

## Metabolomic profiling of urine: response to a randomised, controlled feeding study of select fruits and vegetables, and application to an observational study

Damon H. May<sup>1</sup>†, Sandi L. Navarro<sup>1</sup>†, Ingo Ruczinski<sup>2</sup>, Jason Hogan<sup>3</sup>, Yuko Ogata<sup>3</sup>, Yvonne Schwarz<sup>1</sup>, Lisa Levy<sup>1</sup>, Ted Holzman<sup>1</sup>, Martin W. McIntosh<sup>1</sup> and Johanna W. Lampe<sup>1</sup>\*

<sup>1</sup>Fred Hutchinson Cancer Research Center, Division of Public Health Sciences, 1100 Fairview Avenue North, M4-B402 Seattle, WA 98109, USA

<sup>2</sup>Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA

<sup>3</sup>Fred Hutchinson Cancer Research Center, Proteomics Facility, Seattle, WA, USA

(Submitted 3 October 2012 – Final revision received 1 February 2013 – Accepted 21 March 2013 – First published online 9 May 2013)

### Abstract

Metabolomic profiles were used to characterise the effects of consuming a high-phytochemical diet compared with a diet devoid of fruits and vegetables (F&V) in a randomised trial and cross-sectional study. In the trial, 8 h fasting urine from healthy men (*n* 5) and women (*n* 5) was collected after a 2-week randomised, controlled trial of two diet periods: a diet rich in cruciferous vegetables, citrus and soya (F&V), and a fruit- and vegetable-free (basal) diet. Among the ions found to differentiate the diets, 176 were putatively annotated with compound identifications, with forty-six supported by MS/MS fragment evidence. Metabolites more abundant in the F&V diet included markers of the dietary intervention (e.g. crucifers, citrus and soya), fatty acids and niacin metabolites. Ions more abundant in the basal diet included riboflavin, several acylcarnitines and amino acid metabolites. In the cross-sectional study, we compared the participants based on the tertiles of crucifers, citrus and soya from 3 d food records (*n* 36) and FFQ (*n* 57); intake was separately divided into the tertiles of total fruit and vegetable intake for FFQ. As a group, ions individually differential between the experimental diets differentiated the observational study participants. However, only four ions were significant individually, differentiating the third *v.* first tertile of crucifer, citrus and soya intake based on 3 d food records. One of these ions was putatively annotated: proline betaine, a marker of citrus consumption. There were no ions significantly distinguishing tertiles by FFQ. The metabolomic assessment of controlled dietary interventions provides a more accurate and stronger characterisation of the diet than observational data.

**Key words:** Metabolomics: Citrus fruit: Soya: Cruciferous vegetables: Dietary interventions: FFQ: Mass spectrometry

A higher consumption of fruits and vegetables is associated with a reduced risk of several chronic diseases, including CVD and certain cancers. This association is attributed, in part, to the phytochemical content of plant foods (e.g. isoflavones in soya, flavonols in citrus fruits and isothiocyanates in cruciferous vegetables)<sup>(1)</sup>. By necessity, the majority of dietary intervention studies have been restricted to the evaluation of known or hypothesised pathways involved in the disease process being studied and rely on a limited number of biomarkers and disease-related outcomes. Metabolomic profiling, measuring large numbers of metabolites not selected *a priori*, is an alternative approach to characterising dietary response<sup>(2–4)</sup>. Untargeted metabolomic profiling of urine by

liquid chromatography (LC)–tandem MS (LC–MS/MS) can be used to assess differences in a wide variety of chemical species to describe detailed biochemical responses of cellular systems in response to dietary exposures<sup>(5–7)</sup>. In addition to proposed pathways, these profiles, in which a broad range of metabolites in a biosample are measured, allow for the discovery of previously unrecognised mechanisms through which dietary factors affect disease risk. Application of metabolomics in the context of controlled dietary studies can also be used to identify new biomarkers for the intake of specific foods (e.g. cruciferous vegetables), which can then be used for the evaluation of dietary intake and the association with disease risk in an observational setting.

**Abbreviations:** 3DFR, 3 d food records; F&V, fruits and vegetables; HMDB, Human Metabolome Database; LC, liquid chromatography; PC, principal component.

\* **Corresponding author:** Dr J. W. Lampe, fax +1 206 667 7850, email jlampe@fhcrc.org

† Both authors contributed equally to this work.

The aims of the present study were twofold. First, we intended to characterise intervention-induced changes in untargeted urinary metabolomic patterns in response to a 2-week controlled high-phytochemical diet including specific fruits and vegetables compared with a fruit- and vegetable-free basal diet, and to identify metabolites that differed between the two interventions. Second, we compared the metabolites affected by the dietary intervention with the metabolites observed in high- and low-fruit and vegetable diets based on 3 d food records (3DFR) and FFQ in a cross-sectional, observational study of free-living individuals. All raw LC-MS/MS data, datasets derived from the present analyses and computational tools for creating metabolomic profiles using the msInspect software suite (Fred Hutchinson Cancer Research Center)<sup>(8)</sup> are provided (for access instructions, see the Supplementary material, available online) to allow the research community to further mine the data from the present experiments.

**Methods**

*Study design*

Biological samples and participant information were from a completed study, ‘Dietary Influences on Glucuronidation Study’. The Dietary Influences on Glucuronidation Study was a two-tiered study comprising a cross-sectional study from which subjects were recruited for participation in a randomised, controlled cross-over feeding trial comparing 2 weeks of a diet rich in fruits and vegetables (F&V) with 2 weeks of a fruit- and vegetable-free (basal) diet<sup>(9,10)</sup>. For the present analysis, 8 h urine samples from both the

cross-sectional and the feeding intervention studies were used. The Institutional Review Board at the Fred Hutchinson Cancer Research Center, Seattle, Washington, approved the study, and all participants gave informed written consent. The study design is summarised in Fig. 1.

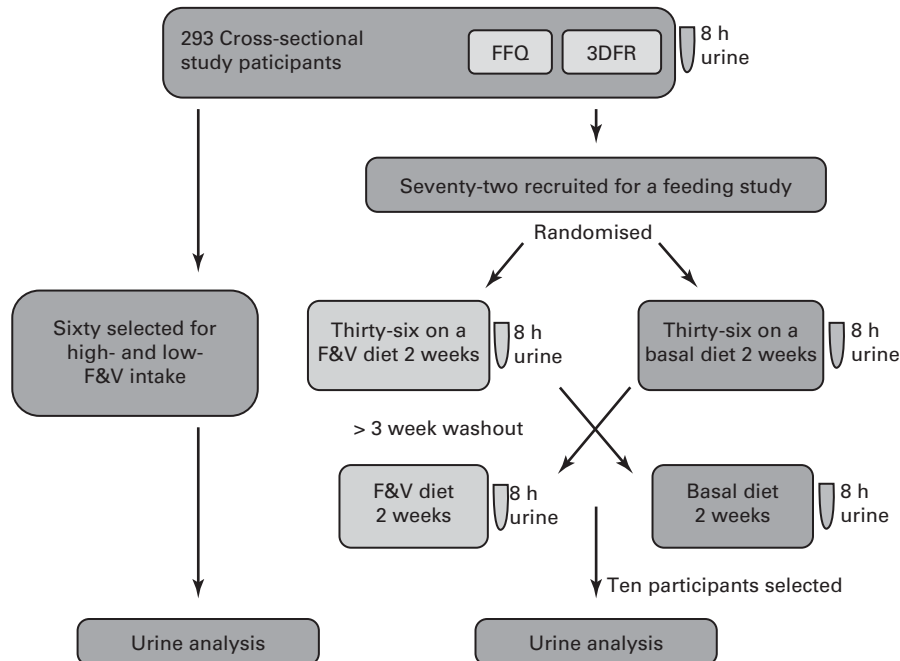
*Participants*

In the cross-sectional study, healthy, non-smoking men and women aged 21–45 years were recruited from the greater Seattle area as described previously<sup>(10)</sup>. Briefly, exclusion criteria included chronic disease, current use of over-the-counter or prescription drugs or alcohol intake >2 drinks/d. Participants were asked to discontinue use of all dietary supplements 1 week before the start of the study. For the present analysis, urine samples for sixty of the 293 individuals in the cross-sectional study (thirty from the highest tertile of total fruit and vegetable intake and thirty from the lowest tertile, based on FFQ data) were selected for LC-MS/MS analysis.

