Metabolomic based approach to identify biomarkers of apple intake

Aoife E. McNamara[1,2], Cassandra Collins[1,2], Pedapati S. C. Sri Harsha [1,2], Diana González-Peña [1,2], Helena Gibbons[1,2], Breige A. McNulty[1], Anne P Nugent[1,4], Janette Walton[3,5], Albert Flynn[3], Lorraine Brennan[1,2]

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

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

[3] School of Food and Nutritional Sciences, University College Cork, Cork, Ireland.

[4] Institute for Global Food Security, School of Biological Sciences, Queens University Belfast, Northern Ireland.

[5] Dept. Biological Sciences, Cork Institute of Technology, Cork, Ireland

Corresponding Author:

Professor Lorraine Brennan

UCD School of Agriculture and Food Science,

UCD Institute of Food and Health,

UCD, Belfield, Dublin 4, Ireland.

Email: lorraine.brennan@ucd.ie

Phone: 00 353 1 7166815

Short Title: Biomarkers of apple intake

Keywords: apples, biomarkers, dietary assessment, food intake, metabolomics

Abstract

Scope

There is an increased interest in developing biomarkers of food intake to address some of the limitations associated with self-reported data. The objective was to identify biomarkers of apple intake, examine dose-response relationships and agreement with self-reported data.

Methods and Results

Metabolomic data from three studies were examined: an acute intervention, a short-term intervention and a free-living cohort study. Fasting and postprandial urine samples were collected for analysis by ^{1}H-NMR and LC-MS. Calibration curves were developed to determine apple intake and classify individuals into categories of intake. Multivariate analysis of data revealed that levels of multiple metabolites increased significantly post-apple consumption, compared to the control food- broccoli. In the dose-response study, urinary xylose, epicatechin sulfate and 2, 6-dimethyl-2-(2-hydroxyethyl)-3,4-dihydro-2H-1-benzopyran increased as apple intake increased. Urinary xylose concentrations in a free-living cohort performed poorly at an individual level but were capable of ranking individuals in categories of intake.

Conclusion

Urinary xylose exhibited a dose-response relationship with apple intake and performed well as a ranking biomarker in the population study. Other potential biomarkers were identified and future work will combine these with xylose in a biomarker panel which may allow for a more objective determination of individual intake

1. Introduction

It is well established that dietary intake and habits have an influence on both health and disease outcomes. [1] Elucidating this relationship is challenging due to a variety of essential nutrients and bioactive compounds available in foods and the diversity of interactions they may have within the body. [2] Furthermore, accurate assessment of dietary intake is difficult with several well-established limitations associated with traditional methods such as weighed food diaries, food frequencies questionnaires (FFQs) and 24-hour dietary recalls. Examples of such limitations include subjective estimation of portion sizes, recall bias and misreporting. [3–5] As a result there is increased interest in the use of food intake biomarkers which offer the potential of more objective and accurate measures of dietary intake. [6,7] Food intake biomarkers are single metabolites, or a combination of metabolites, reflecting the consumption of either a specific food or food group, displaying a clear time- and dose-response after intake. [8] Currently there are well-established biomarkers for salt, protein, sucrose/fructose intake and energy expenditure, [7,9–11] however, there is growing interest in identifying and using biomarkers associated with food intake. Metabolomics is a key emerging area in the discovery of food intake biomarkers. The human metabolome is influenced by an array of factors such as genetics, the microbiome, [12] as well as environmental factors including diet and lifestyle. [13] Application of metabolomics to the study of dietary biomarkers has resulted in several putative biomarkers that are related to food intake. To date, putative biomarkers have been discovered, for foods such as red meat, fish, citrus fruit, cruciferous vegetables and coffee. [14–18] Following discovery, it is critical that biomarkers are assessed for validity. Recently a set of criteria have been identified for validation of food intake biomarkers. [19] Furthermore, recent work has demonstrated the utility of biomarkers: proline betaine was used to determine intake of citrus fruit in a cross sectional study using calibration curves developed in a well-controlled intervention study. [20]

Fruits and vegetables are sources of multiple nutrients, bioactives [21] and non-nutritive compounds [22] and their consumption has been associated with multiple health benefits. [23–28] Apples are one of the most commonly consumed fruits in Europe [29] and they are a rich source of polyphenols but only a few studies have investigated biomarkers that reflect their intake. A frequently reported potential biomarker for apple consumption is phloretin, a dihydrochalcone whose glycosides are thought to be nearly unique in apples. [30] Previously phloretin has been significantly correlated with various levels of apple intake. [31–33] The polyphenol epicatechin, its conjugates and metabolites, have also been associated with apple intake [31,33] but this polyphenol is commonly found in multiple foods. Many other compounds have been identified as potential markers of apple consumption, mainly in studies looking at consumption of multiple fruits, including rhamnitol [34] and metabolic products of quercetin. [32] While some interesting biomarkers for apple intake have emerged in the literature, it is clear that more work is needed in order to discover robust biomarkers of apple intake, and steps taken to validate these biomarkers. The objective of the present research was to identify novel biomarkers of apple intake and examine the dose-response relationships in a habitual diet environment. To achieve this we employed a number of different study designs including a short acute intervention study to identify the biomarkers, a dose-response study to examine the response to intake of different amounts of apple and a cross-sectional study to examine the relationships in a free living population.

2. Materials and Methods

2.1 A-Diet Discovery study

Ethical approval for the A-DIET Discovery study was granted by the UCD Sciences Human Research Ethics Committee (LS-15-69-Brennan). The study was designed to identify novel biomarkers of nine commonly consumed foods (apples, broccoli, peppers, oranges, white bread, wholemeal bread, spaghetti, cheese and madeira cake.). The primary test food of interest for this research was apples (Supplementary Figure 1) and we report only the data with respect to the discovery of apple biomarkers. Recruitment was achieved through a variety of methods including posters around campus and surrounding businesses/areas and radio adverts. An outline of the recruitment response is shown in Supplementary Figure 2. The inclusion criteria included healthy, non-pregnant/lactating, non-smokers, an age range between 18 and 60 years old, and a body mass index (BMI) between 18.5 and 30 kg/m^2. Exclusion criteria included any diagnosed health condition (chronic or infectious diseases), consumption of medications/nutritional supplements or any allergies/intolerances to the test foods. Once informed consent was obtained, participants were invited to take part in a nine week study, where each week a different test food was consumed. The data obtained following consumption of apples and broccoli are presented in this study. Participants consumed 360 g of raw apples and 135 g of cooked broccoli.

Participants were asked to avoid consuming alcohol, medication and foods related to specific test food (apples or broccoli) for 24 hours prior to biofluid collection. On the morning of the study day participants provided fasting first void urine (in a chill bag on ice) and fasting blood samples. The test foods were consumed with 100mL of water in the morning. Urine and blood samples were collected at two and four hours following apple/broccoli consumption. During this four hour period participants refrained from eating, however water was provided. Following the four hour timepoint participants were allowed to consume their diet as normal with the exception of

continuing to avoid alcohol, medication and foods related to specific test food. The following day (24 hours post consumption) fasting first void urine and blood samples were collected.

Postprandial urine samples were collected in the test centre and placed on ice immediately. All blood samples were collected by venepuncture by trained phlebotomists in the test centre. A total of 12 mL of blood was collected at each time point into two tubes: one serum tube and one lithium heparin tube for plasma collection. Following collection all blood sample tubes were inverted gently. The serum tubes were allowed to clot at room temperature for 30 minutes while the lithium heparin tubes were placed on ice immediately. Once transferred to the laboratory urine and blood samples were centrifuged at 1800 x g for 10 min at 4 °C within one hour of collection. Samples were then aliquoted and stored at -80 °C until further analysis. Participants' baseline plasma samples were analysed for cholesterol, triglycerides and glucose on the Randox RX Daytona Platform.

2.2 A-DIET Dose-response Study

Inclusion criteria for the dose-response study were the same as for the Discovery study. Ethical approval was granted by the UCD Sciences Human Research Ethics Committee (LS-17-16-Brennan). Participants were randomly assigned to either a lunch (N = 27) or dinner (N = 34) test meal group and asked to partake in a five week study (Supplementary Figure 3). From the dinner group a total of two participants dropped out. Weeks 1, 3 and 5 were test weeks and Weeks 2 and 4 were rest periods. Each test week participants were provided with four portions of a test meal and asked to consume this test meal for four consecutive days. During these four days, participants were also asked to avoid consuming any other foods related to the test meal ingredients. Participants maintained their habitual diet (in compliance with food avoidance instructions) for all meals outside test meal and outside of test weeks. Participants were also asked to keep a four day food diary for each test week, to ensure compliance with food avoidance.

The test lunch comprised oranges, white bread, and a sugar-sweetened beverage. The test dinner consisted of broccoli, red bell peppers, apples, Madeira sponge cake, pasta, and cheddar cheese. Each week test meals were designed to deliver similar intakes of dietary energy and fibre, ~590 kcal and ~5 g fibre for lunches, and ~1285 kcal and ~16 g fibre for dinners, but portion sizes of each test food changed from week to week (high, medium or low portion), e.g. high orange and low bread portions in the same meal. Participants received the meals in random order. The high, medium and low apple portions were 300 g, 100 g and 50 g, respectively and broccoli portions were 153 g, 101 g, and 49 g. All test meals were provided for the duration of the study, on two occasions each test week, once at the beginning (Monday) and once in the middle (Wednesday). Fasting first void urine and fasting blood samples, were collected prior to commencement of the study and at the end of each test week.

Fasting first void urine and fasting blood samples were collected as described for the Discovery study. Centrifugation and storage of the samples also followed the protocol for the Discovery study.

2.3 National Adult Nutrition Survey

Dietary data and fasting urinary metabolomic profiles from the National Adult Nutrition Survey (NANS) were used to investigate the ability of candidate biomarkers identified in the Discovery study and characterized in the Dose-response study (e.g. xylose) to estimate apple intake. Details of the NANS study have been published elsewhere (https://www.iuna.net/). [35] Ethical approval for this study was granted by the University College Cork Clinical Research Ethics Committee of the Cork Teaching Hospitals (ECM 3 (p) 4 September 2008) and recruitment began in May 2008. Briefly, NANS collected data on habitual food and beverage consumption, lifestyle, health

indicators and attitudes to food and health in 1,500 adults, representative of the population during 2008-2010 in Republic of Ireland. A subset of this population (N = 565) was randomly selected, to ensure equal numbers of men and women across the age range (18-90 years) for metabolomic analysis as previously described. [20] A four day semi-weighed food diary was used to collect dietary data over four consecutive days. Detailed information on the type and amount of all foods, drinks and nutritional supplements consumed over the four days was recorded by participants in the food diary.

2.4 Metabolomic analysis of urine samples

Metabolomic analysis was performed using nuclear magnetic resonance (NMR) spectroscopy and liquid chromatography mass spectrometry (LC-MS) based techniques. For NMR analysis, urine samples were first defrosted and then prepared by addition of 250 µL phosphate buffer (0.2 mol KH_2PO_4/L, 0.8 mol K_2HPO_4/L) to 500 µL urine. After centrifugation at 5360 x g for 5 minutes at 4 °C, 10 µL sodium trimethylsilyl [2,2,3,3-$2H_4$]proprionate (TSP) and 50 µL deuterium oxide (D_2O) were added to 540 µL of the supernatant. Spectra were acquired on a 600 MHz Varian Spectrometer (Varian Limited, Oxford, United Kingdom) by using the first increment of a nuclear Overhauser enhancement spectroscopy pulse sequence at 25 °C. Spectra were acquired with 16,384 data points and 128 scans. Water suppression was achieved during the relaxation delay (2.5 s) and the mixing time (100 ms). All ^{1}H-NMR urine spectra were referenced to TSP at 0.0 parts per million (ppm) and processed manually with the Chenomx NMR Suite (version 7.7) by using a line broadening of 0.2 Hz, followed by phase and baseline correction. Data were normalized to the total area of the spectra integral. NMR spectra from the Discovery study were exported at high resolution using 7500 spectral regions. The water region was excluded. Identification of metabolites was achieved using the Chenomx library. To confirm correct assignment, where possible, a urine sample was spiked with an analytical standard and a ^{1}H NMR spectrum acquired.

