

## METABOLOMICS IN THE IDENTIFICATION OF BIOMARKERS OF DIETARY INTAKE

Aoife O'Gorman<sup>a,b</sup>, Helena Gibbons<sup>a,b</sup>, Lorraine Brennan<sup>a,b,\*</sup>

**Abstract:** Traditional methods for assessing dietary exposure can be unreliable, with under reporting one of the main problems. In an attempt to overcome such problems there is increasing interest in identifying biomarkers of dietary intake to provide a more accurate measurement. Metabolomics is an analytical technique that aims to identify and quantify small metabolites. Recently, there has been an increased interest in the application of metabolomics coupled with statistical analysis for the identification of dietary biomarkers, with a number of putative biomarkers identified. This minireview focuses on metabolomics based approaches and highlights some of the key successes.

### MINI REVIEW ARTICLE

#### Introduction

In today's modern world nutrition research is focused on improving population and individual health through diet [1]. Nutrition and health related research are beginning to understand that in addition to their essential functions, nutrients and non-nutrient components of foods interact with numerous metabolic pathways and influence health reducing or increasing the risk of disease. Diet is considered one of the major factors contributing to the rapid increase in the incidence of metabolic disorders such as obesity, diabetes and cardiovascular disease [2].

Reliable dietary assessment methods are vital when attempting to understand the links between diet and chronic disease profiles. Conventional tools for collecting quantitative information on dietary exposure such as food diaries, 24-h recalls and food frequency questionnaires (FFQ) can be unreliable for characterising and quantifying eating behaviour and are all subject to possible reporting and other biases [3,4]. In addition, these methods are unreliable for certain groups such as the obese or elderly people, whose self-reported energy intakes tend to be underestimated, as assessed by energy expenditure measurements using the doubly labelled water method [5,6]. A full critical review of limitations associated with the current techniques is beyond the scope of the present review and the reader is referred to the following papers [3,4,7-9].

In an attempt to overcome the problems with measuring dietary exposure with self-reported methods, nutritional epidemiologists started examining biomarkers as measures of dietary intake and nutrient status [10,11]. The use of dietary biomarkers provides a more objective and accurate measure of intake in comparison to traditional questionnaires as they take into account the nutrient bioavailability and metabolism [12,13]. One of the main applications of these dietary biomarkers is to use them as reference measurements

to assess the validity of dietary assessment measures [11,14,15]. So far ideal biomarkers exist for salt and protein intake (sodium/nitrogen measure in a 24 h urine sample) and energy expenditure (double labelled water technique) [11]. Other biomarkers exist that do not provide information on the exact dietary intake but which are highly correlated with intake for example the measurement of serum carotenoids and vitamin C as biomarkers of fruit and vegetable intake [16].

The development of robust food biomarkers will help in better classifying a person's dietary intake and in turn will improve the assessment of the relationship between diet and chronic disease [17]. In recent years there has been an increased interest in applying metabolomics for the discovery of biomarkers of dietary intake. This review will focus on metabolomics and its use in assessing dietary biomarkers.

#### Metabolomic Technologies

Metabolomics refers to comprehensive and non-selective analytical chemistry approaches aiming to provide a global description of all metabolites present in a biofluid at a given time [18-21]. The two main approaches employed in metabolomics are nuclear magnetic resonance (NMR) spectroscopy and mass spectroscopy (MS). These techniques both have their advantages and disadvantages and at present there is no single analytical technique capable of measuring and identifying all metabolites in a single sample simultaneously and therefore comprehensive metabolomic data needs to be assessed by bringing together data from different platforms [22].

For instance, NMR spectroscopy uses an untargeted approach where all molecules are interrogated simultaneously by properties that they all share (NMR active hydrogen or carbon) [23]. <sup>1</sup>H NMR requires little or no pre-treatment, it is quantitative (absolute), non-destructive, reproducible and unbiased [24], although is not as sensitive as MS techniques. There are a range of MS based approaches, with the most recent advances leading to the use of hyphenated techniques such as liquid chromatography-mass spectroscopy (LC-MS), capillary electrophoresis-mass spectroscopy (CE-MS) and gas chromatography-mass spectroscopy (GC-MS). The chromatographic step allows the separation of metabolites before

<sup>a</sup>UCD Conway Institute, Belfield, Dublin, Ireland

<sup>b</sup>UCD Institute of Food and Health, Belfield, Dublin, Ireland

\* Corresponding author.

E-mail address: [lorraine.brennan@ucd.ie](mailto:lorraine.brennan@ucd.ie) (Lorraine Brennan)

detection takes place. One of the main advantages of these techniques is the associated high sensitivity and therefore may detect metabolites that are present in a concentration below the detection limit of  $^1\text{H}$  NMR spectroscopy. A disadvantage is the necessity for sample preparation before analysis. GC-MS also requires derivatization, which lengthens the sample preparation time. As stated above, a comprehensive overview of the metabolome is achieved by using the multiple platforms.

