NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH Public Access
Author Manuscript
*Br J Nutr*. Author manuscript; available in PMC 2014 November 01.

Published in final edited form as:
*Br J Nutr*. 2013 November ; 110(10): . doi:10.1017/S000711451300127X.

# Metabolomic Profiling of Urine: Response to a Randomized, Controlled Feeding Study of Select Fruits and Vegetables, and Application to an Observational Study [1,,2]

**Damon H. May**[3,*], **Sandi L. Navarro**[3,*], **Ingo Ruczinski**[4], **Jason Hogan**[5], **Yuko Ogata**[5], **Yvonne Schwarz**[3], **Lisa Levy**[3], **Ted Holzman**[3], **Martin W. McIntosh**[3], and **Johanna W. Lampe**[3]

[3]Fred Hutchinson Cancer Research Center, Division of Public Health Sciences, Seattle, WA, 98109

[4]Johns Hopkins Bloomberg School of Public Health, Department of Biostatistics

[5]Fred Hutchinson Cancer Research Center, Proteomics Facility

## Abstract

Metabolomic profiles were used to characterize the effects of consuming a high-phytochemical diet compared to a diet devoid of fruits and vegetables in a randomized 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 randomized, controlled trial of 2 diet periods: a diet rich in cruciferous vegetables, citrus and soy (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 46 supported by MS/MS fragment evidence. Metabolites more abundant in the F&V diet included markers of dietary intervention (e.g., crucifers, citrus and soy), 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 participants based on tertiles of crucifers, citrus and soy from 3 d food records (3DFR; n=36) and food frequency questionnaires (FFQ; n=57); intake was separately divided into 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 4 ions were significant individually, differentiating the third vs. first tertile of crucifer, citrus and soy intake based on 3FDR. One of these was putatively annotated: proline betaine, a marker of citrus consumption. There were no ions significantly distinguishing tertiles by FFQ. Metabolomics assessment of controlled dietary interventions provides a more accurate and stronger characterization of diet than observational data.

## Introduction

Higher consumption of fruits and vegetables is associated with reduced risk of several chronic diseases, including cardiovascular disease and certain cancers. This association is

[1]This work was supported by grants R01CA142695, R56CA70913, R25CA94880, and P30CA015704 from the National Institutes of Health, National Cancer Institute. Partial funding for the LTQ-FT mass spectrometer used in this work was provided by the M.J. Murdock Charitable Trust.

[2]Author disclosures: D.H. May, S.L. Navarro, I. Ruczinski, J. Hogan, Y. Ogata, Y. Schwarz, L. Levy, T. Holzman, M.W. McIntosh, J.W. Lampe have no conflicts of interest.

Corresponding author: Johanna W. Lampe, 1100 Fairview Ave. N, M4-B402, Seattle, WA, 98109; Phone: (206) 667 6580; Fax: (206) 667 7850; jlampe@fhcrc.org.
*These authors contributed equally

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

attributed, in part, to the phytochemical content of plant foods (e.g., isoflavones in soy, flavonols in citrus fruits, and isothiocyanates in cruciferous vegetables) [1]. By necessity, the majority of dietary intervention studies have been restricted to evaluation of known or hypothesized 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 characterizing dietary response [2–4]. Untargeted metabolomic profiling of urine by liquid chromatography tandem mass spectrometry (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 [5–7]. In addition to proposed pathways, these profiles, in which a broad range of metabolites in a biosample are measured, allow for discovery of previously unrecognized 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 of intake of specific foods (e.g., cruciferous vegetables), which can then be used for evaluation of dietary intake and association with disease risk in an observational setting.

The aims of this study were two-fold. First, we intended to characterize the intervention-induced changes in untargeted urinary metabolomic patterns in response to a 2-week controlled high-phytochemical diet including specific fruits and vegetables compared to 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 metabolites observed in high- and low-fruit and vegetable diets based on 3-d food records (3DFR) and food frequency questionnaires (FFQ) in a cross-sectional, observational study of free-living individuals. All raw LC-MS/MS data, data sets derived from our analyses, and computational tools for creating metabolomics profiles using the msInspect software suite [8] are provided (access instructions in supplementary material) to allow the research community to further mine the data from our experiments.

## Methods

### Study Design

Biological samples and participant information are from a completed study, "Dietary Influences on Glucuronidation Study" (DIGEST). DIGEST was a two-tiered study comprised of a cross-sectional study from which participants were recruited for participation in a randomized, controlled crossover feeding trial comparing 2 weeks of a diet rich in fruits and vegetables (F&V) to 2 weeks of a fruit- and vegetable-free, basal diet [9;10]. 8-h urine samples from both the cross-sectional and the feeding intervention studies were used for the present analysis. The Institutional Review Board at the Fred Hutchinson Cancer Research Center (FHCRC), Seattle, Washington, approved the study and all participants gave informed written consent. Study design is summarized in Figure 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 [10]. 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 prior to the start of the study. For the present analysis, urine samples for 60 of the 293 individuals in the cross-sectional study (30 from the highest tertile of total fruit and vegetable intake and 30 from the lowest tertile, based on FFQ data) were selected for LC-MS/MS analysis

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

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 (UGT) genotype in relation to the aims of the parent study [10]. In all, 72 participants were randomized. Day 11 urine samples from 10 individuals in the feeding study (5 men and 5 women, chosen at random) were used to characterize differential metabolite abundances between the basal and F&V diets (Figure 1).

## Study Diets

Participants consumed two different diets: a basal, low-phytochemical diet, devoid of fruit and vegetables, and the basal diet supplemented with cruciferous vegetables (broccoli, cabbage, and daikon radish sprouts), soy foods (soy milk, soy cheese slices, tofu, and roasted soy nuts), and citrus fruits (grapefruit and orange juices, orange/grapefruit segments, and dried orange peel), dosed on a g-per-kg body weight basis to minimize confounding by body weight between the sexes. Weight was checked daily and kcal adjusted so that participants remained weight-stable throughout the study. Each diet was consumed for 14 d with at least a 3-week washout period between 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 use any type of medication, vitamins or other dietary supplements. Based on 24-h urinary analysis of total isothiocyanates and isoflavone excretion and daily food check-off forms, participant compliance to the study was excellent, with consumption of non-study food items on fewer than 1% of the study d [9]. Additional diet details have been published elsewhere [9].