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

For LC-MS analysis, a suite of biologically relevant urinary metabolites (arginine, taurine, 1-methyl-L-histidine, citrulline, creatinine, malic acid, methionine, citric acid, succinic acid, ketoglutaric acid, isoleucine, fumaric acid, leucine, methylmalonic acid, glutaric acid, adipic acid, hippuric acid, and pimelic acid, each at 10 µg/ml in 20% (v/v) EtOH/ Millipore H_2O) was used as a metabolite standard mix for quality control before and after each batch analysis. Five compounds (malic acid d_3, methionine d_3, mysristic acid ^{13}C, adipic acid d_4 and succinic acid d_4 10 µg/ml in 20% (v/v) EtOH/Millipore H_2O) were used as internal standards for quality control throughout each sample analysis. Test urine and pooled urine samples (Quality Control (QC) samples) were thawed on a roller and centrifuged at 5500 x g for 5 min, 4 °C. Urine samples (100 µL) were added to internal standard (100 µL) which had been placed on ice, vortexed 35 Hz for 10 s and centrifuged 2000 x g for 2 min. The supernatant was then transferred to vials with 250 µL inserts and placed into autosampler at 4 °C.

The sample analysis was performed using an Agilent LC-QTOF-MS, consisting of a 1290 Infinity II LC system and an Agilent Jetstream (AJS) Electrospray ionization (ESI) source coupled to a 6545 QTOF mass spectrometer. The chromatography was performed in reverse phase mode using Zorbax eclipse plus C18 (2.1 x 50 mm, 1.8 µm) column and Zorbax eclipse plus C18 2.1 x 5 mm 1.8 µm was used as a guard column. The sample run order was as follows: 3 blanks- 1 metabolite

standard- 2 conditioning QC samples- 3 pooled QC samples- 12 randomised test samples- 1 pooled QC sample- 1 metabolite standard. The LC parameters used for the analysis: Column temperature, 35 °C; Injection volume, 5 µL; Flow rate, 0.4 mL/min; Mobile phase, 0.1% formic acid in water (Eluent A) and 0.1% formic acid in acetonitrile/water (80:20) (Eluent B); Gradient conditions, 1% B (0–1.5 min), 11% B (1.5–9 min), 25% B (9–15 min), 50% B (15–18 min), 99% B (18–18.05min), 99% B (18.05–21 min), 1% B (21–21.05 min) and 1% B (21.05–23 min). The MS parameters used for the analysis: drying gas temperature, 325 °C; drying gas flow rate, 10 L/min; sheath gas temperature, 350 °C; sheath gas flow rate, 11 L/min; nebulizer pressure, 45 gauge pressure (pounds per square inch); capillary voltage, 3500 V; nozzle voltage, 1000 V; fragmentor voltage, 100 V; skimmer, 45 V. An analysis mass range of mass-to-charge-ratio (*m/z*) 50-1600 using 2-GHz extended dynamic range mode was used. Application of centroid mode at scan rate of 1 spectra/s was used to collect data. Both positive and negative ionization modes were selected for analysis.

Data were acquired using MassHunter acquisition B.08.00 software (B.08.00.8058.3 Sp1 Agilent Technologies) and were processed using MassHunter Qualitative Analysis (B.07.00 Sp2 Agilent Technologies) software. The molecular feature extractor (MFE) algorithm was performed to extract features characterized by retention time and *m/z*, representing adducts or isotopes of compounds, providing details of signal intensity and accurate mass. The target data type was set to small molecules and the peak height threshold value was set to 5000 counts. The isotopic peak spacing tolerance was 0.0025 *m/z* with a maximum charge state of 1 and isotope model was set to common organic molecules. The data output was then transformed to Compound Exchange Format (.cef) files. These .cef files were imported into MassProfiler (Version B.07.01; build 99.0 Agilent Technologies) which was used to align compounds and filter data. The alignment parameters were set with a retention time tolerance ± 0.3 min and a mass tolerance of ± 15 ppm +

2.0 mDa. The sample occurrence frequency was set to ≥25%, across all samples. Features with missing values, were replaced with half the minimum abundance value for that particular feature. Features abundances were normalised to osmolality before performing multivariate data analysis. Initial identification of interesting features was performed in MassHunter ID Browser (B 7.0.799.2 Agilent Technologies) by putative formula generation using accurate mass data and optional retention time with a mass tolerance of 5 ppm, positive ions +H, neutral loss H_2O and negative ions -H and these formulas compared against a local METLIN (METabolite LINk, https://metlin.scripps.edu) database. Urine samples were further analysed by LC-MS/MS using a targeted approach with the same chromatographic conditions as described above and with three collision energies 10 eV, 20 eV and 40 eV. MS/MS compound identification efforts included comparisons with authentic reference standards, fragmentation modelling using CFM ID [36] and the Human Metabolome Database (HMDB) [37] was used as a source Metabolite Identification levels are reported according to Metabolomics standards initiative (MSI): with four levels of confidence in metabolite identification: Level I (Identified compounds); Level II (putatively annotated compounds); Level III (putatively characterized compound classes); Level IV (Unknowns) [48].

2.5 Statistical analysis

Multivariate statistical analyses of both NMR and LC-MS data were performed using SIMCA 13 (SIMCA Version 13.0.3.0 Umetrics, AB). Data sets were scaled using Pareto scaling. Principal Component Analysis (PCA) was performed to examine trends and outliers in the data. Partial Least-Squares Discriminant Analysis (PLS-DA) was used to examine differences between spectral data and were validated using permutation testing. These models were used to generate variable

importance of projection (VIP) scores to identify the most influential variables with a cut-off of 2.5. One-way repeated measures analysis of variance (ANOVA) was performed on NMR data from the Dose-response study using SPSS 24.0 to compare average urinary xylose concentrations across the three different apple portions. A p-value < 0.05 was considered to indicate significance. Concentration curves were determined based on NMR data from the Dose-response study and apple intakes were estimated in the NANS cohort based on xylose concentrations in urine using curve-fitting software (WinCurveFit).

3. Results

3.1 Identification of urinary biomarkers of apple intake.

In total eight males and 12 females were recruited to the A-Diet Discovery study. One participant attended for baseline measurements only. From the remaining volunteers, 17 completed the apple test visit (Supplementary Figure 2). The demographics are presented in Table 1. The participants' mean age was 34 years old, with a mean BMI of 24.01 kg/m^2.

PCA (Figure 1a) was performed on the urinary dataset from ^{1}H-NMR spectra and displayed separation between the fasting samples and the postprandial samples after apple consumption. Robust PLS-DA models were constructed between fasting and postprandial samples: the fasting and four hour postprandial comparison model had the highest Q^2 value (0.79) (Figure 1b). Permutation tests were performed to check for overfitting of the model and indicated a robust model (intercepts R^2 = (0.0, 0.85), Q^2 = (0.0, -0.30)). Using VIP scores generated from PLS-DA models, a number of the spectral regions of potential interest, score ≥ 2.5, were selected for identification as potential biomarkers, a list of these regions can be seen in Table 2. Time-course analysis was performed for these spectral regions, comparing peak intensities following apple consumption across the four time points to intensities following consumption of the control food, broccoli. The analysis revealed multiple spectral regions which exhibited an increased peak intensity with time following apple consumption only (Figure 2). A number of these spectral regions of interest were assigned to xylose and identification confirmed through comparison with an authentic analytical standard (Supplementary Figure 4). Other potential biomarkers were identified as 3-hydroxyisovalerate, acetylsalicylate and glycine (Figure 2). Profiling of urinary xylose concentrations revealed that there was a significant increase in the postprandial samples (Figure 3).

Using the LC-MS data a total of 4,023 features were obtained from urine samples following apple and broccoli consumption in positive mode and 4,695 features were obtained in negative mode. Analysis of datasets from fasting and four hour post consumption in positive mode was carried out in SIMCA. PCA revealed separation of the timepoints. Robust PLS-DA models were obtained (PLS-DA score $R^2X = 0.37$, $Q^2 = 0.71$) (Supplementary Figure 5). Similarly, analysis of fasting and four hour post-consumption datasets from negative mode also revealed robust models of discriminating features: PLS-DA (score $R^2X = 0.35$, $Q^2 = 0.73$).

From the PLS-DA models a VIP list was generated, and 214 features were selected for examination (84 from positive mode and 130 from negative mode). Features with errors (at 95% confidence) crossing zero were removed resulting in 179 features (70 positive and 109 negative) remaining for further investigation. Time-course analysis was also performed for these features, comparing peak intensities after apple and broccoli consumption across the four time points (Figures 4 and 5). Thus, 40 features in positive mode and 56 in negative mode with differential kinetic profiles were identified from LC-MS data. Features were sorted by VIP score and many features with high VIP scores also displayed discriminating time profiles when compared to the control food (Figure 4 and 5). A molecular formula for each feature was generated by MassHunter using single MS accurate mass data for the molecular ion and its isotopes. Five and 10 features of interest, in positive and negative mode respectively, were selected for LC-MS/MS for more complete metabolite identification (Table 3).

Mass spectral data generated from MS/MS was compared with candidate spectra from CFM-ID database for fragmentation matches and putative identifications were achieved for 10 of the 15 features. Metabolite identification was confirmed for two features using authentic standards, 4-

pyridoxic acid and epicatechin sulfate. Matching of MS/MS fragments confirmed the identification of 4-pyridoxic acid (Supplementary Figure 6). Epicatechin sulfate with *m/z* 369.02869 was identified in negative ionization mode. The identification was based on the MS/MS fragmentation match in CFM-ID and Metlin. An authentic standard was available for the parent compound, epicatechin, and the fragments obtained from the standard (*m/z* 289.0715, 245.0815, 203.0708, 151.0397 and 109.0293) matched with those of the sample in terms of percentage of intensity and fragmentation pattern. The difference in mass between epicatechin (~290 Da) and epicatechin sulfate (~370 Da) is ~80 Da, which corresponds to the mass of a sulfate group (SO_3) (Supplementary Figure 7). The excretion kinetics of epicatechin sulfate displayed a peak excretion at four hours and exhibited a differential time course when compared with the test food broccoli (Figure 4). An interesting compound from the positive ionisation mode was 2,6-dimethyl-2-(2-hydroxyethyl)-3,4-dihydro-2H-1-benzopyran with *m/z* 207.13795. The identification was based on the input MS/MS spectra scored against candidate spectra generated and calculated from the HMDB database using CFM-ID (Score = 0.4445, based on DotProduct+Metadata function). The major fragments *m/z* 207.1385, 189.1279, 161.0966, and 121.0653 matched well between the candidate spectra and the input spectra at low collision energy (10 eV). The compound identification was also supported by comparison with LC-MS/MS spectra of a core compound (3,4-dihydrocoumarin), from the HMDB spectral library, obtained at 10, 20 and 40 eV (HMDB ID: HMDB0036626). The excretion kinetics of 2,6-dimethyl-2-(2-hydroxyethyl)-3,4-dihydro-2H-1-benzopyran displayed a peak excretion at four hours and exhibited a differential time course when compared with the test food broccoli (Figure 5).

3.2 Confirmation of identified biomarkers of apple intake.

In order to confirm a dose-response for the putative biomarkers, metabolites were examined in urine samples following consumption of different portions of apples. The study population

demographics of the Dose-response study are reported in Table 1. The participants' mean age was approximately 28 years old. The average participant BMI and waist-to-hip ratio measurements were 23.96 kg/m2 and 0.84 respectively.

Fasting urinary xylose concentrations were determined, using NMR, following intake of low (50 g), medium (100 g) and high (300 g) portions of apple. The average urinary xylose concentrations increased as apple intake increased (from 0.65 to 1.34 μm/mOsm/kg) (Supplementary Table 1). Repeated measures ANOVA indicated that xylose exhibits a dose-response relationship to apple intake in fasting urine samples ($p < 0.001$). The fit of the data resulted in the following equation: $y = 0.0028x + 0.4997$; $R^2 = 0.9997$. Features which exhibited an increased and differential time-course following apple consumption using LC-MS analysis were also examined for a dose-response relationship (Figure 6). Extract ion chromatograms (EIC) were examined for the features of interest (epicatechin sulfate and 2,6-dimethyl-2-(2-hydroxyethyl)-3,4-dihydro-2H-1-benzopyran) and peak area values for each feature determined following consumption of the different apple portions. Both 2,6-dimethyl-2-(2-hydroxyethyl)-3,4-dihydro-2H-1-benzopyran and epicatechin sulfate exhibited a dose-response, with peak area increasing as apple consumption increased.

