



Cite this: *Food Funct.*, 2023, **14**, 8586

## Metabolomic based approach to identify biomarkers of broccoli intake†

Aoife E. McNamara,<sup>‡a,b</sup> Xiaofei Yin,<sup>‡a,b</sup> Cassandra Collins<sup>‡a,b</sup> and Lorraine Brennan<sup>‡a,b</sup>

It is well-established that consumption of cruciferous and brassica vegetables has a correlation with reduced rates of many negative health outcomes. There is an increased interest in identifying food intake biomarkers to address limitations related to self-reported dietary assessment. The study aims to identify biomarkers of broccoli intake using metabolomic approaches, examine the dose–response relationship, and predict the intake by multimarker panel. Eighteen volunteers consumed cooked broccoli in A-Diet Discovery study and fasting and postprandial urine samples were collected at 2, 4 and 24 hours. Subsequently the A-Diet Dose–response study was performed where volunteers consumed different portions of broccoli (49, 101 or 153 g) and urine samples were collected at the end of each intervention week. Urine samples were analysed by <sup>1</sup>H-NMR and LC-MS. Multivariate data analysis and one-way ANOVA were performed to identify discriminating biomarkers. A panel of putative biomarkers was examined for its ability to predict intake through a multiMarker model. Multivariate analysis revealed discriminatory spectral regions between fasting and fed metabolic profiles. Subsequent time-series plots revealed multiple features increased in concentration following the consumption. Urinary *S*-methyl cysteine sulf-oxide (SMCSO) increased as broccoli intake increased (0.17–0.24 μM per mOSM per kg, *p* < 0.001). Similarly from LC-MS data genipin, dihydro-β-tubaic acid and sinapic acid increased with increasing portions of intake. A panel of 8 features displayed good ability to predict intake from biomarker data only. In conclusion, urinary SMCSO and several LC-MS features appeared as potentially promising biomarkers of broccoli consumption and demonstrated dose–response relationship. Future work should focus on validating these compounds as food intake biomarkers.

Received 22nd December 2022,  
Accepted 27th August 2023

DOI: 10.1039/d2fo03988e

rsc.li/food-function

## 1 Introduction

Broccoli is a member of a group known as cruciferous or brassica vegetables which also includes kale, cabbage, Brussels sprouts, cauliflower and others.<sup>1</sup> Brassica are among the most important vegetables in Europe as they are a popular food, cheap and easily accessible.<sup>2</sup> Brassica are rich in beneficial plant metabolites and many of the associated health benefits are linked to their glucosinolate-derived phytochemical content.<sup>3</sup> A recent meta-analysis found cruciferous consumption was inversely associated with stroke, cardiovascular disease (CVD), cancer and all-cause mortality, with greatest

effects at intakes >200 g day<sup>-1</sup>.<sup>4</sup> However, levels of heterogeneity between studies were moderate to high for stroke, cardiovascular disease and total cancer and high for all-cause mortality. One of the factors contributing to the inconsistent results could be the difficulty in obtaining accurate dietary exposure data.

Traditional commonly used dietary assessment techniques are self-reported methods which are subject to well documented limitations such as inaccurate portion size estimation, recall bias and misreporting.<sup>5–8</sup> These limitations highlight the need to develop more accurate and objective dietary assessment measures, for example, the identification of food intake biomarkers. 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.<sup>9</sup>

Metabolomics, the study of small metabolites, plays a key role in the discovery of food intake biomarkers. Multiple studies examining cruciferous vegetable intake assessment are present in literature.<sup>10–15</sup> The most promising biomarkers of cruciferous vegetables discovered to date are sulphur-contain-

<sup>a</sup>UCD School of Agriculture and Food Science, Institute of Food and Health, University College Dublin, Belfield, Dublin 4, Ireland.

E-mail: lorraine.brennan@ucd.ie; Tel: +353 1 7166815

<sup>b</sup>UCD Conway Institute of Biomolecular and Biomedical Research, University College Dublin, Dublin, Ireland

† Electronic supplementary information (ESI) available. See DOI: <https://doi.org/10.1039/d2fo03988e>

‡ These authors contributed equally.



ing compounds such as: sulforaphane,<sup>10,11</sup> SMCSO (*S*-methylcysteine sulfoxide)<sup>12</sup> and other compounds from the isothiocyanate class.<sup>10,11,15</sup> Sulforaphane, an isocyanate, is abundantly present in cruciferous vegetables,<sup>16</sup> and sulforaphane *N*-acetylcysteine and sulforaphane *N*-cysteine were identified by Andersen *et al.*, (2013)<sup>11</sup> as urinary markers of brassica containing meals in a cross-over meal study representative of the new Nordic diet. Upon further analysis, sulforaphane *N*-acetylcysteine was discovered to be a strong brassica intake biomarker (sensitivity and specificity >80%). SMCSO and 3 of its metabolites were identified as urinary biomarkers of cruciferous intake following a high cruciferous intake dietary intervention (500 g d<sup>-1</sup> of broccoli and Brussels sprouts for 14 days).<sup>12</sup> A range of acetyl-cysteine derivatives of isothiocyanates have been identified as biomarkers of brassica-containing diets including *N*-Acetyl-(*N'*-benzylthiocarbonyl)-cysteine, Iberin *N*-acetyl cysteine,<sup>10,11</sup> and iberin.<sup>15</sup> However there has been no studies attempting to validate the compounds as broccoli biomarkers or to assess their ability to predict intake. To this end the objectives of the present study were as follows: (1) to use a metabolomics approach to identify and confirm biomarkers related to broccoli consumption; (2) to examine the dose–response relationship and (3) to assess the ability of a panel of potential biomarkers for predicting food intake.

## 2 Materials and methods

### 2.1 A-diet discovery study

Ethical approval for the A-Diet Discovery study was granted by the UCD Sciences Human Research Ethics Committee (LS-15-69-Brennan). The study design has been previously been described in detail,<sup>17</sup> to identify novel biomarkers of 9 commonly consumed foods (apples, broccoli, peppers, oranges, white bread, wholemeal bread, spaghetti, cheese and Madeira cake.). The primary test food of interest for this research was broccoli and apple intake was used as a control food and we reported only the data with respect to the discovery of broccoli biomarkers. The inclusion criteria included healthy, non-pregnant/lactating, non-smokers, an age range between 18 and 60 years old, and a body mass index (BMI) between 18.5 and 30 kg m<sup>-2</sup>. Exclusion criteria included any diagnosed health condition (chronic or infectious diseases), consumption of medications/nutritional supplements or any allergies/intolerances to the test foods. Furthermore, participants were asked to avoid consuming alcohol, medication, and foods related to specific food (apples or broccoli) for 24 h prior to biofluid collection. Participants consumed 135 g of cooked broccoli and 360 g of raw apples, and provided fasting first void and post-prandial (2 h and 4 h) urine samples. Following 24 h a fasting first void urine sample was collected.

### 2.2 A-diet dose–response study

Ethical approval for the A-Diet Dose–response study was granted by the UCD Sciences Human Research Ethics

Committee (LS-17-16-Brennan). Inclusion criteria for the dose–response study were the same as for the discovery study above. Details of the study design are published elsewhere.<sup>17</sup> Participants were assigned to either a lunch (*N* = 27) or dinner (*N* = 34) test meal group and invited to partake in a 5-week study. Each week the portion sizes of each test food changed (high, medium or low portion). Participants received the meals in random order. The high, medium and low broccoli portions were 153 g, 101 g, and 49 g and apple portions were 300 g, 100 g and 50 g, respectively. During the test week, participants were also asked to avoid consuming any other foods related to the test meal ingredients.