## 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 [1;10]. Total fruit and vegetable intake was used for tertile summation. Line items in each botanical category used for crucifer, citrus and soy 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, soy cheese, and miso soup for soy. FFQ with incomplete information or data that suggested biologically implausible daily energy intakes (<600 or >4,000 kcal for women or <800 or >5,000 kcal for men) were excluded [a modification from ref.[11]] providing a total of 57 samples for analysis of total fruit and vegetable intake, and 48 for combined crucifer, citrus and soy 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 3 consecutive d after receiving training by a registered dietitian [10]. The 3DFR were analyzed to estimate intake by botanical family based on standard serving sizes [12]. Botanical families used for tertile summation included Cruciferae (i.e., cruciferous vegetables), Rutaceae (i.e., citrus fruits) and Leguminosae (i.e., soy, beans and pulses). Although this last category contained beans and pulses in addition to soy, intake from soy alone tracked similarly (data not shown). 3DFR completed >10 d from the time of urine collection were excluded, providing a total of 36 participants in the 3DFR analyses.

## Urine Collection

8-h fasting urine collected overnight after a standard aspirin dose (650 mg) was used for the present analysis [1]. The protocol was the same across all treatments (both diet periods in the intervention study and 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 d 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 the 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 FHCRC in the morning after collection. 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 *x g* in an Eppendorf 5415D table-top centrifuge, 10 μL 1% trifluoroacetic acid was added to the supernatant and dried in a speed vac. The sample was dissolved in 25 μL 2% acetonitrile with 0.1% formic acid; 5 μL was used for each analysis.

### LC-MS/MS Sample Analysis

LC-MS/MS analysis was performed using an LTQ-FT mass spectrometer (Thermo Fisher Scientific, San Jose, CA) with an electrospray ion source. The LC system was a NanoLC 2D (Eksigent, Dublin, CA) connected to an Inertsil® ODS-SP reverse phase column, 0.5 × 150mm, 3μm particle size (GL Sciences Inc., Torrance, CA). The urine samples were loaded onto the column, and chromatographic separation was performed using a 2-mobile-phase solvent system consisting of 0.1% formic acid in water (A) and 0.1% formic acid in acetonitrile (B) over 42-min gradient from 5% to 95% solvent B at 8 μL/min. The mass spectrometer operated in 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 3 most abundant ions from each MS scan were selected for MS/MS analysis using 35% normalized collision energy. Resolution and automatic gain control target were 100,000 and 2,000,000, respectively. Selected ions were dynamically excluded for 30s. Though LC-MS and MS/MS scans were acquired for the entire run time, including the column washing and equilibration, all data from the high-organic phase (after 40 min) were excluded from analysis. Samples were run in random order in duplicate, with a control every 8 samples.

### LC-MS/MS Data Extraction and Analytic Data Set Creation

Complete details of all data analysis steps, data, software and instructions necessary to recreate the analytic data sets described below are provided in Supplemental Text. In brief, msInspect [8], a freely available suite of software and algorithms for analyzing profiles from high-resolution LC-MS/MS data, was extended to accurately detect and quantify small-molecule ion features. These improvements, which include optimized 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 [13]. Duplicate sample runs were combined by aligning ions by *m/z* and retention time [8;14], and the union of features was retained. Ion intensities were quantile normalized [8;15] and then aligned by *m/z* and retention time to create an analytic data set (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 metabolite profile we used Principal Components Analysis (PCA) on the ln intensities of features observed in at least 3 participants. PCA was



NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

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 the ln ratios (F&V/basal) for each participant. Missing data are characteristic of LC-MS/MS profiling data [8;15], 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 1$^{st}$ 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 toward 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 [16] for all ions having 4 or more ratios across samples. Analyses were carried out using the R statistical environment [17].

To investigate the cross-sectional study data, a separate ion array was created using the method described above. Analyses were performed separately to evaluate combined crucifer, citrus and soy consumption defined by 3DFR and by 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 5 samples in both tertiles, ln intensities were compared, adjusting for participant BMI using the Limma R package for differential expression analysis [18], and p- and q-values were calculated.

### Metabolite Annotation

As others have noted [13; 19], identifying metabolites in 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 Metabolite Database (HMDB) version 2.5 [19], a separate theoretical mass value was calculated for 4 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 mean mass value to these ion masses; all matches within 5ppm 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 Supplemental Text). For ions annotated as dimeric compounds, abundant fragment ions matching $[M+H]^+$ 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 positive ion mode (references in Supplemental Tables). 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[20] Metabolite functions and pathway information were determined by searching HMDB, Kyoto Encyclopedia of Genes and Genomes (KEGG), 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 tertiles of crucifer, citrus and soy intake; however, overall energy intake was higher among individuals in the highest tertile of intake in the cross-sectional study. On average, 3,763 ions (range: 2,378–5,065) passing filtering criteria were located in each of the 160 LC-MS/MS runs of the 80 samples run (20 feeding study and 60 cross-sectional). After combination of ions across duplicate runs, an average of 4,790 ions (range: 3,155–6,575) were detected per sample. The median CV of the ln ion intensities between duplicate runs was 0.104 (range: 0.082–0.195).

### Feeding Study

**Data Summary—**A total of 7,382 ions were observed in at least 3 samples across both diets. To analyze global differences in ion abundance between diets (without background-level estimation of unobserved ions), we used PCA analysis on these ions. Ln ion intensity variation in PC dimensions 1 and 2 explained 26% of the variance between samples. PC1 was not associated with diet, but PC2 systematically separated the diets with high accuracy: PC2>0 correctly classifies 9 of 10 samples from each diet (Figure 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. Analyzed by t-test on an ion-by-ion level, a total of 2,857 ions were determined to be significantly higher or lower in the F&V vs. basal diet with $q < 0.1$ (1,360 more abundant in F&V, 1,497 in basal).