3.3 Estimation of apple intake in an independent cohort.

Fasting urinary xylose concentrations were determined for the NANS participants (N=565) using NMR and ranged from 0.07 to 2.19 mM. The calibration curve determined using A-Diet Dose-response participants fasting urine xylose concentrations was used to estimate apple intake for the NANS participants. Biomarker-estimated intakes of apples ranged from 0.2 g to 1.32 kg per day for consumers.

Comparison of the biomarker-estimated intakes to the self-reported mean daily total apple intake data revealed a low but significant correlation between the two measures of intake (Pearson correlation coefficient $r = 0.141$, $p = 0.001$). Participants were classified into categories of apple intake based on their urinary xylose concentrations. Comparison of this classification to that achieved with self-reported data revealed good agreement between the two methods (Table 4). At the population level similar numbers of individuals were classified into the different categories: for example 379 individuals were classified as consuming < 50 g using self-reported data and 381 individuals was classified into the same category using the biomarker data. However, agreement at an individual level was low as indicated by the low correlation coefficient and low agreement at the individual level.

4. Discussion

The identification of validated biomarkers of specific food intake is a research area of increasing interest. The current study used a metabolomics-based approach to identify three potential urinary biomarkers of apple intake, xylose, 2,6-dimethyl-2-(2-hydroxyethyl)-3,4-dihydro-2H-1-benzopyran, and epicatechin sulfate. Importantly, the potential biomarkers were confirmed in a dose-response study, demonstrating that the biomarkers are both robust and responsive to increasing portions. The urinary biomarker xylose was confirmed in an independent cross-sectional cohort. Although the low correlation coefficients indicate that xylose did not perform well at estimating individual intake, xylose was capable of ranking individuals into quartiles of apple intake.

This study applied an untargeted multi-platform metabolomic approach for the discovery of novel food intake biomarkers identifying many discriminating compounds between fasting and four hour postprandial metabolomic profiles. Xylose was identified as a candidate apple biomarker using an NMR discovery approach. Xylose is a sugar found in edible fruits, it is among the main sugars found in apples [38] and is a proposed urinary biomarker for apples and fruit consumption. [39] Our literature search did not reveal any other studies which examined xylose solely as a biomarker of apple intake but it is well known that apples are a source of xylose. A potential limitation of this biomarker is that there are multiple fruit and vegetable sources of xylose. [39] This may explain the poorer agreement with self-reported data at an individual level. However, xylose performed well as an intake ranking biomarker. Furthermore, although agreement with self-reported intake was varied there is the potential to use xylose as a biomarker associated with apple intake for examining health relationships.

2,6-dimethyl-2-(2-hydroxyethyl)-3,4-dihydro-2H-1-benzopyran, and epicatechin sulfate were both identified as potential biomarkers of apple intake using an LC-MS approach. 2,6-dimethyl-2-(2-hydroxyethyl)-3,4-dihydro-2H-1-benzopyran belongs to a class of compounds called 1-benzopyrans, a core structure of which is 2H-1-benzopyran-2-one (coumarin). Coumarins are lactones derived from hydroxycinnamic acids, a flavonoid, [40,41], which are well-documented in many plant-based foods including apples. [42,43] A previous study reported that plasma concentrations of hydroxycinnamic acids (including phloretin and coumaric acid) significantly increased after consumption of one kilogram of apple, peaking between 1-3 hours following apple consumption, in a short-term intervention study. [43] In this study, the urinary levels of 2,6-dimethyl-2-(2-hydroxyethyl)-3,4-dihydro-2H-1-benzopyran peaked at four hours following apple consumption. The difference in excretion peaks is likely due to the different biofluids used but also because this metabolite is a hydroxycinnamic-related compound and not a pure hydroxycinnamic acid. Urinary metabolites of dietary polyphenols, such as hydroxycinnamic acid, are difficult to identify as they undergo extensive modification during absorption and their production is influenced by composition of intestine flora. [44] Another study also identified m-coumaric acid in 24 hour urine samples and found urinary concentrations to be correlated, along with other flavonoids, with apple intake from two day dietary records. [32]

Epicatechin sulfate is the sulfate form of epicatechin, a flavonoid, which has multiple plant sources. [31,42] Methyl(epi)catechin sulfate has previously been associated with apple and pear intake. [31] When comparing urinary metabolome profiles with data from a 24 hour dietary recall, Edmands *et al.* [31] found methyl(epi)catechin sulfate to be associated with intake of apples and pears (ROC (receiver operator characteristic) AUC (area under the curve) for sensitivity and specificity to food = 70.7%). However this compound was also associated with intake of tea and chocolate (sensitivity and specificity ROC AUCs = 66.9% and 74.4% respectively), which is

expected as there are multiple food sources of this flavonoid. A study performed by Saenger *et al.* [33] found that epicatechin increased following apple intake, peaking after three hours postprandial and was capable of differentiating between low/medium and high apple intake (200/400 g versus 790 g respectively). Our study is in line with previous research as urinary epicatechin sulfate levels peaked at four hours following apple consumption. Interestingly our data demonstrated that a dose response was evident from lower intakes of apple. Further work is warranted in examining the potential of epicatechin and its metabolites as part of a larger panel of biomarkers of apple intake. Previous literature has also identified biomarkers such as phloretin and rhamnitol as candidate biomarkers of apple intake. [31-34] Due to different study designs, analytical procedures and apple variety it is not surprising to obtain different biomarkers. Previous work has demonstrated that rhamnitol levels in apples vary due to geographic location and highlights the need for validation of candidate biomarkers. [45]

The importance of having detailed assessments of potential biomarkers is paramount to establishing them not only as measures of dietary compliance but also their usefulness in studying diet-disease relationship. Although many potential biomarkers have been reported there is a lack of research demonstrating their validity and subsequently their utility. The evaluation of a dose-response relationship is essential to demonstrating the utility of a biomarker. [19] Recent studies by Gibbons *et al.* [20] in citrus fruit, Yin *et al*. [46] in chicken intake and Garcia-Perez *et al.* [47] in grape intake have shown that well-defined biomarkers exhibiting dose-response relationships can be used to estimate dietary intake. Our present study results demonstrate that urinary xylose, epicatechin sulfate and 2,6-dimethyl-2-(2-hydroxyethyl)-3,4-dihydro-2H-1-benzopyran levels all increased as the portion of apple increased, illustrating a dose-response relationship between actual food intake and dietary biomarkers in a mixed meal setting. The calibration curves developed as a result of this research were capable of grouping individuals into categories of apple

intake based on urinary xylose concentrations. However, the agreement at an individual level was lower than we had previously observed with other biomarkers indicating that xylose alone is not sufficient to determine individual intake. [20] Future work will examine the potential of xylose in combination with the other biomarkers that demonstrated dose response relationships to determine apple intake.

A potential limitation of this research is that the discovery, dose-response and independent cohort studies were carried out only in the Republic of Ireland and therefore confirmation of these findings would be required in other ethnic populations. This study has multiple strengths associated with it: a range of different study designs were used to examine the biomarkers in different scenarios. Confirmation of a dose-response in the background of habitual diet is a strength worth highlighting. The use of a multi-platform approach allowed a diverse range of metabolites to be measured.

Conclusion

The present study has illustrated the successful implementation of an untargeted metabolomics approach in search of dietary biomarkers of apple intake. Urinary xylose exhibited a dose response relationship with apple intake and correlated with self reported intake in a cross sectional study. Importantly it performed excellently as a ranking biomarker in the population study. Other potential biomarkers of apple intake were identified and future work will examine if combining these with xylose into a panel of biomarkers will work as an objective measure of apple intake. The work presented here is an important step in the development of more objective measures of intake and will pave the way for the use of biomarkers in nutritional epidemiology.

Figure Legends

Figure 1. (A) PCA of 1H-NMR urine data of time point 0 compared with the 4 h post-consumption time point for the apple visit. R^2X, 0.38; Q^2, 0.15. $R^2X(1)$, 0.16, $R^2X(2)$, 0.12. t[1], Principal component 1; t[2], Principal component 2; (B) PLS-DA of 1H-NMR urine data of time point 0 compared with the time point 4 h post-consumption for the apple visit. R^2X, 0.31; Q^2, 0.79;

Figure 2. Kinetics plots of the urinary NMR spectral intensities of identified potential biomarkers over the four time-points (0, 2, 4, 24h) following consumption of apples and broccoli in separate test sessions. Values are means ± SEMs

Figure 3. Xylose concentration in urine across the four time-points (0, 2, 4, 24 hours), following consumption of apples and broccoli in separate test sessions from NMR analysis. Values are mean ±SEM.

Figure 4. Urinary profiles of compounds selected for further identification by LC-MS/MS-negative mode. Timeline plot showing features of interest following apple consumption compared to the control food, broccoli. Values are means ± SEM.

Figure 5. Urinary profiles of compounds selected for further identification by LC-MS/MS-positive mode. Timeline plot showing features of interest following apple consumption compared to the control food, broccoli. Values are means ± SEM

Figure 6. Dose-response relationships of urinary glucodistylin (A); epicatechin sulfate (B); [(4-{3-[2-(2,4-dihydroxyphenyl)-2-oxoethyl]-DHMPMB-SA (C) and 2,6-dimethyl-2-(2-hydroxyethyl)-3,4-dihydro-2H-1-benzopyran (G) dictaquinazol C (D); 1-(malonylamino)-cyclopropanecarboxylic acid ((1-malonylamino)-CCA) (E); L-suberyl carnitine (H) and 4-Pyridoxic acid (F) peak areas across the different portions of apple intake. Values are means ± SEMs (n = 37). X-axis values represent apple intake (g/d); Y-axis values represent peak area. [(4-{3-[2-(2,4-dihydroxyphenyl)-2-oxoethyl]-DHMPMB-SA;[(4-{3-[2-(2,4-dihydroxyphenyl)-2-oxoethyl]-4,6-dihydroxy-2-methyoxyphenyl}-2-methylbut-2-en-1-yl)oxy] sulfonic acid.

Author Contributions

A.E.M. conducted the discovery and dose-response studies, analyzed data, and prepared the manuscript. C.C. and P.S.C.S.H. acquired and analysed data and prepared the manuscript. D.G.P. acquired part of the metabolomics data. L.B. designed research, analysed data and prepared the manuscript. B.A.M., A.N, J.W and A.F contributed data from the NANS study. All authors read and accepted the final version of the manuscript.

Acknowledgements

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

The authors would like to thank all volunteers for their commitment and patience during the study.

Conflict of interest

The authors have no conflict of interest.

References

[1]. SA. Bingham, *Public Health Nutr*. **2002**, *5,* 821.

[2]. A. Scalbert, L. Brennan, C. Manach, C. Andres-Lacueva, LO. Dragsted, J. Draper, SM. Rappaport, JJJ. Van Der Hooft, DS. Wishart, *Am J Clin Nutr.* **2014**, *99*, 1286.

[3]. V. Kipnis, D. Midthune, L. Freedman, SA. Bingham, NE. Day, E. Riboli, P. Ferrari, RJ. Carroll, *J Public Health Nutr.* **2002**, *5*, 915.

[4]. JS. Shim, K. Oh, HC. Kim, *Epidemiol Health.* **2014,** *22*, doi: 10.4178/epih/e2014009.

[5]. EJ. Dhurandhar, J. Dawson, A. Alcorn, LH. Larsen, EA. Thomas, M. Cardel, AC. Bourland, A. Astrup, MP. St-Onge, JO. Hill, CM. Apovian, JM. Shikany, DB. Allison, *Am J Clin Nutr.* **2014**, *100,* 507.

[6]. M. Jenab, N. Slimani, M. Bictash, P. Ferrari, SA. Bingham, *Hum Genet.* **2009**, *125*, 507.

[7]. H. Gibbons, BA. McNulty, AP. Nugent, J. Walton, A. Flynn, MJ. Gibney, L. Brennan, *Am J Clin Nutr.* **2015**, *101,* 471.

