Role of metabolomics in identification of biomarkers related to food intake

Cassandra Collins¹,², Aoife E. McNamara¹,² and Lorraine Brennan¹,²*

¹UCD School of Agriculture and Food Science, Institute of Food and Health, UCD, Belfield, Dublin 4, Ireland
²UCD Conway Institute, UCD, Belfield, Dublin 4, Ireland

Dietary assessment methods including FFQ and food diaries are associated with many measurement errors including energy under-reporting and incorrect estimation of portion sizes. Such errors can lead to inconsistent results especially when investigating the relationship between food intake and disease causation. To improve the classification of a person’s dietary intake and therefore clarify proposed links between diet and disease, reliable and accurate dietary assessment methods are essential. Dietary biomarkers have emerged as a complementary approach to the traditional methods, and in recent years, metabolomics has developed as a key technology for the identification of new dietary biomarkers. The objective of this review is to give an overview of the approaches used for the identification of biomarkers and potential use of the biomarkers. Over the years, a number of strategies have emerged for the discovery of dietary biomarkers including acute and medium term interventions and cross-sectional/cohorts study approaches. Examples of the different approaches will be presented. Concomitant with the focus on single biomarkers of specific foods, there is an interest in the development of biomarker signatures for the identification of dietary patterns. In the present review, we present an overview of the techniques used in food intake biomarker discovery, including the experimental approaches used and challenges faced in the field. While significant progress has been achieved in the field of dietary biomarkers in recent years, a number of challenges remain. Addressing these challenges will be key to ensure success in implementing use of dietary biomarkers.

Metabolomics: Biomarkers: Food intake: Dietary patterns

In recent years, there has been growing interest in the potential of biomarkers in nutrition research. One of the areas with great expectations is the field of dietary biomarkers or food intake biomarkers. The interest in these biomarkers stems from the need for objective measures of dietary intake. The traditional methods such as FFQ, 24 h recalls and food diaries are all associated with a number of well-defined limitations including under-reporting, recall errors and difficulty in assessment of portion sizes⁴⁻³. Currently dietary biomarkers include 24 h urinary sodium, nitrogen and sucrose/fructose for estimation of salt, protein and sugar intake⁴⁻⁷.

In recent years, the concept of biomarkers reflecting specific food intake has emerged. To date a number of putative biomarkers exist for the intake of a range of foods including, but not limited to, red meat, coffee, nuts, wine, vegetables, legumes, citrus fruit, tea and sugar-sweetened beverages⁷⁻¹¹. While some confusion exists in the literature over classification of biomarkers into recovery or concentration biomarkers, we prefer to use the newly defined flex classification scheme for biomarkers related to food intake¹². Food intake biomarkers are single metabolites, or a combination of metabolites, reflecting the consumption of either a specific food.

Abbreviation: CV, cruciferous vegetables.
*Corresponding author: Lorraine Brennan, email lorraine.brennan@ucd.ie
or food group, displaying a clear time–dose–response after intake(12). With this in mind, we present here an overview of the techniques used in food intake biomarker discovery, the experimental approaches used for biomarker discovery and challenges faced in the field.

**Metabolomics: role in biomarker discovery**

Metabolomics is the study of endogenous or exogenous metabolites in an organism. Metabolites are found in tissues and bio-fluids and are influenced by a number of factors including genetics(13), the microbiome(14) and environmental exposures such as food, exercise and pollutants(15,16). Metabolomics has emerged as a key tool in biomarker studies and in particular for biomarkers related to food intake. The sensitivity of modern instrumentation used in metabolomics can detect metabolite concentrations as low as 0.1 ng/ml in plasma(17). Metabolites by their nature have a prodigious range of structures which can inhibit identification as they can be transitory intermediates or end products of biological processes. Identification of the vast array of possible metabolites is currently the limiting factor in biomarker discovery. To aid the identification of metabolites, a number of databases have emerged. The human metabolite database (HMDB – http://www.hmdb.ca/) includes 114 100 empirical and in silico compounds and is readily searchable. Other databases include MyCompoundID, a library of 8 021 endogenous human metabolites with 10 583 901 predicted products of these metabolites (http://www.mycompoundid.org/mycompoundid_IsoMS);(19) the METLIN database (http://metlin.scripps.edu);(20) and MassBank of North America (MoNA; http://mona.fiehnlab.ucdavis.edu/).

