Role of metabolomics in identification of biomarkers related to food intake

Cassandra Collins[1,2], Aoife McNamara[1,2] and Lorraine Brennan[1,2]

[1] UCD School of Agriculture and Food Science, Institute of Food and Health, UCD, Belfield, Dublin 4, Ireland

[2] UCD Conway Institute, UCD, Belfield, Dublin 4, Ireland

Corresponding Author:
Professor Lorraine Brennan
UCD School of Agriculture and Food Science,
UCD Institute of Food and Health,
UCD, Belfield, Dublin 4, Ireland.

Email: lorraine.brennan@ucd.ie
Phone: 00 353 1 7162811

Short Title: Biomarkers of dietary intake

Keywords: metabolomics, biomarkers, food intake, dietary patterns

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

Abstract

Dietary assessment methods including food-frequency questionnaires 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 complimentary 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/cohort 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 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 discover and the experimental approaches used for biomarker discovery 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.

Introduction

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 food frequency questionnaires (FFQs), 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 sizes [1–3]. Currently dietary biomarkers include 24h urinary sodium, nitrogen and sucrose/fructose for estimation of salt, protein and sugar intake [4–7]. 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, sugar sweetened beverages [7–11]. While some confusion exists in the literature over classification of biomarkers into recovery or concentration biomarkers we prefer to use the newly defined flexible classification scheme for biomarkers related to food intake [12]. Food intake biomarkers are single metabolites, or a combination of metabolites, reflecting the consumption of either a specific food or food group, displaying a clear time- and 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/) [18]

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 the 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 with Nuclear Magnetic resonance (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 mass spectrometry based approaches are extremely sensitive and are often coupled with a chromatography step to help with separation of the metabolites. Gas chromatography mass spectrometry (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 mass spectrometry (LC-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 ionized, 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 (CE) 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 (ESI) [27]. For high through-put techniques where it is desirable to have low run time per sample direct infusion mass spectrometry (DIMS) 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 mentioned above, 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,…..MS^n. 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 above 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* (2011) employed this approach when examining 24h urine samples for biomarkers of meat consumption. Participants were asked to consume 4 different diets for 14 days each containing a low (60g/d), medium (120g/d)-, high-portion of red meat (420g/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 to high CV consumption and remained elevated above baseline at 48h post consumption. The peaks were identified as S-methyl cysteine sulfoxide, a sulfur containing amino acid ubiquitous in CV, and its metabolites.

Parallel group intervention studies have also been successful in food intake biomarker discovery. Hanhineva and colleagues randomised participants to follow one of three diets over a twelve week period including a healthy diet (wholegrain enriched diet, fatty fish and bilberries), a wholegrain-enriched diet or a control diet (avoiding whole grain cereals and bilberries, consuming low-fibre products, limiting fatty fish intake to one portion per week)[(34)]. Plasma metabolomics revealed that CMPF (3-carboxy-4-methyl-5-propyl-2-furanpropionic acid) was associated with fatty fish intake and alkylresorcinol metabolites were associated with wholegrain intake.

Using samples from epidemiology studies one examines correlations between self-reported food intake and biomarkers measured in urine or blood samples. Guertin et al (2014), applied an UPLC (ultra high pressure liquid chromatography)- and GC-MS metabolomics approach when examining serum samples from a subset of the Prostate, Lung, Colorectal, and Ovarian (PLCO) Cancer Screening Trial to identify biomarkers related to intake of 36 food groups [(8)]. The data revealed that 39 biomarkers were significantly associated with intake of food groups such as citrus, green vegetables, red meat, fish, shellfish, butter, peanuts, rice, coffee, beer, liquor, total alcohol, and multivitamins. Other approaches have compared