[8]. Q. Gao, G. Praticò, A. Scalbert, G. Vergères, M. Kolehmainen, C. Manach, L. Brennan, LA. Afman, DS. Wishart, C. Andres-Lacueva, M. Garcia-Aloy, H. Verhagen, EJM. Feskens, LO. Dragsted, *Genes Nutr.* **2017**, *12*, 34.

[9]. LS. Freedman, JM. Commins, JE. Moler, W. Willett, LF. Tinker, AF. Subar, D. Spiegelman, D. Rhodes, N. Potischman, ML. Neuhouser, AJ. Moshfegh, V. Kipnis, L. Arab, RL. Prentice, *Am J Epidemiol.* **2014**, *180,* 172.

[10]. N. Tasevska, SA. Runswick, A. McTaggart, SA. Bingham, *Cancer Epidemiol Biomarkers Prev.* **2005**, *14,* 1287.

[11]. L. Ovesen, H. Boeing, *Eur J Clin Nutr.* **2002**, *56,* 12.

[12]. S. Fujisaka, J. Avila-Pacheco, M. Soto, A. Kostic, JM. Dreyfuss, H. Pan, S. Ussar, E. Altindis, N. Li, L. Bry, CB. Clish, CR. Kahn, *Cell Rep.* **2018**, *22*, 3072.

[13]. R. Zamora-Ros, D. Achaintre, JA. Rothwell, S. Rinaldi, N. Assi, P. Ferrari, M. Leitzmann, MC. Boutron-Ruault, G. Fagherazzi, A. Auffret, T. Kühn, V. Katzke, H. Boeing, A. Trichopoulou, A. Naska, E. Vasilopoulou, D. Palli, S.Grioni, A. Mattiello, R. Tumino, F. Ricceri, N. Slimani, I. Romieu, A. Scalbert, *Sci Rep.* **2016**, *6*, 1.

[14]. AJ. Cross, JM. Major, R. Sinha, *Cancer Epidemiol Biomarkers Prev.* **2011,** *20*, 1107.

[15]. AJ. Lloyd, G. Favé, M. Beckmann, W. Lin, K. Tailliart, L. Xie, JC. Mathers, J. Draper, *Am J Clin Nutr.* **2011**, *94,* 981.

[16]. SS. Heinzmann, IJ. Brown, Q. Chan, M. Bictash, M. Dumas, S. Kochhar, J. Stamler, E. Holmes, P. Elliott, JK. Nicholson, *Am J Clin Nutr.* **2010**, *92*, 436.

[17]. WMB. Edmands, OP. Beckonert, C. Stella, A. Campbell, BG. Lake, JC. Lindon, E. Holmes, NJ. Gooderham, *J Proteome Res.* **2011,** *10,* 4513.

[18]. JA. Rothwell, Y. Fillâtre, JF. Martin, B. Lyan, E. Pujos-Guillot, L. Fezeu, S. Hercberg, B. Comte, P. Galan, M. Touvier, C. Manach, *PLoS One.* **2014**, *9*, doi: 10.1371/journal.pone.0093474.

[19]. LO. Dragsted, Q. Gao, A. Scalbert, G. Vergères, M. Kolehmainen, C. Manach, L. Brennan, LA. Afman, DC. Wishart, G. Praticò, *Genes Nutr.* **2018**, *13,* 1.

[20]. H. Gibbons, CJR. Michielsen, M. Rundle, G. Frost, BA. McNulty, AP. Nugent, J. Walton, A. Flynn, MJ. Gibney, L. Brennan, *Mol Nutr Food Res.* **2017**, *61,* 1.

[21]. SE. Nielsen, R. Freese, P. Kleemola, M. Mutanen. *Cancer Epidemiol Biomarkers Prev* **2002,** *11*, 459.

[22]. MC. Walsh, L. Brennan, E. Pujos-Guillot, JL. Sébédio, A. Scalbert, A. Fagan, DG. Higgins, MJ. Gibney, *Am J Clin Nutr.* **2007**, *86*, 1687.

[23]. D. Panagiotakos, C. Pitsavos, P. Kokkino, C. Chrysohoou, M. Vavuranakis, C. Stefanadis, P. Toutouzas, *Nutr J.* **2003**, *2*, 2.

[24]. U. Nothlings, MB. Schulze, C. Weikert, H. Boeing, YT. van der Schouw, C. Bamia, V. Benetou, P. Lagiou, V. Krogh, JWJ. Beulens, PHM. Peeters, J. Halkjaer, A. Tjonneland, R. Tumino, S. Panico, G. Masala, F. Clavel-Chapelon, B. de Lauzon, MC. Boutron-Ruault, MN. Vercambre, R. Kaaks, J. Linseisen, K. Overvad, L. Arriola, E. Ardanaz, CA. Gonzalez, MJ. Tormo, SA. Bingham, KT. Khaw, TJA. Key, P. Vineis, E. Riboli, P. Ferrari, P. Boffetta, HB. Bueno-de-Mesquita, DL. van der A, G. Berglund, E. Wirfalt, G. Hallmans, I. Johansson, E. Lund, A. Trichopoulo, *J Nutr.* **2008**, *138,* 775.

[25]. JM. Nuñez-Cordoba, A. Alonso, JJ. Beunza, S. Palma, E. Gomez-Gracia, MA. Martinez-Gonzalez, *Eur J Clin Nutr.* **2009,** *63,* 605.

[26]. EM. Alissa, GM. Ferns, *Crit Rev Food Sci Nutr*. **2017**, *57*, 1950.

[27]. AR. Vieira, S. Vingeliene, DSM. Chan, D. Aune, L, Abar, D, Navarro Rosenblatt D, DC. Greenwood, T. Norat, *Cancer Med.* **2015**, *4*, 136.

[28]. NP. Gullett, ARM. Ruhul Amin, S. Bayraktar, JM. Pezzuto, DM. Shin, FR. Khuri, BB. Aggarwal, YJ. Surh, O. Kucuk, *Semin Oncol.* **2010**, *37*, 258.

[29]. Konopacka D, Jesionkowska K, Kruczyńska D, Stehr R, Schoorl F, Buehler A, S. Egger, S. Codarin, C. Hilaire, I. Höller, W. Guerra, A. Liverani, F. Donati, S. Sansavini, A. Martinelli, C. Petiot, J. Carbó, G. Echeverria, I. Iglesias, J. Bonany, *Appetite.* **2010**, *55,* 478.

[30]. SC. Marks, W. Mullen, G. Borges, A. Crozier, *J Agric Food Chem.* **2009**, *57*, 2009.

[31]. WMB. Edmands, P. Ferrari, JA. Rothwell, S. Rinaldi, N. Slimani, DK. Barupal, C. Biessy, M. Jenab, F. Clavel-Chapelon, G. Fagherazzi, MC. Boutron-Ruault, VA. Katzke, T. Kühn, H. Boeing, A. Trichopoulou, P. Lagiou, D. Trichopoulos, D. Palli, S. Grioni, R. Tumino, P. Vineis, A. Mattiello, I. Romieu, A. Scalbert, *Am J Clin Nutr*. **2015**, *102*, 905.

[32]. L. Mennen, D. Sapinho, H. Ito, S. Bertrais, P. Galan, S. Hercberg, A. Scalbert, *Br J Nutr*. **2006**, *96*, 191.

[33]. T. Saenger, F. Hübner, HUU. Humpf, *Mol Nutr Food Res*. **2017**, *61,* 1.

[34]. JM. Posma, I. Garcia-Perez, JC. Heaton, P. Burdisso, JC. Mathers, J. Draper, M. Lewis, JC. Lindon, G. Frost, E. Holmes, JK. Nicholson, *Anal Chem.* **2017**, *89*, 3300.

[35]. KD. Cashman, S. Muldowney, BA. McNulty, A. Nugent, AP. Fitzgerald, M. Kiely, J. Walton, MJ. Gibney, A. Flynn, *Br J Nutr.* **2013**, *109,* 1248.

[36]. Y. Djoumbou-Feunang, A. Pon, N. Karu, J. Zheng, C. Li, D. Arndt, M. Gautam, F. Allen, DS. Wishart, *Metabolites*. **2019**, *9*, 72.

[37]. DS. Wishart,YD. Feunang, A. Marcu, AC. Guo, K. Liang, R. Vázquez-Fresno, T. Sajed, D. Johnson, C. Li, N. Karu, Z. Sayeeda, E. Lo, N. Assempour, M. Berjanskii, S. Singhal, D. Arndt, Y. Liang, H. Badran, J. Grant, A. Serra-Cayuela, Y. Liu, R. Mandal, V. Neveu, A. Pon, C. Knox, M. Wilson, C. Manach, A. Scalbert, *Nucleic Acids Res*. **2018**, *46*, 608.

[38]. Y. Zhang, P. Li, L. Cheng, *Food Chem.* **2010**, *123*, 1013.

[39]. CI. Mack, CH. Weinert, B. Egert, PG. Ferrario, A. Bub, I. Hoffmann, B. Watzl, *Am J Clin Nutr*. **2018**, *108*, 502.

[40]. Y. Yu, L. Wu, X. Zou, X. Dai, K. Liu, H. Su, *J Phys Chem A*. **2013,** *117*, 7767.

[41]. LMR. da Silva, EAT. de Figueiredo, NMPS. Ricardo, IGP. Vieira, RW. de Figueiredo, IM. Brasil, CL. Gomes, *Food Chem.* **2014**, *143*, 398.

[42]. M. Ulaszewska, N. Vázquez-Manjarrez, M. Garcia-Aloy, R. Llorach, F. Mattivi, LO. Dragsted, G. Praticò, C. Manach, *Genes Nutr*. **2018**, *13*, 29.

[43]. BA. Stracke, CE. Rufer, A. Bub, S. Seifert, FP. Weibel, C. Kunz, B. Watzl, *Eur J Nutr*. **2010**, *49*, 301.

[44]. MR. Olthof, PCH. Hollman, MB. Katan, *J Nutr*, **2001**, *131*, 66.

[45] S. Tomita, T. Nemoto, Y. Matsuo, T. Shoji, F. Tanaka, H. Nakagawa, H. Ono, J. Kikuchi, M. Ohnishi-Kameyama,Y. Sekiyama*, Food Chem*. **2015**;174, 163.

[46]. X. Yin, H. Gibbons, M. Rundle, G Frost, BA McNulty, AP Nugent, J Walton, A Flynn, MJ Gibney, L Brennan, *J Nutr*. **2017**, 147, 10.

[47]. I. Garcia-Perez, JM. Posma, ES. Chambers, JK. Nicholson, JC. Mathers, M. Beckmann, J. Draper, E. Holmes, G. Frost, *J Agric Food Chem*. **2016**, *64*, 2423.

[48] L.W. Sumner, A, Amberg, D. Barrett, MH. Beale, R. Beger, CA, Daykin, TW. Fan, O. Fiehn, R. Goodacre, JL. Griffin, T. Hankemeier, N. hardy, J. Harnly, R. Higashi, J. Kopka, AN. Lane, JC. Lindon, P. Marriott, AW. Nicholls, MD. Reily, JJ. Thaden, MR. Viant *Metabolomics* **2007**, *3*, 211-221.

Table 1 Participant demographics for both A-DIET Discovery and Dose-response Studies.

Characteristics	Discovery Study (N=17)	Dose-response Study (N=32)
Gender	8M 9F	16M 16F
Age (yrs.)	34 ± 12	29 ± 10
Anthropometrics		
BMI (kg/m^2)	24.01 ± 2.98	23.96 ± 2.99
W:H	0.77 ± 0.07	0.84 ± 0.06

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

Table 2. Spectral regions of interest, their VIP scores and potential identifications

Spectral Region(ppm)[1]	0h vs 4h VIP Score	0h vs 2h VIP Score	Potential Identification
3.554	7.57	-	Glycine
3.643	5.78	4.88	Xylose
3.312	5.56	5.73	Xylose
3.311	5.28	4.68	Xylose
3.553	5.28	_	Glycine
3.642	5.13	4.56	Xylose
2.336	5.07	_	Acetylsalicylate
2.337	5.03	_	Acetylsalicylate
3.623	4.98	4.20	Xylose
3.624	4.98	4.71	Xylose
3.430	4.84	_	Xylose
3.225	4.81	4.31	Xylose
3.684	4.70	_	Xylose
3.705	4.64	5.80	Xylose
3.313	4.48	5.75	Xylose
3.704	4.93	5.58	Xylose
1.262	4.65	5.28	3-Hydroxyisovalerate
3.688	2.64	5.04	Xylose
1.261	4.51	4.97	3-Hydroxyisovalerate
3.706	3.57	4.92	Xylose
3.635	4.87	_	Xylose

[1] A selection of top-ranking regions are presented. These regions represent peaks and one metabolite can have multiple peaks depending on the chemical structure of the metabolite.
VIP; Variable Importance of Projection

Table 3. LC-MS/MS of compounds of interest.

