2D) NMR experiments such as TOCSY, ^{1}H J-RES and ^{1}H–^{13}C HSQC are important to use in conjunction with one-dimensional spectra for metabolite identification[21] and to compare with reference databases such as the Human Metabolome Database[22] and Biological Magnetic Resonance Databank[23]. In addition, one-dimensional and 2D NMR experiments offer the potential to identify previously unknown metabolites[24].

NMR has the capability of not only measuring solution states (e.g. urine, plasma and serum), but can also measure tissue samples directly through a technique called magic angle spinning[25]. Its application has been particularly utilised in the area of tumour monitoring[26]. For example, it has been used to robustly determine the differences between benign and malignant tissue from patients with breast and colon cancer with a high degree of sensitivity and specificity[27]. This technique can also be used for real-time monitoring as it is conceivable that magic angle spinning–NMR spectroscopy can be performed within a 10–20 min time frame, highlighting the translational potential as a clinical resource for rapid diagnostics[5].

As mentioned previously there are some disadvantages associated with NMR, the major one being its limited sensitivity. To address this issue a number of recent developments have been seen. For example, the use of labelled compounds can lead to significant improvements in sensitivity (e.g. using ^{13}C or ^{15}N NMR)[28] and instrumental advances such as the use of cryogenically cooled probes and microcoil probes have also enhanced sensitivity, allowing the detection of metabolites at low abundance[21,29].

https://doi.org/10.1017/S0029665116002974 Published online by Cambridge University Press

Table 1. Overview of the metabolite coverage achievable by the different metabolomic technologies

Metabolomic technique	Metabolite coverage	References
GC–MS	Amino acids Organic acids Fatty acids Phosphates Sugars Alcohols Steroids Bile acids Nucleotides	Wishart *et al.*[22]; Halket *et al.*[34]; Wishart *et al.*[80]; Bouatra *et al.*[81]; Schauer *et al.*[82], Psychogios *et al.*[83]
LC–MS	Amino acids Organic acids Fatty acids Sugars Sterols Steroids Glycerophospholipids Glycerolipids Sphingolipids Eicosanoids Prenol lipids Oxylipins Polyketides Saccharolipids Bile acids Metal ions Neurotransmitters Biogenic amines Nucleotides	Wishart *et al.*[22]; Halket *et al.*[34]; Wishart *et al.*[80]; Bouatra *et al.*[81]; Psychogios *et al.*[83]
NMR spectroscopy	Amino acids Organic acids Keto acids Sugars Alcohols Lipids (HDL, LDL, lipoproteins particles) Nucleotides	Wishart *et al.*[22]; Kobayashi *et al.*[23]; Wishart *et al.*[80]; Bouatra *et al.*[81]; Psychogios *et al.*[83]

LC–MS, liquid chromatography–MS.

In recent years, a number of software packages have been developed that support semi-automatic NMR spectral profiling of one-dimensional and 2D ^{1}H NMR spectra[30,31]. Of note is an open-source software called BAYESIL (http://www.bayesil.ca), which provides a fully automated system, allowing analysis of complex mixtures quickly and accurately. The development of such tools provides quantitative and accurate NMR profiling effectively, without the need for trained experts, enabling a wealth of new applications of NMR in clinical settings[20].

MS-based technologies

MS-based techniques are coupled with a chromatographic step, most commonly GC or liquid chromatography. This chromatographic step allows the separation of metabolites prior to detection, which reduces the complexity of the mass spectra and enhances resolution, sensitivity and selectivity[32,33].

GC–MS has long been used in metabolomics for the comprehensive analysis of metabolites due to its high selectivity and sensitivity. It is most suitable for volatile metabolites and those that can become volatile following chemical derivatisation. Although, chemical derivatisation can improve significantly the GC separation of compounds, it is also a disadvantage as it can introduce unwanted artefacts[16,34]. The development of comprehensive 2D GC further enhances separation performance by coupling two columns coated with different stationary phases, which greatly increases separation capacity[32], chemical selectivity and sensitivity, thus providing more accurate information about metabolite retention times and mass spectra[35]. A recent study by Rocha *et al.*[36] used a 2D-GC–MS untargeted approach to analyse human urine. They identified 700 compounds from a diverse number of chemical families (e.g. ketones, alcohols, aldehydes, thiols, amines, etc.), providing the most complete information available on the volatile components of human urine.