A subset of individuals from the cross-sectional study were contacted and invited to participate in the feeding study on the basis of their UDP-glucuronosyl transferase genotype in relation to the aims of the parent study<sup>(10)</sup>. In all, seventy-two participants were randomised. Day 11 urine samples from ten individuals in the feeding study (five men and five women, chosen at random) were used to characterise differential metabolite abundances between the basal and F&V diets (Fig. 1).

*Study diets*

Participants consumed two different diets: a basal, low-phytochemical diet, devoid of fruit and vegetables, and the



**Fig. 1.** Study design and general analysis workflow. 8 h urine, FFQ and 3 d food records (3DFR) were collected from 293 cross-sectional study participants. Of these, sixty were selected for analysis based on the high- and low-fruit and vegetable (F&V) intake. A total of seventy-two were recruited for a feeding study; half of these were randomised to a F&V intervention diet and half to a basal (fruit- and vegetable-free) diet for 2 weeks and 8 h urine collected. After a 3-week washout, participants switched the diets and a second 8 h urine was collected. Samples from ten of these participants were used in the present analysis.

basal diet supplemented with cruciferous vegetables (broccoli, cabbage and daikon radish sprouts), soya foods (soya milk, soya cheese slices, tofu and roasted soya nuts) and citrus fruits (grapefruit and orange juices, orange/grapefruit segments and dried orange peel), dosed on a g/kg body-weight basis to minimise confounding by body weight between the sexes. Weight was checked daily and energy adjusted so that participants remained weight stable throughout the study. Each diet was consumed for 14 d with at least a 3-week wash-out period between the diets. Participants were instructed to consume only the food and beverages provided to them during both diet periods, maintain their usual physical activity and not to use any type of medication, vitamins or other dietary supplements. Based on the 24 h urinary analysis of total isothiocyanates and isoflavone excretion and daily food check-off forms, participant compliance to the study was excellent, with the consumption of non-study food items being less than 1% of the study design<sup>(9)</sup>. Additional diet details have been published elsewhere<sup>(9)</sup>.

### Dietary assessment

**FFQ.** Participants in the cross-sectional study completed a FFQ reporting their dietary intake within the past 3 months. FFQ details have been described previously<sup>(1,10)</sup>. Total fruit and vegetable intake was used for tertile summation. Line items in each botanical category used for crucifer, citrus and soya tertile summation included: broccoli, cauliflower, cabbage, Brussels sprouts and coleslaw, for cruciferous vegetables; oranges and orange juice, grapefruit and grapefruit juice, and tangerines for citrus; and tofu, tempeh, products such as soy hot dogs and burgers, soya cheese and miso soup for soya. FFQ with incomplete information or data that suggested biologically implausible daily energy intakes (<2510 kJ (600 kcal) or >16736 kJ (4000 kcal) for women or <3347 kJ (800 kcal) or >20920 kJ (5000 kcal) for men) were excluded (a modification from Willett<sup>(11)</sup>), providing a total of fifty-seven samples for the analysis of total fruit and vegetable intake, and forty-eight for combined crucifer, citrus and soya intake (fewer samples were available for analysis by botanical families, as samples were chosen for LC-MS/MS analysis based on total fruit and vegetable intake).

**Food records.** Participants also completed 3DFR on three consecutive days after receiving training by a registered dietitian<sup>(10)</sup>. The 3DFR were analysed to estimate intake by botanical families based on standard serving sizes<sup>(12)</sup>. Botanical families used for tertile summation included Cruciferae (i.e. cruciferous vegetables), Rutaceae (i.e. citrus fruits) and Leguminosae (i.e. soya, beans and pulses). Although this last category contained beans and pulses in addition to soya, intake from soya alone tracked similarly (data not shown). 3DFR completed >10 d from the time of urine collection were excluded, providing a total of thirty-six participants in the 3DFR analyses.

### Urine collection

For the present analyses, 8 h fasting urine collected overnight after a standard aspirin dose (650 mg) was used<sup>(4)</sup>.

The protocol was the same across all treatments (both diet periods in the intervention study and the cross-sectional study). Urine was collected within an average of 5.5 d (range 0–48 d) of completing the FFQ and 3.3 d (range 0–10 d) of completing the 3DFR for the cross-sectional study, and on day 11 of each 14 d controlled diet period for the feeding study. On the evening of the aspirin challenge, participants were instructed to void and consume aspirin before retiring, and collect all of their urine for the next 8 h (in most cases, this was a single void after waking). Urine specimens were stored at 4°C until delivery to the Fred Hutchinson Cancer Research Center in the morning after collection. The total volume and initial pH of urine samples were recorded; samples were aliquoted and stored at –80°C.

### Urine sample preparation for MS/MS analysis

For each urine sample, 100 µl acetonitrile was added to an equal volume of urine and stored overnight at –20°C. After centrifugation for 10 min at 16 000 g in an Eppendorf 5415D table-top centrifuge, 10 µl of 1% trifluoroacetic acid were added to the supernatant and dried in a speed-vacuum centrifuge. The sample was dissolved in 25 µl of 2% acetonitrile with 0.1% formic acid; 5 µl were used for each analysis.

### LC-MS/MS sample analysis

LC-MS/MS analysis was performed using a linear trap quadrupole Fourier transform mass spectrometer (Thermo Fisher Scientific) with an electrospray ion source. The LC system was a NanoLC 2D (Eksigent) connected to an Inertsil<sup>®</sup> ODS-SP reverse-phase column (0.5 × 150 mm, 3 µm particle size; GL Sciences, Inc.). The urine samples were loaded onto the column, and chromatographic separation was performed using a two-mobile-phase solvent system consisting of 0.1% formic acid in water (A) and 0.1% formic acid in acetonitrile (B) over a 42 min gradient from 5 to 95% solvent B at 8 µl/min. The mass spectrometer operated in a positive-ion mode (chosen based on success in prior urine experiments), in a data-dependent MS/MS mode over the *m/z* range 100–650. For each cycle, the three most abundant ions from each MS scan were selected for MS/MS analysis using 35% normalised collision energy. Resolution and the automatic gain control target were 100 000 and 2 000 000, respectively. The selected ions were dynamically excluded for 30 s. Though LC-MS and MS/MS scans were acquired for the entire run time, including column washing and equilibration, all data from the high-organic phase (after 40 min) were excluded from the analysis. Samples were run in random order in duplicate, with a control every eight samples.