RT	Mass	Prec. m/z	Ion	MS/MS	VIP Score	Putative Name	Putative formula
10.3	400.174	399.1664	$[M-H]^-$	366.037, 334.277, 260.362, 193.014	8.22	Unidentified[a]	$C_{19}H_{28}O_9$
2.2	165.079	166.08642	$[M+H]^+$	120.044, 100.078, 75.044, 57.034	7.16	Ethyl 2-aminobenzoate[b,II]	$C9H_{11}NO_2$
1.8	228.148	229.15491	$[M+H]^+$	229.155, 170.080, 142.086, 114.054, 100.075	6.02	Pro Leu[b,c, II]	$C_{11}H_{20}N_2O_3$
0.6	186.029	185.0221	$[M-H]^-$	169.061, 149.045, 145.061, 131.033	5.47	1-(Malonylamino)cyclopropanecarboxylic acid[a, II]	$C_7H_9NO_5$
11.4	370.036	369.02869	$[M-H]^-$	369.029, 289.072, 245.082, 203.071, 151.039, 123.044	4.98	Epicatechin Sulfate[d, I]	$C_{15}H_{14}O_9S$
10.3	206.131	207.13795	$[M+H]^+$	117.299, 103.051	4.62	2,6-Dimethyl-2-(2-hydroxyethyl)-3,4-dihydro-2H-1-benzopyran[b, II]	$C_{13}H_{18}O_2$
4.6	229.005	227.99752	$[M-H]^-$	227.997, 210.948, 172.991, 148.040, 137.070	3.94	Dopachrome o-semiquinone[b, II]	$C_9H_8NO_4$
1.8	183.053	182.04605	$[M-H]^-$	155.057, 147.066, 129.055, 115.039	3.41	4-Pyridoxic acid[d, I]	$C_8H_9NO_4$
10.4	317.184	318.19132	$[M+H]^+$	221.068, 132.082, 106.069	3.73	L-Suberyl carnitine[a, II]	$C_{15}H_{27}NO_6$
1.1	148.037	147.0301	$[M-H]^-$	147.030, 129.018, 103.039, 87.008,	2.98	D-Xylono-1,5 lactone[b, II]	$C_5H_8O_5$
11.3	466.111	465.104	$[M-H]^-$	409.2206, 328.015, 245.007, 188.912, 162.838	2.97	Glucodistylin[b, II]	$C_{21}H_{22}O_{12}$
17.2	342.122	341.11477	$[M-H]^-$	262.143, 217.756, 183.011, 135.045, 122037	2.92	Dictyoquinazol C[c, II]	$C_{18}H_{18}N_2O_5$
14.7	454.093	453.08611	$[M-H]^-$	-	2.69	[(4-{3-[2-(2,4-dihydroxyphenyl)-2-oxoethyl]-4,6-dihydroxy-2-methyoxyphenyl}-2-methylbut-2-en-1-yl)oxy] sulfonic acid[b, II]	$C_{16}H_{22}O_{10}$
0.6	155.07	154.06235	$[M-H]^-$	154.061, 137.035, 119.034, 109.040, 96.965	2.69	L-2-Amino-3-(1-pyrazolyl) propanoic acid[b, II]	$C_6H_9N_3O_2$
12.5	316.142	317.14955	$[M+H]^+$	281.144, 170.116, 130.049, 97.101	2.62	Glucosyl (E)-2,6-Dimethyl-2,5-heptadienoate[b, II]	$C_{15}H_{24}O_7$

RT, retention time: Prec.m/z, precision mass to charge ratio; MS/MS, tandem mass spectrometry, VIP, variable importance of projection. Putative identification from HMDB after analysis by CFM. LC-MS/MS in positive mode[M+H]+ and negative mode [M-H]-. Analysis shows putative formula and results of interrogation of the METLIN and HMDB databases showing CAS-ID where found. Note: a) MassHunter database, b) CFM-ID, c) METLIN database, d) authentic standard, I: level I identification, II: level II identification

Table 4. NANS participants grouped into intake categories according to self-reported apple intake (g/d) and biomarker estimated intake.

Apple (g/d)	Self-Reported Apple Intake (N)	Biomarker-Estimated Apple Intake (N)	Agreement (N)
0-50 g	379	381	276
51-300 g	154	142	47
>300 g	13	23	9

Agreement is the number of individuals classified by both methods into the same category of intake using self-reported data as the reference.

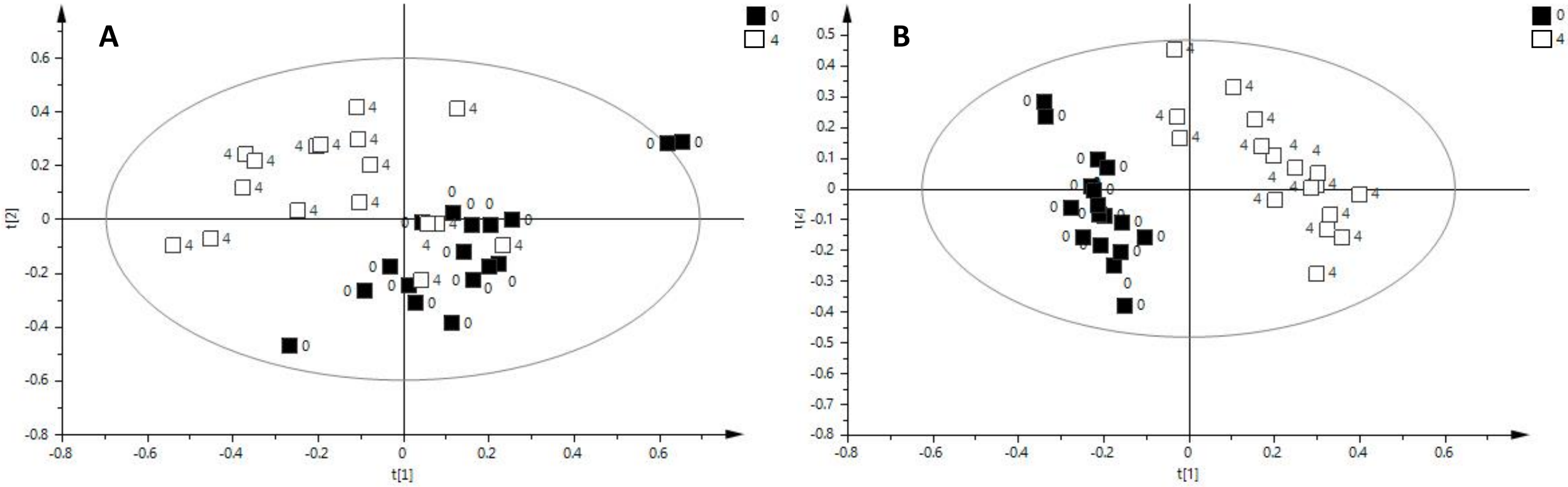
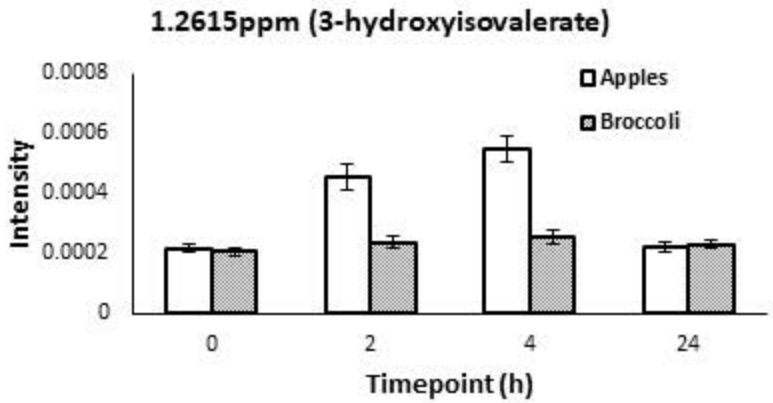
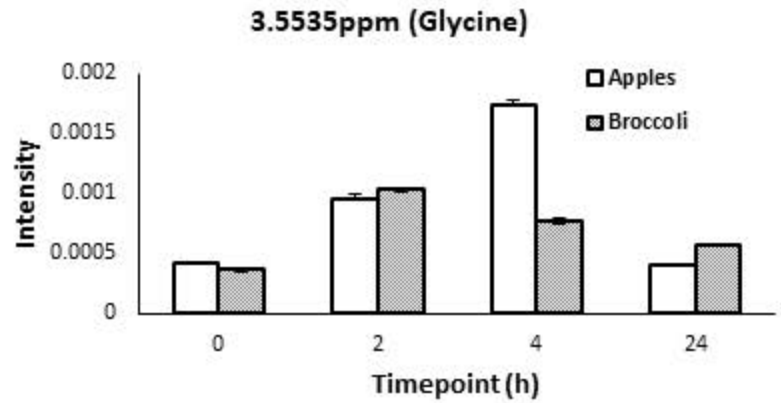
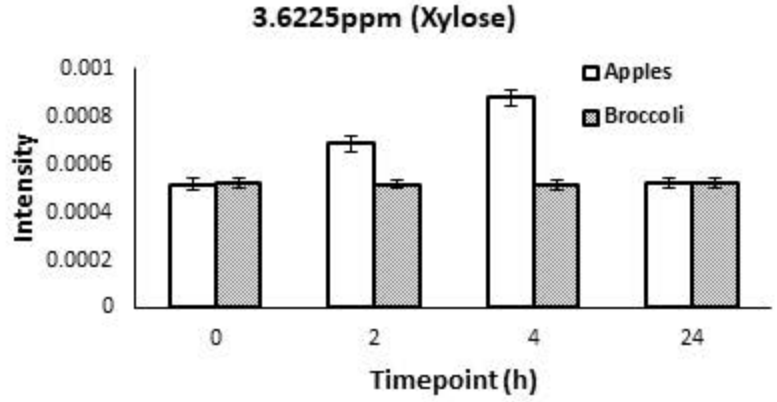
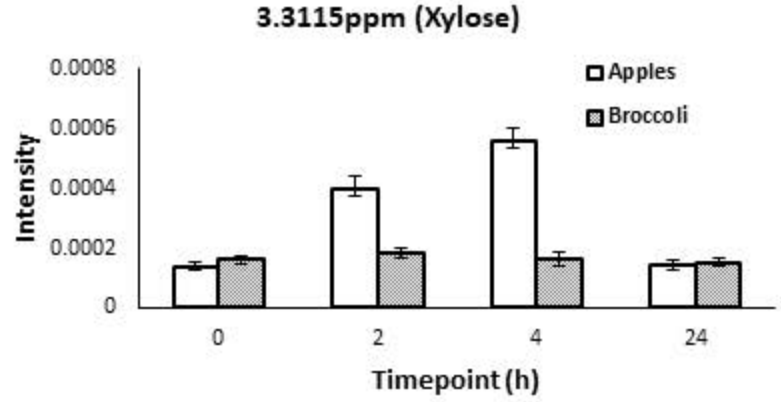


Figure 1


1.2615ppm (3-hydroxyisovalerate)
Intensity
0.0008
0.0006
0.0004
0.0002
0
Apples
Broccoli
0
2
4
24
Timepoint (h)


3.5535ppm (Glycine)
Intensity
0.002
0.0015
0.001
0.0005
0
Apples
Broccoli
0
2
4
24
Timepoint (h)


3.6225ppm (Xylose)
Intensity
0.001
0.0008
0.0006
0.0004
0.0002
0
Apples
Broccoli
0
2
4
24
Timepoint (h)


3.3115ppm (Xylose)
Intensity
0.0008
0.0006
0.0004
0.0002
0
Apples
Broccoli
0
2
4
24
Timepoint (h)