Liquid chromatography–MS requires no need for chemical derivatisation of metabolites, which is an advantage over GC–MS. Analysis of a wide range of metabolites ranging from low to high molecular weight and from hydrophilic to hydrophobic can be carried out through selection of the appropriate column and mobile phases. HPLC separations are best suited for the analysis of labile and non-volatile polar and non-polar compounds in their native form[37]. Column technology has also improved greatly in terms of metabolite coverage and for reducing analysis time. The introduction of ultra-high pressure liquid chromatography, using porous particles with internal diameter smaller than 2 μm, in conjunction with MS[37] has substantially increased chromatographic resolution and peak capacity compared with conventional HPLC columns[38]. Overall, the ultra-high pressure liquid chromatography–MS technique shows great promise as a hyphenated micro-separation tool in metabolomics as the majority of primary metabolites are intrinsically polar[37].

During the last decade liquid chromatography–MS techniques have developed, which employ soft ionisation approaches such as electrospray ionisation and atmospheric pressure chemical ionisation, making MS more sophisticated and more robust for daily use[16].

A number of different types of MS detectors exist, which vary in terms of cost, selectivity, sensitivity and accuracy. These include ion traps, single quadrupoles, triple quadrupoles, time-of-flight, quadrupole time-of-flight, orbitraps and Fourier transform ion cyclotron resonance. Triple quadrupoles are particularly good for quantification of biomarkers because of their high selectivity and robustness, whereas quadrupole time-of-flights are particularly suited for untargeted metabolomic analysis because of their higher mass accuracy (detect the difference between m/z 300·00 and 300·003) and scan rate. They are also increasingly used for quantification, much like the triple quadrupole instruments[39].

https://doi.org/10.1017/S0029665116002974 Published online by Cambridge University Press

MS-based metabolomics offers quantitative analyses with the ability to quantify very low concentrations of potential biomarkers[33]. Metabolites can be fully quantified through the use of internal standards, ideally using stable isotope labelled standards and spiking them into the sample or semi-quantitatively which is often the case when a range of metabolites are measured. Semi-quantification normalises the metabolite signal intensity to that of an internal standard or another relative metabolite, whereas full quantification or absolute quantification determines the absolute metabolite quantity. Stable isotopes are critical for absolute quantification; however, the availability of commercial isotope labelled standards is limited and costs can be prohibitive to large-scale use[38].

A drawback of these conventional separation methods is the low sample throughput, so a number of complementary approaches are emerging in order to reduce analysis time[40]. One such approach is direct infusion MS, which allows the direct introduction of a sample into the spectrophotometer, bypassing the conventional chromatographic separation step. Advantages include rapid analysis, reproducibility and consequently high-throughput screening ability[41]; however, it does suffer from metabolite interferences, particularly in complex matrices such as serum[40].

Overall, one of the main advantages of these MS techniques is the associated high sensitivity, allowing the detection of metabolites that are below the detection limit of ^{1}H NMR spectroscopy. Exciting developments in automation and quantitation for NMR and MS-based metabolomics have been recently described[42,43]. Shifts towards commercial kits, automation and better standardisation will ultimately reduce costs, increase throughput, allow greater reproducibility and substantially reduce sample-handling errors[42].

Metabolomics and dietary biomarkers

The diet is an important environmental exposure and therefore its measurement is a vital part of health-related research. Measuring habitual dietary intake should be both accurate and applicable to large numbers of free-living individuals, hence measuring dietary exposure is one of the greatest challenges in nutritional research[44]. Traditional tools for collecting information on dietary intake including FFQ, food diaries and 24-h recalls are often unreliable and are subject to possible underreporting and recall errors[45,46]. Due to these well-documented problems, there has been a growing appreciation to improve methods to assess dietary intake[47,48]. The use of dietary biomarkers provides a more objective and accurate measure of intake and if used in combination with traditional methods will improve the ability to assess dietary intake. Currently, ideal biomarkers exist for salt and protein intake (sodium/nitrogen measure in a 24-h urine) and energy expenditure (double-labelled water technique)[45]. The further development of robust dietary biomarkers will improve the assessment of the relationship between diet and chronic disease[49]. In recent years, metabolomics has emerged as a key tool in dietary biomarkers discovery.

Applications of metabolomics to identify novel biomarkers of dietary intake have in general taken three approaches: (1) specific acute intervention studies to identify specific biomarkers of intake; (2) searching for biomarkers in cohort studies by correlating to self-reported intake of a specific food/food group(s); (3) analysis of dietary patterns in conjunction with metabolomic profiles to identify nutritypes and biomarkers. These biomarker discovery studies have in general applied untargeted metabolomic approaches[50].