**Measurement of the metabolites**

Metabolites in biofluid samples represent a wide range of molecules with diverse chemical nature and dynamic range. As a result, a number of platforms have emerged as key players in terms of measuring metabolites for biomarker discovery. A complete detailed review of all the techniques is beyond the scope of this review but an overview is given below and readers are referred to the following review for technical details on each approach(21). In the initial years of emergence of metabolomics, the literature was dominated by NMR-based applications. NMR spectroscopy is a technique which has comparatively low sensitivity compared with other techniques(22). However, it is useful as it is non-destructive, reproducible, quantitative and furnishes structural information. Little sample preparation is required, and results are consistent between different laboratories(23).

The MS-based approaches are extremely sensitive and are often coupled with a chromatography step to help with separation of the metabolites. GC–MS is a technique particularly suited to compounds of low polarity such as fatty acids, amino acids and sterols. Preparation of samples is somewhat complicated as samples must undergo chemical derivatisation prior to analysis to ensure that they are volatile. Compounds are separated on a column by their chemical properties causing them to elute at specific times (retention time). The eluted compounds are ionised and their mass-to-charge ratio (m/z) is determined(24). This technique is particularly suited to lipids and all non-polar compounds(25).

Liquid chromatography MS is suitable for analysis of a broad range of metabolites. Its advantages over GC–MS include simple sample preparation and ability to analyse highly polar compounds(26). Metabolites are separated on a column and the eluted compounds are ionised, and their m/z and retention time is detected as output. For analysis of large batches (greater than 100 samples), one must include the necessary controls to account for instrument instability over time and batch-to-batch variation(21). Capillary electrophoresis separates compounds by their mobility in an electric field, based on their charge, viscosity and size. It is well suited to highly charged polar metabolites such as organics acids, nucleotides, peptides and their conjugates. It is coupled to MS instruments using electrospray ionisation(27). For high-throughput techniques where it is desirable to have low run time per sample direct infusion MS is often employed. In this approach, metabolites are analysed by nano-electrospray ion source after infusion directly into the ion source without prior separation. A high-resolution, high-accuracy instrument such as a Q-Exactive Orbitrap can identify individual metabolites based on their m/z ratios(28).

As afore-mentioned, a key bottleneck in employing any of these techniques is the identification of the compounds. Tandem MS or MS/MS is a powerful technique which enables identification of compounds. Using this approach, initial ionised analytes are fragmented to produce smaller product ions from a parent ion. The ions can undergo several rounds of fragmentation, depending on the instrument. The first round (MS) is known as MS1 and the subsequent fragmentation is MS2, MS3,…MSn. As modern instruments have high mass accuracy, m/z of the fragments are used to build up a profile of a compound enabling identification which can then be confirmed with original standards(29,30). Finally, it is worth noting that all these techniques can be run in either a targeted or un-targeted mode. In the targeted mode, a predefined list of metabolites are measured, whereas in an un-targeted mode, as many features as possible are measured. Depending on the research question, one can decide to operate in either mode or use a combination of both.

**Food intake biomarkers**

There are multiple study designs in which metabolomics can be applied to identify food intake biomarkers. Previous research study designs have employed one of two approaches either conducting an intervention study or using samples from a cross-sectional or epidemiology study to identify metabolites associated with food intake(31,32). Human intervention study designs involve requesting participants to consume specific food(s) over...
a defined period of time and biofluids, such as blood and urine, are collected at specific time-points depending on research interests. Once biofluids are collected, a range of metabolomic techniques as described earlier can be used to identify metabolites associated with the food intake. The time period involved in intervention studies varies depending on the research aims and can range from acute (single-day food challenge), to short- (days) or medium- (weeks) term interventions. Within the umbrella term of intervention studies, there are multiple designs and considerations. When implementing a cross-over design, participants are asked to follow specific dietary instructions, i.e. consuming a specific amount of a food of interest for a set time and changing to a diet with different amounts of, or completely lacking, the food of interest, thereby acting as their own control. Cross et al. employed this approach when examining 24 h urine samples for biomarkers of meat consumption. Participants were asked to consume four different diets for 14 d each containing a low- (60 g/d), medium- (120 g/d), high-portions of red meat (420 g/d) or a protein equivalent vegetarian diet(32). Targeted metabolic analyses were performed for four known meat-specific urinary metabolites, creatine, taurine, 1-methylhistidine and 3-methylhistidine. All four metabolites increased in concentration with increased meat consumption but only 1- and 3-methylhistidine concentrations were statistically different for each meat dose. In these cross-over studies, it is often necessary to consider a washout period: in this period certain dietary restrictions are in place, for example, avoiding specific foods/food groups for a time prior to consuming a high ‘food of interest’ diet. In a study related to cruciferous vegetables (CV) participants avoided CV and alliums for 12 days either side of a high CV diet intervention, containing broccoli and Brussel sprouts(33). Clear urinary metabolic differentiation was seen between high and low CV diets, as signified in NMR spectra by four singlet peaks which were exclusive between high and low CV diets, as significant discriminating biomarkers in the urinary metabolome of twenty high coffee consumers and nineteen non-consumers in a subset of the SU.VI.MAX2 cohort(35). Many other examples using this approach have emerged in recent years and readers are referred to Guisch-Ferré et al., for an overview of such studies(36).