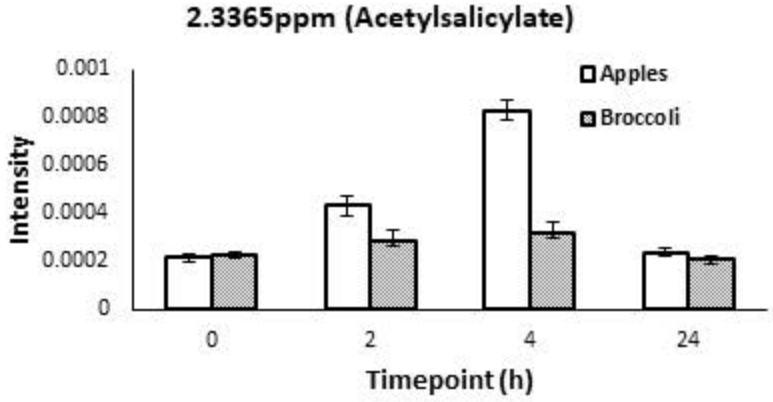

2.3365ppm (Acetylsalicylate)
Intensity
0.001
0.0008
0.0006
0.0004
0.0002
0
Apples
Broccoli
0
2
4
24
Timepoint (h)

Figure 2

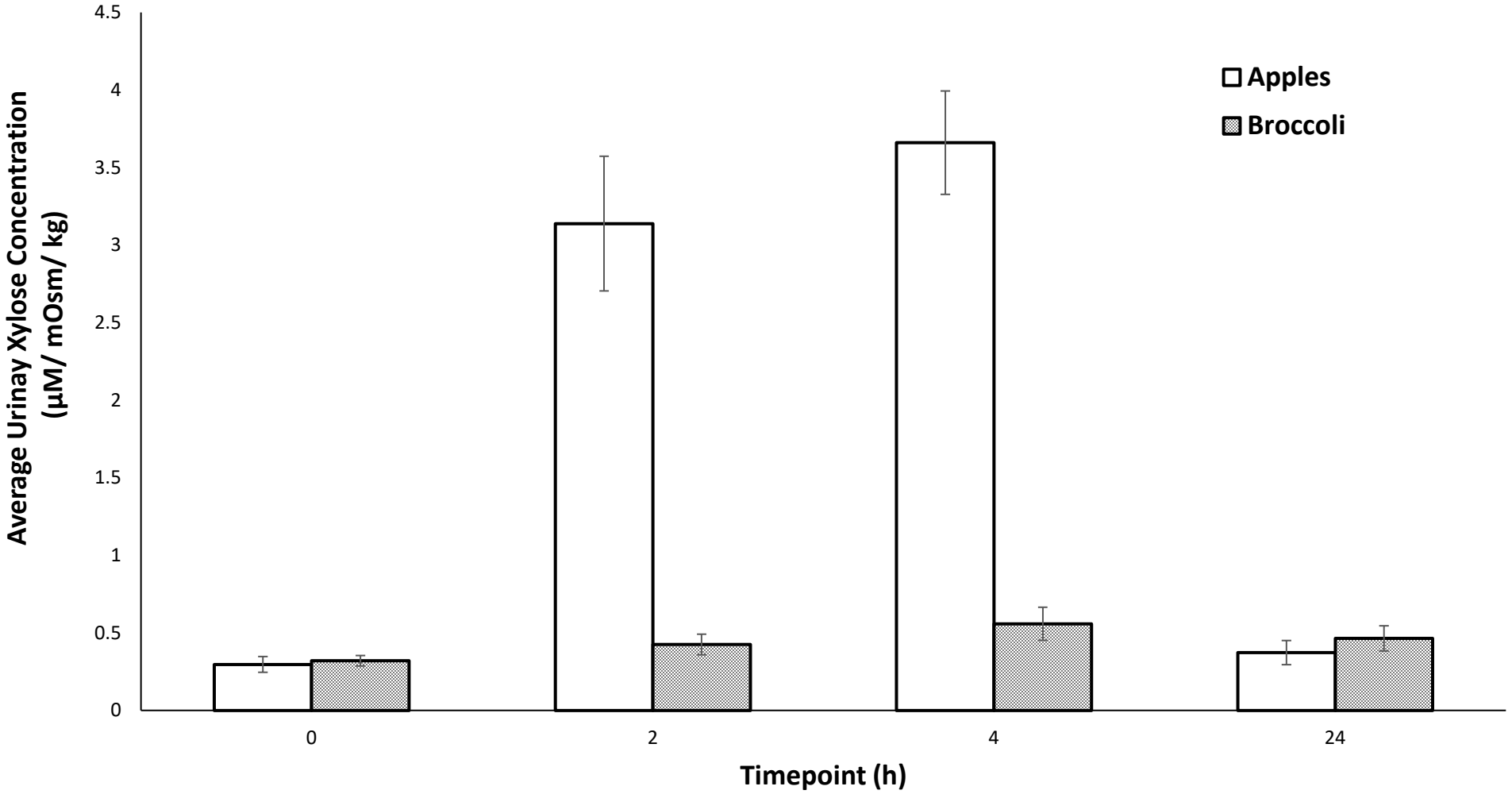
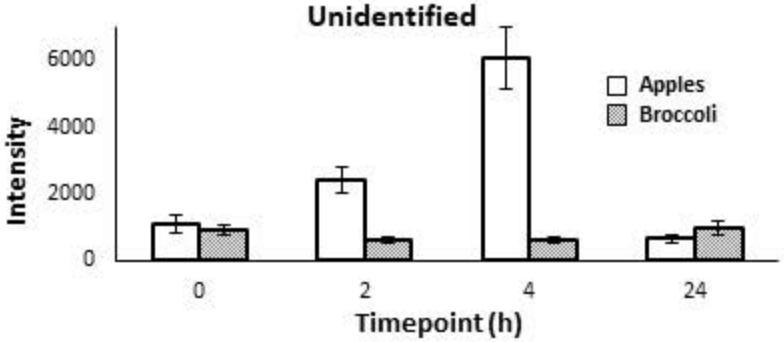
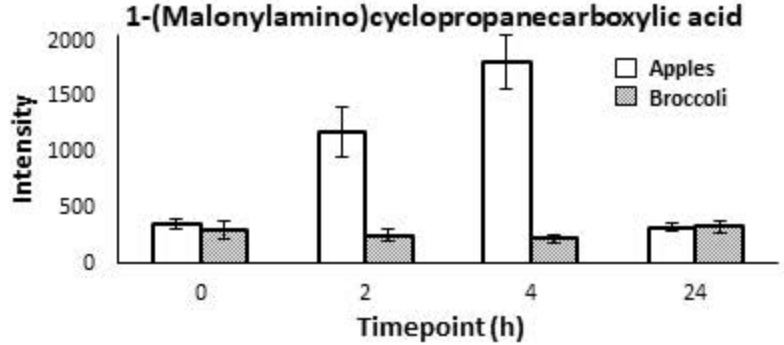
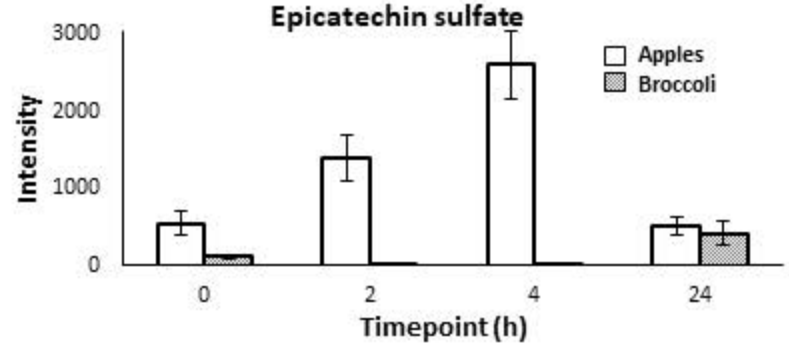
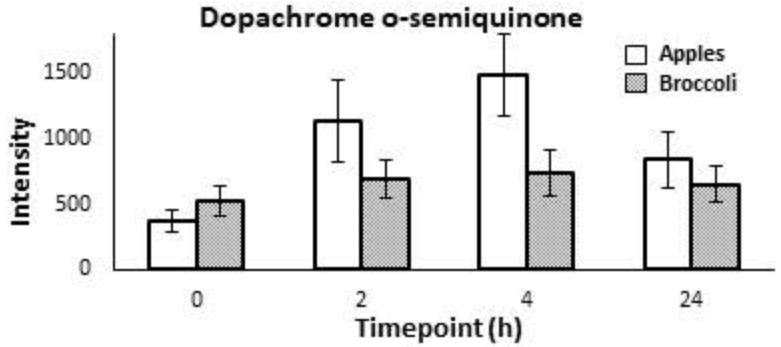

Average Urinay Xylose Concentration (µM/ mOsm/ kg)
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
0
2
4
24
Timepoint (h)
Apples
Broccoli

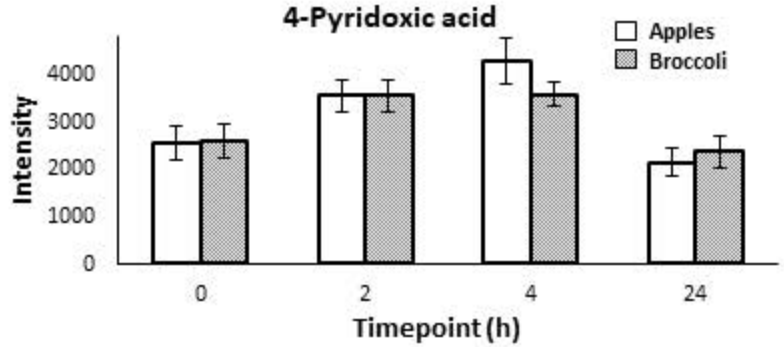
Figure 3

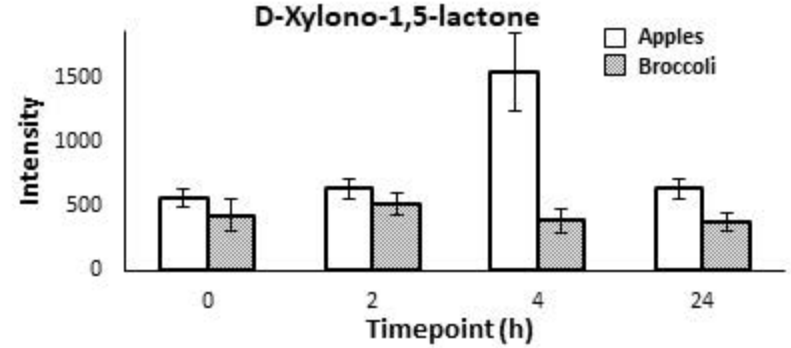

Unidentified
Intensity
Timepoint (h)
Apples
Broccoli

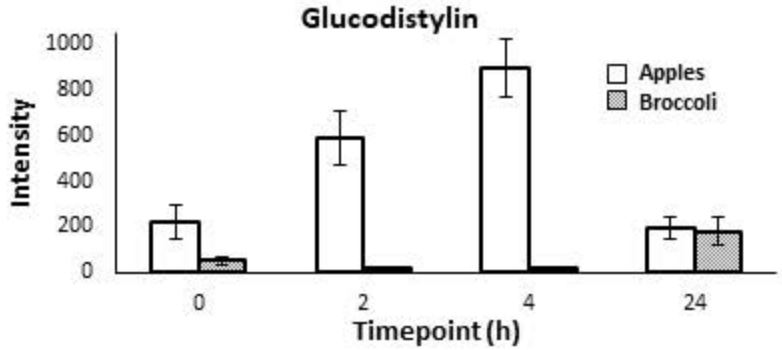

1-(Malonylamino)cyclopropanecarboxylic acid
Intensity
Timepoint (h)
Apples
Broccoli