## Data Analysis

Metabolomics like other 'omic' technologies generates large and complex datasets and because of this data analysis using multivariate statistics has become an important part of metabolomics. There are a number of statistical methods available for metabolomic data, with principal component analysis (PCA), partial least squares discriminant analysis (PLS-DA) and orthogonal PLS-DA being the most commonly used. PCA is probably the best known method, it is an unsupervised method which assesses the natural grouping of sample classes and can be used to identify extreme outliers [25].

Despite its widespread use in metabolomics, PCA has a number of shortcomings. Mainly, PCA does not have an associated probabilistic model, which makes assessing the fit of PCA to the data difficult, limiting the scope of its application. In addition PCA can fail to reveal underlying groups of subjects in the data, thereby providing a false view of the underlying data structure [26,27]. Probabilistic principal component and covariates analysis (PPCCA) is a novel extension of probabilistic principal component analysis (PPCA) [28] which has recently been introduced to analyse metabolomic data. PPCCA incorporates covariates into the model and facilitates joint modelling of metabolomic data and covariates, meaning that the PPCA model directly models any variation due to the covariates, thus ensuring that the principal components provide a clear picture of the underlying data. This method has great potential for use within the metabolomics field [29].

Supervised techniques require prior knowledge of the class of a sample and examples frequently employed in metabolomics studies include, PLS-DA or O-PLS-DA, combining a data filtering step. PLS-DA provides a way to filter out metabolic information which is not correlated to the predefined classes and the loadings plots provide information on the spectral signals associated with the observed trends giving a means to interpret the metabolic information. Despite its powerful ability to separate classes, care must be taken during fitting of PLS-DA to the training datasets, which exaggerate generalisation ability. Generally, cross-validation or permutation tests are required to assess the ability of the trained PLS-DA model [30]. For further information on these techniques and other analysis tools such as random forests (RF) [31], support vector machines (SVM) [32] and artificial neural networks (ANN) [33], please see the following recent reviews [34,35].

## Metabolomic & Dietary Biomarker Studies

Applications of metabolomics to identify novel dietary biomarkers have in general terms taken three approaches (i) specific acute intervention to identify food markers (ii) searching for biomarkers in cohort studies and (iii) analysis of dietary patterns in conjunction with metabolomic profiles to identify nutritypes and biomarkers. Approaches (i) and (ii) form the basis of the studies described under biomarkers of specific foods while approach (iii) is discussed under dietary patterns.

### (i) Biomarkers of specific foods

Over the past few years a number of studies have emerged where specific acute interventions have been used to identify the presence of food specific biomarkers or to monitor concentration changes in diet related metabolites. To date, application of metabolomics has identified a number of putative biomarkers of intake of certain foods including salmon, broccoli, wholegrain wheat cereal, raspberry [36], cruciferous vegetables [37], citrus fruits [17], coffee [38,39], onions [40] and red meat [41]. An initial literature search identified a number of biomarkers associated with a broad range of foods (meat, fish, wholegrains, cocoa etc.). However, for the purpose of this minireview we focused on the following foods for table 1: Meat, Fish, Vegetables, Citrus Fruits, Coffee and Tea.

#### *Biomarkers of fish intake*

Many metabolomic studies have reported high levels of trimethylamine-*N*-oxide (TMAO) in urine samples following fish consumption 24 h prior to sample collection [36,55,65]. Lloyd and colleagues specifically searched for biomarkers of salmon in a study where subjects ( $n=24$ ) consumed a breakfast with either one of four test foods, salmon being one of the test foods, 6 times over an 8 month period. Postprandial urine samples were collected at 3 different time points (1.5-, 3-, and 4.5-h) and analysed by flow infusion electrospray-ionisation mass spectrometry, followed by supervised data analysis in order to identify signals resulting from consumption of each test food. A combination of TMAO and 1-methylhistidine were found to be associated with salmon consumption with higher levels found after consuming the fish when compared to the standard breakfast [36].

#### *Biomarkers of meat intake*

Meat intake is an important contributor to dietary protein in omnivorous populations and therefore has a potential impact on a range of nutritional and health outcomes [66]. As a result numerous studies, both metabolomics and non-metabolomics based, throughout the years have proposed the following metabolites as biomarkers of meat intake; creatinine, creatine, carnitine, carnosine, taurine, 1-methylhistidine and 3-methylhistidine and TMAO. A fully dietary-controlled study was analyzed by Stella and colleagues using  $^1\text{H}$  NMR spectroscopy in combination with multivariate statistical analysis to characterize the effects of three diets; 'vegetarian', 'low meat' and 'high meat' [41]. Twelve healthy male participants (24-74 years) consumed each of these diets in a randomized order for continuous 15-day periods with an intervening wash out period between each diet of 7 days duration. Three consecutive urine samples were collected from days 10-12 during each intervention period. The following metabolites were found to be increased in the high meat consumption period; creatine, carnitine, acetyl-carnitine and TMAO. Creatine is known to be influenced by a number of factors such as muscle mass hence its reliability as a biomarker needs to be further investigated. With respect to carnitine, the dietary matrix is known to have an influence on excretion so its use as a quantitative biomarker may be limited.