**Annotated Compounds—**3,666 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 2,857 ions significantly differentially abundant in the F&V vs. basal diets with $q < 0.1$ (179 more abundant in F&V, 244 in basal). Of these, 195 had at least 1 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 12 such MS/MS scans contained the expected major fragment ion representing the ionized 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 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. 46 of these putative ion annotations distinguishing the 2 diets had some support available from MS/MS fragment ions; the remaining 177 annotations of significantly differentially abundant ions (76 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 urines collected after a standard aspirin dose, in relation to the parent study aims, were used for the present analysis, acetyl salicylic acid and other aspirin metabolites were evaluated to ensure that they did not differ between diet treatments or between tertiles of intake in the cross-sectional comparison. Two aspirin metabolites (salicylic acid and salicyluric acid) were detected in urine from all groups, both in the feeding study and in the cross-sectional study, but were not statistically significantly different between groups (e.g., F&V vs. basal, and third vs. first tertile of intake via the 3DFR and FFQ, data not shown).



NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

**Compounds more abundant in the F&V diet—**Of the 93 annotated compounds significantly more abundant in the F&V diet (q<0.1), 17 were supported by MS/MS fragment ions from literature (Table 2). These metabolites were mainly markers associated with F&V consumption, e.g., metabolites of cruciferous vegetables, citrus and soy. The 76 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), and a number of miscellaneous compounds (n=42; Supplemental Table 1).

**Compounds more abundant in the basal diet—**Of the 130 annotated compounds more abundant in the basal diet, 29 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 prostaglandins (n=10); vitamins and vitamin metabolites (n=6); plant compounds (n=4); and a number of miscellaneous compounds (n=40). 21 ions were putatively annotated as carnitines, 17 of which were more abundant in the basal diet. We investigated the abundance ratios of these 21 carnitines via 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 9 of the 10 basal samples but was not present in any of the F&V samples.

## Cross-Sectional Study

**3DFR—**The ion array created using 36 samples from the first and third tertiles of crucifer, citrus and soy consumption based on 3DFR data contained 6,580 ions observed in at least 2 samples in 1 or both of the tertiles compared. There were 4 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 third tertile; p=0.0001, q=0.050).

Putative annotations of 13 carnitines were made in at least 10 samples in the first tertile and 10 samples in the third tertile of crucifer, citrus and soy consumption. All 13 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 soy consumption than in the third tertile (p<0.0001 via t-test on ln ratios of mean abundance within each tertile).

We compared the behavior of ions across both the feeding and cross-sectional studies, examining cross-sectional study ions observed in at least 5 samples from each tertile. 51 feeding study ions (some with and some without putative annotations) that were significantly more abundant in the F&V diet vs. the basal diet (q < 0.1) fit these criteria in the cross-sectional study (Figure 3); as a group, these ions were significantly more abundant in the third vs. first tertile of crucifer, citrus and soy consumption based on 3DFR (p<0.0001; two-sample t-test). The ions significantly more abundant in the basal diet also had higher mean intensity in the first vs. third tertile based on 3DFR, but this difference was not statistically significant (p=0.188).

**FFQ—**The ion array created using 57 samples from the first and third tertiles of total fruit and vegetable intake based on FFQ data contained 12,524 ions observed in at least 2

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

samples in each of the tertiles compared; the array using 48 samples from the first and third tertiles of crucifers, citrus and soy contained 10,898. Although many of these ions in both analyses were differentially abundant in the third vs. first tertile with p-value < 0.01 by individual testing, there were no ions with q < 0.1 in either analysis.

Putative annotations of 18 carnitines were made in at least 10 samples in the first and 10 samples in the third tertile of total fruit and vegetable intake, and 19 carnitines when using crucifer, citrus and soy tertiles. 15 of 18 carnitines, and 16 of 19 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 (p=0.0017 and p=0.0002, respectively).

As a group, the 53 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 5 samples in both the first and third FFQ-based tertiles of total fruit and vegetable intake, were significantly more abundant in the third vs. first vs. the third tertile (p<0.0001; two-sample t-test). For the equivalent comparison based on crucifer, citrus and soy tertiles, the 53 ions were also significantly more abundant in the third vs. first tertile (p<0.0001).

## DISCUSSION

In this untargeted metabolomics analysis, we found 46 putatively annotated ions, with MS/MS fragment ion support, that were differentially abundant between the two 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 complete list). For example, proline betaine, a marker of citrus consumption [21], was observed in 4 different ionic forms; 3 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 the glucosinolate glucoraphinin found in cruciferous vegetables [22]. Several isoflavones and their metabolites were also more abundant in the F&V diet [23]. Another ion mass-matched to two different metabolites, 7C-aglycone and enterolactone, with 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 fiber intake [12].

Other compounds more abundant in the F&V diet relative to the basal diet matched to fatty acid metabolites (isovalerylglycine or valerylglycine, hydroxyphenylacetyglycine); nicotinuric 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 our intervention. It is not clear whether the presence of these urinary metabolites indicate 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 the 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 vs. 24 mg/d for basal and F&V diets, respectively), but tryptophan, a precursor of niacin, was slightly higher (94 vs. 84 mg/d for basal and F&V diets, respectively).

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 [24]. As these compounds were putatively annotated in both diets, their relationship with our 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 (Supplemental Table 1).

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.

Five ions putatively annotated as carnitines or acylcarnitines were significantly more abundant (4 supported with MS/MS fragment information, and a fifth that was not) after consumption of the basal diet compared to the F&V diet. Further, the 21 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 β-oxidation [25]. Higher concentrations of carnitines are found in meat and dairy foods [26]. While meat was not a part of our protocol for either diet, dairy products were fed on the basal diet in place of soy 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 [25;27]. The higher relative abundance of acylcarnitines may be in response to a buildup of acyl-Coenzyme A (CoA) intermediates and 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 [27]. Strikingly, all but 1 of the 17 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 β-oxidation and, thus, an accumulation of acyl-group byproducts. 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 the acylcarnitines suggests that there may be a shift in energy metabolism toward β-oxidation, and potentially an overall reduction of energy production via the TCA cycle with consumption of a diet devoid of fruit and vegetables compared to 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 perturbation of any particular pathways.