### LC-MS/MS data extraction and analytic dataset creation

Complete details of all data analysis steps, data, software and instructions necessary to recreate the analytic datasets described below are provided in the Supplementary material (available online). In brief, msInspect<sup>(8)</sup>, a freely available suite of software and algorithms for analysing profiles from high-resolution LC-MS/MS data, was extended to accurately



detect and quantify small-molecule ion features. These improvements, which include optimised handling of relative peak abundances, among others, are now released as part of the open-source distribution and provided in conjunction with this publication. Ion features were extracted from each LC–MS/MS run, and a mass filter was applied to eliminate ions with fractional masses, indicating that they were unlikely to represent organic molecules<sup>(13)</sup>. Duplicate sample runs were combined by aligning ions by  $m/z$  and retention time<sup>(8,14)</sup>, and the union of features was retained. Ion intensities were quantile normalised<sup>(8,15)</sup> and then aligned by  $m/z$  and retention time to create an analytic dataset (an ion array) whose columns represented samples and whose rows represented ions; values were ion intensities (averaged across duplicate runs where applicable) or missing values for ions not detected in a particular sample.

To evaluate the overall effect of diet on the metabolite profile, we used principal components (PC) analysis on the  $\ln$  intensities of features observed in at least three participants. The PC analysis was applied without regard to the pairing of samples; this provides a conservative approach to evaluating the global effect of diet, as the diet effect must be strong enough to overcome the similarity of the two profiles from each individual. Downstream analyses were conducted on  $\ln$  ratios (F&V:basal) for each participant. Missing data are characteristic of LC–MS/MS profiling data<sup>(8,15)</sup>, and missing values in a sample can result from either an ion being low abundance (i.e. falling below the detection limit of the instrument) or high abundance but missing owing to analytic errors. Under the assumption of missingness due to falling below the level of detection, the following statistically conservative procedures were used to calculate  $\ln$  ratios when an ion was observed in only one of the diets from an individual: a pseudo-background level was estimated separately for each sample run as the first percentile of ion intensities in the run, and  $\ln$  pseudo-ratios were calculated using this background for missing values. Because these imputed  $\ln$  pseudo-ratios naturally contained a disproportionate number of large absolute values (i.e. outliers), we reduced their influence on the analysis by shrinking their quantiles towards the quantiles of the ratios calculated from fully observed ions. This strategy was based on an assumption that true ratios from ions with one missing value were not systematically different from ratios from ions with no missing values. A one-sample  $t$  test applied to the  $\ln$  ratios was used to calculate  $P$  values based on the null hypothesis that mean  $\ln$  ratio=0, and false discovery rates ( $q$  values) were derived<sup>(16)</sup> for all ions having four or more ratios across the samples. Analyses were carried out using the R statistical environment (R Foundation for Statistical Computing)<sup>(17)</sup>.

To investigate the cross-sectional study data, a separate ion array was created using the method described previously. Analyses were performed separately to evaluate combined crucifer, citrus and soya consumption defined by 3DFR and FFQ, and also, separately, on combined fruit and vegetable consumption defined by FFQ, by comparing the participants in the first and third tertiles of consumption. For each ion observed in at least five samples in both tertiles,  $\ln$  intensities

were compared, adjusting for participants' BMI using the Limma R package (Bioconductor) for differential expression analysis<sup>(18)</sup>, and  $P$  and  $q$  values were calculated.

### Metabolite annotation

As others have noted<sup>(13,19)</sup>, identifying metabolites in the untargeted profiling of biosamples without reference standards is a difficult process. We took a multi-step approach, assigning putative annotations to ions by mass-matching, and then, for ions of interest that separated the two groups under investigation, supporting those annotations with MS/MS evidence where possible. For each compound in the Human Metabolome Database (HMDB) version 2.5<sup>(19)</sup>, a separate theoretical mass value was calculated for four ion types: the  $[M + H]^+$  and  $[M + Na]^+$  ions and the dimeric  $[2M + H]^+$  and  $[2M + Na]^+$  ions. Each row of every ion array was matched by the mean mass value to these ion masses; all matches within 5 parts per million were retained. For each ion of interest ( $q < 0.1$ ), further evidence of molecular identity was garnered by examining MS/MS scans. In each sample in which the ion was observed, the most abundant fragment ions from all nearby MS/MS scans were combined into a single list of the most abundant fragments (see the Supplementary material, available online). For ions annotated as dimeric compounds, abundant fragment ions matching the  $[M + H]^+$  or  $[M + Na]^+$  monomeric ions, appropriately, provide a small amount of additional support. For monomers, the fragment list was compared with the available literature documenting the fragmentation of the same compounds by MS/MS in a positive-ion mode (for references see Tables S1 and S2, available online). Ion annotations both with and without this MS/MS support are 'putatively annotated compounds', as described by the Metabolomics Standards Initiative Chemical Analysis Working Group<sup>(20)</sup>. Metabolite functions and pathway information were determined by searching the HMDB, Kyoto Encyclopedia of Genes and Genomes, PubChem and PubMed.

### Results

Participant and dietary characteristics for the feeding and cross-sectional studies are given in Table 1. There were no significant differences in participant characteristics between the tertiles of crucifer, citrus and soya intake; however, the overall energy intake was higher among individuals in the highest tertile of intake in the cross-sectional study. On average, 3763 ions (range 2378–5065) passing the filtering criteria were located in each of the 160 LC–MS/MS runs of the eighty samples run (twenty for the feeding study and sixty for the cross-sectional study). After the combination of ions across the duplicate runs, an average of 4790 ions (range 3155–6575) were detected per sample. The median CV of  $\ln$  ion intensities between the duplicate runs was 0.104 (range 0.082–0.195).

### Feeding study

**Data summary.** A total of 7382 ions were observed in at least three samples across both diets. To analyse global

**Table 1.** Characteristics of the randomised, cross-over, dietary intervention and cross-sectional study participants and diet components (Mean values and standard deviations)

	Feeding study (basal, <i>n</i> 10)		Feeding study (F&V, <i>n</i> 10)		Tertile 1 (3DFR, <i>n</i> 18)†		Tertile 3 (3DFR, <i>n</i> 18)†		<i>P</i> * (3DFR)	Tertile 1 (FFQ, <i>n</i> 22)‡		Tertile 3 (FFQ, <i>n</i> 26)‡		<i>P</i> * (FFQ)
	Mean	SD	Mean	SD	Mean	SD	Mean	SD		Mean	SD	Mean	SD	
Age (years)	29	4.9			26	5.1	27	5.8	0.47	27	5.9	29	5.9	0.30
Height (cm)	173	6.7			173	10.4	172	10.9	0.34	172	10.4	174	10.4	0.99
Weight (kg)	70	12.4			71	12.9	66	12.7	0.29	71	13.9	69	12.5	0.67
BMI (kg/m <sup>2</sup> )	24	3.6			24	2.9	22	3.4	0.08	24	2.9	23	3.0	0.21
Female (%)			50		44		61		0.50	41		46		0.94
Ethnicity (%)														
Caucasian			60		61		56		1.00	68		65		1.00
Asian			40		17		39		0.26	18		31		0.50
Other/unknown			0		22		6		0.34	14		4		0.52
Energy (kcal/d)	2482	567	2481	663	1986	529	2387	716	0.05	1482	614	2154	612	< 0.01
Energy (kJ/d)	10 385	2372	10 381	2774	8309	2213	9987	2996	0.05	6201	2569	9012	2561	< 0.01
F&V (servings/d)		0	9	1.3	0	0.7	3	1.8	< 0.01	0	0.1	3	1.5	< 0.01
Protein (% energy)§	15	3.5	17	3.7	15	3.3	15	3.6	0.58	16	4.0	16	1.8	0.83
Fat (total; % energy)	30	6.7	28	7.8	33	9.9	32	7.6	0.96	34	9.2	30	5.6	0.13
SFA (% energy)	14	3.2	9	3.3	11	3.7	10	3.9	0.26	11	3.4	9	2.7	0.15
MUFA (% energy)	10	2.4	8	2.7	13	5.6	13	2.9	0.54	12	3.7	11	2.2	0.08
PUFA (% energy)	3	0.8	7	1.2	7	2.0	8	2.8	0.24	8	2.6	7	1.5	0.74
Carbohydrate (% energy)	55	12.8	57	15.4	52	11.3	56	6.8	0.52	50	10.3	56	6.4	0.02
Fibre (g/d)	10	3.0	29	5.1	13	3.4	32	14.1	< 0.01	13	8.7	29	10.8	< 0.01