3-methylhistidine and 1-methylhistidine have also been proposed as biomarkers of dietary intake [42,47,67]. A recent study investigated both of these metabolites in conjunction with taurine and creatinine as biomarkers of meat intake [42]. This targeted analysis of urine samples following consumption of increasing amounts of red meat indicated that both 3-methylhistidine and 1-methylhistidine are good markers of red meat intake but also highlighted that 1-

methylhistidine may be more useful as its excretion is independent of muscle mass and catabolism.

Previous metabolomic studies have shown TMAO to be elevated after consumption of high-meat diets [24,41], although it has also been reported to be found in higher levels after fish consumption [36], indicating that TMAO may be used as a dietary biomarker of protein as opposed to a specific food i.e. meat/fish. A recent study assessing the effect of high or low protein diets found that the TMAO signal in the NMR spectra of urine was highly correlated to daily urinary nitrogen excretion ( $r=0.89$ ) and thereby consumed protein [68].

#### *Biomarkers of fruit and vegetable intake*

In recent years two groups have independently identified proline betaine as a marker of citrus fruit consumption [17,59]. Heinzmann performed an acute intervention involving 8 volunteers where they consumed a standardised breakfast, lunch and dinner meal from day 0 until lunch on day 3. In addition to the standard dinner a supplementary mixed-fruit meal (apple, orange, grapefruit and grapes) was introduced on the evening of day 2. Urine was collected 4 times/day from the morning of day 1 until the evening of day 3.  $^1\text{H}$  NMR and PLS-DA analysis identified the urinary excretion of proline betaine as a biomarker of citrus fruit intake. Following on from this the authors quantified the relative concentrations of proline betaine in citrus products and evaluated the urinary excretion profile after orange juice consumption. Finally, validation was carried out on the biomarker proline betaine by using urinary NMR spectra from participants of the INTERMAP UK cohort [69]. A receiver operating characteristic (ROC) curve resulted with an AUC of 92.3% with a sensitivity and specificity of 90.6% and 86.3% respectively.

In the study performed by Lloyd and colleagues proline betaine was identified as a biomarker of citrus intake using an acute breakfast challenge. Acute exposure of volunteers to orange juice resulted in the appearance of proline betaine and a number of biotransformed products in postprandial urine samples. In addition, a process of validation showed sensitivities of 80.8-92.2% and specificities of 74.2-94.1% for elevated levels of proline betaine in those volunteers who reported a high consumption [59].

Applications of a metabolomics strategy for the identification of biomarkers of cruciferous vegetable consumption has recently identified S-methyl-L-cysteine sulphoxide (SMCSO) and metabolic derivatives as putative biomarkers [37]. Twenty healthy male subjects ( $n=20$ ) were recruited to a three period dietary intervention study with each period lasting 14 days. For phases I and III a low cruciferous vegetable intake was consumed, whereas phase II consisted of a high cruciferous vegetable intake. On day 13 of each phase, following an overnight fast, a time zero spot urine sample was obtained from each participant in the study. Participants were then maintained on a standardised diet and urine sample collections were obtained for the periods 0-10, 10-24 and 24-48 h. Analysis of the NMR spectra showed clear differentiation between the high and low cruciferous vegetable consumption and was attributed to SMCSO and metabolites derived from it.

Other candidate biomarkers for fruit and vegetable intake include antioxidant vitamins such as vitamin C [70,71] and flavanoids [58], including quercetin [72]. Mennen *et al.* examined associations between dietary intake and the concentrations of selected urinary polyphenols and metabolites in free living subjects [58]. In this study 13 polyphenols and metabolites were measured in urine samples using HPLC-ESI-MS-MS along with two day food diaries which estimated habitual intake. In spot urine samples, significant correlations were reported for different fruits and beverages and several polyphenol compounds, for example apple consumption was positively correlated

to phloretin, grapefruit consumption to naringenin, orange to hesperetin, citrus fruit consumption to both naringenin and hesperetin, with  $r$  coefficients ranging from 0.31 to 0.57 ( $P<0.05$ ).

#### *Biomarkers of tea consumption*

Tea is a widely consumed beverage and is a major dietary source of polyphenolic compounds, including phenolic acids and flavanoids. Several potential biomarkers of exposure to tea derived polyphenols have been identified [73-75]. These include specific *O*-methylated polyphenols derived from *in vivo* polyphenol metabolism, such as 4-*O*-methylgallic acid (4OMGA) [73,76,77]. One such study explored the relationship of 24 h urinary excretion of 4OMGA with usual ( $n=111$ ) and current ( $n=344$ ) tea intake in human subjects using a GC-MS approach [75]. The authors found that urinary excretion was significantly related to both usual tea intake ( $r$  0.50,  $P<0.001$ ) and current tea intake ( $r$  0.57,  $P<0.001$ ) and that a cut-off concentration for 4OMGA excretion of 25  $\mu\text{g}/\text{mmol}$  creatinine had 82% sensitivity and 81% specificity for prediction of tea drinking status.

#### *(ii) Dietary patterns and metabolomic profiles of habitual diet to identify nutritypes and biomarkers*

The studies described above have efficiently identified biomarkers of certain foods. In recent years there has been an interest in dietary patterns and their use as a method of studying relationships between diet and disease.