Once identified it is critical that the biomarkers are assessed for validity as biomarkers of food intake. Recently a validation procedure was put forward as part of the FoodBall consortium which included plausibility, dose–response, time–response, robustness, reliability, stability, analytical performance and inter-laboratory reproducibility as the eight criteria for assessment of validation(37). While assessment of all these criteria may not be possible in a single study, it is important that they are considered and that at least the plausibility and dose–response are assessed. Using the aforesaid study designs, a number of putative biomarkers have emerged in the literature; a full review of such markers is beyond the scope of this review and readers are referred to work by the FoodBall consortium which has performed a series of systematic reviews for commonly consumed foods. The foods covered to date in the systematic reviews include (1) apples, pears and stone fruit, (2) legumes, (3) dairy and egg products and (4) non-alcoholic beverages(38–41). Other reviews which cover the commonly consumed foods in Europe are underway. From the presently published reviews, it is obvious that a number of putative markers exist; however, there are no fully validated markers of these foods. This highlights the urgency in developing strategies to ensure that we have fully validated biomarkers.

Use of food intake biomarkers in quantifying intake

The ultimate goal of a food intake biomarker is to quantify intake of the specific food. Despite the proliferation in the number of putative biomarkers of food intake, there is paucity of data demonstrating the quantitative ability of food intake biomarkers. Notwithstanding this, there are two examples in the literature that demonstrate the potential.

Examining the potential of the well-established marker of citrus intake, our previous work demonstrated that proline betaine could be used to determine citrus intake. Using a controlled dietary intervention approach, participants consumed standardised breakfasts for 3 consecutive days over 3 weeks where orange juice intake was decreased over the 3-week period(42). Using the urinary proline betaine concentrations, calibration curves were established. Using these calibration curves, the citrus intake was determined in an independent cross-sectional...
In nutrition research, there has been an increased interest in examining the diet as a whole instead of examining intake of single foods or nutrients. With this in mind, the concept of dietary patterns has emerged and the potential of using biomarkers to classify individuals into different dietary patterns is of interest. For the present review, we focus on the studies that have used a metabolomics-based approach to classify individuals into dietary patterns.

Andersen et al. used an untargeted metabolic phenotyping approach to distinguish between two dietary patterns with the purpose of developing a compliance measure for adherence to the new Nordic diet or an average Danish diet (44) (see Table 1). Using the urinary metabolome, a multivariate model was established that could distinguish the two dietary patterns with a low misclassification error rate (19 %) clearly indicating that this approach could be used for the examination of compliance to a certain dietary pattern. A follow-up paper also demonstrated that a classification model could be built using plasma metabolites to assess compliance to the new Nordic diet and average Danish diet diets (11). Esko et al. used a controlled feeding study to examine three different dietary patterns. These dietary patterns differed in macronutrient composition: low fat (60% carbohydrate, 20% fat, 20% protein), low glycaemic index (40% carbohydrate, 40% fat, 20% protein) and very-low carbohydrate (10% carbohydrate, 60% fat, 30% protein) (45). A classification model was built that could distinguish the three dietary patterns using plasma metabolites. These results support the concept that a metabolite-based model could be used in checking for adherence to specific diets and for the examination of relationship between dietary patterns and health outcomes in large epidemiological studies. Garcia-Perez et al. used a controlled intervention to develop a urinary metabolomics model that could classify individuals into dietary patterns (46). The four diets were based on the WHO healthy eating guidelines for the prevention of non-communicable diseases. Work from our laboratory used a cross-sectional study to develop a model based on urinary metabolomic data which could classify subjects into either a healthy or an unhealthy dietary pattern (47). The classification into the dietary patterns was supported by significant differences in blood parameters such as higher folate and 25(OH)-vitamin D in the healthy dietary pattern. The work presented by these examples demonstrates the potential of metabolomics-based approaches to identify dietary patterns and study the relationships with health outcomes. However, further work is needed to refine and develop these concepts further so that metabolomics-based biomarkers can be used for rapid and objective classification of individuals into dietary patterns.