F&V, fruits and vegetables; 3DFR, 3 d food records.

\* *P* value using Mann–Whitney tests for significant differences between the first and third tertiles in the cross-sectional study based on 3DFR and FFQ; sex and ethnicity were tested with a proportion test.

† First and third tertiles of crucifer, citrus and soya intake among the cross-sectional study participants, as calculated based on 3DFR.

‡ First and third tertiles of crucifer, citrus and soya intake among the cross-sectional study participants, as calculated based on FFQ.

§ Percentages may not add up to 100 due to rounding.

differences in ion abundance between the diets (without the background-level estimation of unobserved ions), we used the PC analysis on these ions. The  $\ln$  ion intensity variation in PC dimensions 1 and 2 explained 26% of the variance between the samples. PC1 was not associated with the diet, but PC2 systematically separated the diets with high accuracy:  $PC2 > 0$  correctly classifies nine of ten samples from each diet (Fig. 2). Experimental measurements fell into two distinct groups along PC1; interestingly, pairs of measurements from the same individual cluster together, with both members of each pair always having  $PC1 < 0$  or  $PC1 > 0$ . The meaning of these clusters has not been identified; they are not associated with age, diet order, sex, BMI or other known factors. Analysed by a  $t$  test on an ion-by-ion level, a total of 2857 ions were determined to be significantly higher or lower in the F&V *v.* basal diet with  $q < 0.1$  (1360 more abundant in the F&V diet; 1497 in the basal diet).

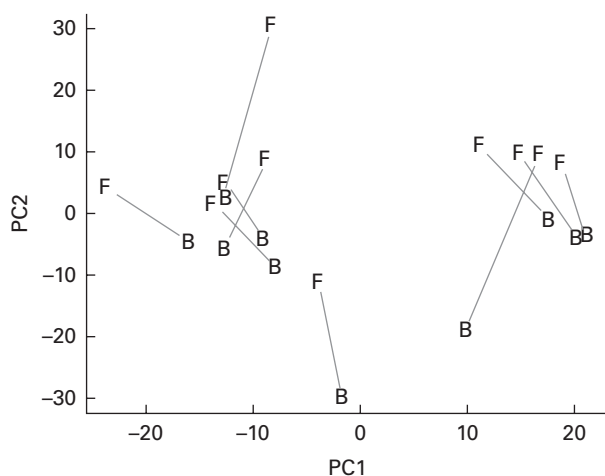
**Annotated compounds.** A total of 3666 ions observed in one or more samples were assigned putative annotations by mass-matching to the HMDB. Of these mass-matched ions, 423 were among the above-mentioned 2857 ions significantly differentially abundant in the F&V *v.* basal diet with  $q < 0.1$  (179 more abundant in the F&V diet; 244 in the basal diet). Of these, 195 had at least one match to a dimeric ion of a compound ( $[2M + H]^+$  or  $[2M + Na]^+$ ). Although 102 dimeric ions had MS/MS scans nearby in both retention time and  $m/z$ , only twelve such MS/MS scans contained the expected major fragment ion representing the ionised monomeric compound ( $[M + H]^+$  or  $[M + Na]^+$ ). Due to this high proportion of presumed false dimeric ion mass-matches, dimeric annotations were discarded for ions with no available MS/MS scans. A literature survey, a necessarily subjective process, was performed for each remaining putatively annotated, differential ion with a nearby MS/MS scan. Putative annotations were discarded when the literature was available that indicated consistent observation of fragment ions, using similar instrumentation, that conflicted with our observed fragment

ions, leaving 223 annotations of differentially abundant ions. Of these putative ion annotations distinguishing the two diets, forty-six had some support available from MS/MS fragment ions; the remaining 177 annotations of significantly differentially abundant ions (seventy-six more abundant in the F&V diet; 101 in the basal diet) were unsupported, lacking adjacent MS/MS scans or available literature describing the expected MS/MS fragment ions.

Because urine samples collected after a standard aspirin dose, in relation to the parent study's aims, were used for the present analysis, acetyl salicylic acid and other aspirin metabolites were evaluated to ensure that they did not differ between the diet treatments or between the tertiles of intake in the cross-sectional comparison. From all groups, both in the feeding study and in the cross-sectional study, two aspirin metabolites (salicylic acid and salicyluric acid) were detected in the urine, but were not statistically significantly different between the groups (e.g. F&V *v.* basal, and third *v.* first tertile of intake via the 3DFR and FFQ; data not shown).

**Compounds more abundant in the fruit and vegetable diet.** Of the ninety-three annotated compounds significantly more abundant in the F&V diet ( $q < 0.1$ ), seventeen were supported by MS/MS fragment ions from the literature (Table 2). These metabolites were mainly markers associated with F&V consumption, e.g. metabolites of cruciferous vegetables, citrus and soya. The seventy-six unsupported metabolites, although in many cases putatively matched to several compounds, can be roughly grouped into the following categories: plant-derived metabolites ( $n$  10); vitamins ( $n$  8); steroids and steroid conjugates ( $n$  8); hormones ( $n$  8); a number of miscellaneous compounds ( $n$  42; Table S1, available online).

**Compounds more abundant in the basal diet.** Of the 130 annotated compounds more abundant in the basal diet, twenty-nine had some level of MS/MS fragment ion support (Table 3). These included a number of vitamins and their metabolites ( $n$  6), compounds related to fatty acid metabolism ( $n$  6), amino acid metabolism ( $n$  5) and several other miscellaneous compounds ( $n$  12). The 101 unsupported compounds can be roughly classified into the following categories: compounds involved in fatty acid ( $n$  13), amino acid ( $n$  22) and carbohydrate ( $n$  6) metabolism; eicosanoids and PG ( $n$  10); vitamins and vitamin metabolites ( $n$  6); plant compounds ( $n$  4) and a number of miscellaneous compounds ( $n$  40). A total of twenty-one ions were putatively annotated as carnitines, seventeen of which were more abundant in the basal diet. We investigated the abundance ratios of these twenty-one carnitines via a  $t$  test; as a group, the observed carnitines were significantly more abundant in the basal diet ( $P = 0.002$ ). The  $[M + H]^+$  ion of L-acetylcarnitine (an annotation strongly supported by MS/MS fragment ion evidence) was observed in all samples in both diets, but was significantly more abundant in the basal diet ( $q = 0.019$ ). A putatively annotated dimeric  $[2M + H]^+$  ion for L-acetylcarnitine was also observed in nine of the ten basal samples, but was not present in any of the F&V samples.