Work in our laboratory applied dietary pattern analysis to 125 subjects for which dietary data was recorded using 3 day food diaries [43]. The identification of dietary clusters was performed using *k*-means clustering and resulted in three cluster groups which were associated with unique food intakes and differed in aspects of their nutrient intake profiles. Assessment of the metabolomic profiles revealed that the cluster groups were reflected in the urinary metabolomic profiles. Further analysis using PLS-DA identified metabolites associated with the different dietary patterns. Cluster 3 was defined by high intakes of meat products, white bread, butter and preserves and had significantly higher levels of *O*-acetylcarnitine. The novelty of this work lies in the fact that identification of nutritypes (i.e. metabolic profiles that reflect dietary intake) has the potential to aid dietary assessment by unobjectively classifying people into certain dietary patterns.

Peré-Trepát and colleagues also developed a strategy for assessing links between dietary data (FFQ) and metabolomic profiles [78]. In this work dietary patterns were defined by PCA and then re-coded and regressed against NMR metabolic profiles to obtain loadings and identify metabolites associated with dietary patterns. While this study was a method development study it successfully linked dietary patterns with certain metabolites and further supports the concept of nutritypes.

Using the KORA (Cooperative Health Research in the Region of Augsburg) study population, Altmaier *et al.* identified seven dietary patterns [79]. Metabolomic analysis was performed on plasma samples using electrospray ionization (ESI) tandem mass spectrometry (MS-MS). Statistical analysis revealed that certain dietary patterns were highly associated with serum metabolite concentrations.

Overall, these studies provide good evidence for the potential of metabolomics to be used to define a profile of markers that are reflective of a habitual dietary pattern. Further studies will be necessary to develop this concept further and establish its robustness across different populations.

**Table I.** Examples of dietary biomarkers identified using metabolomic based approaches

Food	Sample	Metabolic Approach	Biomarker	Study	
Red Meat	Urine	Ion exchange chromatography	1- and 3-methylhistidine	[42]	
	Urine	<sup>1</sup> H NMR spectroscopy	O-acetylcarnitine, N,N-dimethylglycine	[43]	
	Urine	<sup>1</sup> H NMR spectroscopy	O-acetylcarnitine	[44]	
	Serum and urine	<sup>1</sup> H NMR spectroscopy	Creatine, histidine, urea	[45]	
	Urine	<sup>1</sup> H NMR spectroscopy	Carnitine, Creatinine, TMAO, acetyl-carnitine, taurine, 1- and 3-methylhistidine	[41]	
	Plasma	HPLC	Carnosine	[46]	
	Urine	Ion exchange chromatography	Creatinine, taurine, 1- and 3-methylhistidine	[47]	
Cooked Meats	Hair	LC-MS	PhIP	[48]	
	Urine	LC-MS-MS	PhIP metabolites	[49]	
	Urine	GC-MS	PhIP	[50]	
	Urine	NCI-GC-MS <sup>a</sup>	PhIP	[51]	
	Urine	GC-MS	4'-OH-PhIP	[52]	
Fish	Urine	FIE-MS	TMAO, anserine, 1- and 3- methylhistidine	[36]	
	Plasma	LC-MS-MS	Proline-hydroxyproline	[53]	
	Urine	<sup>1</sup> H NMR spectroscopy	TMAO	[54-57]	
Vegetables	Cruciferous Vegetables	Urine	<sup>1</sup> H NMR spectroscopy	SMCSO	[37]
		Urine	<sup>1</sup> H NMR spectroscopy	Phenylacetylglutamine and glycine	[43]
	Vegetarian diet	Urine	<sup>1</sup> H NMR spectroscopy	p-hydroxyphenylacetate	[41]
		Urine	<sup>1</sup> H NMR spectroscopy	Hippurate, N-acetyl glycoprotein and succinate	[44]
		Urine	HPLC-ESI-MS-MS	Enterolactone and kaempferol	[58]
Citrus Fruits	Urine	<sup>1</sup> H NMR spectroscopy	Proline betaine	[17]	
	Urine	FIE-MS	proline betaine and conjugates-sulphate	[59]	
	Urine	HPLC-ESI-MS-MS	Naringenin, hesperetin and sulphonated derivatives of caffeic acid	[58]	
Coffee	Urine	HPLC-ESI-MS-MS	Chlorogenic acid	[58]	
	Serum	ESI-MS-MS	Sphingomyelins	[38]	
	Plasma & urine	HPLC	Urinary dihydrocaffeic acid-3-O-sulphate & feruloylglycine	[39]	
	Plasma	LC-MS-MS	3,4-Dimethoxycinnamic acid, 3,4-Dimethoxy dihydrocaffeic acid	[60]	
Black/Green Tea	Urine	HPLC-MS-MS	Hippuric acid	[61]	
	Urine	HPLC-FTMS <sup>b</sup> and HPLC-TOFMS-SPE-NMR <sup>c</sup>	Hippuric acid and a structurally related hydroxybenzoic glycine conjugate, vanilloylglycine, and pyrogallol-2-O-sulfate	[62]	
	Urine	<sup>1</sup> H NMR spectroscopy	Hippuric acid and 1,3-dihydroxyphenyl-2-O-sulfate	[63]	
	Urine	<sup>1</sup> H NMR spectroscopy	Hippuric acid, 1,3-Dihydroxyphenyl-2-O-sulphate and 4-O-methylgallic acids	[64]	
	Urine	HPLC-ESI-MS-MS	4-O-methylgallic acids	[58]	