consumer and non-consumers of certain foods to identify biomarkers increased in the consumers. Using this approach Rothwell et al. identified discriminating biomarkers in the urinary metabolome of 20 high coffee consumers and 19 non-consumers in a subset of the SU.VI.MAX2 cohort [35]. Many other examples using this approach have emerged in recent years and the readers are referred to Guasch-Ferré et al. (2018), 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 above 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 the 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 makers of these foods. This highlight 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 standardized breakfasts for three consecutive days over three weeks where orange juice intake was decreased over the three 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 study of 560 individuals. There was excellent agreement between the self-report intake (estimated from a 4 day semi-weighed food diary) and the estimated intake from the biomarker with a low mean bias of 4.3g between the methods. This study clearly demonstrates the potential of well validated food intake biomarkers. In a separate study Garcia-Perez and colleagues examined the ability of tartaric acid to determine grape intake [43]. A dose response relationship was established between grape intake and urinary tartaric acid levels. The agreement between estimated intake and actual intake was good and a correlation coefficient of R^2=0.9 was reported. Overall, these two examples provide strong evidence of the potential of food intake biomarkers and demonstrate the importance of assessing dose response relationships on identified biomarkers. However, it is also worth noting that not all biomarkers will be fully quantitative but will still yield useful information for examining relationships with health outcomes (Figure 1).

Biomarkers of Dietary patterns

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 and colleagues 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 (NND) or an Average Danish Diet (ADD) [44] (see Table 1). Using the urinary metabolic profile 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 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 NND and ADD diets (11). Esko and colleagues 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 glycemic 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 and colleagues 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 (NCDs). 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 (16). 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 demonstrate 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 above 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 above 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 last 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 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 identification of metabolites to a high degree of certainty is challenging and many of the current databases lack metabolites that are related to food intake. International collaborative efforts are needed to try optimise the identification process. To ensure that the food intake biomarkers are functional in different ethnic groups it will be 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 longterm 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.

Financial Support

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

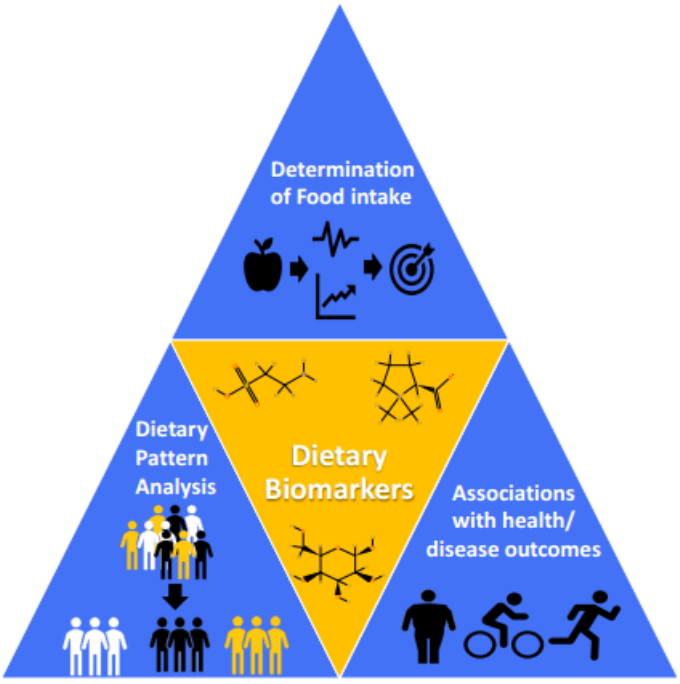
Acknowledgements: The authors acknowledge the support from Dr Aoife O'Gorman.

Conflict of Interest

The authors have conflict of interest.

Figure Legend

Figure 1. An overview of the applications of Dietary biomarkers. Biomarkers can give information on (1) food intake (2) dietary patterns and (3) relationships with health outcomes.