**Fig. 2.** Principal component 1 (PC1) and 2 (PC2) scores for ten basal diet ('B') and ten fruit and vegetable intervention diet ('F') samples, calculated using the observed  $\ln$  ion intensities from all features observed in at least three of twenty samples. Grey lines connect the basal and intervention diet samples from the same participant.

**Table 2.** Metabolites with supporting MS/MS fragment ions more abundant in the fruit and vegetable (F&V) diet

HMDB ID†	Metabolite	Description†	Ion type‡	q§
05 792	Sulforaphane	Isothiocyanate hydrolysis product from cruciferous vegetables <sup>(22)</sup>	[M + H] <sup>+</sup>	<0.001
04 827	Proline betaine	Compound in citrus fruits and juices <sup>(21)</sup>	[M + H] <sup>+</sup> [2M + H] <sup>+</sup> [2M + Na] <sup>+</sup>	<0.001* <0.001* 0.0584
00 714	Hippuric acid	Microbial polyphenolic metabolite <sup>(23)</sup>	[M + H] <sup>+</sup>	0.006
03 217	Genistein	Isoflavone in soya <sup>(23)</sup>	[M + H] <sup>+</sup>	0.004
03 312	Daidzein	Isoflavone in soya <sup>(23)</sup>	[M + H] <sup>+</sup>	<0.001*
02 209	Equol	Microbially derived isoflavone metabolite <sup>(23)</sup>	[M + H] <sup>+</sup>	<0.001*
05 781	Glycitein	Isoflavone in soya <sup>(23)</sup>	[M + H] <sup>+</sup>	<0.001*
04 629	O-Desmethylangolensin	Microbially derived isoflavone metabolite <sup>(23)</sup>	[M + H] <sup>+</sup>	<0.001*
04 808, 06 101	7C-aglycone; enterolactone	Vitamin K metabolite; microbially derived lignan metabolite, a marker of fibre intake <sup>(12)</sup>	[M + H] <sup>+</sup>	<0.001
00 875	Trigonelline	Metabolite of niacin; also found in coffee and other plant foods	[M + H] <sup>+</sup>	0.005
00 678, 00 927	Isovalerylglycine; valerylglycine	Fatty acid metabolites	[M + H] <sup>+</sup>	<0.001*
00 735	Hydroxyphenylacetyl-glycine	Fatty acid metabolite	[M + H] <sup>+</sup>	0.007
03 269	Nicotinuric acid	Fatty acid metabolite	[M + H] <sup>+</sup>	0.098
00 050	Adenosine	Nucleoside	[M + H] <sup>+</sup>	0.081
00 982	5-Methylcytidine	Modified nucleoside	[2M + H] <sup>+</sup>	0.082

HMDB, Human Metabolome Database; ID, identity.

\* Ions that were differentially abundant at  $\alpha = 0.1$  with Bonferroni correction.

† ID numbers and descriptions are from the HMDB if not otherwise referenced; more than one ID indicates that multiple compounds match to the observed ion.

‡ Multiple ion types indicate that more than one ion type was observed with  $q < 0.1$ . Ion types with  $q$  values  $> 0.1$  are not reported here, but are included in Table S1 (available online).

§  $q$  values represent false discovery rate.

### Cross-sectional study

**3d Food records.** The ion array created using thirty-six samples from the first and third tertiles of crucifer, citrus and soya consumption based on the 3DFR data contained 6580 ions observed in at least two samples in one or both of the tertiles compared. There were four ions with putative annotations whose abundance distinguished the third from the first tertile with  $q < 0.1$ ; the only such ion supported by MS/MS fragments was proline betaine (more abundant in the third tertile;  $P=0.0001$ ,  $q=0.050$ ).

Putative annotations of thirteen carnitines were made in at least ten samples in the first tertile and ten samples in the third tertile of crucifer, citrus and soya consumption. All thirteen of these carnitines were more abundant in the first tertile, and, as a group, were significantly more abundant in the first tertile of crucifer, citrus and soya consumption than in the third tertile ( $P<0.0001$  via a  $t$  test on the ln ratios of mean abundance within each tertile).

We compared the behaviour of ions across both the feeding and cross-sectional studies, examining the cross-sectional study ions observed in at least five samples from each tertile. A total of fifty-one feeding study ions (some with and without putative annotations) that were significantly more abundant in the F&V diet *v.* the basal diet ( $q < 0.1$ ) fit these criteria in the cross-sectional study (Fig. 3); as a group, these ions were significantly more abundant in the third *v.* first tertile of crucifer, citrus and soya consumption based on the 3DFR ( $P<0.0001$ ; two-sample  $t$  test). The ions significantly more abundant in the basal diet also had a higher mean intensity in the first *v.* third tertile based on the 3DFR, but this difference was not statistically significant ( $P=0.188$ ).

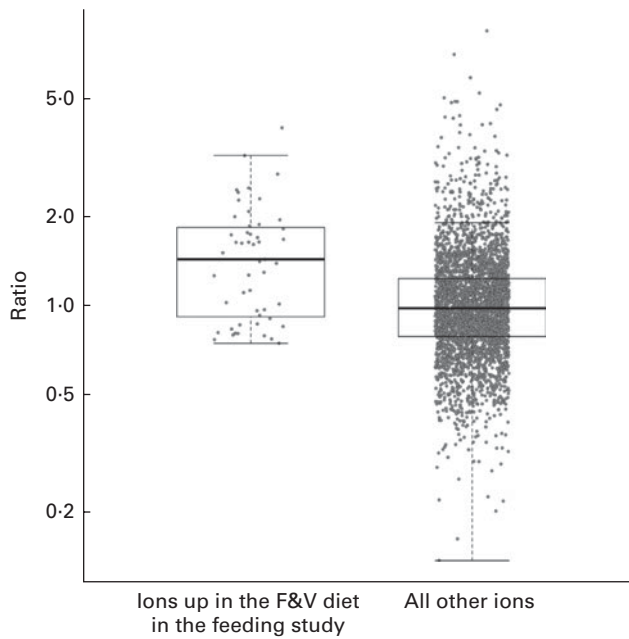
**FFQ.** The ion array created using fifty-seven samples from the first and third tertiles of total fruit and vegetable intake based on the FFQ data contained 12 524 ions observed in at least two samples in each of the tertiles compared; the array using forty-eight samples from the first and third tertiles of crucifers, citrus and soya contained 10 898. Although many of these ions in both analyses were differentially abundant in the third *v.* first tertile with a  $P$  value  $< 0.01$  by individual testing, there were no ions with  $q < 0.1$  in either analysis.

Putative annotations of eighteen carnitines were made in at least ten samples in the first tertile and ten samples in the third tertile of total fruit and vegetable intake, and nineteen carnitines when using crucifer, citrus and soya tertiles. Overall, fifteen of the eighteen carnitines, and sixteen of the nineteen carnitines, respectively, were more abundant in the first tertile. As a group, in both comparisons, carnitines were significantly more abundant in the first tertile of intake than in the third tertile ( $P=0.0017$  and  $P=0.0002$ , respectively).