Ion arrays and putatively annotated compounds (e.g., biomarkers) from the intervention study were used to make comparisons to metabolite excretion in the cross-sectional study. When looking at 3FDR, 4 ions statistically significantly distinguished the highest from the lowest tertile of crucifer, citrus and soy intake; however, proline betaine was the only compound annotated with MS/MS support. Proline betaine, a biomarker of citrus intake [21], was also found among individuals after consumption of the F&V diet in the feeding study. Similarly to what we found in the intervention study, 13 carnitines were observed in greater relative abundance among individuals in the lowest tertile of crucifer, citrus and soy intake. No ions were significantly differentially abundant between the first and third tertiles of total fruit and vegetable intake, or crucifer, citrus and soy, specifically, based on FFQs. However, as a group, the carnitines were again significantly more abundant among individuals in the



NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

lowest versus 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 versus 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 data sets derived from our analyses. In addition, we have released a new version of the freely-available, open source msInspect software suite containing the optimizations for small-molecule analysis that we have described. This will allow the community to freely mine our data and reproduce our results. Access instructions for all data and software are provided in Supplemental Text.

To our knowledge, this is the first metabolomics study to assess untargeted urinary excretion of compounds after longer-term consumption of a high- compared to a low- (fruit- and vegetable-free) phytochemical diet in humans. Strengths of this study include the controlled feeding study design, evaluation of components from 3 botanical families associated with decreased cancer and other chronic disease risk, the 2-week duration of each study period, and the crossover study design which 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-per-kg body weight basis, our data provide variability that would be typical with a range of intakes, providing greater generalizability to the general population. A novel aspect of the study is the comparison of the compounds annotated in the intervention study to 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 analyzing only LC-MS/MS positive-mode ion 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 metabolomics profile of urine would need to include both positive- and negative-mode LC-MS/MS analysis, as well as gas chromatography mass spectrometry (GC-MS) and nuclear magnetic resonance (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 analyze the reproducibility of the 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 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 urine are not present in HMDB or other databases, or are modified versions of compounds in 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 metabolites observed.

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

In summary, proline-betaine, sulforaphane, and several isoflavones were robust biomarkers of intake in our feeding study for citrus, crucifers and soy, respectively. However, only proline-betaine was annotated in urine based on 3DFR from individuals who were high versus 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 metabolized and excreted, to adequately distinguish high- versus 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 urinary excretion of shorter-chain acylcarnitines and TCA cycle-intermediates suggests that there may be a change in energy utilization 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 diets of lower fruit and vegetable intake is novel and warrants further evaluation.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

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. analyzed 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 final content. All authors read and approved the final manuscript.

## References

1. Navarro SL, Saracino MR, Makar KW, et al. Determinants of aspirin metabolism in healthy men and women: Effects of dietary inducers of UGT. J Nutrigenet Nutrigenomics. 2011; 4:110–118. [PubMed: 21625173]
2. Li Q, Wacholder S, Hunter DJ, et al. Genetic background comparison using distance-based regression, with applications in population stratification evaluation and adjustment. Genet Epidemiol. 2009; 33:432–441. [PubMed: 19140130]
3. Theodoridis GA, Gika HG, Want EJ, et al. Liquid chromatography-mass spectrometry based global metabolite profiling: a review. Anal Chim Acta. 2012; 711:7–16. [PubMed: 22152789]
4. Edmands WM, Beckonert OP, Stella C, et al. Identification of human urinary biomarkers of cruciferous vegetable consumption by metabonomic profiling. J Proteome Res. 2011; 10:4513–4521. [PubMed: 21770373]
5. Brennan L. Session 2: Personalised nutrition. Metabolomic applications in nutritional research. Proc Nutr Soc. 2008; 67:404–408. [PubMed: 18847517]
6. Lodge JK. Symposium 2: Modern approaches to nutritional research challenges: Targeted and non-targeted approaches for metabolite profiling in nutritional research. Proc Nutr Soc. 2010; 69:95–102. [PubMed: 19954566]
7. Walsh MC, Brennan L, Pujos-Guillot E, et al. Influence of acute phytochemical intake on human urinary metabolomic profiles. Am J Clin Nutr. 2007; 86:1687–1693. [PubMed: 18065587]
8. Bellew M, Coram M, Fitzgibbon M, et al. A suite of algorithms for the comprehensive analysis of complex protein mixtures using high-resolution LC-MS. Bioinformatics. 2006; 22:1902–1909. [PubMed: 16766559]
9. Chang JL, Bigler J, Schwarz Y, et al. UGT1A1 polymorphism is associated with serum bilirubin concentrations in a randomized, controlled, fruit and vegetable feeding trial. J Nutr. 2007; 137:890–897. [PubMed: 17374650]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