While the afore-mentioned papers have developed the concept of examination of dietary patterns using metabolite biomarkers, there is also a large interest in examining the relationship between the metabolomic profile and known predefined dietary patterns such as the Mediterranean diet. The potential of such approaches is that it will allow the examination of the impact of dietary patterns on metabolic processes and pathways (47). Collectively, the studies presented earlier provide compelling evidence for the potential of metabolite biomarkers as a method for objectively assigning individuals into dietary patterns and for studying the effects of the certain dietary patterns on metabolic pathways.

**Future challenges and outlook**

While significant progress has been made in the past 5 years in the area of dietary biomarkers, there remain a number of challenges that need to be addressed. The validation of putative biomarkers is often overlooked and
<table>
<thead>
<tr>
<th>Dietary pattern</th>
<th>Study type; participants (n)</th>
<th>Dietary assessment tool</th>
<th>Biofluid</th>
<th>Analytic technique</th>
<th>Results</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>New Nordic diet (NND) or average Danish diet (ADD)</td>
<td>Six-month parallel intervention study (181)</td>
<td>Weighed dietary records</td>
<td>24 h urine samples</td>
<td>UPLC-qTOF-MS</td>
<td>Identified metabolite markers of individual foods such as citrus, cocoa-containing products and fish as well as more general dietary traits such as high fruit and vegetable intake or high intake of heat-treated foods. Misclassification rate for two dietary patterns in a validation set with 139 samples was 19% based on 67 selected features in urine.</td>
<td>(44)</td>
</tr>
<tr>
<td>NND or ADD</td>
<td>26-week parallel intervention study (146)</td>
<td>N/A had control of food provided</td>
<td>Fasting plasma samples at 0, 12 and 26 weeks</td>
<td>UPLC-qTOF-MS</td>
<td>Demonstrated that supervised machine learning with feature selection can separate NND and ADD samples (average test set performance AUC = 0.88). NND plasma metabolome characterised by diet-related metabolites, such as piperolic acid (whole grain), trimethylamine oxide, and prolyl hydroxyproline (both fish intake), theobromine (chocolate). Metabolites of amino acid (i.e. indolelactic acid and hydroxy-3-methylbutyrate) and fat metabolism (butyryl carnitine) characterised ADD whereas NND was associated with higher concentrations of polyunsaturated phosphatidylcholines.</td>
<td>(11)</td>
</tr>
<tr>
<td>Low fat (60 % CHO, 20 % fat, 20 % protein), low GI (40 % CHO, 40 % fat, 20 % protein), or very-low CHO (10 % CHO, 60 % fat, 30 % protein)</td>
<td>Three test diets, each for a 4-week period cross-over design (21)</td>
<td>N/A observed consumption</td>
<td>Fasting plasma samples at baseline and end of each 4-week period</td>
<td>LC-MS/MS</td>
<td>Identified 152 metabolites whose concentrations differed for ≥1 diet compared with the others, including DAG and TAGS, BCAA, and markers reflecting metabolic status. A classifier model was constructed to identify each diet.</td>
<td>(45)</td>
</tr>
<tr>
<td>Four dietary interventions in concordance with the WHO healthy eating guidelines</td>
<td>RCT crossover 4 × 72 h study stays (19) Cohort studies: INTERMAP UK (225) Healthy eating Danish (66)</td>
<td>N/A observed consumption</td>
<td>24 h pooled urine samples</td>
<td>¹H-NMR</td>
<td>Developed urinary metabolite models for each diet and identified the associated metabolic profiles. Validated the models using data and samples from the cohort studies. Significant stepwise differences in metabolite concentrations were seen between diets with the lowest and highest metabolic risks. Application of metabolite models to the validation datasets confirmed the association between urinary metabolic and dietary profiles in the cohort studies: INTERMAP UK (P &lt; 0.0001) and Danish (P &lt; 0.0001).</td>
<td>(46)</td>
</tr>
<tr>
<td>Dietary pattern</td>
<td>Study type; participants (n)</td>
<td>Dietary assessment tool</td>
<td>Biofluid</td>
<td>Analytic technique</td>
<td>Results</td>
<td>Reference</td>
</tr>
<tr>
<td>----------------</td>
<td>-----------------------------</td>
<td>------------------------</td>
<td>----------</td>
<td>-------------------</td>
<td>---------</td>
<td>-----------</td>
</tr>
<tr>
<td>Healthy eating index (HEI) 2010, Alternate Mediterranean diet score (aMED), WHO healthy diet indicator (HDI), and Baltic Sea diet (BSD)</td>
<td>Alpha- tocopherol, β-carotene cancer prevention study cohort (1336)</td>
<td>12-month validated FFQ</td>
<td>Fasting serum samples</td>
<td>LC-MS, UHPLC-MS/MS and GC-MS</td>
<td>The HEI-2010, aMED, HDI and BSD were associated with 23, 46, 23 and 33 metabolites, respectively (17, 21, 11 and 10 metabolites, respectively, were chemically identified; r-range: −0.30 to 0.20; ( P = 6 \times 10^{-15} ) to ( 8 \times 10^{-6} )). Food-based diet indexes (HEI-2010, aMED and BSD) were associated with metabolites correlated with most components used to score adherence (e.g. fruit, vegetables, wholegrains, fish and unsaturated fat). HDI correlated with metabolites related to polyunsaturated fat and fibre components, but not other macro- or micronutrients (e.g. percentages of protein and cholesterol). The lysolipid and food and plant xenobiotic pathways were most strongly associated with diet quality.</td>
<td>(47)</td>
</tr>
</tbody>
</table>