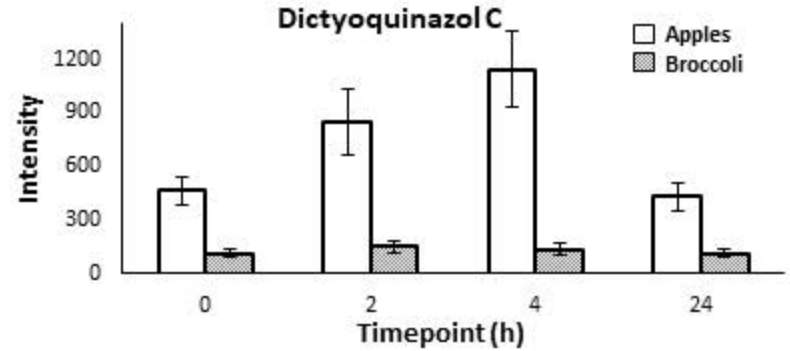

Epicatechin sulfate
Intensity
Timepoint (h)
Apples
Broccoli

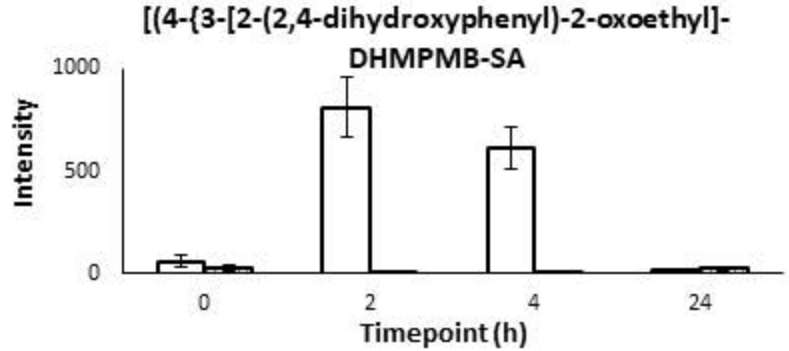

Dopachrome o-semiquinone
Intensity
Timepoint (h)
Apples
Broccoli


4-Pyridoxic acid
Intensity
Timepoint (h)
Apples
Broccoli


D-Xylono-1,5-lactone
Intensity
Timepoint (h)
Apples
Broccoli


Glucodistylin
Intensity
Timepoint (h)
Apples
Broccoli


Dictyoquinazol C
Intensity
Timepoint (h)
Apples
Broccoli


[(4-{3-[2-(2,4-dihydroxyphenyl)-2-oxoethyl]-
DHMPMB-SA
Intensity
Timepoint (h)

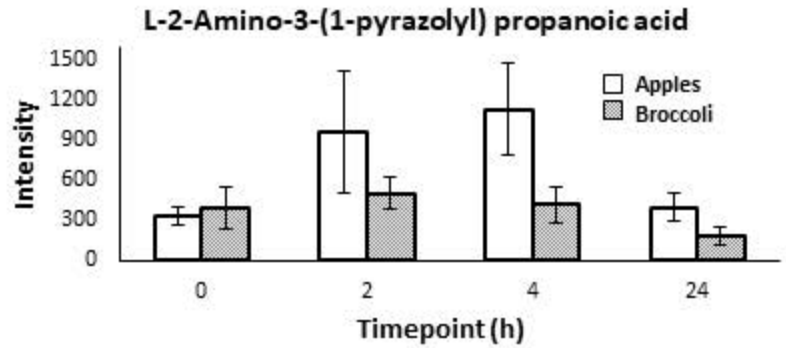
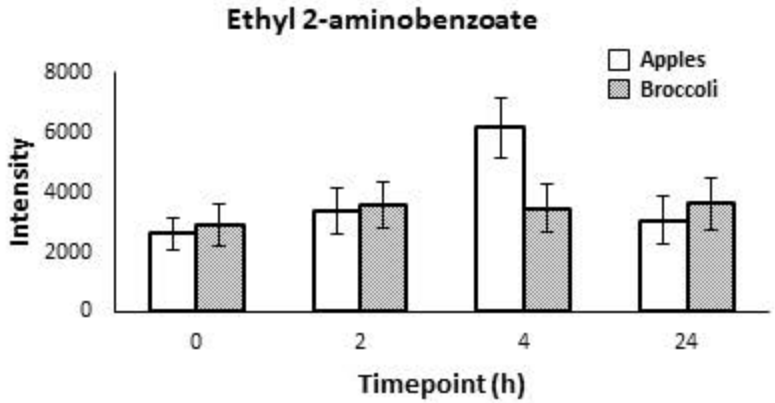
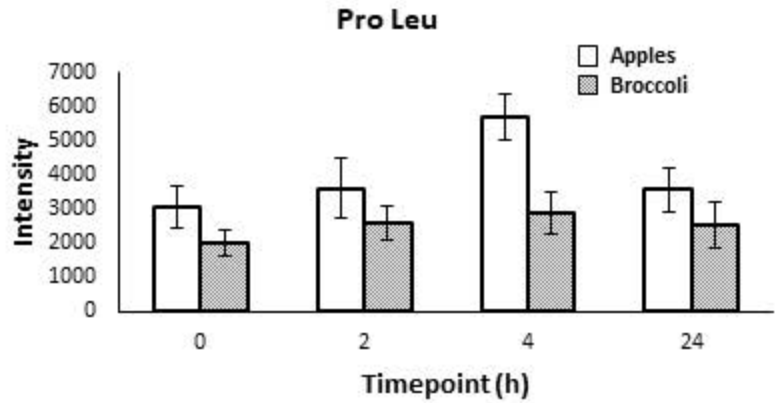
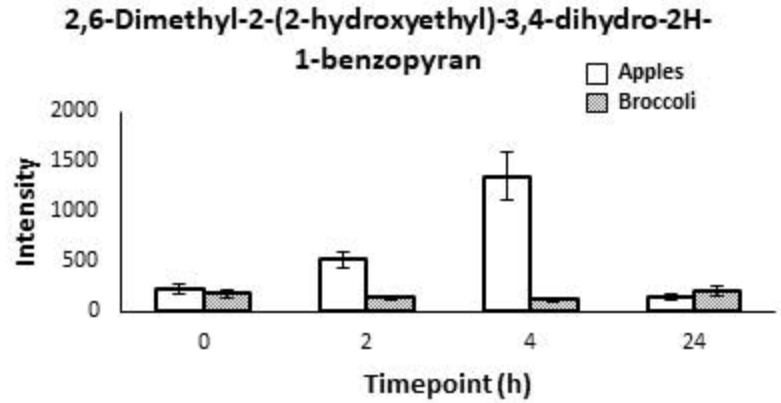
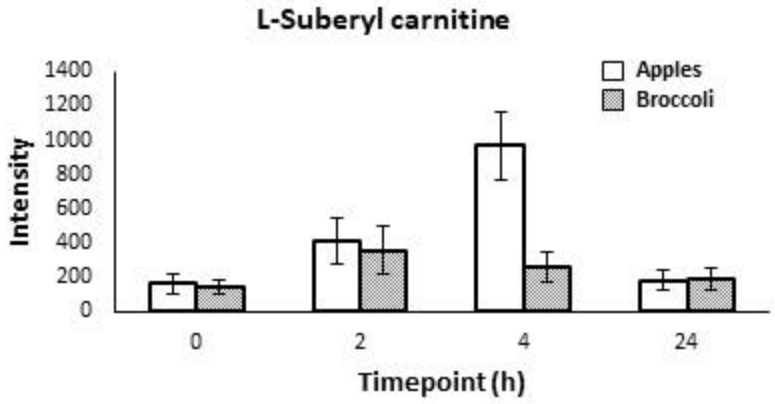

L-2-Amino-3-(1-pyrazolyl) propanoic acid
Intensity
Timepoint (h)
Apples
Broccoli

Figure 4


Ethyl 2-aminobenzoate
Apples
Broccoli
Intensity
8000
6000
4000
2000
0
0
2
4
24
Timepoint (h)


Pro Leu
Apples
Broccoli
Intensity
7000
6000
5000
4000
3000
2000
1000
0
0
2
4
24
Timepoint (h)


2,6-Dimethyl-2-(2-hydroxyethyl)-3,4-dihydro-2H-1-benzopyran
Apples
Broccoli
Intensity
2000
1500
1000
500
0
0
2
4
24
Timepoint (h)


L-Suberyl carnitine
Apples
Broccoli
Intensity
1400
1200
1000
800
600
400
200
0
0
2
4
24
Timepoint (h)

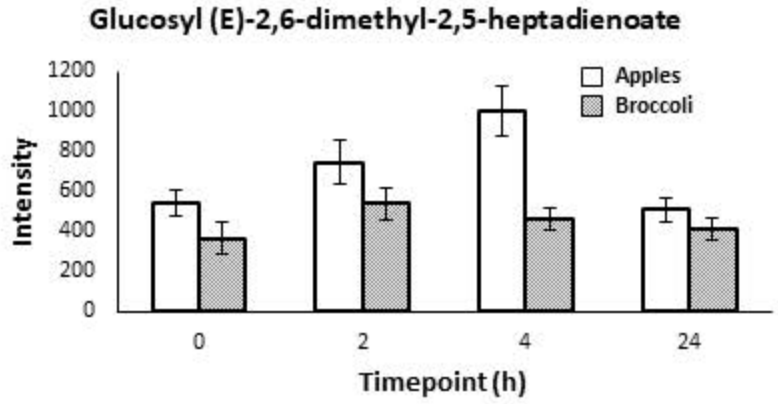

Glucosyl (E)-2,6-dimethyl-2,5-heptadienoate
Apples
Broccoli
Intensity
1200
1000
800
600
400
200
0
0
2
4
24
Timepoint (h)

Figure 5

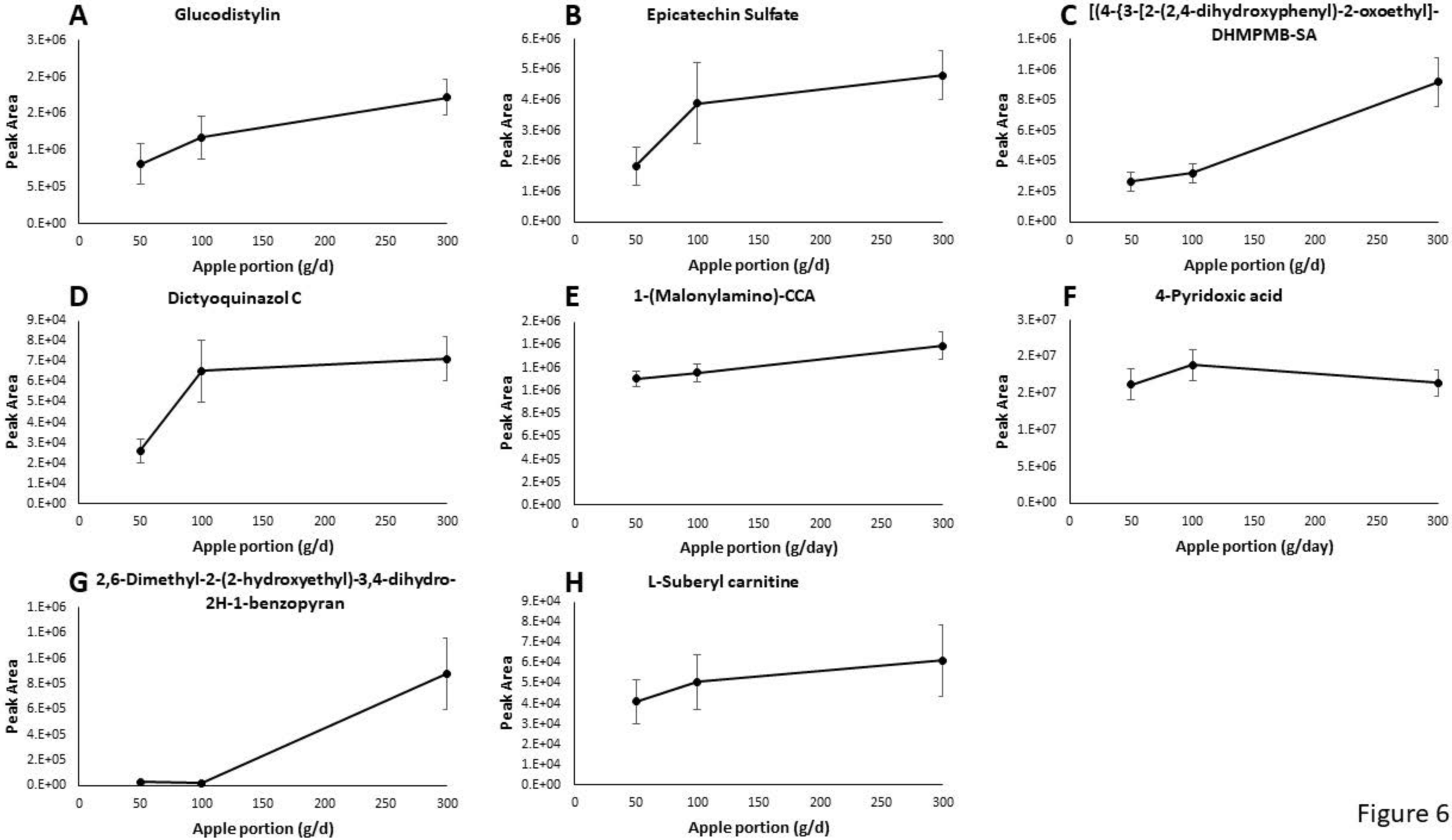

A
Glucodistylin
Peak Area
Apple portion (g/d)
B
Epicatechin Sulfate
Peak Area
Apple portion (g/d)
C
[(4-{3-[2-(2,4-dihydroxyphenyl)-2-oxoethyl]-
DHMPMB-SA
Peak Area
Apple portion (g/d)
D
Dictyoquinazol C
Peak Area
Apple portion (g/d)
E
1-(Malonylamino)-CCA
Peak Area
Apple portion (g/day)
F
4-Pyridoxic acid
Peak Area
Apple portion (g/day)
G
2,6-Dimethyl-2-(2-hydroxyethyl)-3,4-dihydro-
2H-1-benzopyran
Peak Area
Apple portion (g/d)
H
L-Suberyl carnitine
Peak Area
Apple portion (g/d)