REFERENCES

1. Bingham SA. Biomarkers in nutritional epidemiology. Public Health Nutr [Internet]. 2002;5(6a):821–7. Available from: http://www.journals.cambridge.org/abstract_S136898000200109X

2. Kipnis V, Midthune D, Freedman L et al. Bias in dietary-report instruments and its implications for nutritional epidemiology. Public Health Nutr [Internet]. 2002;5(6a):915–23. Available from: http://www.journals.cambridge.org/abstract_S1368980002001234

3. Dhurandhar N, Schoeller D, Brown AW et al. Energy Balance Measurement:When Something is Not Better than Nothing. Int J Obes. 2015;39(7):1109–13.

4. Ovesen L, Boeing H. The use of biomarkers in multicentric studies with particular consideration of iodine, sodium, iron, folate and vitamin D. Eur J Clin Nutr. 2002;56(SUPPL. 2):12–7.

5. Freedman LS, Commins JM, Moler JE et al. Pooled Results From 5 Validation Studies of Dietary Self-Report Instruments Using Recovery Biomarkers for Energy and Protein Intake Laurence. Am J Epidemiol. 2014;180(2):172–88.

6. Tasevska N, Runswick SA, McTaggart A et al. Urinary sucrose and fructose as biomarkers for sugar consumption. Cancer Epidemiol Biomarkers Prev. 2005;14(5):1287–94.

7. Gibbons H, Mcnulty BA, Nugent AP et al. A metabolomics approach to the identification of biomarkers of sugar-sweetened beverage intake. Am J Clin Nutr. 2015;101(3):471–7.

8. Guertin KA, Moore SC, Sampson JN et al. Metabolomics in nutritional epidemiology: Identifying metabolites associated with diet and quantifying their potential to uncover diet-disease relations in populations. Am J Clin Nutr. 2014;100(1):208–17.

9. Schmidt JA, Rinaldi S, Ferrari P et al. Metabolic profiles of male meat eaters, fish eaters, vegetarians, and vegans from the EPIC-Oxford cohort. Am J Clin Nutr. 2015;102(6):1518–26.

10. Rådjursöga M, Karlsson GB, Lindqvist HM et al. Metabolic profiles from two different breakfast meals characterized by 1H NMR-based metabolomics. Food Chem. 2017;231:267–74.

11. Acar E, Gürdeniz G, Khakimov B et al. Biomarkers of Individual Foods, and Separation of Diets Using Untargeted LC-MS Based Plasma Metabolomics in a Randomized Controlled Trial. Mol Nutr Food Res [Internet]. 2018;1800215:1800215. Available from: http://doi.wiley.com/10.1002/mnfr.201800215

12. Gao Q, Praticò G, Scalbert A et al. A scheme for a flexible classification of dietary and health biomarkers. Genes Nutr. 2017;12(1):1–15.

13. Falcke JM, Bose N, Artyukhin AB et al. Linking Genomic and Metabolomic Natural Variation Uncovers Nematode Pheromone Biosynthesis. Cell Chem Biol. 2018;787–96.

14. Schuijt TJ, Lankelma JM, Scicluna BP et al. The gut microbiota plays a protective role in the host defence against pneumococcal pneumonia. Gut. 2016;65(4):575–83.

15. Lampe JW, Huang Y, Neuhouser ML et al. Dietary biomarker evaluation in a controlled feeding study in women from the Women's Health Initiative cohort. Am J Clin Nutr. 2017;105(April):466–75.

16. O'Sullivan A, Gibney MJ, Brennan L. Dietary intake patterns are reflected in metabolomic profiles : potential role in dietary assessment studies 1 – 3. Am J Clin Nutr. 2011;(July):314–21.

17. Li W, Yang H, Buckley B, Wang L et al. A Novel Triple Stage Ion Trap MS method validated for curcumin pharmacokinetics application: A comparison summary of the latest validated curcumin LC/MS methods. J Pharm Biomed Anal [Internet]. 2018;156:116–24. Available from: https://doi.org/10.1016/j.jpba.2018.04.022

18. Wishart DS, Feunang YD, Marcu A et al. HMDB 4.0: The human metabolome database for 2018. Nucleic Acids Res. 2018;46(D1):D608–17.