10. Saracino MR, Bigler J, Schwarz Y, et al. Citrus fruit intake is associated with lower serum bilirubin concentration among women with the UGT1a1*28 polymorphism. J Nutr. 2009; 139:555–560. [PubMed: 19141701]
11. Willett, W. Nutritional Epidemiology. 2nd ed.. New York: Oxford University Press; 1998.
12. Horner NK, Kristal AR, Prunty J, et al. Dietary determinants of plasma enterolactone. Cancer Epidemiol Biomarkers Prev. 2002; 11:121–126. [PubMed: 11815409]
13. Dettmer K, Aronov PA, Hammock BD. Mass spectrometry-based metabolomics. Mass Spectrom Rev. 2007; 26:51–78. [PubMed: 16921475]
14. May D, Pan S, Crispin DA, et al. Investigating neoplastic progression of ulcerative colitis with label-free comparative proteomics. J Proteome Res. 2011; 10:200–209. [PubMed: 20828217]
15. Wang P, Tang H, Fitzgibbon MP, et al. A statistical method for chromatographic alignment of LC-MS data. Biostatistics. 2007; 8:357–367. [PubMed: 16880200]
16. Storey JD, Tibshirani R. Statistical significance for genomewide studies. Proc Natl Acad Sci U S A. 2003; 100:9440–9445. [PubMed: 12883005]
17. R Development Core Team. R: A language and environment for statistical computing, R Foundation for Statistical Computering. Vienna, Australia:
18. Gentleman, R.; Carey, V.; Dudoit, S.; Irizarry, R.; Huber, IW. sLimma: linear models for microarray data. Bioinformatics and Computational Biology Solutions using R and Bioconductor. New York: Springer; 2005.
19. Wishart DS, Knox C, Guo AC, et al. HMDB: a knowledgebase for the human metabolome. Nucleic Acids Res. 2009; 37:D603–D610. [PubMed: 18953024]
20. Sumner LW, Amberg A, Barrett D, et al. Proposed minimum reporting standards for chemical analysis. Metabolomics. 2007; 3:211–221. [PubMed: 24039616]
21. Lloyd AJ, Beckmann M, Fave G, et al. Proline betaine and its biotransformation products in fasting urine samples are potential biomarkers of habitual citrus fruit consumption. Br J Nutr. 2011; 106:812–824. [PubMed: 21736852]
22. Navarro SL, Li F, Lampe JW. Mechanisms of action of isothiocyanates in cancer chemoprevention: an update. Food Funct. 2011; 2:579–587. [PubMed: 21935537]
23. Atkinson C, Newton KM, Stanczyk FZ, et al. 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. 2008; 19:1085–1093. [PubMed: 18478336]
24. Loft S, Hogh Danielsen P, Mikkelsen L, et al. Biomarkers of oxidative damage to DNA and repair. Biochem Soc Trans. 2008; 36:1071–1076. [PubMed: 18793191]
25. Corso G, D'Apolito O, Garofalo D, et al. 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. 2011; 1811:669–679. [PubMed: 21683155]
26. Rebouche, CJ. Carnitine. Modern Nutrition in Health and Disease. 10th ed.. Philadelphia: Lippincott, Williams & Wilkins; 2006.
27. Brass EP, Hiatt WR. The role of carnitine and carnitine supplementation during exercise in man and in individuals with special needs. J Am Coll Nutr. 1998; 17:207–215. [PubMed: 9627906]
28. Manach C, Hubert J, Llorach R, et al. The complex links between dietary phytochemicals and human health deciphered by metabolomics. Mol Nutr Food Res. 2009; 53:1303–1315. [PubMed: 19764066]