Fasting first void urine was collected after an overnight 12 h fast at the end of each test week after participants had consumed the test meal for 4 days and chilled until processing. All urine samples were centrifuged at 1800g for 10 min at 4 °C. Aliquots of 1 mL were stored at –80 °C for later analysis.

### 2.3 Metabolomic analysis of urine samples

The <sup>1</sup>H nuclear magnetic resonance (<sup>1</sup>H-NMR) spectroscopy and liquid chromatography mass spectrometry (LC-MS) based techniques were applied for urine metabolomic analysis. Urine samples from participants were prepared for <sup>1</sup>H-NMR spectroscopy analysis first defrosted and then prepared by addition of 250 μL phosphate buffer (0.2 mol KH<sub>2</sub>PO<sub>4</sub> per L, 0.8 mol K<sub>2</sub>HPO<sub>4</sub> per L) to 500 μL urine, and the mixed sample was then centrifuged at 5360g for 5 min at 4 °C. The collected supernatant (540 μL) were mixed with 10 μL sodium trimethylsilyl [2,2,3,3-2H<sub>4</sub>] propionate (TSP) and 50 μL deuterium oxide (D<sub>2</sub>O). Subsequently, the samples were transferred into NMR tubes for analysis. Spectra were acquired on a 600 MHz Varian Spectrometer (Varian Limited, Oxford, United Kingdom) by using the first increment of a nuclear Overhauser enhancement spectroscopy pulse sequence at 25 °C. A total of 16 384 data points and 128 scans was used, and water suppression was achieved during the 2.5 s of relaxation delay and the 100 ms of mixing time. All <sup>1</sup>H-NMR urine spectra were referenced to TSP at 0.0 parts per million (ppm) and then processed manually in the Chenomx NMR Suite software (version 7.7) by using a line broadening of 0.2 Hz, followed by phase and baseline correction. Data were normalized to the total area of the spectra integral. NMR spectra from the Discovery study were exported at high resolution using 7500 spectral regions, furthermore the water region was excluded. The exported spectral region dataset was used for multivariate data analysis to extract the most discriminated molecule features. Metabolites were identified using the Chenomx library. To confirm the correct peak assignment of metabolites a pure standard of SMCSO was prepared and analysed by <sup>1</sup>H-NMR. A 50 μL solution of pure standard and phosphate buffer (0.0025 mol L<sup>-1</sup>) was added to a urine sample. The <sup>1</sup>H-NMR spectra were acquired prior to and after the addition of the pure compound using the same parameters as study samples. The spectra of the pure SMCSO standard, the urine and urine spiked with SMCSO were compared to confirm the identification of SMCSO.



Osmolality was measured using an Advanced Micro Osmometer model 3300 (Advanced Instruments) employing freezing point depression. Values are reported as the number of solute particles, in moles, dissolved in a kilogram of urine (mOsm per kg). Profiled urinary metabolite concentrations from NMR analysis were normalized to osmolality.

Urine samples were analysed by LC-MS as previously described.<sup>18</sup> A mixture of five isotope labelled compounds (malic acid d<sub>3</sub>, methionine d<sub>3</sub>, myristic acid <sup>13</sup>C, adipic acid d<sub>4</sub> and succinic acid d<sub>4</sub>, each at 10 µg mL<sup>-1</sup> in 20% (v/v) EtOH/Millipore H<sub>2</sub>O) as internal standards were used. Test urine and pooled urine samples (used as Quality Control (QC) samples) were thawed on a roller for 30 min and centrifuged at 1800g for 5 min at 4 °C. Urine samples (100 µL) were mixed with 100 µL internal standards, and then vortexed at 35 Hz for 10 s and centrifuged at 2000g for 2 min. The supernatant was subsequently transferred to LC-MS vials with 250 µL inserts and placed into LC autosampler at 4 °C.

An Agilent LC-QTOF-MS, consisting of a 1290 Infinity II LC system and an Agilent Jetstream Electrospray ionization (ESI) source coupled to a 6545 QTOF mass spectrometer were applied to perform urine metabolomic analysis. The chromatography was performed in reverse phase mode with a Zorbax eclipse plus C18 column (2.1 × 50 mm, 1.8 µm) coupled to a Zorbax eclipse plus C18 guard column (2.1 × 5 mm, 1.8 µm). The column temperature was set as 30 °C, and injection volume was 5 µL. The mobile phase A and B were water and 80% of acetonitrile, respectively, both with 0.1% formic acid. The flow rate was 0.4 mL min<sup>-1</sup>, and the gradient conditions were set as follows: 1% B (0–1.5 min), 11% B (1.5–9 min), 25% B (9–15 min), 50% B (15–18 min), 99% B (18–18.05 min), 99% B (18.05–21 min), 1% B (21–21.05 min) and 1% B (21.05–23 min). The MS parameters used for the analysis: drying gas temperature, 325 °C; drying gas flow rate, 10 L min<sup>-1</sup>; sheath gas temperature, 350 °C; sheath gas flow rate, 11 L min<sup>-1</sup>; nebulizer pressure, 45 gauge pressure (pounds per square inch); capillary voltage, 3500 V; nozzle voltage, 1000 V; fragment or voltage, 100 V; skimmer, 45 V. The mass range of mass-to-charge-ratio (*m/z*) was between 50 and 1600, and the 2 GHz extended dynamic range mode was used. Application of centroid mode at scan rate of 1 spectra per s was used to collect data. Samples were run by using both positive and negative ionization modes.

Data were acquired using MassHunter acquisition B.08.00 software (B.08.00.8058.3 Sp1 Agilent Technologies) and were further processed in MassHunter qualitative analysis software (B.07.00 Sp2 Agilent Technologies). Molecular features were extracted using the molecular feature extractor (MFE) algorithm, and a list of features were generated with retention time, *m/z*, adducts or isotopes of compounds, signal intensity and accurate mass. The target data type was set to small molecules and the threshold value for peak height was set to 5000 counts. The isotopic peak spacing tolerance was 0.0025 *m/z* with a maximum charge state of 1 and isotope model was set to common organic molecules. The

Compound Exchange Format (.cef) files, transformed from the original spectral data, were generated and imported into MassProfiler (Version B.07.01; build 99.0 Agilent Technologies) in order to align compounds and filter data. The alignment parameters were set with a retention time tolerance ±0.3 min and a mass tolerance of ±15 ppm + 2.0 mDa. The final dataset was exported including molecule features with retention time, mass and feature abundance. For features with missing values, they were replaced with half the minimum abundance value for that particular feature. Feature abundances were normalised to osmolality before performing multivariate data analysis.