19. Huan T, Tang C, Li R et al. MyCompoundID MS/MS Search: Metabolite Identification Using a Library of Predicted Fragment-Ion-Spectra of 383,830 Possible Human Metabolites. Anal Chem. 2015;87(20):10619–26.

20. Tautenhahn R, Patti GJ, Rinehart D et al. XCMS Online: a web-based platform to process untargeted metabolomic data. Anal Chem. 2012;84(11):5035–9.

21. Ulaszewska MM, Weinert CH, Trimigno A, et al. Nutrimetabolomics: An Integrative Action for Metabolomic Analyses in Human Nutritional Studies. Mol Nutr Food Res [Internet]. 2018;1800384. Available from: http://doi.wiley.com/10.1002/mnfr.201800384

22. Pan Z, Raftery D. Comparing and combining NMR spectroscopy and mass spectrometry in metabolomics. Anal Bioanal Chem. 2007;387(2):525–7.

23. Gibbons H, O'Gorman A, Brennan L. Metabolomics as a tool in nutritional research. Curr Opin Lipidol. 2015;26(1):30–4.

24. Sparkman OD, Penton ZE, Kitson FG. Gas Chromatography and Mass Spectrometry, A Practical Guide [Internet]. Gas Chromatography and Mass Spectrometry (Second edition). Academic Press; 2011 [cited 2018 Jul 24]. 285-292 p. Available from: http://www.sciencedirect.com/science/article/pii/B978012373628400053 8%5Cn

25. Halket JM, Waterman D, Przyborowska AM et al. Chemical derivatization and mass spectral libraries in metabolic profiling by GC/MS and LC/MS/MS. J Exp Bot [Internet]. 2005 Jan 1 [cited 2018 Mar 7];56(410):219–43. Available from: https://academic.oup.com/jxb/article-lookup/doi/10.1093/jxb/eri069

26. Marquet P. LC-MS vs. GC-MS, online extraction systems, advantages of technology for drug screening assays. Methods Mol Biol [Internet]. 2012 [cited 2018 Sep 25];902:15–27. Available from: http://link.springer.com/10.1007/978-1-61779-934-1_2

27. Mikšík I. Coupling of CE-MS for protein and peptide analysis. J Sep Sci [Internet]. 2018;1–37. Available from: http://doi.wiley.com/10.1002/jssc.201800817

28. Ren JL, Zhang AH, Kong L et al. Advances in mass spectrometry-based metabolomics for investigation of metabolites. RSC Adv. 2018;8(40):22335–50.

29. Sumner LW, Amberg A, Barrett D et al. Proposed minimum reporting standards for

chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). Metabolomics. 2007;3(3):211–21.

30. Blaženović I, Kind T, Ji J et al. Software tools and approaches for compound identification of LC-MS/MS data in metabolomics. Metabolites. 2018;8(2).

31. O'Gorman A, Gibbons H, Brennan L. Metabolomics in the identification of biomarkers of dietary intake. Comput Struct Biotechnol [Internet]. 2013;4(5):1–7. Available from: http://dx.doi.org/10.5936/csbj.201301004

32. Cross AJ, Major JM, Sinha R. Urinary Biomarkers of Meat Consumption. Cancer Epidemiol Biomarkers Prev. 2011;20(6):1107–11.

33. Edmands WMB, Beckonert OP, Stella C, et al. Identification of human urinary biomarkers of cruciferous vegetable consumption by metabonomic profiling. J Proteome Res. 2011;10(10):4513–21.

34. Hanhineva K, Lankinen MA, Pedret A et al. Nontargeted Metabolite Profiling Discriminates Diet-Specific Biomarkers for Consumption of Whole Grains, Fatty Fish, and Bilberries in a Randomized Controlled Trial. J Nutr [Internet]. 2015;145(1):7–17. Available from: http://jn.nutrition.org/cgi/doi/10.3945/jn.114.196840

35. Rothwell JA, Fillâtre Y, Martin JF et al. New biomarkers of coffee consumption identified by the non-targeted metabolomic profiling of cohort study subjects. PLoS One. 2014;9(4).