NIH-PA Author Manuscript



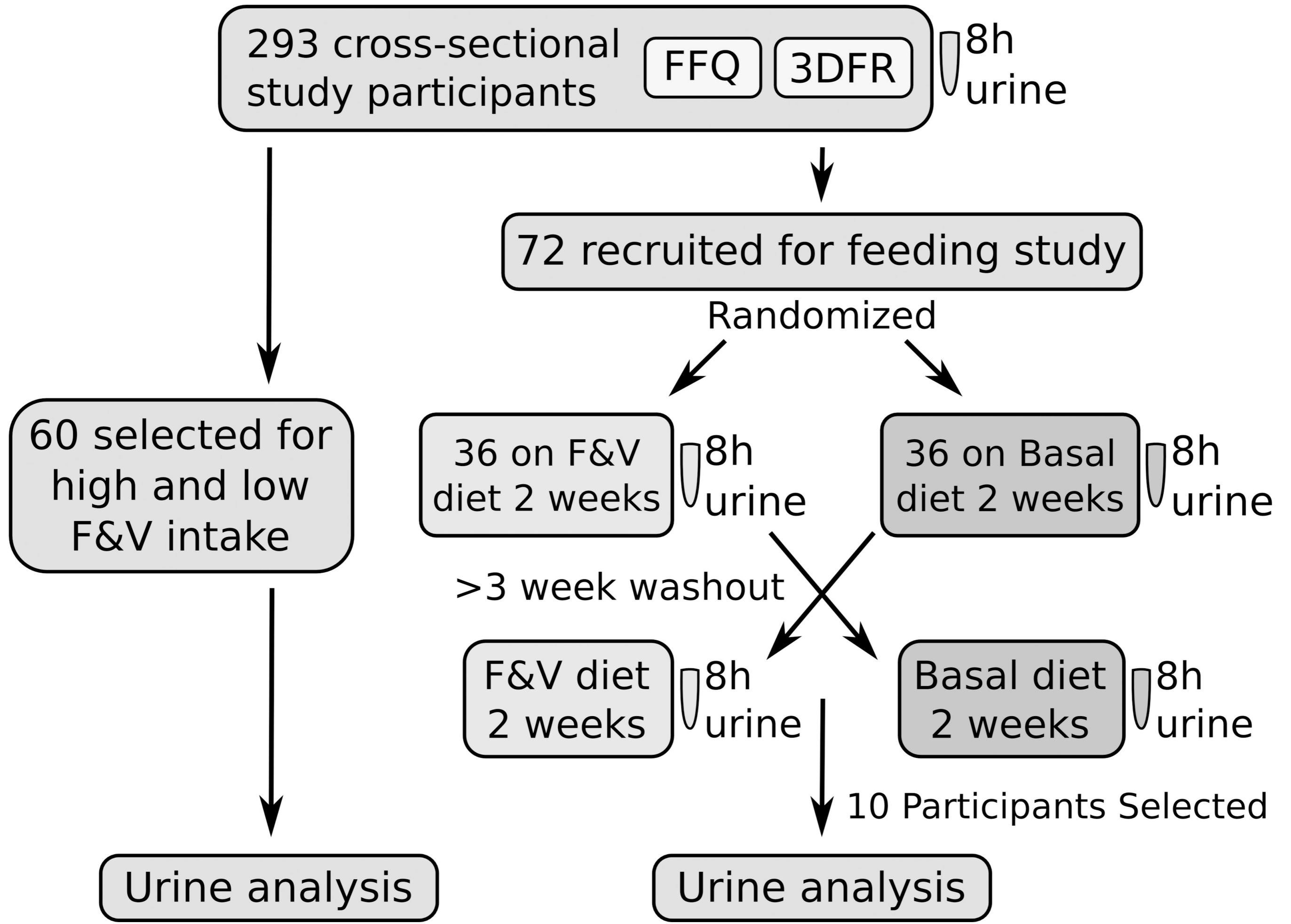


**Figure 1.**

Study Design and General Analysis Workflow. 8h urine, FFQ (food frequency questionnaire) and 3DFR (3d food record) were collected from 293 cross-sectional study participants. 60 of these were selected for analysis based on high and low F&V (fruit and vegetable) intake. 72 were recruited for a feeding study; half of these were randomized to a F&V intervention diet and half to a basal (fruit- and vegetable-free) diet for 2 weeks and 8h urine collected. After 3 week washout, participants switched diets and a second 8h urine was collected. Samples from 10 of these participants were used in the current analysis.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

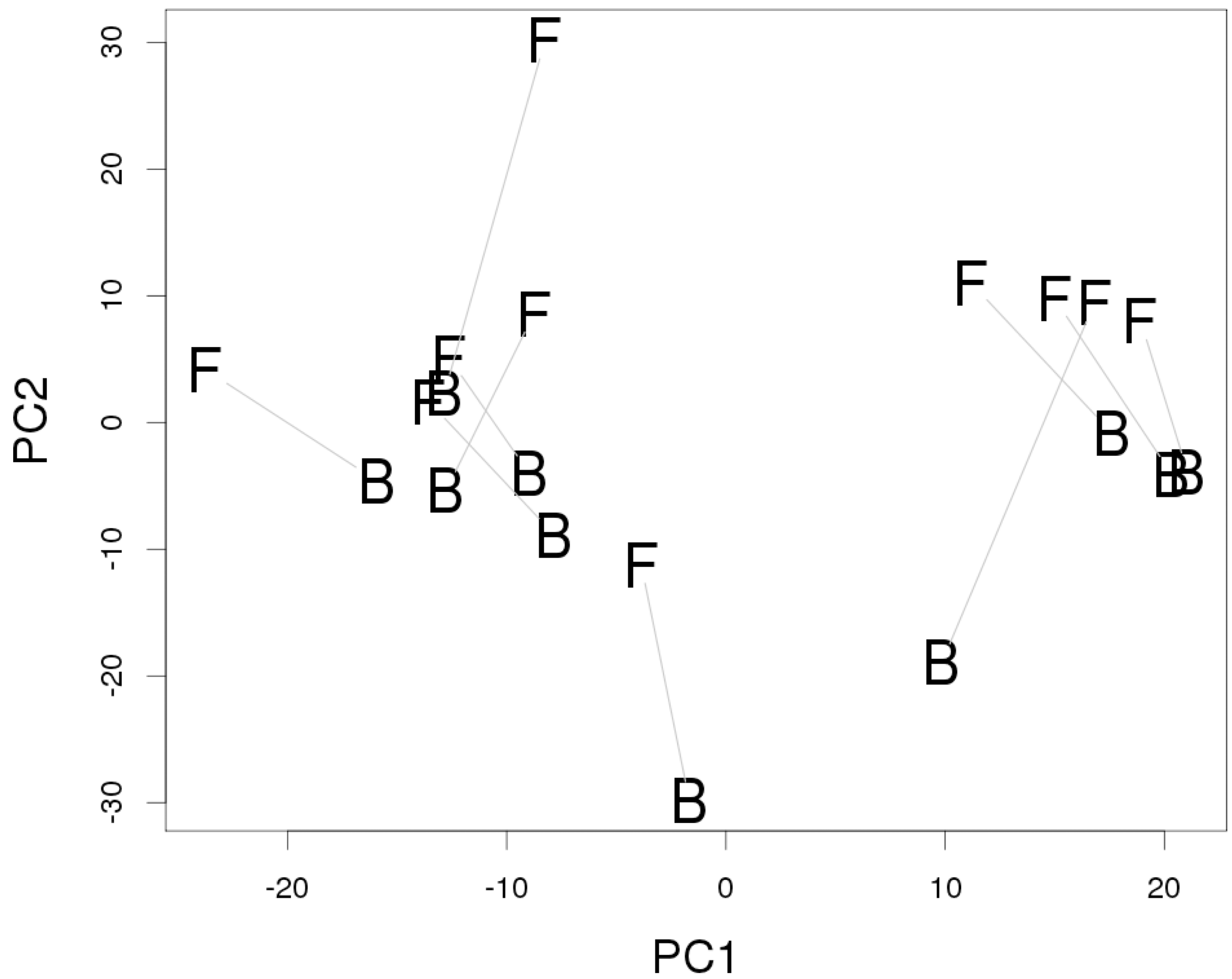


**Figure 2.**
Principal Component 1 (PC1) and 2 (PC2) scores for 10 basal diet ('B') and 10 F&V intervention diet ('F') samples, calculated using observed ln ion intensities from all features observed in at least 3 of 20 samples. Grey lines connect basal and intervention diet samples from the same participant.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