As a group, the fifty-three ions observed in the feeding study to be significantly more abundant in the F&V diet than in the basal diet, and also observed in at least five samples in both the first and third FFQ-based tertiles of total fruit and vegetable intake, were significantly more abundant in the third *v.* first *v.* third tertile ( $P<0.0001$ ; two-sample  $t$  test). For the equivalent comparison based on crucifer, citrus and soya tertiles, the fifty-three ions were also significantly more abundant in the third *v.* first tertile ( $P<0.0001$ ).

### Discussion

In the present untargeted metabolomic analysis, we found forty-six putatively annotated ions, with MS/MS fragment ion support, that were differentially abundant between the two



**Fig. 3.** Strip charts of geometric mean intensity ratios (third tertile plant-derived fruit and vegetables (F&V) based on 3 d food records (3DFR): first tertile) in the cross-sectional study, for ions observed in at least five samples in each tertile, y-axis on the log scale. Charts show, respectively, fifty-one ions significantly more abundant in the F&V diet in the feeding study ( $q < 0.1$ ), and all other ions. As a group, ions significantly more abundant in the F&V diet were significantly more abundant in the cross-sectional 3DFR-based third tertile v. first tertile ( $P < 0.0001$ ). Box plots indicate interquartile range and extremes.

intervention diets (F&V and basal). As expected, many of the metabolites found in greater abundance in the F&V (high-phytochemical) diet were associated with the intervention foods consumed (see Table 2 for a complete list). For example, proline betaine, a marker of citrus consumption<sup>(21)</sup>, was observed in four different ionic forms; three supported by MS/MS fragment ions ( $[M + H]^+$ ,  $[2M + H]^+$  and  $[2M + Na]^+$ ) and a fourth (putatively  $[M + Na]^+$ ) unsupported. Sulforaphane is a hydrolysis product of glucosinolate glucoraphanin found in cruciferous vegetables<sup>(22)</sup>. Several isoflavones and their metabolites were also more abundant in the F&V diet<sup>(23)</sup>. Another ion mass-matched to two different metabolites, 7C-aglycone and enterolactone, with the literature supporting the fragment ions observed as derived from 7C-aglycone and no available literature for enterolactone. Either compound is feasible in a high-plant food diet. 7C-aglycone is a vitamin K metabolite, while enterolactone is a microbial metabolite of plant lignans and is associated with fibre intake<sup>(12)</sup>.

Other compounds more abundant in the F&V diet relative to the basal diet matched to fatty acid metabolites (isovalerylglycine or valerylglycine and hydroxyphenylacetylglycine); nicotinic acid and trigonelline, compounds involved in niacin metabolism; adenosine, involved in energy transfer; and 5-methylcytidine, a modified base. These metabolites have not previously been associated with the dietary components used in the present intervention. It is not clear whether the presence of these urinary metabolites indicates differential

metabolism between the diets or not. The higher abundance of particular fatty acids may simply reflect the difference in the major fat sources between the diets. Although the overall total fat content of the diets was very similar, the F&V diet was higher in polyunsaturated fats, whereas the basal diet was higher in saturated fat. The higher concentration of niacin metabolites does not appear to be related to dietary niacin content, as the mean intake between the two diets was only marginally different (23 v. 24 mg/d for the basal and F&V diets, respectively), but tryptophan, a precursor of niacin, was slightly higher (94 v. 84 mg/d for the basal and F&V diets, respectively).

A greater relative abundance of 5-methylcytidine was observed in the F&V diet relative to the basal diet, whereas more abundant 1-methyladenosine and methylguanosine were observed in the basal diet. Modified bases are typically considered markers of DNA damage<sup>(24)</sup>. As these compounds were putatively annotated in both diets, their relationship with the present dietary intervention, if there is one, is not clear. In addition to the compounds listed above, there were several other differentially abundant ions with annotations lacking MS/MS fragment ion support (Table S1, available online).

Many ions more abundant in the basal diet, with some level of annotation support, mass-matched to more than one metabolite. For the most part, they fell into similar categories, e.g. amino acid or pterin metabolites, etc. Only one of the metabolites appears to reflect the specific diet components consumed. Riboflavin, which is fortified in refined grains and therefore higher in the refined basal diet, was relatively more abundant in the urine after consumption of the basal diet.

Compared with the F&V diet, five ions putatively annotated as carnitines or acylcarnitines were significantly more abundant (four supported with MS/MS fragment information, and a fifth that was not) after consumption of the basal diet. Further, the twenty-one putatively annotated carnitines and derivatives, as a group (both those that were significantly different between the diets and those that were not), were significantly more abundant in the basal diet overall. Carnitines serve as shuttles for long-chain fatty acids into the mitochondria for energy production via  $\beta$ -oxidation<sup>(25)</sup>. Higher concentrations of carnitines are found in meat and dairy foods<sup>(26)</sup>. While meat was not a part of our protocol for either diet, dairy products were fed on the basal diet in place of soya given on the F&V diet, and may partly explain the higher excretion of carnitines.

Carnitines also prevent potentially toxic accumulation of fatty acyl moieties by removing these metabolites from the mitochondria, followed by urinary excretion as carnitine conjugates<sup>(25,27)</sup>. The higher relative abundance of acylcarnitines may be in response to a build-up of acyl-CoA intermediates and the incomplete oxidation of fatty acids. Whereas long-chain acylcarnitines transport fatty acids into the mitochondria, short- to medium-chain acylcarnitines move acyl groups out of the cell<sup>(27)</sup>. Strikingly, all but one of the seventeen acylcarnitines observed in higher relative concentrations after consumption of the basal diet were short- and medium-chain species, suggesting an increase in energy production via  $\beta$ -oxidation and, thus, an accumulation of acyl-group