The interesting features were initially identified in MassHunter ID Browser (B 7.0.799.2 Agilent Technologies) by putative formula generation and an accurate mass data search against a local METLIN (METabolite LINK, <https://metlin.scripps.edu>) database. Tandem MS (MS/MS) with 3 collision energies 10 eV, 20 eV and 40 eV under the same chromatographic conditions were further performed to attain structural confidence for the metabolites identified by the accurate mass. MS/MS compound identification efforts included fragmentation modelling using CFM ID<sup>19</sup> and fragment matches with the Human Metabolome Database (HMDB).<sup>20</sup> Fragment data generated from MS/MS was also input into METLIN MS/MS Spectrum Match Search (precursor *m/z* and top 30 most intense fragments *m/z*) in order to confirm identification. Furthermore, some authentic standards were analysed by the same LC-MS/MS method to confirm the identification. All metabolites were attributed to one of four levels of confidence in metabolite identification according to Metabolomics standards initiative: Level I for identified compounds; Level II for putatively annotated compounds; Level III for putatively characterized compound classes; Level IV for unknowns.<sup>21</sup>

#### 2.4 Statistical analysis

Data from NMR and LC-MS from the A-Diet Discovery study were analysed using Multivariate Empirical Bayes Analysis of Variance (MEBA) for Time Series in Metaboanalyst (<https://www.metaboanalyst.ca/>). Features with significant changes over time (using the Hotelling-T<sub>2</sub> values) were examined for their dose–response in the A-Diet Dose–response study. One-way repeated measures analysis of variance (ANOVA) was performed on the dose–response data. A *p* value < 0.05 was considered to indicate significance.

Using the multiMarker web application a panel of putative biomarkers was examined for its ability to predict intake.<sup>22</sup> The multiMarker approach is a factor analytic model that expresses the relationship between the biomarkers and the latent intake. The model is built within a Bayesian framework which allows estimation of uncertainties in the predicted intake. The training set was established from randomly selecting 90% of the data to build the model and the rest of data were used to predict intake using biomarker data. This was repeated in an iterative fashion to investigate the agreement between the predicted and actual intake.



### 3 Results

#### 3.1 Identification of biomarkers of broccoli intake

In total, 8 males and 12 females were recruited for the Discovery Study. One participant attended for baseline measurements only. From the remaining volunteers, 18 completed the broccoli test visit (ESI Fig. 1†). The participants' mean age was 34 years old, with a mean body mass index (BMI) of 23.88 kg m<sup>-2</sup> (Table 1).

After conducting MEBA for Time Series on <sup>1</sup>H-NMR data, the top 30 spectral regions were selected and examined against the control food at 0, 2, and 4 h time points. The analysis revealed multiple spectral regions which exhibited an increased peak intensity with time following broccoli consumption only (ESI Fig. 2†) and these spectral regions were listed in Table 2 with their potential identification.

A number of these spectral regions of interest were assigned to SMCSO and identification confirmed through comparison with an authentic analytical standard (ESI Fig. 3†). Other regions were identified as acetate, levulinate, 2-hydroxyisobutyrate, creatinine and sarcosine (Table 2). Profiling of urinary SMCSO revealed that there was a significant increase in the

postprandial samples at 2 and 4 hours compared to fasting and 24 hours (Fig. 1A). After 24 hours, the level of SMCSO was back to the similar level as 0 hour. On the contrary, the level of SMCSO after apple intake remained the same across 4 time points.

Using the LC-MS data a total of 1741 features were obtained from urine samples in positive mode and 2086 features were obtained in negative mode. Following MEBA analysis the top 30 features were selected and examined for increases following broccoli consumption compared to control food. A total of 14 features in positive and 7 features in negative indicated an increase after the consumption and also displayed discriminating time profiles when compared to the control food. The results displayed in Fig. 1B–O and Fig. 2A–G show the increased intensity following broccoli intake, whereas the peak intensities didn't change following consumption of the control food, apple.

A molecular formula for each feature was generated by MassHunter using single MS accurate mass data for the molecular ion and its isotopes. Fourteen features of interest in positive, and seven in negative mode, were selected for LC-MS/MS for more complete metabolite identification (Table 3). Matching of MS/MS fragments between urine sample and standard confirmed the identification of Sinapic acid (ESI Fig. 4†). The identification of 3-hydroxy-sebacic acid (*m/z* 219.1206) was based on comparison with an authentic standard of the functional parent compound Sebacic acid. The fragmentation pattern obtained from the standard reflected key fragments from the urine sample with a spectral shift of 2 Da (*m/z* 97.1010 and 95.0853, 121.1012 and 119.0854, 139.1119 and 137.0960). The difference in mass between sebacic acid (~202 Da) and 3-hydroxysebacic acid (~218 Da) is ~16 Da which corresponds to the mass of the addition of a hydroxy group (–OH) and loss of a double bond (ESI Fig. 5†). Matching of MS/MS fragments between urine sample and standard confirmed the identification of Genipin: the main fragment in the standard was *m/z* 149.0595 however in the urine sample *m/z* 167.0703 was a prominent fragment, which would indicate addition of a hydroxy group (ESI Fig. 6†).

#### 3.2 Confirmation of identified biomarkers of broccoli intake

In order to confirm a dose–response for the putative biomarkers, metabolites were examined in urine samples following consumption of different portions of broccoli. The final population consisted of 32 participants, 5 of which did not complete all 3 of the test dinners but have data available for at least 1 meal. The study population demographics of the Dose–response study are reported in Table 1. The participants' mean age was 29 years old, with a mean BMI of 23.96 kg m<sup>-2</sup>.

Urinary SMCSO concentrations were determined following intake of low (49 g), medium (101 g) and high (153 g) portions of broccoli. The average urinary SMCSO concentrations increased as broccoli intake increased (from 0.17 to 0.24 μM per mOsm per kg) (Fig. 3A). Repeated measures ANOVA indicated that SMCSO exhibits a dose–response relationship to broccoli intake in fasting urine samples (*p* < 0.001).

**Table 1** Demographics for A-Diet Discovery and Dose–response studies for broccoli intake

Characteristics	Discovery study ( <i>N</i> = 18)	Dose–response study ( <i>N</i> = 32)
Gender	8M 10F	16M 16F
Age (y)	34 ± 12	29 ± 10
Anthropometrics		
BMI (kg m <sup>-2</sup> )	23.84 ± 2.98	23.96 ± 2.99
W : H	0.77 ± 0.07	0.84 ± 0.06