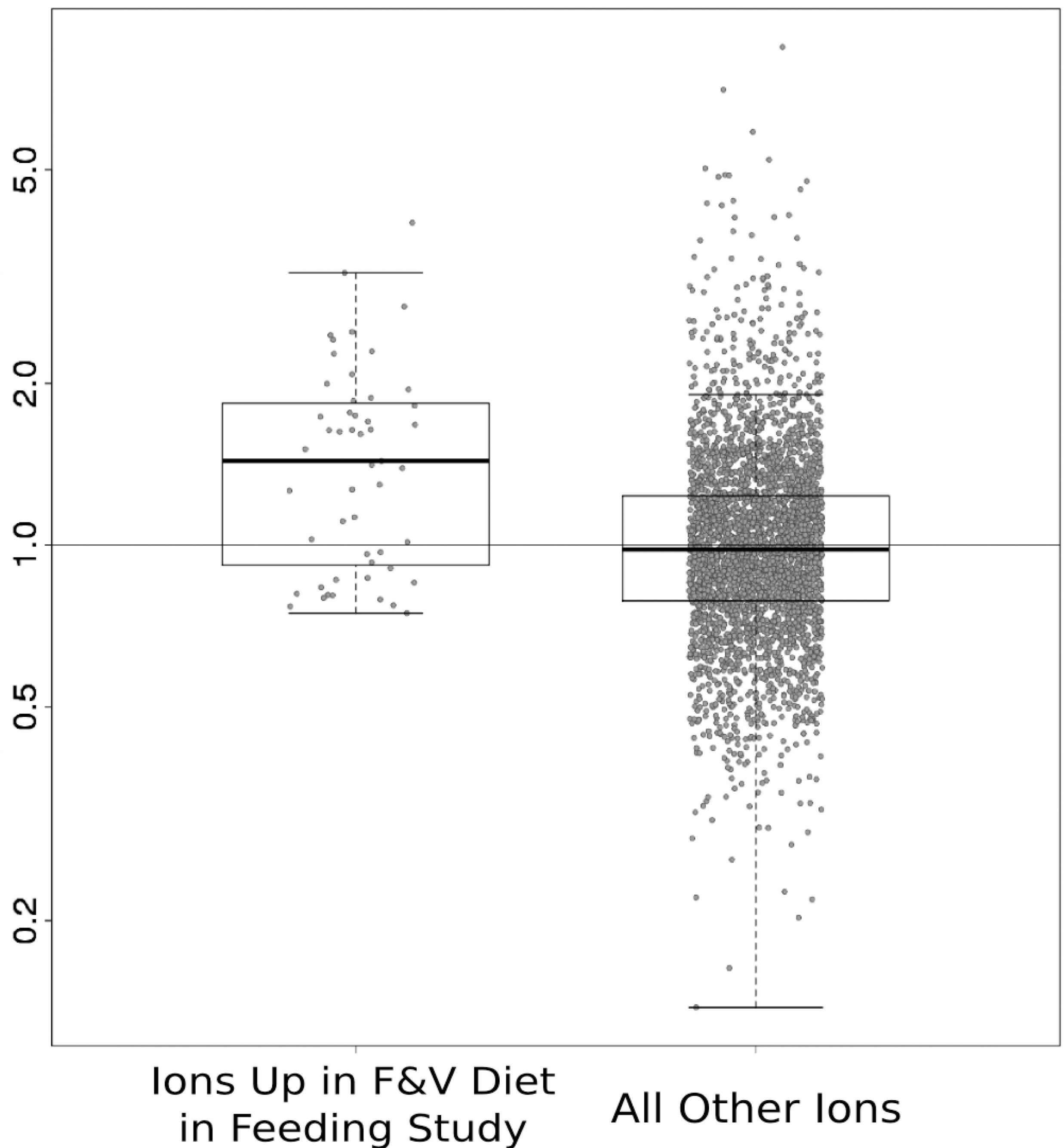


**Figure 3.**

Strip charts of geometric mean intensity ratios (third tertile plant-derived F&V based on 3-DFR: first tertile) in the cross-sectional study, for ions observed in at least 5 samples in each tertile, y axis on log scale. Charts show, respectively, 51 ions significantly more abundant in 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 cross-sectional 3DFR-based 3rd tertile vs. 1st ($p<0.0001$). Box plots indicate interquartile range (IQR) and extremes.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

## Table 1

Characteristics of intervention and cross-sectional study participants and diet components.

| | Feeding Study Basal (n=10) | Feeding Study F&V (n=10) | Tertile 1 (3DFR n=18)[2] | Tertile 3 (3DFR n=18)[2] | *P*[*] *(3DFR)* | Tertile 1 (FFQ n=22)[3] | Tertile 3 (FFQ n=26)[3] | *P*[*] *(FFQ)* |
|---|---|---|---|---|---|---|---|---|
| **Age (yr)** | 29 (4.9)[1] | | 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$^2$)** | 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 |
| **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)**[4] | 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 |
| **Saturated (% energy)** | 14 (3.2) | 9 (3.3) | 11 (3.7) | 10 (3.9) | 0.26 | 11 (3.4) | 9 (2.7) | 0.15 |
| **Mono-Unsaturated (% 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 |
| **Fiber (g/d)** | 10 (3.0) | 29 (5.1) | 13 (3.4) | 32(14.1) | <0.01 | 13 (8.7) | 29 (10.8) | <0.01 |

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

[1] Mean (SD)

[2] First and third tertiles of crucifer, citrus and soy intake among the cross-sectional study participants, as calculated based on 3DFR

[3] First and third tertiles of crucifer, citrus and soy intake among the cross-sectional study participants, as calculated based on FFQ

[4] Percents may not add up to 100 due to rounding

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

## Table 2

Metabolites with supporting MS/MS fragment ions more abundant in the F&V diet.