36. Guasch-Ferré M, Bhupathiraju SN, Hu FB. Use of Metabolomics in Improving Assessment of Dietary Intake. Clin Chem [Internet]. 2018 Jan 16 [cited 2018 Sep 26];64(1):82–98. Available from: http://www.ncbi.nlm.nih.gov/pubmed/29038146

37. Dragsted L, Gao Q, Scalbert A et al. Validation of food intake biomarkers - critical assessment of candidate biomarkers. Genes Nutr. 2018;13(14):1–14.

38. Ulaszewska M, Vázquez-Manjarrez N, Garcia-Aloy M et al. Food intake biomarkers for apple, pear, and stone fruit. Genes Nutr. 2018;1–16.

39. Sri Harsha PSC, Wahab RA, Garcia-Aloy M et al. Biomarkers of legume intake in human intervention and observational studies: a systematic review. Genes Nutr . 2018;13(1):25.

40. Münger LH, Garcia-Aloy M, Vázquez-Fresno R et al. Biomarker of food intake for assessing the consumption of dairy and egg products. Genes Nutr. 2018;13(1):26. Available from: https://genesandnutrition.biomedcentral.com/articles/10.1186/s12263-018-0615-5

41. Rothwell JA, Madrid-Gambin F, Garcia-Aloy M et al. Biomarkers of intake for coffee, tea, and sweetened beverages. Genes Nutr. 2018;13(1):1–18.

42. Gibbons H, Michielsen CJR, Rundle M et al. Demonstration of the utility of biomarkers for dietary intake assessment; proline betaine as an example. Mol Nutr Food Res. 2017;61(10):1–8.

43. Garcia-Perez I, Posma JM, Chambers ES et al. An Analytical Pipeline for Quantitative Characterization of Dietary Intake: Application to Assess Grape Intake. J Agric Food Chem. 2016;64(11):2423–31.

44. Andersen MBS, Rinnan Å, Manach C et al. Untargeted metabolomics as a screening tool for estimating compliance to a dietary pattern. J Proteome Res. 2014;13(3):1405–18.

45. Esko TT, Hirschhorn JN, Feldman HA et al. Metabolomic profiles as reliable biomarkers of dietary composition. Am J Clin Nutr. 2017;105(3):547–54.

46. Garcia-Perez I, Posma JM, Gibson R et al. Objective assessment of dietary patterns by use of metabolic phenotyping: a randomised, controlled, crossover trial. Lancet Diabetes Endocrinol [Internet]. 2017;5(3):184–95. Available from: http://dx.doi.org/10.1016/S2213-8587(16)30419-3

47. Playdon MC, Moore SC, Derkach A et al. Identifying biomarkers of dietary patterns by using metabolomics. Am J Clin Nutr [Internet]. 2017;105(2):ajcn144501. Available from: http://www.ncbi.nlm.nih.gov/pubmed/28031192%0Ahttp://ajcn.nutrition.org/lookup/doi/10.3945/ajcn.116.144501

48. Gibbons H, Carr E, McNulty BA et al. Metabolomic-based identification of clusters that reflect dietary patterns. Mol Nutr Food Res. 2017;61(10):1–11.

Table 1. Overview of studies using biomarkers for determining dietary patterns.