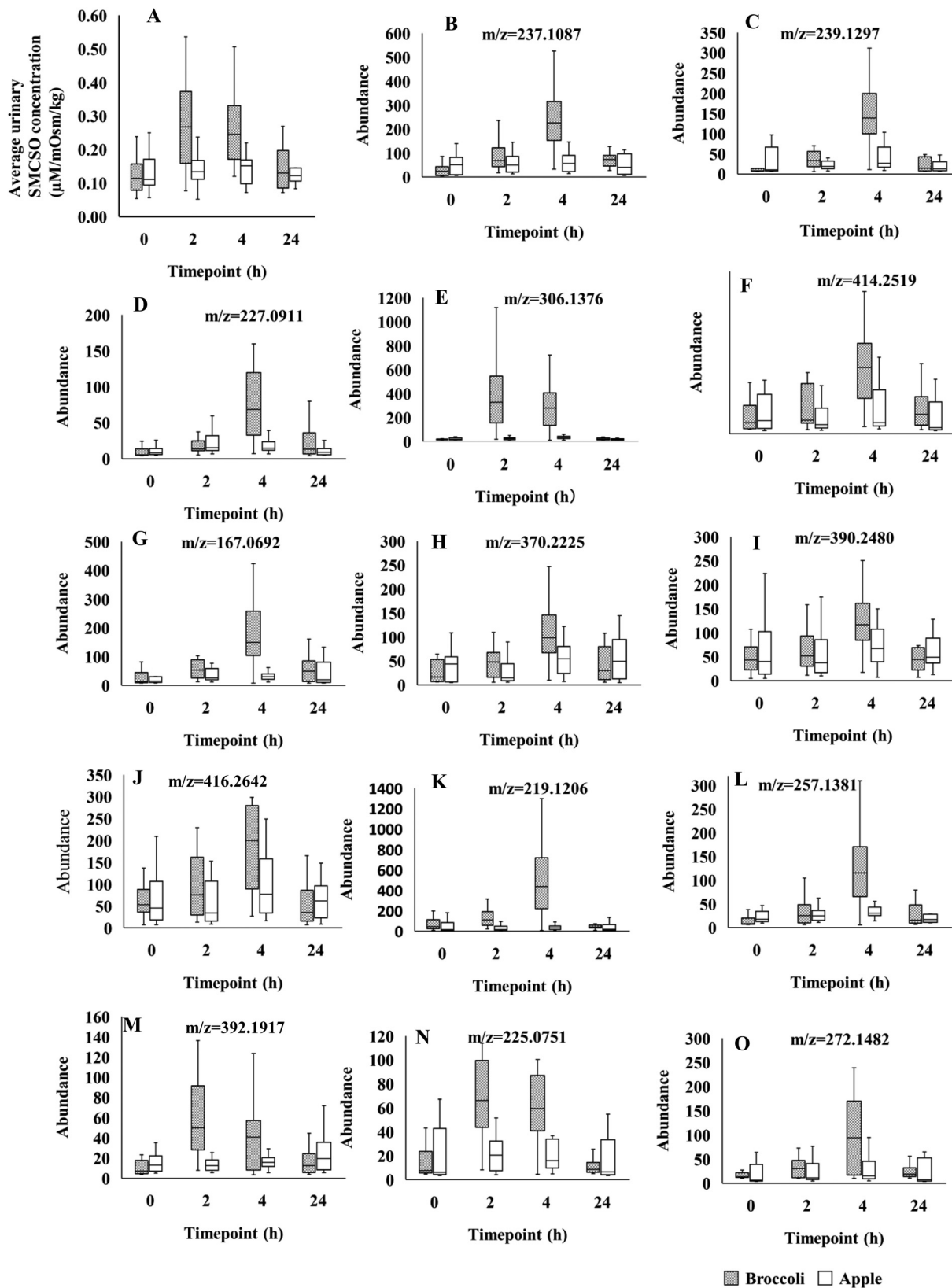
Data are mean ± SD. *N*, number of subjects; M, male; F, female; *y*, years; SD, standard deviation; BMI, body mass index; W : H; waist to hip ratio.

**Table 2** Spectral regions of interest and their potential identifications

Spectral region (ppm) <sup>a</sup>	Hotelling-T2	Potential identification
2.7675	39.066	Levulinate
2.7685	30.431	Levulinate
1.3455	27.932	2-Hydroxyisobutyrate
2.7925	27.835	Unknown 1
2.7935	27.657	Unknown 1
2.7695	25.597	Levulinate
2.7395	24.35	Sarcosine
3.0365	23.95	Creatinine
2.8195	22.839	SMCSO
1.8925	22.575	Acetate
1.8915	22.125	Acetate
2.8205	21.969	SMCSO

<sup>a</sup> A selection of top-ranking regions from MEBA timecourse analysis are presented. These regions represent peaks and one metabolite can have multiple peaks depending on the chemical structure of the metabolite. ppm, parts per million; SMCSO, *s*-methyl cysteine sulfoxide.





**Fig. 1** (A) Box plots of *S*-methyl cysteine sulfoxide (SMCSO) concentrations in NMR analysis across the 4 time-points (0, 2, 4, 24 h), following consumption of broccoli and apples in separate test sessions in A-Diet Discovery study. (B–O) Box plots of fourteen LC-MS feature intensities across the 4 time-points (0, 2, 4, 24 h). Grey and white boxes represent broccoli and apple group, respectively. The solid black line denotes the median of the group. Figure titles indicated the  $m/z$  of features. Those features were identified in positive mode.



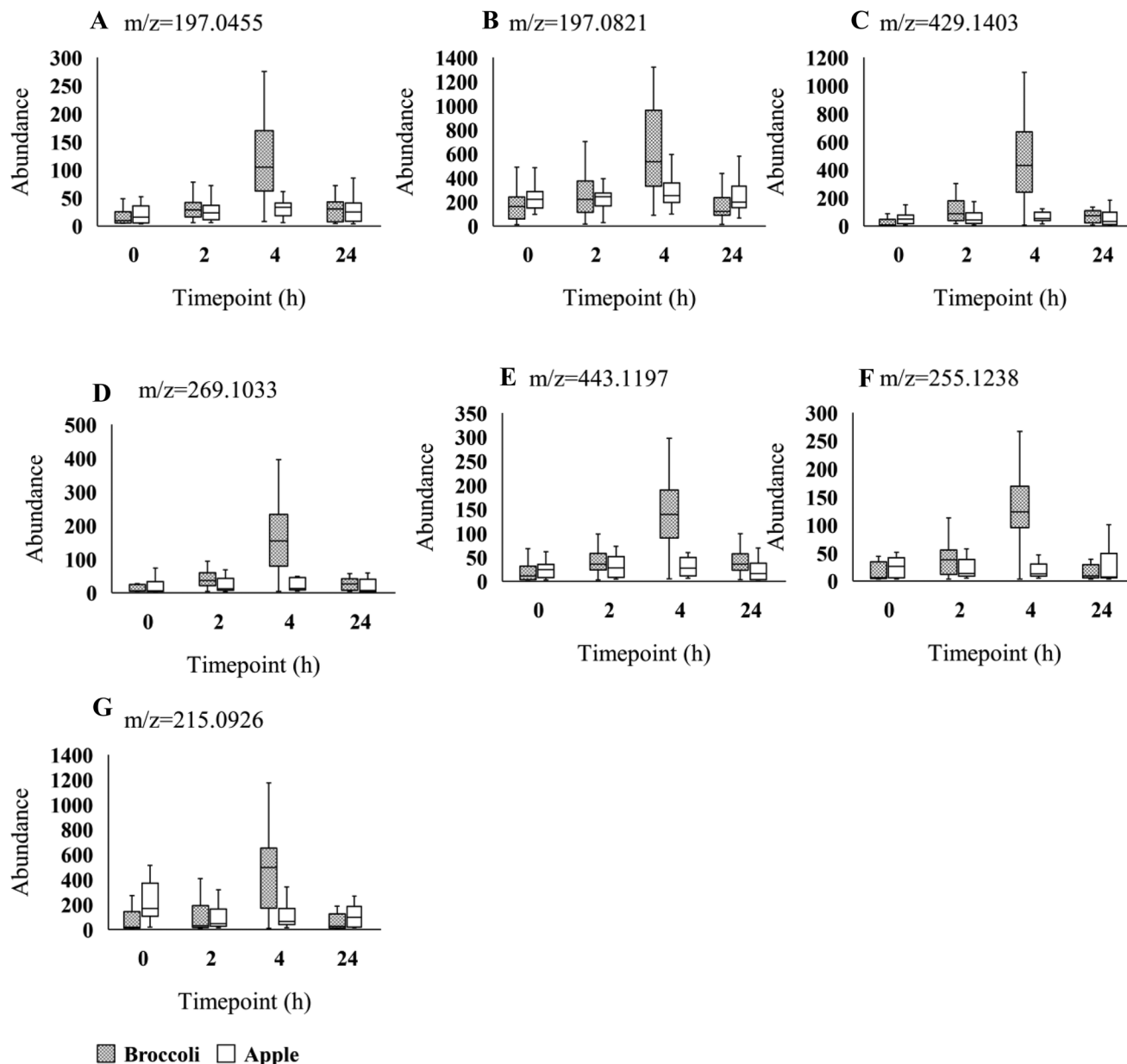


Fig. 2 Box plots of seven LC-MS feature intensities (A-G) across the 4 time-points (0, 2, 4, 24 h). Grey and white boxes represent broccoli and apple group, respectively. The solid black line denotes the median of the group. Figure titles indicated the  $m/z$  of features. Those features were identified in negative mode.