Healthy cluster
unhealthy cluster

<table>
<thead>
<tr>
<th>Study type; participants (n)</th>
<th>Dietary assessment tool</th>
<th>Biofluid</th>
<th>Analytic technique</th>
<th>Results</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>National Adult Nutrition Survey (NANS) (567)</td>
<td>Four-day semi-weighed food diaries</td>
<td>50 ml first void urine sample fasting spot urine samples</td>
<td>(^1)H-NMR</td>
<td>Two-step cluster analysis applied to the urinary data to identify clusters. The subsequent model was used to classify an independent cohort into dietary patterns. Classification was supported by significant differences in nutrient status (( P &lt; 0.05 )). Validation in an independent group revealed that 94 % of subjects were correctly classified.</td>
<td>(48)</td>
</tr>
</tbody>
</table>

UPLC-qTOF-MS, ultra-high-performance liquid chromatography quadrupole time of flight MS; AUC, area under the curve; CHO, carbohydrate; GI, glycaemic index; DAG, diacylglycerols; BCAA, branched chain amino acids; RCT, randomised control trial; \(^1\)H-NMR, proton NMR.
confusion thus arises as to the validity of biomarkers. It is essential in moving forward that all food intake biomarkers are validated and a suggested validation scheme now exists. In many metabolomics studies, the biomarkers are validated and a suggested validation is essential in moving forward that all food intake confusion thus arises as to the validity of biomarkers. It is essential to develop quantitative methods for biomarker measurement to ensure reliable cross-cohort comparison. Examples of other challenges include the potential use of multiple biomarkers for single foods: optimal methods for their use to estimate intake will need to be developed. Furthermore, many biomarkers will be indicators of short-term intake and defining strategies to obtain measures of long-term intake still remains a challenge. While multiple challenges exist for the field, it is also worth noting that considerable advances have been made in recent years, and with global consolidated efforts, it remains a possibility that objective biomarkers will improve our methods for assessing dietary intake.

Acknowledgements
The authors acknowledge the support from Dr Aoife O’Gorman.

Financial Support
This work was supported by a H2020 European Research Council (647783).

Conflict of Interest
None.

Authorship
L. B. conceived and oversaw the writing of the paper. All authors contributed to writing of the paper. C. C. and A. E. McN. contributed equally.

References