<sup>a</sup> Negative chemical ionization Gas chromatography–mass spectrometry

<sup>b</sup> accurate mass fragmentation

<sup>c</sup> Mass-guided SPE-trapping of selected compounds for nuclear magnetic resonance spectroscopy measurements

TMAO: Trimethylamine-N-oxide; PhIP: 2-Amino-1-methyl-6-phenylimidazo[4,5-b]pyridine; SMCSO: S-methyl-L-cysteine sulfoxide

## Summary and Outlook

Although biomarkers cannot replace traditional dietary assessment methods, the use of metabolomics in identifying novel and robust biomarkers of dietary exposure and intake can enhance and validate such methods. Additionally the use of metabolomics in characterising habitual dietary exposure and the identification of nutritypes is an interesting and emerging field with potential applications in nutrition epidemiology.

For metabolomics to reach its full potential in this field a number of challenges need to be addressed. Examples of these challenges include a requirement for technology advancement to enhance our metabolite coverage and advancement in the identification of unknown metabolites to allow novel biomarker discovery. Finally cooperation across disciplines is required to ensure optimal usage of dietary biomarkers.

## Acknowledgements

We kindly acknowledge financial support from Nutritech project number 289511.

### Citation

O’Gorman A, Gibbons H, Brennan L (2013) Metabolomics in the identification of biomarkers of dietary intake. *Computational and Structural Biotechnology Journal*. 4 (5): e201301004. doi: <http://dx.doi.org/10.5936/csbj.201301004>

## References

- Kussmann M, Raymond F, and Affolter M (2006) OMICS-driven biomarker discovery in nutrition and health. *J Biotech* 124(4): 758-787.
- Swinburn BA, Caterson I, Seidell JC, and James WPT (2004) Diet, nutrition and the prevention of excess weight gain and obesity. *Public Health Nutr* 7(1A): 122-146.
- Kipnis V, Midthune D, Freedman L, Bingham S, Day NE, et al. (2002) Bias in dietary-report instruments and its implications for nutritional epidemiology. *Public Health Nutr* 5(6A): 915-923.
- Bingham SA (1991) Limitations of the various methods for collecting dietary intake data. *Ann Nutr Metab* 35(3): 117-127.
- Seale JL, Klein G, Friedman J, Jensen GL, Mitchell DC, et al. (2002) Energy expenditure measured by doubly labeled water, activity recall, and diet records in the rural elderly. *Nutrition* 18(7-8): 568-573.
- Weber JL, Reid PM, Greaves KA, DeLany JP, Stanford VA, et al. (2001) Validity of self-reported energy intake in lean and obese young women, using two nutrient databases, compared with total energy expenditure assessed by doubly labeled water. *Eur J Clin Nutr* 55(11): 940-950.
- Bingham SA, Gill C, Welch A, Day K, Cassidy A, et al. (1994) Comparison of dietary assessment methods in nutritional epidemiology: weighed records v. 24 h recalls, food-frequency questionnaires and estimated-diet records. *Br J Nutr* 72(04): 619-643.
- Penn L, Boeing H, Boushey CJ, Dragsted LO, Kaput J, et al. (2010) Assessment of dietary intake: NuGO symposium report. *Genes & Nutr* 5(3): 205-213.
- Spencer JPE, Abd El Mohsen MM, Minihane A, and Mathers JC (2008) Biomarkers of the intake of dietary polyphenols: strengths, limitations and application in nutrition research. *Br J Nutr* 99(01): 12-22.
- Jenab M, Slimani N, Bictash M, Ferarri P, and Bingham SA (2009) Biomarkers in nutritional epidemiology: applications, needs and new horizons. *Hum Genet* 125(5-6): 507-525.
- Bingham SA (2002) Biomarkers in nutritional epidemiology. *Public Health Nutr* 5(6A): 821-827.
- Potischman N (2003) Biologic and methodologic issues for nutritional biomarkers. *J Nutr* 133(3): 875S-880S.
- Zamora-Ros R, Rabassa M, Llorach R, González CA, and Andres-Lacueva C (2012) Application of phenolic biomarkers in epidemiology: past, present, and future. *J Agric Food Chem* 60(27): 6648-6657.
- Potischman N and Freudenheim JL (2003) Biomarkers of nutritional exposure and nutritional status: An overview. *J Nutr* 133(3): 873S-874S.
- Tasevska N, Runswick SA, McTaggart A, and Bingham SA (2005) Urinary sucrose and fructose as biomarkers for sugar consumption. *Cancer Epidemiol Biomarkers Prevent* 14: 1287-1294.
- Baldrick FR, Woodside JV, Elborn JS, Young IS, and McKinley MC (2011) Biomarkers of fruit and vegetable intake in human intervention studies: A systematic review. *Crit Rev Food Sci & Nutr* 51(9): 795-815.
- Heinzmann SS, Brown IJ, Chan Q, Bictash M, Dumas M-E, et al. (2010) Metabolic profiling strategy for discovery of nutritional biomarkers: proline and betaine as a marker of citrus consumption. *Am J Clin Nutr* 92(2): 436-443.
- Bino RJ, Hall RD, Fiehn O, Kopka J, Saito K, et al. (2004) Potential of metabolomics as a functional genomics tool. *Trends Plant Sci* 9(9): 418-425.
- Fiehn O (2002) Metabolomics-the link between genotypes and phenotypes. *Plant Mol Biol* 48(1-2): 155-171.
- Kell DB (2004) Metabolomics and systems biology: making sense of the soup. *Curr Opin Microbiol* 7(3): 296-307.
- Sumner LW, Mendes P, and Dixon RA (2003) Plant metabolomics: large-scale phytochemistry in the functional genomics era. *Phytochemistry* 62(6): 817-836.
- German JB, Hammock BD, and Watkins SM (2005) Metabolomics: building on a century of biochemistry to guide human health. *Metabolomics* 1(1): 3-9.
- Nicholson JK and Wilson ID (2003) Opinion: understanding 'global' systems biology: metabolomics and the continuum of metabolism. *Nat Rev Drug Discov* 2(8): 668-776.
- Dumas ME, Maibaum EC, Teague C, Ueshima H, Zhou B, et al. (2006) Assessment of analytical reproducibility of <sup>1</sup>H NMR spectroscopy based metabolomics for large scale epidemiological research: the INTERMAP study. *Anal Chem* 78(7): 2199-2208.
- Favé G, Beckmann ME, Draper JH, and Mathers JC (2009) Measurement of dietary exposure a challenging problem which may be overcome thanks to metabolomics? *Genes & Nutr* 4(2): 135-141.
- Chang D (1983) On using principal components before separating a mixture of two multivariate normal distributions. *Applied Statistics* 32(3): 267-275.
- McLachlan GJ and Peel D (2000) *Finite Mixture Models*. New York: Wiley.
- Tipping ME and Bishop CM (1999) Probabilistic principal component analysis *Journal of the Royal Statistical Society, Series B* 61(3): 611-622.
- Nyamundanda G, Brennan L, and Gormley IC (2010) Probabilistic principal component analysis for metabolomic data *BMC Bioinformatics* 11: 571-582.
- Westerhuis JA, Hoefsloot HCJ, Smit S, Vis DJ, Smilde AK, et al. (2008) Assessment of PLS-DA cross validation. *Metabolomics* 4(1): 81-89.
- Breiman L (2001) Random forests. *Mach Learn* 45(1): 5-32.
- Henneges C, Bullinger D, Fux R, Friese N, Seeger H, et al. (2009) Prediction of breast cancer by profiling urinary RNA metabolites using Support Vector Machine-based feature selection. *BMC Cancer* 9:104.
- Izquierdo-García JL, Rodríguez I, Kyriazis A, Villa P, Barreiro P, et al. (2009) A novel R-package graphic user for the analysis of metabolomic profiles. *BMC Bioinformatics* 10:363.
- Sugimoto M, Kawakami M, Robert M, Soga T, and Tomita M (2012) Bioinformatics tools for mass spectrometry-based metabolomic data processing and analysis. *Curr Bioinformatics* 7(1): 96-108.