Features which exhibited an increased and differential time-course following broccoli consumption using LC-MS analysis were also examined for a dose-response relationship. Peak areas were examined for the 14 features of interest in positive and 7 features in negative following consumption of the different broccoli portions. Genipin ( $m/z = 227.0911$ ), dihydro- $\beta$ -tubaic acid ( $m/z = 237.1087$ ), sinapic acid ( $m/z = 225.0751$ ) and 5 unidentified compounds ( $m/z = 239.1297$ ,  $370.2225$ ,  $257.1381$ ,  $255.1238$  and  $306.1376$ ) increased in normalised peak area as portion of broccoli increased (Fig. 3B-I). All were significantly increased with the exception of one feature ( $m/z = 306.1376$ ).

### 3.3 Prediction of broccoli intake using a panel of biomarkers

One feature, SMSCO, from NMR analysis and 7 LC-MS features ( $m/z = 237.1087$ ,  $239.1297$ ,  $225.0751$ ,  $227.0911$ ,  $370.2225$ ,  $257.1381$ , and  $255.1238$ ) indicated a significant difference across the three portions and were selected into a panel of biomarkers. Using the training set approach a total of 11 out of 15 observations (73%) revealed good agreement between the predicted and actual intake for the low portion (Table 4). The agreement for medium portion intake and high portion intake were 80% and 67%, respectively (ESI Tables 1 and 2†).



Table 3 LC-MS/MS of compounds of interest

RT	Mass	Prec. <i>m/z</i>	Ion	MS/MS	Hotelling-T2	Putative name	Putative formula
14.2	236.1048	237.1087	[M + H] <sup>+</sup>	134.0961, 133.0645, 197.1283, 84.0802, 43.0173	47.14	Dihydro-beta-tubaic Acid <sup>c,f</sup>	C <sub>13</sub> H <sub>16</sub> O <sub>4</sub>
14.4	238.1215	239.1297	[M + H] <sup>+</sup>	69.0695, 134.0961, 57.0699, 123.0438, 43.0535	44.826	Unidentified <sup>d,g</sup>	C <sub>16</sub> H <sub>14</sub> O <sub>2</sub>
16.4	226.0844	227.0911	[M + H] <sup>+</sup>	167.0703, 121.0645, 168.0736, 122.0678, 85.0283	40.61	Genipin <sup>d,e</sup>	C <sub>11</sub> H <sub>14</sub> O <sub>5</sub>
14.9	305.1300	306.1376	[M + H] <sup>+</sup>	149.096, 122.0267, 167.1059, 185.118, 85.0278	37.144	Unidentified <sup>d,g</sup>	C <sub>14</sub> H <sub>19</sub> N <sub>5</sub> O <sub>5</sub>
16.6	413.2418	414.2519	[M + H] <sup>+</sup>	414.2482, 416.3141, 416.2666, 415.2508, 223.1690	36.066	Unidentified <sup>d,g</sup>	C <sub>20</sub> H <sub>29</sub> N <sub>8</sub> O <sub>2</sub>
15.8	166.0627	167.0692	[M + H] <sup>+</sup>		35.47	3-(4-Hydroxyphenyl) propionic acid <sup>a</sup>	C <sub>9</sub> H <sub>10</sub> O <sub>3</sub>
16.9	369.2152	370.2225	[M + H] <sup>+</sup>	370.2219, 85.0282, 191.1060, 374.252, 372.2363	35.321	Unidentified <sup>d,g</sup>	C <sub>18</sub> H <sub>15</sub> N <sub>8</sub> O
15.3	389.2411	390.2480	[M + H] <sup>+</sup>	113.0611, 67.0546, 287.9454, 43.0173, 114.0670	34.664	Unidentified <sup>d,g</sup>	C <sub>19</sub> H <sub>35</sub> NO <sub>7</sub>
17.5	415.2567	416.2642	[M + H] <sup>+</sup>		34.55	Unidentified <sup>d</sup>	C <sub>21</sub> H <sub>37</sub> NO <sub>7</sub>
11.2	218.1156	219.1206	[M + H] <sup>+</sup>	119.0854, 137.0960, 95.0853, 109.1014, 91.0542	34.416	3-Hydroxy-sebacic acid <sup>b,d,e</sup>	C <sub>10</sub> H <sub>18</sub> O <sub>5</sub>
16.0	256.1311	257.1381	[M + H] <sup>+</sup>		34.142	1-Methoxy-1-(2,4,5-trimethoxyphenyl)-2-propanol <sup>d</sup>	C <sub>13</sub> H <sub>20</sub> O <sub>5</sub>
18.0	391.1852	392.1917	[M + H] <sup>+</sup>		33.728	Oryzastrobins	C <sub>18</sub> H <sub>25</sub> N <sub>5</sub> O <sub>5</sub>
9.5	224.0688	225.0751	[M + H] <sup>+</sup>	55.0537, 69.0692, 147.0444, 57.0686, 119.0487	33.706	Sinapic acid <sup>c,d,e</sup>	C <sub>11</sub> H <sub>12</sub> O <sub>5</sub>
14.2	271.1427	272.1482	[M + H] <sup>+</sup>	55.0539, 60.0806, 69.0342, 69.0688, 83.0845	28.647	Unidentified <sup>d,g</sup>	C <sub>14</sub> H <sub>25</sub> NS <sub>2</sub>
11.8	198.0532	197.0455	[M - H] <sup>-</sup>		50.044	Ethyl gallate <sup>a</sup>	C <sub>9</sub> H <sub>10</sub> O <sub>5</sub>
15.1	198.0892	197.0821	[M - H] <sup>-</sup>		41.914	Guaifenesin <sup>a</sup>	C <sub>10</sub> H <sub>14</sub> O <sub>4</sub>
17.3	430.1481	429.1403	[M - H] <sup>-</sup>		41.559	Unidentified <sup>a</sup>	C <sub>19</sub> H <sub>26</sub> O <sub>11</sub>
14.6	270.1109	269.1033	[M - H] <sup>-</sup>		40.696	Unidentified <sup>d</sup>	C <sub>13</sub> H <sub>18</sub> O <sub>6</sub>
13.9	444.1268	443.1197	[M - H] <sup>-</sup>		34.745	Unidentified <sup>a</sup>	C <sub>19</sub> H <sub>24</sub> O <sub>12</sub>
14.4	256.1310	255.1238	[M - H] <sup>-</sup>		30.969	1-Methoxy-1-(2,4,5-trimethoxyphenyl)-2-propanol <sup>d</sup>	C <sub>13</sub> H <sub>20</sub> O <sub>5</sub>
11.6	216.0998	215.0926	[M - H] <sup>-</sup>		30.047	Unidentified <sup>a</sup>	C <sub>10</sub> H <sub>16</sub> O <sub>5</sub>

RT, retention time; Prec. *m/z*, precision mass to charge ratio; MS/MS, tandem mass spectrometry; putative identification from HMDB after analysis by CFM-ID. LC-MS/MS in positive mode [M + H]<sup>+</sup> and negative mode [M - H]<sup>-</sup>. Analysis shows putative formula and results of interrogation of the METLIN and HMDB databases. <sup>a</sup> METLIN formula database. <sup>b</sup> CFM-ID. <sup>c</sup> METLIN MSMS spectrum match. <sup>d</sup> Authentic standard. <sup>e</sup> Level I identification. <sup>f</sup> Level II identification. <sup>g</sup> Level IV identification. For the molecule features without identification level indicated the formula and identification from single MS library search.