Use of specific acute intervention studies to identify specific biomarkers of intake

Dietary intervention studies generally involve participants consuming specific food(s) followed by the collection of biofluids either postprandially or following a short-term intervention[2]. The most popular choice of biofluid for these types of studies is urine, due to it being non-invasive; however, plasma and serum are also possible and useful. This approach has resulted in the identification of a number of putative biomarkers of specific foods and drinks such as citrus fruit[49,51,52], cruciferous vegetables[53,54], red meat[55,56], coffee[57,58], tea[59,60], sugar-sweetened beverages[61] and wine[62]. A good example of a robust biomarker of citrus fruit is proline betaine, which was identified initially by Atkinson *et al.*[52] and subsequently validated by independent research groups[49,51,63]. The biofluid used in all of these studies was urine; however, different metabolomic analytical strategies were applied to measure proline betaine, which included NMR and MS. One such study was performed by Heinzmann *et al.*[49], which involved eight volunteers consuming standardised meals over a 3 d period. On the second day of the study, a mixed-fruit meal (apple, orange, grapefruit and grapes) was introduced. Urine samples were collected and subsequent analysis via NMR spectroscopy and partial least-squares-discriminant analysis identified proline betaine as a potential biomarker of citrus fruit intake. To confirm the findings, proline betaine was measured in citrus fruits and the urinary excretion kinetics were evaluated. Furthermore, the biomarker was further validated in a large cohort study (INTERMAP, UK), with excellent sensitivity and specificity (90·6 and 86·3 %, respectively) for discriminating between consumers and non-consumers of citrus fruit[49].

Although a number of biomarkers of specific foods have been reported, it is worth noting that few have been validated in large separate cohorts, making it difficult to translate these biomarkers into practice. It should also be noted that these acute biomarkers of intake are often short-term biomarkers that are rapidly excreted in urine, almost completely over a period of 24 h[64]. Therefore, searching for longer-term biomarkers of habitual intake is required.

Searching for biomarkers in cohort studies by correlating to self-reported intake of a specific food/food group(s)

Searching for biomarkers of specific food(s) can also be carried out through the use of cohort studies. In this

https://doi.org/10.1017/S0029665116002974 Published online by Cambridge University Press

approach, dietary data are collected using a traditional method (e.g. FFQ, food diaries) to identify low and high consumers or consumers and non-consumers of a specific food. Metabolomic profiles are then compared between these groups in order to identify potential biomarkers. These cohort studies tend to be larger in terms of study participants in comparison with acute intervention studies, but rely on self-reported dietary assessment methods which are prone to error. It also needs to be highlighted that biomarkers identified in cohort studies do not assess the direct relationships of food amounts consumed, they are simply correlations between the food and the metabolite(s) and therefore the relationship is only an association[65]. Confirmation of these associations would be required in an intervention study in order to validate the metabolite as a specific biomarker of intake.

To date, a number of biomarkers of food intake have been identified using cohort studies for e.g. fish[66], red meat[67,68], whole-grain bread[69] and walnuts[70]. An important recent study applied metabolomics to serum samples to identify biomarkers of red meat intake and identified a relationship between a number of those identified biomarkers (ferritin, glycine, diacyl phosphatidylcholines 36 : 4 and 38 : 4, lysophosphatidylcholine 17 : 0 and hydroxy-sphingomyelin 14 : 1) with risk of type-2 diabetes. The authors found that high levels of ferritin, low glycine and altered hepatic derived lipids in the circulation were associated with both total red meat consumption and diabetes risk. The findings are consistent with the hypothesis that metabolic processes reflected in the circulating concentrations of these biomarkers take part in linking red meat consumption to type-2 diabetes risk[67].

This was the first reported study to evaluate a large set of metabolites as potential mediators linking exposure and disease, which is an important next step in biomarker discovery. Identifying these links between diet and disease will provide an insight into which metabolites and metabolic pathways are potential disease mediators, which could then be targeted or modulated through dietary interventions to improve health outcomes. In order to achieve this much work is required, particularly in the area of biomarker validation.