**Table 3.** Metabolites with supporting MS/MS fragment ions more abundant in the basal diet

HMDB ID†	Metabolite	Description†	Ion type‡	q*§
00 468; 00 633; 01 195; 00 817; 02 263; 00 238 04 194	Biopterin; D-biopterin; dyspropterin; orinapterin; primapterin; sepiapterin N1-Methyl-4-pyridone-3-carboxamide	Coenzymes involved in redox reactions and the production of neurotransmitters End product of nicotinamide adenine dinucleotide degradation	[2M + H] <sup>+</sup> [M + H] <sup>+</sup> [M + Na] <sup>+</sup>	0-004 0-081 0-029
00 245	Porphobilinogen	Pyrrrole involved in porphyrin/vitamin B <sub>12</sub> synthesis	[2M + H] <sup>+</sup>	0-052
00 244	Riboflavin	Vitamin B <sub>2</sub>	[M + H] <sup>+</sup> [M + Na] <sup>+</sup>	0-032 0-032
02 234; 01 843; 00 267; 00 805	1-Pyrroline-4-hydroxy-2-carboxylate; N-acryloylglycine; pyroglutamic acid; pyrrolidonecarboxylic acid	Amino acid metabolism; fatty acid metabolite; glutamic acid derivative; glutamic acid derivative	[2M + Na] <sup>+</sup>	0-053
00 378; 00 678	2-Methylbutyrylcarnitine; isovalerylcarnitine	Short-chain acylcarnitines	[M + H] <sup>+</sup>	0-031
06 320	2,6-Dimethylheptanoyl carnitine	Short-chain acylcarnitine	[M + H] <sup>+</sup>	0-013
00 062	L-Carnitine	Amino acid involved mainly in fatty acid metabolism	[M + H] <sup>+</sup>	0-016
00 201	L-Acetylcarnitine	Carnitine metabolite involved mainly in fatty acid metabolism	[M + H] <sup>+</sup>	0-004
00 684	L-Kynurenine	Tryptophan metabolite	[M + H] <sup>+</sup>	0-078
00 715	Kynurenic acid	Tryptophan metabolite	[M + H] <sup>+</sup>	0-024
00 881	Xanthurenic acid	Tryptophan metabolite	[M + H] <sup>+</sup>	0-061
06 116	3-Hydroxyhippuric acid; salicylic acid	Polyphenolic microbial metabolite; aspirin metabolite	[M + Na] <sup>+</sup>	0-092
00 052	Argininosuccinic acid	Amino acid	[M + H] <sup>+</sup>	0-002
00 177	L-His	Amino acid	[M + H] <sup>+</sup>	0-020
03 331	1-Methyladenosine	Modified nucleoside	[M + H] <sup>+</sup>	0-044
01 563; 05 862; 06 038	1-Methylguanosine; 2-methylguanosine; 3'-O-methylguanosine	Modified nucleosides	[M + Na] <sup>+</sup> [2M + Na] <sup>+</sup>	0-049 0-029
01 067	N-Acetylaspartylglutamic acid	Neuropeptide	[M + H] <sup>+</sup>	0-010
01 119; 03 831; 02 184; 00 158	4-Hydroxy-4-(3-pyridyl)-butanoic acid; β-Tyr; L-threo-3-phenylserine; L-Tyr; o-Tyr	Nicotine metabolite; amino acid; amino acid derivative; amino acid; amino acid	[2M + H] <sup>+</sup>	0-046
00 296; 00 767	Uridine; pseudouridine	Nucleoside; isomer of uridine	[M + H] <sup>+</sup> [2M + Na] <sup>+</sup>	0-051 0-008
00 562	Creatinine	Protein catabolite	[2M + H] <sup>+</sup>	0-040
06 005	Indolylacryloylglycine	Amino acid derivative	[2M + Na] <sup>+</sup>	0-002
00 331; 00 138	3a,7b,12a-Trihydroxyoxocholanyl-glycine; glycocholic acid	Secondary bile acid; secondary bile acid	[M + H] <sup>+</sup>	0-035

HMDB, Human Metabolome Database; ID, identity.

\* No ions were differentially abundant at  $\alpha = 0.1$  with Bonferroni correction.

† ID numbers and descriptions are from the HMDB; more than one ID indicates that multiple compounds match to the observed ion.

‡ Multiple ion types indicate that more than one ion type was observed with  $q < 0.1$ . Ion types with  $q$  values  $> 0.1$  are not reported here, but are included in Table S1 (available online).

§  $q$  values represent false discovery rate.

by-products. Several amino acids and their derivatives, as well as TCA-cycle substrates and intermediates, were also observed in higher abundance after consumption of the basal diet. The greater accumulation of these metabolites coupled with acylcarnitines suggests that there may be a shift in energy metabolism towards  $\beta$ -oxidation, and potentially an overall reduction of energy production via the TCA cycle with consumption of a diet devoid of fruit and vegetables compared with a diet high in these plant foods. The remainder of urinary metabolites that were relatively more abundant after consumption of the basal diet does not appear to reflect the perturbation of any particular pathways.

Ion arrays and putatively annotated compounds (e.g. biomarkers) from the intervention study were used to make comparisons with metabolite excretion in the cross-sectional study. When looking at 3DFR, four ions statistically significantly distinguished the highest from the lowest tertile of crucifer, citrus and soya intake; however, proline betaine was the only compound annotated with MS/MS support. Proline betaine, a biomarker of citrus intake<sup>(21)</sup>, was also found among individuals after consumption of the F&V diet in the feeding study. Similarly to what we have found in the intervention study, thirteen carnitines were observed in greater relative abundance among individuals in the lowest tertile of crucifer, citrus and soya intake. No ions were significantly differentially abundant between the first and third tertiles of total fruit and vegetable intake, or crucifer, citrus and soya, specifically, based on FFQ. However, as a group, carnitines were again significantly more abundant among individuals in the lowest *v.* the highest tertile of total fruit and vegetable intake. Overall, these findings – specific compounds in the feeding study, but only general trends in the cross-sectional study – are in agreement with what is commonly observed for biomarkers and other validation methods of dietary intake in a controlled setting *v.* recall of habitual diet.

In addition to reporting these results, we have provided all raw LC–MS/MS data, extracted ions from each machine run and datasets derived from our analyses. In addition, we have released a new version of the freely available, open-source msInspect software suite containing the optimisations for small-molecule analysis that we have described. This will allow the community to freely mine our data and reproduce the present results. Access instructions for all data and software are provided in the Supplementary material (available online).

To our knowledge, this is the first metabolomic study to assess the untargeted urinary excretion of compounds after a longer-term consumption of a high compared with a low (fruit- and vegetable-free) phytochemical diet in human subjects. The strengths of the present study include the controlled feeding study design, the evaluation of components from three botanical families associated with decreased cancer and other chronic disease risk, the 2-week duration of each study period and the cross-over study design that allowed each participant to act as his/her own control. Additionally, as the fruits and vegetables on the intervention diet were given on a g/kg body-weight basis, the present data provide variability that would be typical with a range of intakes, providing greater generalisability to the general population.

A novel aspect of the study is the comparison of the compounds annotated in the intervention study with the cross-sectional data via 3DFR and FFQ, providing real-world context for the potential biomarkers identified.

There are also several limitations that should be noted. In analysing only the LC–MS/MS positive-ion mode data, we are necessarily missing groups of compounds that are unlikely to be observed in this manner. In order to be as complete as currently possible, a metabolomic profile of urine would need to include both positive- and negative-mode LC–MS/MS analysis, as well as GC–MS and NMR analysis. However, running the samples in duplicate and keeping the union of ions detected across replicates allowed us to recover more low-abundance ions than a single run would have allowed, and to analyse the reproducibility of ion intensity measurements. Although values were missing for a large percentage of ions, imputation of these values allowed for the discovery of many more differentially abundant ions, without biasing the analysis. Additionally, while the detection and annotation of many plant-food metabolites in the F&V diet lends confidence that many of the annotations in the present study are accurate, standards were not run in order to generate definitive MS/MS fragment spectra for these compounds on our instrument. As a result, the identification of any single compound, with or without MS/MS fragment ion evidence, must be treated as a putative annotation.

One of the most challenging issues in metabolomics remains in the identification of metabolites and the determination of specific pathways influenced by a particular set of metabolites. Although the HMDB is currently one of the richest resources for metabolite information, there are some shortcomings associated with its use in the context of nutrition research. First, many ions present in the urine are not present in the HMDB or other databases, or are modified versions of compounds in the HMDB, making identification difficult. Second, the database is enriched for drugs and other exogenous compounds, leading to some improbable annotations. In several instances, ions mass-matched to several potential metabolites, in very disparate classes of compounds. Finally, although some information pertaining to pathway involvement for each metabolite is provided, it is broad and not intended to make connections between the metabolites observed.