## 4 Discussion

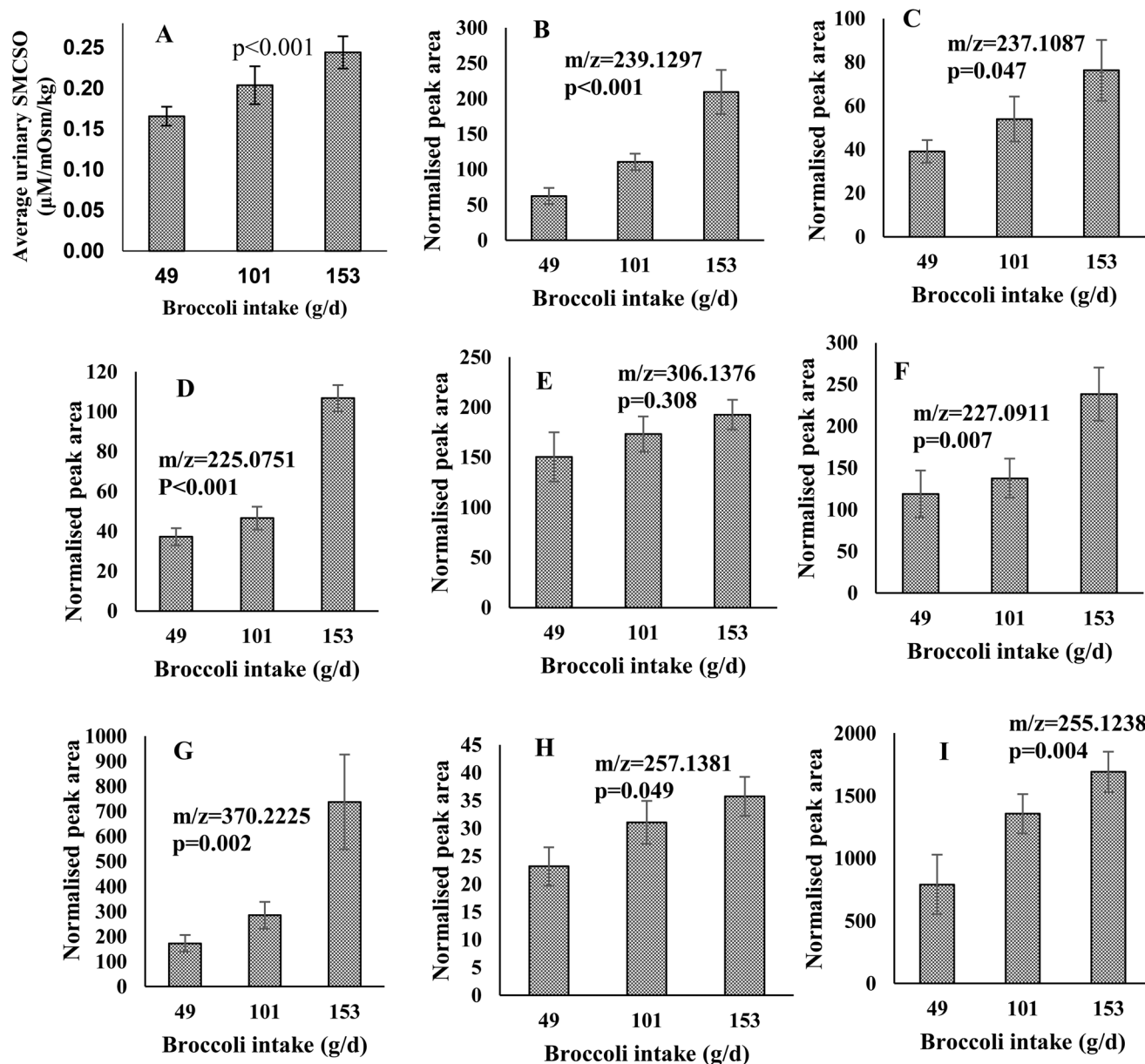
The current study confirmed SMCSO as an NMR-based urinary biomarker of broccoli intake and suggested several potential LC-MS-based biomarkers including genipin, dihydro-β-tubaic acid and sinapic acid. Importantly the work demonstrated a dose response relationship with increased urinary excretion as the amount consumed increased. Furthermore, a panel of biomarkers was capable of estimating broccoli intake using the multiMarker web application. Overall, the present study illustrates the application of metabolomic-based approaches for identifying food intake biomarkers and importantly highlight the potential of such biomarkers in determination of intake.

The panel of biomarkers were combined using the multiMarker model, a latent variable model, which allows prediction of food intake when only biomarker data are available. Using the data from the A-Diet study, the model was successfully applied to estimate the relationship between biomarkers and apple intake and to subsequently determine apple intake from biomarker data.<sup>21</sup> In the current study, the biomarker panel performed excellently to identify low and medium levels

of broccoli intake. For high level of intake the model had difficulties discriminating between 101 and 153 g day<sup>-1</sup>.

This study employed an untargeted metabolomic approach for the discovery of broccoli intake biomarkers identifying discriminating compounds between fasting and 4 hours post-prandial metabolomic profiles. Agreeing with previous literature we identified SMCSO as a biomarker of brassica intake. SMCSO is an organosulfur non-protein amino acid found naturally in brassica.<sup>23</sup> SMCSO, and 3 of its metabolites, were identified in the urine samples of 20 men following 14 day consumption of a high cruciferous vegetable diet (500 g day<sup>-1</sup> of both broccoli and Brussel sprouts).<sup>12</sup> These NMR-identified metabolites were able to differentiate urinary metabolomes after a high cruciferous diet from urinary metabolomes observed after a diet excluding cruciferous vegetables and alliums in the same participants. The present study demonstrated that urinary SMCSO was detectable after a smaller, more achievable intake of cruciferous vegetables, 135 g compared to the 500 g day<sup>-1</sup> used by Edmands *et al.*,<sup>12</sup> which supports the use of SMCSO when examining cruciferous vegetable intake after experimental and habitual intakes. Our study has





**Fig. 3** (A) Dose–response relationship between the fasting urine samples (mean and error bars (SEM) are presented) and amount of broccoli consumed. The X-axis, average broccoli intake (grams) during the Dose–response study; the Y-axis, SMCSO concentrations measured in urine ( $\mu\text{M}$ ) normalised by osmolality (mOsm per kg). Repeated measure ANOVA was performed and SMCSO significantly increased responding to elevated broccoli intake in fasting urine samples ( $p < 0.001$ ). (B–I) Dose–response relationship of the peak areas of fasting urine samples across the different portions of broccoli consumed. Values are means  $\pm$  SEMs. The X-axis, average broccoli intake (grams per day) during the Dose–response study; the Y-axis, normalised peak area measured in urine normalised by osmolality (mOsm per kg). Figure titles indicated the  $m/z$  of features. The feature of  $m/z$  255.1238 were identified in negative mode and the rest of features were identified in positive mode. One way ANOVA was performed to indicate the significant difference across three portions.

added to the research area by confirming that urinary concentration of SMCSO demonstrated a dose–response relationship, increasing as broccoli intake increased.