Use of dietary patterns to identify biomarkers of intake

The concept of identifying biomarkers using dietary patterns in combination with metabolomic patterns was pioneered in our research group[71]. Since then, this approach has been used in a number of studies[72–76], where patterns of intake are related to metabolomic patterns[2]. This approach generally involves applying a multivariate statistical strategy such as principal component (PC) analysis or *k*-means cluster analysis to dietary data to identify dietary patterns and then through the use of regression (or other statistical method) linking these to metabolomic profiles in order to identify dietary biomarkers and/or nutritypes.

A recent study used this novel approach to distinguish between two dietary patterns in an attempt to develop a compliance tool[77]. An untargeted approach was applied using ultra-high pressure liquid chromatography/quadrupole time-of-flight/MS to analyse urine samples from 181 participants as part of a parallel intervention study, who were randomly assigned to follow either a New Nordic Diet or an Average Danish Diet for 6 months. Partial least-squares-discriminant analysis was applied to the urinary metabolomic data to develop a compliance model for the New Nordic Diet and Average Danish Diet based on the metabolites that were identified as being the most discriminatory between the two diets. This resulted in a model with a misclassification error rate of 19 %, showing good promise as a compliance measure for different dietary patterns, which could be used to identify non-compliant subjects or groups of individuals with certain dietary responses[77]. This study demonstrates that metabolomics can be used to discover which metabolites are the strongest predictors of compliance to complex diets; however, these metabolites should be followed up by quantitative measurements to further enhance and validate the model.

The identification of dietary patterns may also be important for studying relationships between diet and disease. For example, Bouchard-Mercier *et al.*[76] investigated the metabolic signatures associated with the Western and Prudent dietary patterns using a targeted approach to profile participants (*n* 37) plasma. Applying PC analysis to the metabolic plasma profiles resulted in the identification of two PC. The first, PC1 was mainly composed of medium- to long-chain acylcarnitines, whereas PC2 was dominated by short-chain acylcarnitines and amino acids, including the branched-chain amino acids. The authors found that PC1 was not correlated to any food groups; however, PC2 was negatively correlated with fruit intake and positively associated with desserts ($r = -0{\cdot}38$, $P = 0{\cdot}03$; $r = 0{\cdot}37$, $P = 0{\cdot}04$, respectively). In addition, PC2 also had a significant positive correlation with saturated fat intakes ($r = 0{\cdot}39$, $P = 0{\cdot}02$). The Western dietary pattern had an inverse relationship with PC1 and a positive relationship with PC2, suggesting that people eating a Western diet are potentially at risk of increasing their long-term risk of cardiometabolic diseases[76].

In summary, these studies indicate the potential of metabolomics as a tool for not only evaluating compliance to a dietary pattern, but also to identify and evaluate relationships between diet and disease.

Challenges and future outlooks

The application of metabolomics for dietary biomarker identification has grown significantly over the past 5 years[78] and in general this approach has produced a number of robust biomarkers of dietary intake. However, a number of challenges exist that need to be overcome in order to advance this field of research. One of the main challenges is in the area of biomarker validation, which is often omitted in study design. Following biomarker identification whether it is via an acute intervention study or cohort study, subsequent

https://doi.org/10.1017/S0029665116002974 Published online by Cambridge University Press

validation in an independent cohort(s) is critical. The independent cohort should not only be large, but also diverse (e.g. multicultural populations) and a number of considerations need to included such as age, sex, ethnicity, as well as sample processing, chromatographic separation and analytical instrumental settings. Another important aspect is the validation of the biomarkers performance across different laboratories. It can be a challenge to obtain identical metabolomic profiles as often laboratories use different protocols and therefore inter-laboratory validation is important[43].

Ideally, a well-validated biomarker should demonstrate a dose–response, which would confirm its suitability for use over a range of intakes[78]; however, in the majority of studies this important step is often missing. Indeed biomarkers identified solely from cohort studies fail to assess the direct relationships of foods consumed and do not demonstrate responsiveness to intake, identifying merely an association between the biomarker and food[65]. A recent review by Gibbons highlighted the need to combine such studies with acute intervention studies to examine direct relationships and dose–response relationships[64].

Another challenge is in the area of metabolite identification, which is one of the main limiting factors, particularly when using MS-based metabolomics. The reason for this is the huge chemical diversity that is present in biological samples[79]. To progress this field the metabolomics community are actively developing databases; examples include the Human Metabolome Database, Metlin and MassBank, which are publicly available. The Human Metabolome Database contains over 41 000 metabolites and is the most comprehensive collection of human metabolism data in the world[80]. A similar database containing food constituents and food additive metabolites has also been created and is updated regularly, which is called the FoodDB. The database contains over 28 000 metabolites, a valuable tool for food and nutrition researchers[78], although caution needs to be exerted as identifying metabolites originating from foods remains difficult.