35. Boccard J, Veuthey JL, and Rudaz S (2010) Knowledge discovery in metabolomics: an overview of MS data handling. *J Sep Sci* 33(3): 290-304.
36. Lloyd AJ, Favé G, Beckmann M, Lin W, Tailliant K, et al. (2011) Use of mass spectrometry fingerprinting to identify urinary metabolites after consumption of specific foods. *Am J Clin Nutr* 94(4): 981-991.
37. Edmonds WM, Beckonert OP, Stella C, Campbell A, Lake BG, et al. (2011) Identification of human urinary biomarkers of cruciferous vegetable consumption by metabolomic profiling. *J Proteome Res* 10(10): 4513-4521.
38. Altmaier E, Kastenmüller G, Römisch-Margl W, Thorand B, Weinberger KM, et al. (2009) Variation in the human lipidome associated with coffee consumption as revealed by quantitative targeted metabolomics. *Mole Nutr & Food Res* 53(11): 1357-1365.
39. Stalmach S, Mullen W, Barron D, Uchida K, Yokota T, et al. (2009) Metabolite profiling of hydroxycinnamate derivatives in plasma and urine after the ingestion of coffee by humans: identification of biomarkers of coffee consumption. *Drug Metab Dispos* 37(8): 1749-1758.
40. Mullen W, Boitier A, Stewart AJ, and Crozier A (2004) Flavanoid metabolites in human plasma and urine after the consumption of red onions; analysis by liquid chromatography with photodiode array and full scan tandem mass spectrometric detection. *J Chromatogr A* 1058(1-2): 163-168.
41. Stella C, Beckwith-Hall BM, Cloarec O, Holmes E, Lindon JC, et al. (2006) Susceptibility of human metabolic phenotypes to dietary modulation. *J Proteome Res* 5(10): 2780-2788.
42. Cross AJ, Major JM, and Sinha R (2011) Urinary biomarkers of meat consumption. *Cancer Epidemiol Biomarkers Prevent* 20(6): 1107-1111.
43. O'Sullivan A, Gibney MJ, and Brennan L (2011) Dietary intake patterns are reflected in metabolomic profiles: potential role in dietary assessment studies<sup>1-3</sup>. *Am J Clin Nutr* 93(2): 314-321.
44. Xu J, Yang S, Cai J, Dong J, Li X, et al. (2010) Identification of biochemical changes in lactovegetarian urine using <sup>1</sup>H NMR spectroscopy and pattern recognition. *Anal Bioanal Chem* 396(4): 1451-1463.
45. Bertram HC, Hoppe C, Peterson BO, Duus JØ, Mølgaard C, et al. (2007) An NMR-based metabolomic investigation on effects of milk and meat protein diets given to 8-year-old boys. *Br J Nutr* 97(4): 758-763.
46. Park YJ, Volpe SL, and Decker EA (2005) Quantitation of carnosine in humans plasma after dietary consumption of beef. *J Agric Food Chem* 53(12): 4736-4739.
47. Myint T, Fraser GE, Lindsted KD, Knutsen SF, Hubbard RW, et al. (2000) Urinary 1-methylhistidine is a marker of meat consumption in black and white California seventh-day adventists. *Am J Epidemiol* 152(8): 752-755.
48. Kobayashi M, Hanaoka T, Hashimoto H, and Tsugane S (2005) 2-Amino-1-methyl-6-phenylimidazo[4,5-b]pyridine (PhIP) level in human hair as biomarkers for dietary grilled/stir-fried meat and fish intake. *Mutat Res* 588(2): 136-142.
49. Kulp KS, Knize MG, Fowler ND, Salmon CP, and Felton JS (2004) PhIP metabolites in human urine after consumption of well-cooked chicken. *J Chromatogr B Analyt Biomed Life Sci* 802(1): 143-153.
50. Strickland PT, Qian Z, Friesen MD, Rothman N, and Sinha R (2002) Metabolites of 2-amino-1-methyl-6-phenylimidazo(4,5-b)pyridine (PhIP) in human urine after consumption of charbroiled or fried beef. *Mutat Res* 506-507: 163-173.
51. Stillwell WG, Sinha R, and Tannenbaum SR (2002) Excretion of the N2-glucuronide conjugate of 2-hydroxyamino-1-methyl-6-phenylimidazo[4,5-b]pyridine and its relationship to CYP1A2 and NAT2 activity levels in humans. *Carcinogenesis* 23(5): 831-838.