Dietary Pattern	Study Type (N)	Dietary Assessment tool	Biofluid	Analytic technique	Results	Reference
New Nordic Diet (NND) or Average Danish Diet (ADD)	6 month parallel intervention study (181)	Weighed dietary records	24h urine samples	UPLC-qTOF-MS	Identified metabolite markers of individual foods such as citrus, cocoa-containing products, & fish as well as more general dietary traits such as high fruit & 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.	(44)
New Nordic Diet (NND) or Average Danish Diet (ADD)	26 week parallel intervention study (146)	N/A had control of food provided	Fasting plasma samples at 0,12 and 26 weeks	UPLC-qTOF-MS	Demonstrated that supervised machine learning with feature selection can separate NND and ADD samples (average test set performance AUC = 0.88). NND plasma metabolome characterized by diet-related metabolites, such as pipecolic acid betaine (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) characterized ADD whereas NND was associated with higher concentrations of polyunsaturated phosphatidylcholines.	(11)
low fat (60% CHO, 20% fat, 20% protein), low GI (40% CHO, 40% fat, 20% protein),	3 test diets, each for a 4-wk period crossover design (21)	N/A observed consumption	Fasting Plasma samples at baseline & end of	LC-MS/MS	Identified 152 metabolites whose concentrations differed for ≥1 diet compared with the others, including DAGs & TAGSs, BCAAs, & markers reflecting metabolic status. A classifier model was constructed to identify each diet.	(45)

or very-low CHO (10% CHO, 60% fat, 30% protein)			each 4-wk period			
4 dietary interventions in concordance with the WHO healthy eating guidelines	RCT crossover 4 x 72 h study stays (19) Cohort studies: INTERMAP UK (225) Healthy eating Danish (66)	N/A observed consumption	24 h pooled urine samples	^{1}H-NMR	Developed urinary metabolite models for each diet & identified the associated metabolic profiles. Validated the models using data & samples from the cohort studies. Significant stepwise differences in metabolite concentrations were seen between diets with the lowest & highest metabolic risks. Application of metabolite models to the validation datasets confirmed the association between urinary metabolic & dietary profiles in the cohort studies: INTERMAP UK ($p<0{\cdot}0001$) & Danish ($p<0{\cdot}0001$).	(46)
Healthy Eating Index (HEI) 2010, Alternate Mediterranean Diet Score (aMED), WHO Healthy Diet Indicator (HDI), & Baltic Sea Diet (BSD)	Alpha-Tocopherol, Beta Carotene Cancer Prevention Study cohort (1336)	12 month validated FFQ	fasting serum samples	LC-MS, UHPLC-MS/MS, & GC-MS	The HEI-2010, aMED, HDI, & BSD were associated with 23, 46, 23, & 33 metabolites, respectively (17, 21, 11&10 metabolites, respectively, were chemically identified; r-range: -0.30 to 0.20; $P = 6x10^{-15}$ to $8x10^{-6}$). Food-based diet indexes (HEI-2010, aMED, & BSD) were associated with metabolites correlated with most components used to score adherence (e.g. fruit, vegetables, wholegrains, fish, & unsaturated fat). HDI correlated with metabolites related to polyunsaturated fat & fibre components, but not other macro- or micronutrients (e.g., percentages of protein & cholesterol). The lysolipid & food & plant xenobiotic pathways were most strongly associated with diet quality.	(47)
Healthy cluster Unhealthy cluster	National Adult Nutrition	Four day semi-weighed food diaries	50 mL first void urine	^{1}H-NMR	Two-step cluster analysis applied to the urinary data to identify clusters. The subsequent model was used to classify an independent cohort into	(48)

	Survey (NANS) (567)		sample fasting spot urine samples		dietary patterns. Classification was supported by significant differences in nutrient status ($p<0.05$). Validation in an independent group revealed that 94% of subjects were correctly classified	

Note: UPLC-qTOF-MS; ultra high performance liquid chromatography quadrupole time of flight mass spectrometry, AUC; area under the curve, CHO; carbohydrate, GI; glycaemic index, DAGs; diacylglycerols, TAGSs; triacylglycerols, BCAAs; branched chain amino acids, RCT; randomized control trial, ^{1}H-NMR; proton nuclear magnetic resonance, FFQ; food frequency questionnaire, GC-MS; gas chromatography mass spectrometry.