Conclusion

Metabolomics has proven a powerful tool in the area of dietary biomarker discovery and its application has the potential to greatly enhance our ability to assess dietary intake. However, it is pertinent that validated biomarkers of intake are translated into practice. In order for them to meet their full potential cooperation across disciplines is necessary.

Acknowledgements

The authors kindly acknowledge our funding bodies.

Financial Support

The authors received the following funding: ERC (grant no. 647783) and SFI (grant no. JPI-HDHL/B3075, Foodball).

Conflict of Interest

None.

Authorship

A. O. G. conducted the literature search and drafted the manuscript. L. B. critically reviewed the manuscript. Both authors read and approved the final manuscript prior to admission.

References

1. Rochfort S (2005) Metabolomics reviewed: a new 'omics' platform technology for systems biology and implications for natural products research. *J Nat Prod* **68**, 1813–1820.
2. Brennan L (2013) Metabolomics in nutrition research: current status and perspectives. *Biochem Soc Trans* **41**, 670–673.
3. Putri SP, Nakayama Y, Matsuda F *et al.* (2013) Current metabolomics: practical applications. *J Biosci Bioeng* **115**, 579–589.
4. Wang TJ, Larson MG, Vasan RS *et al.* (2011) Metabolite profiles and the risk of developing diabetes. *Nat Med* **17**, 448–453.
5. Nicholson JK, Holmes E, Kinross JM *et al.* (2012) Metabolic phenotyping in clinical and surgical environments. *Nature* **491**, 384–392.
6. Oresic M, Seppanen-Laakso T, Sun D *et al.* (2012) Phospholipids and insulin resistance in psychosis: a lipidomics study of twin pairs discordant for schizophrenia. *Genome Med* **4**, 1.
7. Dunn WB, Bailey NJ & Johnson HE (2005) Measuring the metabolome: current analytical technologies. *Analyst* **130**, 606–625.
8. Goodacre R, Vaidyanathan S, Dunn WB *et al.* (2004) Metabolomics by numbers: acquiring and understanding global metabolite data. *Trends Biotechnol* **22**, 245–252.
9. Lindon JC, Holmes E, Bollard ME *et al.* (2004) Metabonomics technologies and their applications in physiological monitoring, drug safety assessment and disease diagnosis. *Biomarkers* **9**, 1–31.
10. Sumner LW, Mendes P & Dixon RA (2003) Plant metabolomics: large-scale phytochemistry in the functional genomics era. *Phytochemistry* **62**, 817–836.
11. Fiehn O (2002) Metabolomics – the link between genotypes and phenotypes. *Plant Mol Biol* **48**, 155–171.
12. Dunn WB & Ellis DI (2005) Metabolomics: current analytical platforms and methodologies. *Trac-Trends Anal Chem* **24**, 285–294.
13. Alonso A, Marsal S & Julia A (2015) Analytical methods in untargeted metabolomics: state of the art in 2015. *Front Bioeng Biotechnol* **3**, 23.
14. Roberts LD, Souza AL, Gerszten RE *et al.* (2012) Targeted metabolomics. *Curr Protoc Mol Biol* Chapter 30, Unit 30 2 1–24.
15. Vinayavekhin N & Saghatelian A (2010) Untargeted metabolomics. *Curr Protoc Mol Biol* Chapter 30, Unit 30 1 1–24.
16. Shulaev V (2006) Metabolomics technology and bioinformatics. *Brief Bioinformatics* **7**, 128–139.
17. Malet-Martino M & Holzgrabe U (2011) NMR techniques in biomedical and pharmaceutical analysis. *J Pharm Biomed Anal* **55**, 1–15.

https://doi.org/10.1017/S0029665116002974 Published online by Cambridge University Press