From the LC-MS analysis multiple potential broccoli biomarkers were identified. The 3-hydroxysebacic acid is a derivative of 3-hydroxydicarboxylic. It is possible that this potential biomarker is reflective of metabolism of the plant polymers ingested in broccoli and warrant further investigation, however

it is unlikely this compound is specific to broccoli intake. Another feature identified by LC-MS was genipin, a geniposide-derived aglycone present in fruit of *Gardenia jasminoides*. Genipin has multiple food industry applications such as a functional food ingredient, a natural blue colorant, and it is used as part of traditional Chinese medicine.<sup>24</sup> Therefore, it is difficult to determine the exact dietary source of genipin identified in urine samples, however it did demonstrate a



**Table 4** The predicted broccoli intake compared to actual broccoli intake by multimarker application (low portion intake)

Observation	Actual intake (g)	Predicted intake (g)	Standard deviation	2.5% percentile	97.5% percentile
1	<b>49</b>	<b>53.46</b>	24.84	30.66	110.96
2	<b>49</b>	<b>48.96</b>	13.68	28.58	93.52
3	49	91.56	26.64	37.42	124.37
4	<b>49</b>	<b>50.00</b>	18.37	29.91	104.23
5	<b>49</b>	<b>47.16</b>	10.27	25.22	65.4
6	<b>49</b>	<b>48.39</b>	11.06	28.81	73.02
7	<b>49</b>	100.73	15.75	59.14	136.24
8	<b>49</b>	<b>47.47</b>	9.65	25.45	64.85
9	<b>49</b>	<b>50.18</b>	14.27	31.81	98.83
10	<b>49</b>	<b>49.59</b>	11.26	32.19	49.59
11	<b>49</b>	100.39	18.15	49.32	100.39
12	49	94.99	22.01	42.12	94.99
13	<b>49</b>	<b>48.16</b>	10.46	28.57	68.79
14	<b>49</b>	<b>48.63</b>	10.26	29.42	69.18
15	<b>49</b>	<b>48.50</b>	11.22	29.22	75.83

The good agreement between actual and predicted intake was highlighted in bold.

dose–response relationship with broccoli intake so future investigation is necessary to confirm this compound as a broccoli intake biomarker. LC-MS analysis also identified some flavonoid derivatives. Fruit and vegetables are natural sources of flavonoids.<sup>25</sup> As flavonoid derivatives are commonly found in multiple fruit and vegetables, individually, they will not be specific broccoli biomarkers. However, the combination of these compounds into a multi-biomarker panel brings important information.

Even though the majority of previous research has focused on isothiocyanates and their association with brassica consumption, there are a number of limitations associated with verifying these organosulfur compounds as robust markers of intake.<sup>12</sup> Glucosinolates are consumed intact but the bioavailability of the isothiocyanates is affected by food processing techniques, pH of stomach and intestines and microbial digestion.<sup>26,27</sup> As isothiocyanates are highly reactive compounds they bind easily with macromolecules reducing bioavailability. This rapid interaction leads to a huge diversity of isothiocyanates differing structurally based on their side chain.<sup>28</sup> This diversity can affect metabolism and lead to increased concentrations of different conjugated metabolic products excreted as well as the mercapturic acid of parent isothiocyanate. Many studies have limited their analyses to derivatives of specific mercapturates based on the structure of parent glucosinolate and isothiocyanate structures which only count for a proportion of metabolites of original isothiocyanate dose.<sup>28</sup> Hence, there is a need to perform mechanistic studies to identify metabolites of structurally related compounds metabolised through separate pathways to get a full overview of potential cruciferous exposure markers.

A potential limitation of the research is that the discovery and dose–response studies were carried out only in the Republic of Ireland and therefore confirmation of these findings would be required in other ethnic populations. This study has multiple strengths associated: a range of different study designs were used to examine the biomarkers in different scenarios. Confirmation of a dose–response in the background of habitual

diet is a noteworthy strength. To the best of our knowledge this is the first study to examine the dose–response relationship of urinary biomarkers across multiple levels of broccoli intake. Furthermore, the predictive ability of the panel of biomarkers was examined. Further research is needed into biomarkers of broccoli-intake in order to establish their qualitative or quantitative applications for higher quantities of intake. Further work is also needed to fully validate the biomarkers identified to meet the validation criteria previously outlined.<sup>29</sup>

The present study has illustrated the successful implementation of an untargeted metabolomics approach in search of dietary biomarkers of broccoli intake. Urinary SMCSO was identified as a potential intake biomarker of broccoli in NMR analysis and exhibited a dose–response relationship with increasing broccoli intake. Several LC-MS features, a number of which are flavonoid derivatives were identified and dose–response relationships revealed. For both NMR and LC-MS analysis, the panel of molecule features as putative broccoli biomarkers indicate their potential to predict the intake and pave the way forward for its application in supporting and improving objective measurement of broccoli intake. Future work should examine the use of the panel of biomarkers in a range of ethnic groups and in larger observational studies.

## Author contributions

A. E. M. conducted the discovery and dose–response studies, analyzed data, and prepared the manuscript. X. Y. acquired and analysed data and prepared the manuscript. C. C. acquired part of the metabolomics data. L. B. designed research, analysed data and prepared the manuscript. All authors read and accepted the final version of the manuscript.

## Conflicts of interest

The authors have no conflict of interest.



## Acknowledgements

This work was supported by a H2020 European Research Council (647783). The authors would like to thank all volunteers for their commitment and patience during the study.