52. Reistad R, Rossland OJ, Larva-Kala KJ, Rasmussen T, Vikse R, et al. (1997) Heterocyclic aromatic amines in human urine following a fried meat meal. *Food Chem Toxicol* 35(10-11): 945-955.
53. Ichikawa S, Morifuji M, Ohara H, Matsumoto H, Takeuchi Y, et al. (2010) Hydroxyproline-containing dipeptides and tripeptides quantified at high concentration in human blood after oral administration of gelatin hydrolysate. *Int J Food Sci & Nutr* 61(1): 52-60.
54. Holmes E, Loo RL, Stamler J, Bictash M, Yap IKS, et al. (2008) Human metabolic phenotype diversity and its association with diet and blood pressure. *Nature* 453: 396-400.
55. Lee MB, Storer MK, Blunt JW, and Lever M (2006) Validation of <sup>1</sup>H NMR spectroscopy as an analytical tool for methylamine metabolites in urine. *Clin Chim Acta* 365(1-2): 264-269.
56. Lenz EM, Bright J, Wilson ID, Hughes A, Morrison J, et al. (2004) Metabonomics, dietary influences and cultural differences: a <sup>1</sup>H NMR-based study of urine samples obtained from healthy British and Swedish subjects. *J Pharm Biomed Anal* 36(4): 841-849.
57. Zuppi C, Messana I, Forni F, Ferrari F, Rossi C, et al. (1998) Influence of feeding on metabolite excretion evidenced by urine <sup>1</sup>H NMR spectral profiles: a comparison between subjects living in Rome and subjects living at arctic latitudes (Svalbard). *Clin Chim Acta* 278(1): 75-79.
58. Mennen LI, Sapinho D, Ito H, Bertrais S, Galan P, et al. (2006) Urinary flavonoids and phenolic acids as biomarkers of intake for polyphenol-rich foods. *Br J Nutr* 96(01): 191-198.
59. Lloyd AJ, Beckmann M, Favé G, Mathers JC, and Draper J (2011) Proline betaine and its biotransformation products in fasting urine samples are potential biomarkers of habitual citrus fruit consumption. *Br J Nutr* 106(6): 812-824.
60. Nagy K, Redeuil K, Williamson G, Rezzi S, Dionisi F, et al. (2011) First identification of dimethoxycinnamic acids in human plasma after coffee intake by liquid chromatography-mass spectrometry. *J Chromatogr A* 1218(3): 491-491.
61. Mulder TP, Rietveld AG, and van Amelsvoort JM (2005) Consumption of both black tea and green tea results in an increase in the excretion of hippuric acid into urine. *Am J Clin Nutr* 81(1 Suppl): 256S-260S.
62. van der Hoof JJ, de Vos RC, Mihaleva V, Bino RJ, Ridder L, et al. (2012) Structural Elucidation and Quantification of Phenolic Conjugates Present in Human Urine after Tea Intake. *Anal Chem* 84(16): 7263-7271.
63. Van Dorsten FA, Daykin CA, Mulder TPJ, and Van Duynhoven JPM (2006) Metabonomics Approach To Determine Metabolic Differences between Green Tea and Black Tea Consumption. *J Agric Food Chem* 54(18): 6929-6938.
64. Daykin CA, Van Duynhoven JPM, Groenewegen A, Dachtler M, Van Amelsvoort JMM, et al. (2005) Nuclear magnetic resonance spectroscopic based studies of the metabolism of black tea polyphenols in humans. *J Agric Food Chem* 53(5): 1428-1434.
65. Svensson BG, Akesson M, Nilsson A, and Paulsson K (1994) Urinary excretion of methylamines in men with varying intake of fish from the Baltic sea. *J Toxicol Environ Health* 41(4): 411-420.
66. Dragsted LO (2010) Biomarkers of meat intake and the application of nutrigenomics. *Meat Sci* 84(2): 301-307.
67. Sjolín J, Hjort G, Friman G, and Hambraeus L (1987) Urinary excretion of 1-methylhistidine: a qualitative indicator of exogenous 3-methylhistidine and intake of meats from various sources. *Metabolism* 36(12): 1175-1184.