18. Sebedio JL & Brennan L (2014) *Metabolomics as a Tool in Nutrition Research*, 1st ed. Cambridge, UK: Woodhead Publishing.
19. Ramautar R, Demirci A & de Jong GJ (2006) Capillary electrophoresis in metabolomics. *Trac-Trends Anal Chem* **25**, 455–466.
20. Ravanbakhsh S, Liu P, Bjorndahl TC *et al.* (2015) Accurate, fully-automated NMR spectral profiling for metabolomics. *PLoS ONE* **10**, e0124219.
21. Brennan L (2014) NMR-based metabolomics: from sample preparation to applications in nutrition research. *Prog Nucl Magn Reson Spectrosc* **83**, 42–49.
22. Wishart DS, Knox C, Guo AC *et al.* (2009) HMDB: a knowledgebase for the human metabolome. *Nucleic Acids Res* **37**, D603–D610.
23. Kobayashi N, Harano Y, Tochio N *et al.* (2012) An automated system designed for large scale NMR data deposition and annotation: application to over 600 assigned chemical shift data entries to the BioMagResBank from the Riken Structural Genomics/Proteomics Initiative internal database. *J Biomol NMR* **53**, 311–320.
24. Pudakalakatti SM, Dubey A, Jaipuria G *et al.* (2014) A fast NMR method for resonance assignments: application to metabolomics. *J Biomol NMR* **58**, 165–173.
25. Beckonert O, Coen M, Keun HC *et al.* (2010) High-resolution magic-angle-spinning NMR spectroscopy for metabolic profiling of intact tissues. *Nat Protoc* **5**, 019–032.
26. Beger RD (2013) A review of applications of metabolomics in cancer. *Metabolites* **3**, 552–574.
27. Chan EC, Koh PK, Mal M *et al.* (2009) Metabolic profiling of human colorectal cancer using high-resolution magic angle spinning nuclear magnetic resonance (HR-MAS NMR) spectroscopy and gas chromatography mass spectrometry (GC/MS). *J Proteome Res* **8**, 352–361.
28. Putri SP, Yamamoto S, Tsugawa H *et al.* (2013) Current metabolomics: technological advances. *J Biosci Bioeng* **116**, 9–16.
29. Keun HC, Beckonert O, Griffin JL *et al.* (2002) Cryogenic probe ^{13}C NMR spectroscopy of urine for metabonomic studies. *Anal Chem* **74**, 4588–4593.
30. Mercier P, Lewis MJ, Chang D *et al.* (2011) Towards automatic metabolomic profiling of high-resolution one-dimensional proton NMR spectra. *J Biomol NMR* **49**, 307–323.
31. Hao J, Astle W, De Iorio M *et al.* (2012) BATMAN – an R package for the automated quantification of metabolites from nuclear magnetic resonance spectra using a Bayesian model. *Bioinformatics* **28**, 2088–2090.
32. Wang Y, Liu SY, Hu YJ *et al.* (2015) Current state of the art of mass spectrometry-based metabolomics studies – a review focusing on wide coverage, high throughput and easy identification. *RSC Adv* **5**, 78728–78737.
33. Dettmer K, Aronov PA & Hammock BD (2007) Mass spectrometry-based metabolomics. *Mass Spectrom Rev* **26**, 51–78.
34. Halket JM, Waterman D, Przyborowska AM *et al.* (2005) Chemical derivatization and mass spectral libraries in metabolic profiling by GC/MS and LC/MS/MS. *J Exp Bot* **56**, 219–243.
35. Castillo S, Mattila I, Miettinen J *et al.* (2011) Data analysis tool for comprehensive two-dimensional gas chromatography/time-of-flight mass spectrometry. *Anal Chem* **83**, 3058–3067.
36. Rocha SM, Caldeira M, Carrola J *et al.* (2012) Exploring the human urine metabolomic potentialities by comprehensive two-dimensional gas chromatography coupled to time of flight mass spectrometry. *J Chromatogr A* **1252**, 155–163.