| HMDB ID[1] | Metabolite | Description[1] | Ion type[2] | q* |
|---|---|---|---|---|
| 05792 | Sulforaphane | Isothiocyanate hydrolysis product from cruciferous vegetables (22) | $[M+H]^+$ | <0.001 |
| 04827 | Proline betaine | Compound in citrus fruits and juices (21) | $[M+H]^+$<br>$[2M+H]^+$<br>$[2M+Na]^+$ | <0.001 *<br><0.001 *<br>0.0584 |
| 00714 | Hippuric acid | Microbial polyphenolic metabolite (28) | $[M+H]^+$ | 0.006 |
| 03217 | Genistein | Isoflavone in soy (23) | $[M+H]^+$ | 0.004 |
| 03312 | Daidzein | Isoflavone in soy (23) | $[M+H]^+$ | <0.001 * |
| 02209 | Equol | Microbially-derived isoflavone metabolite (23) | $[M+H]^+$ | <0.001 * |
| 05781 | Glycitein | Isoflavone in soy (23) | $[M+H]^+$ | <0.001 * |
| 04629 | O-Desmethylangolensin | Microbially-derived isoflavone metabolite (23) | $[M+H]^+$ | <0.001 * |
| 04808; 06101 | 7C-aglycone; Enterolactone | Vitamin K metabolite; Microbially-derived lignan metabolite, marker of fiber intake (12) | $[M+H]^+$ | <0.001 |
| 00875 | Trigonelline | Metabolite of niacin; also found in coffee and other plant foods | $[M+H]^+$ | 0.005 |
| 00678; 00927 | Isovalerylglycine; Valerylglycine | Fatty acid metabolites | $[M+H]^+$ | <0.001 * |
| 00735 | Hydroxyphenylacetyl-glycine | Fatty acid metabolite | $[M+H]^+$ | 0.007 |
| 03269 | Nicotinuric acid | Fatty acid metabolite | $[M+H]^+$ | 0.098 |
| 00050 | Adenosine | Nucleoside | $[M+H]^+$ | 0.081 |
| 00982 | 5-methylcytidine | Modified nucleoside | $[2M+H]^+$ | 0.082 |

[1] Identity (ID) numbers and descriptions are from the Human Metabolome Database if not otherwise referenced; more than one ID indicates multiple compounds match to the observed ion.

[2] 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 Supplemental Table 1.

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

* q-values represent false discovery rate. q-values with an asterisk indicate ions that were differentially abundant at α=0.1 with Bonferroni correction.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

## Table 3

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

| HMDB ID[1] | Metabolite | Description[1] | Ion type[2] | q* |
|---|---|---|---|---|
| 00468; 00633; 01195; 00817; 02263; 00238 | Biopterin; D-Biopterin; Dyspropterin; Orinapterin; Primapterin; Sepiapterin | Coenzymes involved in redox reactions and production of neurotransmitters | $[2M+H]^+$ | 0.004 |
| 04194 | N1-Methyl-4-pyridone-3-carboxamide | End-product of nicotinamide-adenine dinucleotide (NAD) degradation | $[M+H]^+$<br>$[M+Na]^+$ | 0.081<br>0.029 |
| 00245 | Porphobilinogen | Pyrrole involved in Porphyrin/vitamin B12 synthesis | $[2M+H]^+$ | 0.052 |
| 00244 | Riboflavin | Vitamin B2 | $[M+H]^+$<br>$[M+Na]^+$ | 0.032<br>0.032 |
| 02234; 01843; 00267; 00805 | 1-Pyrroline-4-hydroxy-2-carboxylate; N-Acryloylglycine; Pyroglutamic acid; Pyrrolidonecarboxylic acid | Amino acid metabolism; Fatty acid metabolite; Glutamic acid derivative; Glutamic acid derivative | $[2M+Na]^+$ | 0.053 |
| 00378; 00678 | 2-Methylbutyroylcarnitine; Isovalerylcarnitine | Short-chain acylcarnitines | $[M+H]^+$ | 0.031 |
| 06320 | 2,6 Dimethylheptanoyl carnitine | Short-chain acylcarnitine | $[M+H]^+$ | 0.013 |
| 00062 | L-Carnitine | Amino acid involved mainly in fatty acid metabolism | $[M+H]^+$ | 0.016 |
| 00201 | L-Acetylcarnitine | Carnitine metabolite involved mainly in fatty acid metabolism | $[M+H]^+$ | 0.004 |
| 00684 | L-Kynurenine | Tryptophan metabolite | $[M+H]^+$ | 0.078 |
| 00715 | Kynurenic acid | Tryptophan metabolite | $[M+H]^+$ | 0.024 |
| 00881 | Xanthurenic acid | Tryptophan metabolite | [M+H] + | 0.061 |
| 06116 | 3-Hydroxyhippuric acid; Salicyluric acid | Polyphenolic microbial metabolite; Aspirin metabolite | $[M+Na]^+$ | 0.092 |
| 00052 | Argininosuccinic acid | Amino acid | $[M+H]^+$ | 0.002 |
| 00177 | L-Histidine | Amino acid | $[M+H]^+$ | 0.020 |

NIH-PA Author Manuscript

| HMDB ID[1] | Metabolite | Description[1] | Ion type[2] | q* |
|---|---|---|---|---|
| 03331 | 1-Methyladenosine | Modified nucleoside | $[M+H]^+$ | 0.044 |
| 01563; 05862; 06038 | 1-Methylguanosine; 2-Methylguanosine; 3'-O-Methylguanosine | Modified nucleosides | $[M+Na]^+$<br>$[2M+Na]^+$ | 0.049<br>0.029 |
| 01067 | N-Acetylaspartylglutamic acid | Neuropeptide | $[M+H]^+$ | 0.010 |
| 01119; 03831; 02184; 00158; | 4-Hydroxy-4-(3-pyridyl)-butanoic acid; β-Tyrosine; L-Threo-3-Phenylserine; L-Tyrosine; o-Tyrosine | Nicotine metabolite; Amino acid; Amino acid derivative; Amino acid; Amino acid | $[2M+H]^+$ | 0.046 |
| 00296; 00767 | Uridine; Pseudouridine | Nucleoside; Isomer of Uridine | $[M+H]^+$<br>$[2M+Na]^+$ | 0.051<br>0.008 |
| 00562 | Creatinine | Protein catabolite | $[2M+H]^+$ | 0.040 |
| 06005 | Indolylacryloylglycine | Amino acid derivative | $[2M+Na]^+$ | 0.002 |
| 00331; 00138 | 3a,7b,12a-Trihydroxyoxocholanyl-Glycine; Glycocholic acid | Secondary bile acid; Secondary bile acid | $[M+H]^+$ | 0.035 |

[1] Identity (ID) numbers and descriptions are from the Human Metabolome; more than one ID indicates multiple compounds match to the observed ion.

[2] 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 Supplemental Table 1.

* q-values represent false discovery rate. q-values with an asterisk indicate ions that were differentially abundant at $\alpha=0.1$ with Bonferroni correction.