## References

- 1 M. E. Cartea, M. Francisco, P. Soengas and P. Velasco, Phenolic compounds in Brassica vegetables, *Molecules*, 2011, **16**, 251–280.
- 2 B. Kusznierevicz, A. Bartoszek, L. Wolska, J. Drzewiecki, S. Gorinstein and J. Namieśnik, Partial characterization of white cabbages (*Brassica oleracea* var. *capitata* f. *alba*) from different regions by glucosinolates, bioactive compounds, total antioxidant activities and proteins, *Food Sci. Technol.*, 2008, **41**, 1–9.
- 3 C. Sturm and A. E. Wagner, Brassica-derived plant bioactives as modulators of chemopreventive and inflammatory signaling pathways, *Int. J. Mol. Sci.*, 2017, **18**, 1890.
- 4 D. Aune, E. Giovannucci, P. Boffetta, L. T. Fadnes, N. Keum, T. Norat, D. C. Greenwood, E. Riboli, L. J. Vatten and S. Tonstad, Fruit and vegetable intake and the risk of cardiovascular disease, total cancer and all-cause mortality—a systematic review and dose-response meta-analysis of prospective studies, *Int. J. Epidemiol.*, 2017, **46**, 1029–1056.
- 5 V. Arija, R. Abellana, B. Ribot and J. M. Ramón, Biases and adjustments in nutritional assessments from dietary questionnaires, *Nutr. Hosp.*, 2015, **31**, 113–118.
- 6 N. V. Dhurandhar, D. Schoeller, A. W. Brown, S. B. Heymsfield, D. Thomas, T. I. Sørensen, J. R. Speakman, M. Jeansonne, D. B. Allison and E. B. M. W. Group, Energy balance measurement: when something is not better than nothing, *Int. J. Obes.*, 2015, **39**, 1109–1113.
- 7 V. Kipnis, D. Midthune, L. Freedman, S. Bingham, N. E. Day, E. Riboli, P. Ferrari and R. J. Carroll, Bias in dietary-report instruments and its implications for nutritional epidemiology, *Public Health Nutr.*, 2002, **5**, 915–923.
- 8 J. S. Shim, K. Oh and H. C. Kim, Dietary assessment methods in epidemiologic studies, *Epidemiol. Health*, 2014, **36**, e2014009.
- 9 Q. Gao, G. Praticò, A. Scalbert, G. Vergères, M. Kolehmainen, C. Manach, L. Brennan, L. A. Afman, D. S. Wishart, C. Andres-Lacueva, M. Garcia-Aloy, H. Verhagen, E. J. M. Feskens and L. O. Dragsted, A scheme for a flexible classification of dietary and health biomarkers, *Genes Nutr.*, 2017, **12**, 1–15.
- 10 M. B. Andersen, M. Kristensen, C. Manach, E. Pujos-Guillot, S. K. Poulsen, T. M. Larsen, A. Astrup and L. Dragsted, Discovery and validation of urinary exposure markers for different plant foods by untargeted metabolomics, *Anal. Bioanal. Chem.*, 2014, **406**, 1829–1844.
- 11 M. B. S. Andersen, H. C. Reinbach, A. Rinnan, T. Barri, C. Mithril and L. O. Dragsted, Discovery of exposure markers in urine for Brassica-containing meals served with different protein sources by UPLC-qTOF-MS untargeted metabolomics, *Metabolomics*, 2013, **9**, 984–997.
- 12 W. M. Edmands, O. P. Beckonert, C. Stella, A. Campbell, B. G. Lake, J. C. Lindon, E. Holmes and N. J. Gooderham, Identification of human urinary biomarkers of cruciferous vegetable consumption by metabolomic profiling, *J. Proteome Res.*, 2011, **10**, 4513–4521.
- 13 A. J. Lloyd, G. Fave, M. Beckmann, W. Lin, K. Tailliant, L. Xie, J. C. Mathers and J. Draper, Use of mass spectrometry fingerprinting to identify urinary metabolites after consumption of specific foods, *Am. J. Clin. Nutr.*, 2011, **94**, 981–991.
- 14 D. H. May, S. L. Navarro, I. Ruczinski, J. Hogan, Y. Ogata, Y. Schwarz, L. Levy, T. Holzman, M. W. McIntosh and J. W. Lampe, Metabolomic profiling of urine: response to a randomised, controlled feeding study of select fruits and vegetables, and application to an observational study, *Br. J. Nutr.*, 2013, **110**, 1760–1770.
- 15 M. M. Ulaszewska, K. Trost, J. Stanstrup, K. M. Tuohy, P. Franceschi, M. F.-F. Chong, T. George, A. M. Miniñane, J. A. Lovegrove and F. Mattivi, Urinary metabolomic profiling to identify biomarkers of a flavonoid-rich and flavonoid-poor fruits and vegetables diet in adults: the FLAVURS trial, *Metabolomics*, 2016, **12**, 32.
- 16 H. Kim, S. Yoo, J.-D. Lee, H.-Y. Kim, S. Kim and K.-B. Kim, A Metabolomics approach to sulforaphane efficacy in secondhand smoking-induced pulmonary damage in mice, *Metabolites*, 2022, **12**, 518.
- 17 A. E. McNamara, C. Collins, P. S. C. S. Harsha, D. González-Peña, H. Gibbons, B. A. McNulty, A. P. Nugent, J. Walton, A. Flynn and L. Brennan, Metabolomic-based approach to identify biomarkers of apple intake, *Mol. Nutr. Food Res.*, 2020, **64**, e1901158.
- 18 P. S. C. S. Harsha, R. Abdul Wahab, C. Cuparencu, L. Dragsted and L. Brennan, A metabolomics approach to the identification of urinary biomarkers of pea intake, *Nutrients*, 2018, **10**, 1911.
- 19 Y. Djoumbou-Feunang, A. Pon, N. Karu, J. Zheng, C. Li, D. Arndt, M. Gautam, F. Allen and D. S. Wishart, CFM-ID 3.0: significantly improved ESI-MS/MS prediction and compound identification, *Metabolites*, 2019, **9**, 72.
- 20 D. S. Wishart, Y. D. Feunang, A. Marcu, A. C. Guo, K. Liang, R. Vázquez-Fresno, T. Sajed, D. Johnson, C. Li, N. Karu, Z. Sayeeda, E. Lo, N. Assempour, M. Berjanskii, S. Singhal, D. Arndt, Y. Liang, H. Badran, J. Grant, A. Serra-Cayuela, Y. Liu, R. Mandal, V. Neveu, A. Pon, C. Knox, M. Wilson, C. Manach and A. Scalbert, HMDB 4.0: the human metabolome database for 2018, *Nucleic Acids Res.*, 2018, **46**, D608–D617.
- 21 L. W. Sumner, A. Amberg, D. Barrett, M. H. Beale, R. Beger, C. A. Daykin, T. W.-M. Fan, O. Fiehn, R. Goodacre and J. L. Griffin, Proposed minimum reporting standards for chemical analysis, *Metabolomics*, 2007, **3**, 211–221.
- 22 S. D'Angelo, L. Brennan and I. C. Gormley, Inferring food intake from multiple biomarkers using a latent variable model, *Ann. Appl. Stat.*, 2021, **15**, 2043–2060.



- 23 H. S. Marks, J. A. Hilson, H. C. Leichtweis and G. S. Stoewsand, S-methylcysteine sulfoxide in Brassica vegetables and formation of methyl methanethiosulfinate from brussels sprouts, *J. Agric. Food Chem.*, 1992, **40**, 2098–2101.
- 24 A. M. Ramos-De-La-Pena, C. M. Renard, J. Montañez, M. de la Luz Reyes-Vega and J. C. Contreras-Esquivel, A review through recovery, purification and identification of genipin, *Phytochem. Rev.*, 2016, **15**, 37–49.
- 25 S. Kumar and A. K. Pandey, Chemistry and biological activities of flavonoids: an overview, *Sci. World J.*, 2013, **2013**, 162750.
- 26 B. Holst and G. Williamson, A critical review of the bio-availability of glucosinolates and related compounds, *Nat. Prod. Rep.*, 2004, **21**, 425–447.
- 27 M. Vermeulen, R. van den Berg, A. P. Freidig, P. J. van Bladeren and W. H. Vaes, Association between consumption of cruciferous vegetables and condiments and excretion in urine of isothiocyanate mercapturic acids, *J. Agric. Food Chem.*, 2006, **54**, 5350–5358.
- 28 B. Holst and G. Williamson, A critical review of the bio-availability of glucosinolates and related compounds, *Nat. Prod. Rep.*, 2004, **21**, 425–447.
- 29 L. O. Dragsted, Q. Gao, A. Scalbert, G. Vergères, M. Kolehmainen, C. Manach, L. Brennan, L. A. Afman, D. S. Wishart, C. Andres-Lacueva, M. Garcia-Aloy, H. Verhagen, E. J. M. Feskens and G. Praticò, Validation of biomarkers of food intake—critical assessment of candidate biomarkers, *Genes Nutr.*, 2018, **13**, 14.