37. Zhang A, Sun H, Wang P *et al.* (2012) Modern analytical techniques in metabolomics analysis. *Analyst* **137**, 293–300.
38. Lei ZT, Huhman DV & Sumner LW (2011) Mass spectrometry strategies in metabolomics. *J Biol Chem* **286**, 25435–25442.
39. Lind MV, Savolainen OI & Ross AB (2016) The use of mass spectrometry for analysing metabolite biomarkers in epidemiology: methodological and statistical considerations for application to large numbers of biological samples. *Eur J Epidemiol* **31**, 717–733.
40. Gonzalez-Dominguez R, Garcia-Barrera T & Gomez-Ariza JL (2014) Using direct infusion mass spectrometry for serum metabolomics in Alzheimer's disease. *Anal Bioanal Chem* **406**, 7137–7148.
41. Draper J, Lloyd AJ, Goodacre R *et al.* (2013) Flow infusion electrospray ionisation mass spectrometry for high throughput, non-targeted metabolite fingerprinting: a review. *Metabolomics* **9**, S4–S29.
42. Wishart DS (2016) Emerging applications of metabolomics in drug discovery and precision medicine. *Nat Rev Drug Discov* **15**, 473–484.
43. Gowda GA & Djukovic D (2014) Overview of mass spectrometry-based metabolomics: opportunities and challenges. *Methods Mol Biol* **1198**, 3–12.
44. Fave G, Beckmann ME, Draper JH *et al.* (2009) Measurement of dietary exposure: a challenging problem which may be overcome thanks to metabolomics? *Genes Nutr* **4**, 135–141.
45. Bingham SA (2002) Biomarkers in nutritional epidemiology. *Public Health Nutr* **5**, 821–827.
46. Kipnis V, Midthune D, Freedman L *et al.* (2002) Bias in dietary-report instruments and its implications for nutritional epidemiology. *Public Health Nutr* **5**, 915–923.
47. Jenab M, Slimani N, Bictash M *et al.* (2009) Biomarkers in nutritional epidemiology: applications, needs and new horizons. *Hum Genet* **125**, 507–525.
48. Penn L, Boeing H, Boushey CJ *et al.* (2010) Assessment of dietary intake: NuGO symposium report. *Genes Nutr* **5**, 205–213.
49. Heinzmann SS, Brown IJ, Chan Q *et al.* (2010) Metabolic profiling strategy for discovery of nutritional biomarkers: proline betaine as a marker of citrus consumption. *Am J Clin Nutr* **92**, 436–443.
50. Beckmann M, Lloyd AJ, Haldar S *et al.* (2013) Dietary exposure biomarker-lead discovery based on metabolomics analysis of urine samples. *Proc Nutr Soc* **72**, 352–361.
51. Andersen MB, Kristensen M, Manach C *et al.* (2014) Discovery and validation of urinary exposure markers for different plant foods by untargeted metabolomics. *Anal Bioanal Chem* **406**, 1829–1844.
52. Atkinson W, Downer P, Lever M *et al.* (2007) Effects of orange juice and proline betaine on glycine betaine and homocysteine in healthy male subjects. *Eur J Nutr* **46**, 446–452.
53. Edmands WMB, Beckonert OP, Stella C *et al.* (2011) Identification of human urinary biomarkers of cruciferous vegetable consumption by metabonomic profiling. *J Proteome Res* **10**, 4513–4521.
54. Andersen MBS, Reinbach HC, Rinnan A *et al.* (2013) Discovery of exposure markers in urine for Brassica-containing meals served with different protein sources by UPLC–qTOF–MS untargeted metabolomics. *Metabolomics* **9**, 984–997.