68. Rasmussen LG, Winning H, Savorani F, Toft H, Larsen TM, et al. (2012) Assessment of the effect of high or low protein diet on the human urine metabolome as measured by NMR. *Nutrients* 4(2): 112-131.
69. Stamler J, Elliot P, Dennis B, Dyer AR, Kesteloot H, et al. (2003) INTERMAP: background, aims, design, methods and descriptive statistics (dietary). *J Hum Hypertens* 17(9): 591-608.
70. Dehghan M, Akhtar-Danesh N, McMillan CR, and Thabane L (2007) Is plasma vitamin C an appropriate marker of vitamin C intake? A systematic review and meta-analysis. *Nutr J* 6: 41-53.
71. Padayatty SJ and Levine M (2008) Fruit and vegetables: Think variety, go ahead, eat! *Am J Clin Nutr* 87(1): 5-7.
72. McAnlis GT, McEneny J, Pearce J, and Young IS (1999) Absorption and antioxidant effects of quercetin from onions, in man. *Eur J Clin Nutr* 53(2): 92-96.
73. Hodgson JM, Morton LW, Puddey IB, Beilin LJ, and Croft KD (2000) Gallic acid metabolites are markers of black tea in humans. *J Agric Food Chem* 48(6): 2276-2280.
74. Kivits GAA, Vandersman FJP, and Tijburg LBM (1997) Analysis of catechins from green and black tea in humans: a specific and sensitive colorimetric assay of total catechins in biological fluids. *Int J Food Sci & Nutr* 48(6): 387-392.
75. Hodgson JM, Chan SY, Puddey IB, Devine A, Wattananpenpaiboon N, et al. (2004) Phenolic metabolites as biomarkers for tea- and coffee-derived polyphenol exposure in human subjects. *Br J Nutr* 91(2): 301-306.
76. Rechner AR, Spencer JP, Kuhnle G, Hahn U, and Rice-Evans CA (2001) Novel biomarkers of the metabolism of caffeic acid derivatives in vivo. *Free Radical Bio Med* 30(11): 1213-1222.
77. Shahrzad S, Aoyagi K, Winter A, Koyama A, and Bitsch I (2001) Pharmacokinetics of gallic acid and its relative bioavailability from tea in healthy humans. *J Nutr* 131(4): 1207-1210.
78. Peré-Trepat E, Ross AB, Martin F-P, Rezzi S, Kochhar S, et al. (2010) Chemometric strategies to assess metabonomic imprinting of food habits in epidemiological studies. *Chemom Intell Lab Syst* 104(1): 95-100.
79. Altmaier E, Kastenmüller G, Römisch-Margl W, Thorand B, Weinberger KM, et al. (2011) Questionnaire-based self-reported nutrition habits associate with serum metabolism revealed by quantitative targeted metabolomics. *Eur J Epidemiol* 26(2): 145-156.

**Keywords:**

Biomarkers, habitual diet, metabolomics

**Competing Interests:**

The authors have declared that no competing interests exist.



© 2013 O’Gorman et al.

Licensee: Computational and Structural Biotechnology Journal.

This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are properly cited.

**What is the advantage to you of publishing in *Computational and Structural Biotechnology Journal (CSBJ)* ?**

- ✚ Easy 5 step online submission system & online manuscript tracking
- ✚ Fastest turnaround time with thorough peer review
- ✚ Inclusion in scholarly databases
- ✚ Low Article Processing Charges
- ✚ Author Copyright
- ✚ Open access, available to anyone in the world to download for free

[WWW.CSBJ.ORG](http://WWW.CSBJ.ORG)