In summary, proline betaine, sulforaphane and several isoflavones were robust biomarkers of intake in the present feeding study for citrus, crucifers and soya, respectively. However, only proline betaine was annotated in the urine based on 3DFR from individuals who were high *v.* low consumers of the same three plant foods, and there were no metabolites that significantly separated groups based on FFQ. This speaks to the inability of these biomarkers, which are quickly metabolised and excreted, to adequately distinguish high *v.* low consumers of these plant foods in free-living individuals. Several compounds were putatively annotated that have not previously been associated with fruit and vegetable consumption, highlighting the utility of untargeted metabolomics in nutrition intervention studies. The relative increase in the urinary excretion of shorter-chain acylcarnitines and TCA-cycle intermediates suggests that there may be a change in energy

utilisation from glucose to fat with diets low in fruit and vegetables. Further studies are needed to replicate these findings. The consistent finding of acylcarnitines present in greater abundance in the diets of lower fruit and vegetable intake is novel and warrants further evaluation.

### Supplementary material

To view supplementary material for this article, please visit <http://dx.doi.org/10.1017/S000711451300127X>

### Acknowledgements

The present study was supported by grants R01CA142695, R56CA70913, R25CA94880 and P30CA015704 from the National Institutes of Health, National Cancer Institute. Partial funding for the linear trap quadrupole Fourier transform mass spectrometer used in the study was provided by the M.J. Murdock Charitable Trust.

The authors' contributions are as follows: J. W. L. and M. W. M. designed the research; Y. S., L. L. and J. W. L. conducted the research; J. H. and Y. O. performed the assays; D. H. M., I. R., L. L., T. H. and M. W. M. analysed the data; S. L. N. and J. W. L. interpreted the data; D. H. M., S. L. N., J. W. L. and M. W. M. wrote the manuscript; J. W. L. had primary responsibility for the final content. All authors read and approved the final manuscript.

The authors have no conflict of interest.

### References

1. Navarro SL, Saracino MR, Makar KW, *et al.* (2011) Determinants of aspirin metabolism in healthy men and women: effects of dietary inducers of UGT. *J Nutrigenet Nutrigenomics* **4**, 110–118.
2. Li Q, Wacholder S, Hunter DJ, *et al.* (2009) Genetic background comparison using distance-based regression, with applications in population stratification evaluation and adjustment. *Genet Epidemiol* **33**, 432–441.
3. Theodoridis GA, Gika HG, Want EJ, *et al.* (2012) Liquid chromatography–mass spectrometry based global metabolite profiling: a review. *Anal Chim Acta* **711**, 7–16.
4. Edmands WM, Beckonert OP, Stella C, *et al.* (2011) Identification of human urinary biomarkers of cruciferous vegetable consumption by metabolomic profiling. *J Proteome Res* **10**, 4513–4521.
5. Brennan L (2008) Session 2: personalised nutrition. Metabolomic applications in nutritional research. *Proc Nutr Soc* **67**, 404–408.
6. Lodge JK (2010) Symposium 2: modern approaches to nutritional research challenges: targeted and non-targeted approaches for metabolite profiling in nutritional research. *Proc Nutr Soc* **69**, 95–102.
7. Walsh MC, Brennan L, Pujos-Guillot E, *et al.* (2007) Influence of acute phytochemical intake on human urinary metabolomic profiles. *Am J Clin Nutr* **86**, 1687–1693.
8. Bellew M, Coram M, Fitzgibbon M, *et al.* (2006) A suite of algorithms for the comprehensive analysis of complex protein mixtures using high-resolution LC–MS. *Bioinformatics* **22**, 1902–1909.
9. Chang JL, Bigler J, Schwarz Y, *et al.* (2007) UGT1A1 polymorphism is associated with serum bilirubin concentrations in a randomized, controlled, fruit and vegetable feeding trial. *J Nutr* **137**, 890–897.
10. Saracino MR, Bigler J, Schwarz Y, *et al.* (2009) Citrus fruit intake is associated with lower serum bilirubin concentration among women with the UGT1A1\*28 polymorphism. *J Nutr* **139**, 555–560.
11. Willett W (1998) *Nutritional Epidemiology*, 2nd ed. New York, NY: Oxford University Press.
12. Horner NK, Kristal AR, Prunty J, *et al.* (2002) Dietary determinants of plasma enterolactone. *Cancer Epidemiol Biomarkers Prev* **11**, 121–126.
13. Dettmer K, Aronov PA & Hammock BD (2007) Mass spectrometry-based metabolomics. *Mass Spectrom Rev* **26**, 51–78.
14. May D, Pan S, Crispin DA, *et al.* (2011) Investigating neoplastic progression of ulcerative colitis with label-free comparative proteomics. *J Proteome Res* **10**, 200–209.
15. Wang P, Tang H, Fitzgibbon MP, *et al.* (2007) A statistical method for chromatographic alignment of LC–MS data. *Bio-statistics* **8**, 357–367.
16. Storey JD & Tibshirani R (2003) Statistical significance for genome-wide studies. *Proc Natl Acad Sci U S A* **100**, 9440–9445.
17. R Development Core Team (2010) *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing. Vienna: R Development Core Team.
18. Gentleman R, Carey V, Dudoit S, *et al.* (2005) *Limma: Linear Models for Microarray Data*. *Bioinformatics and Computational Biology Solutions using R and Bioconductor*. New York, NY: Springer.
19. Wishart DS, Knox C, Guo AC, *et al.* (2009) HMDB: a knowledgebase for the human metabolome. *Nucleic Acids Res* **37**, D603–D610.
20. Sumner IW, Amberg A, Barrett D, *et al.* (2007) Proposed minimum reporting standards for chemical analysis. *Metabolomics* **3**, 211–221.
21. Lloyd AJ, Beckmann M, Fave G, *et al.* (2011) Proline betaine and its biotransformation products in fasting urine samples are potential biomarkers of habitual citrus fruit consumption. *Br J Nutr* **106**, 812–824.
22. Navarro SL, Li F & Lampe JW (2011) Mechanisms of action of isothiocyanates in cancer chemoprevention: an update. *Food Funct* **2**, 579–587.
23. Atkinson C, Newton KM, Stanczyk FZ, *et al.* (2008) Daidzein-metabolizing phenotypes in relation to serum hormones and sex hormone binding globulin, and urinary estrogen metabolites in premenopausal women in the United States. *Cancer Causes Control* **19**, 1085–1093.
24. Loft S, Høgh Danielsen P, Mikkelsen L, *et al.* (2008) Biomarkers of oxidative damage to DNA and repair. *Biochem Soc Trans* **36**, 1071–1076.
25. Corso G, D'Apolito O, Garofalo D, *et al.* (2011) Profiling of acylcarnitines and sterols from dried blood or plasma spot by atmospheric pressure thermal desorption chemical ionization (APTDCI) tandem mass spectrometry. *Biochim Biophys Acta* **1811**, 669–679.
26. Rebouche CJ (2006) *Carnitine. Modern Nutrition in Health and Disease*, 10th ed. Philadelphia: Lippincott, Williams & Wilkins.
27. Brass EP & Hiatt WR (1998) The role of carnitine and carnitine supplementation during exercise in man and in individuals with special needs. *J Am Coll Nutr* **17**, 207–215.
28. Manach C, Hubert J, Llorach R, *et al.* (2009) The complex links between dietary phytochemicals and human health deciphered by metabolomics. *Mol Nutr Food Res* **53**, 1303–1315.