https://doi.org/10.1017/S0029665116002974 Published online by Cambridge University Press

55. Stella C, Beckwith-Hall B, Cloarec O *et al.* (2006) Susceptibility of human metabolic phenotypes to dietary modulation. *J Proteome Res* **5**, 2780–2788.
56. Cross AJ, Major JM & Sinha R (2011) Urinary biomarkers of meat consumption. *Cancer Epidemiol Biomarkers Prev* **20**, 1107–1111.
57. Heinzmann SS, Holmes E, Kochhar S *et al.* (2015) 2-Furoylglycine as a candidate biomarker of coffee consumption. *J Agric Food Chem* **63**, 8615–8621.
58. Lang R, Wahl A, Stark T *et al.* (2011) Urinary *N*-methylpyridinium and trigonelline as candidate dietary biomarkers of coffee consumption. *Mol Nutr Food Res* **55**, 1613–1623.
59. van Velzen EJJ, Westerhuis JA, van Duynhoven JPM *et al.* (2009) Phenotyping tea consumers by nutrikinetic analysis of polyphenolic end-metabolites. *J Proteome Res* **8**, 3317–3330.
60. Daykin CA, Van Duynhoven JPM, Groenewegen A *et al.* TPJ (2005) Nuclear magnetic resonance spectroscopic based studies of the metabolism of black tea polyphenols in humans. *J Agric Food Chem* **53**, 1428–1434.
61. Gibbons H, McNulty BA, Nugent AP *et al.* (2015) A metabolomics approach to the identification of biomarkers of sugar-sweetened beverage intake. *Am J Clin Nutr* **101**, 471–477.
62. Jacobs DM, Fuhrmann JC, van Dorsten FA *et al.* (2012) Impact of short-term intake of red wine and grape polyphenol extract on the human metabolome. *J Agric Food Chem* **60**, 3078–3085.
63. Pujos-Guillot E, Hubert J, Martin JF *et al.* (2013) Mass spectrometry-based metabolomics for the discovery of biomarkers of fruit and vegetable intake: citrus fruit as a case study. *J Proteome Res* **12**, 1645–1659.
64. Gibbons H & Brennan L (2016) Metabolomics as a tool in the identification of dietary biomarkers. *Proc Nutr Soc* 1–12.
65. Brennan L, Gibbons H & O'Gorman A (2015) An overview of the role of metabolomics in the identification of dietary biomarkers. *Curr Nutr Reports* **4**, 304–312.
66. Lloyd AJ, Beckmann M, Haldar S *et al.* (2013) Data-driven strategy for the discovery of potential urinary biomarkers of habitual dietary exposure. *Am J Clin Nutr* **97**, 377–389.
67. Wittenbecher C, Muhlenbruch K, Kroger J *et al.* (2015) Amino acids, lipid metabolites, and ferritin as potential mediators linking red meat consumption to type 2 diabetes. *Am J Clin Nutr* **101**, 1241–1250.
68. Myint T, Fraser GE, Lindsted KD *et al.* (2000) Urinary 1-methylhistidine is a marker of meat consumption in black and in white California seventh-day Adventists. *Am J Epidemiol* **152**, 752–755.
69. Garcia-Aloy M, Llorach R, Urpi-Sarda M *et al.* (2015) Nutrimetabolomics fingerprinting to identify biomarkers of bread exposure in a free-living population from the PREDIMED study cohort. *Metabolomics* **11**, 155–165.
70. Garcia-Aloy M, Llorach R, Urpi-Sarda M *et al.* (2014) Novel multimetabolite prediction of walnut consumption by a urinary biomarker model in a free-living population: the PREDIMED study. *J Proteome Res* **13**, 3476–3483.
71. O'Sullivan A, Gibney MJ & Brennan L (2011) Dietary intake patterns are reflected in metabolomic profiles: potential role in dietary assessment studies. *Am J Clin Nutr* **93**, 314–321.
72. Pere-Trepat E, Ross AB, Martin FP *et al.* (2010) Chemometric strategies to assess metabonomic imprinting of food habits in epidemiological studies. *Chemometr Intell Lab Syst* **104**, 95–100.
73. Menni C, Zhai G, Macgregor A *et al.* (2013) Targeted metabolomics profiles are strongly correlated with nutritional patterns in women. *Metabolomics* **9**, 506–514.
74. O'Gorman A, Morris C, Ryan M *et al.* (2014) Habitual dietary intake impacts on the lipidomic profile. *J Chromatogr B, Anal Technol Biomed Life Sci* **966**, 140–146.
75. Floegel A, von Ruesten A, Drogan D *et al.* (2013) Variation of serum metabolites related to habitual diet: a targeted metabolomic approach in EPIC-Potsdam. *Eur J Clin Nutr* **67**, 1100–1108.
76. Bouchard-Mercier A, Rudkowska I, Lemieux S *et al.* (2013) The metabolic signature associated with the Western dietary pattern: a cross-sectional study. *Nutr J* **12**, 158.
77. Andersen MB, Rinnan A, Manach C *et al.* (2014) Untargeted metabolomics as a screening tool for estimating compliance to a dietary pattern. *J Proteome Res* **13**, 1405–1418.
78. Scalbert A, Brennan L, Manach C *et al.* (2014) The food metabolome: a window over dietary exposure. *Am J Clin Nutr* **99**, 1286–1308.
79. Sebedio JL, Martin FP & Pujos E (2008) Nutritional metabolomics: What are the perspective? *OCL* **15**, 341–345.
80. Wishart DS, Tzur D, Knox C *et al.* (2007) HMDB: the human metabolome database. *Nucleic Acids Res* **35** (Database issue), D521–D526.
81. Bouatra S, Aziat F, Mandal R *et al.* (2013) The human urine metabolome. *PLoS ONE* **8**, e73076.
82. Schauer N, Steinhauser D, Strelkov S *et al.* (2005) GC–MS libraries for the rapid identification of metabolites in complex biological samples. *FEBS Lett* **579**, 1332–1337.
83. Psychogios N, Hau DD, Peng J *et al.* (2011) The human serum metabolome. *PLoS ONE* **6**, e16957.

https://doi.org/10.1017/S0029665116002974 Published online by Cambridge University Press