Figure 6

Supporting Information

Metabolomic based approach to identify biomarkers of apple intake

Aoife E. McNamara[1,2], Cassandra Collins[1,2], Pedapati S. C. Sri Harsha [1,2], Diana González-Peña [1,2], Helena Gibbons[1,2], Breige A. McNulty[1], Anne P Nugent[1,4], Janette Walton[3,5], Albert Flynn[3], Lorraine Brennan[1,2]

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

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

[3] School of Food and Nutritional Sciences, University College Cork, Cork, Ireland.

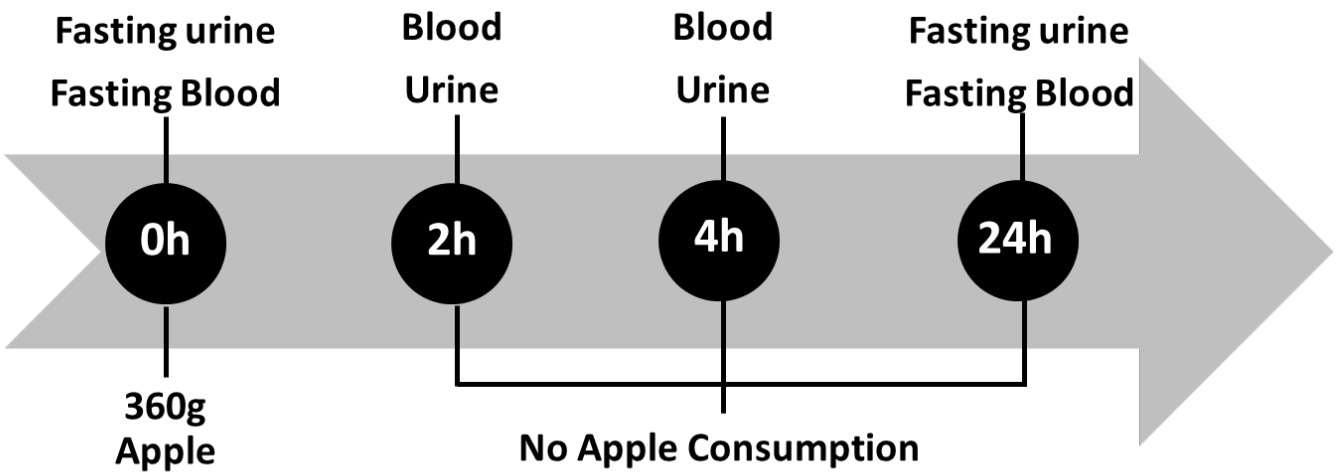
[4] Institute for Global Food Security, School of Biological Sciences, Queens University Belfast, Northern Ireland.

[5] Dept. Biological Sciences, Cork Institute of Technology, Cork, Ireland

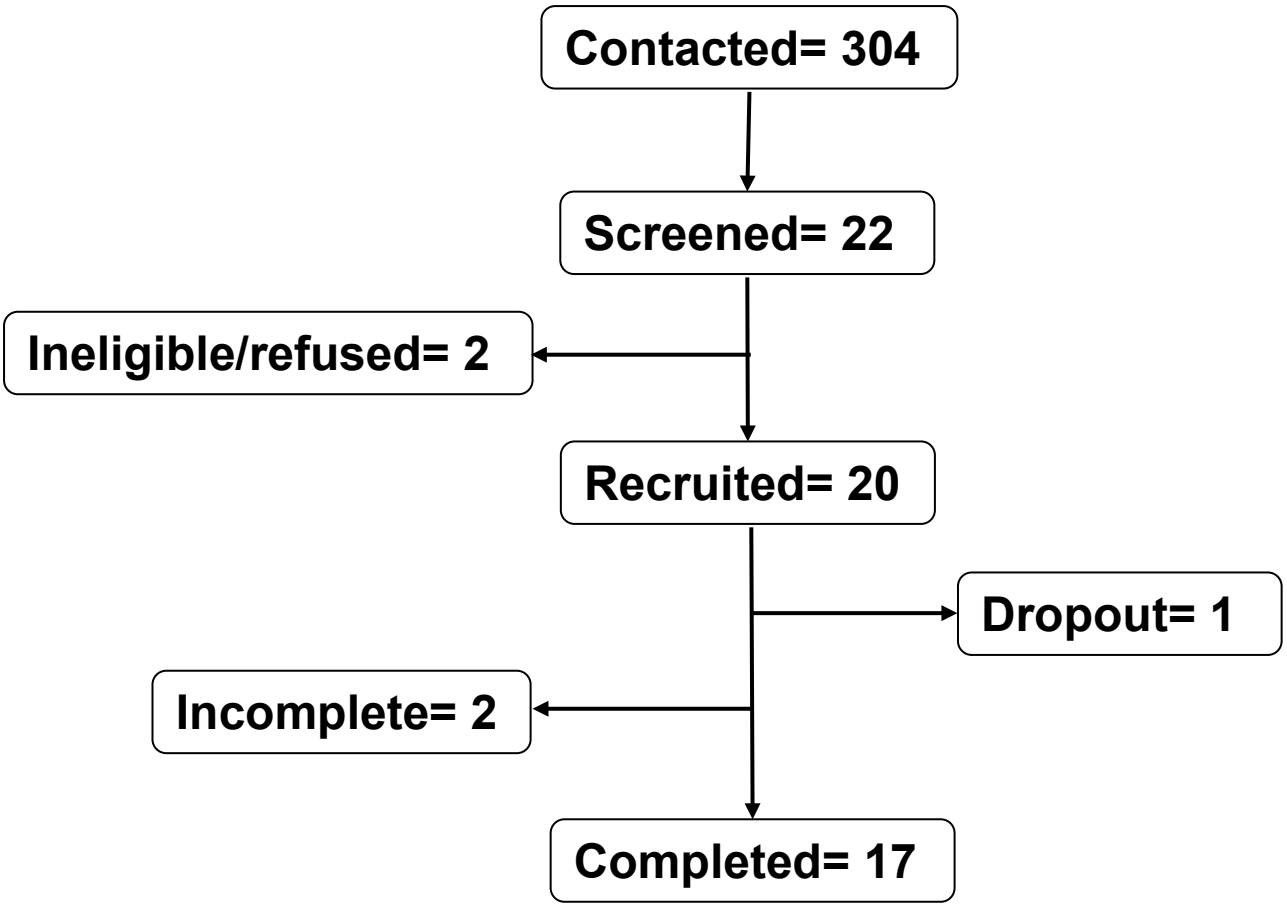
Supplementary Table 1. Average urinary xylose concentration (μM) across the different portions of apple consumed (g/d) in the A-Diet Dose-response study, normalised to osmolality (Osm/ kg).

Apples consumed (g/d)	Average Xylose Concentration (μM/ Osm/ kg)	SD
50	0.65	0.24
100	0.77	0.45
300	1.34	0.79

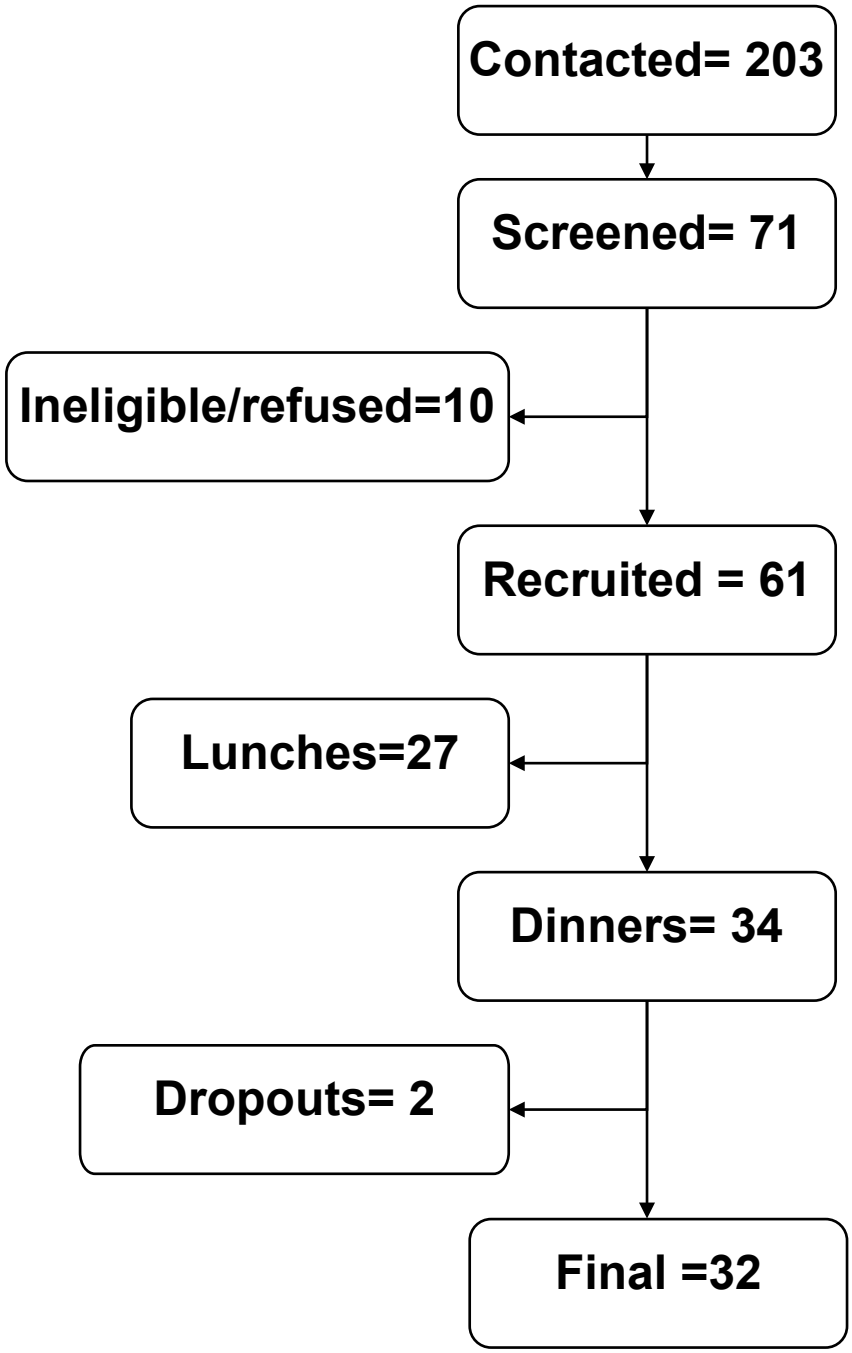
SD; standard deviation.



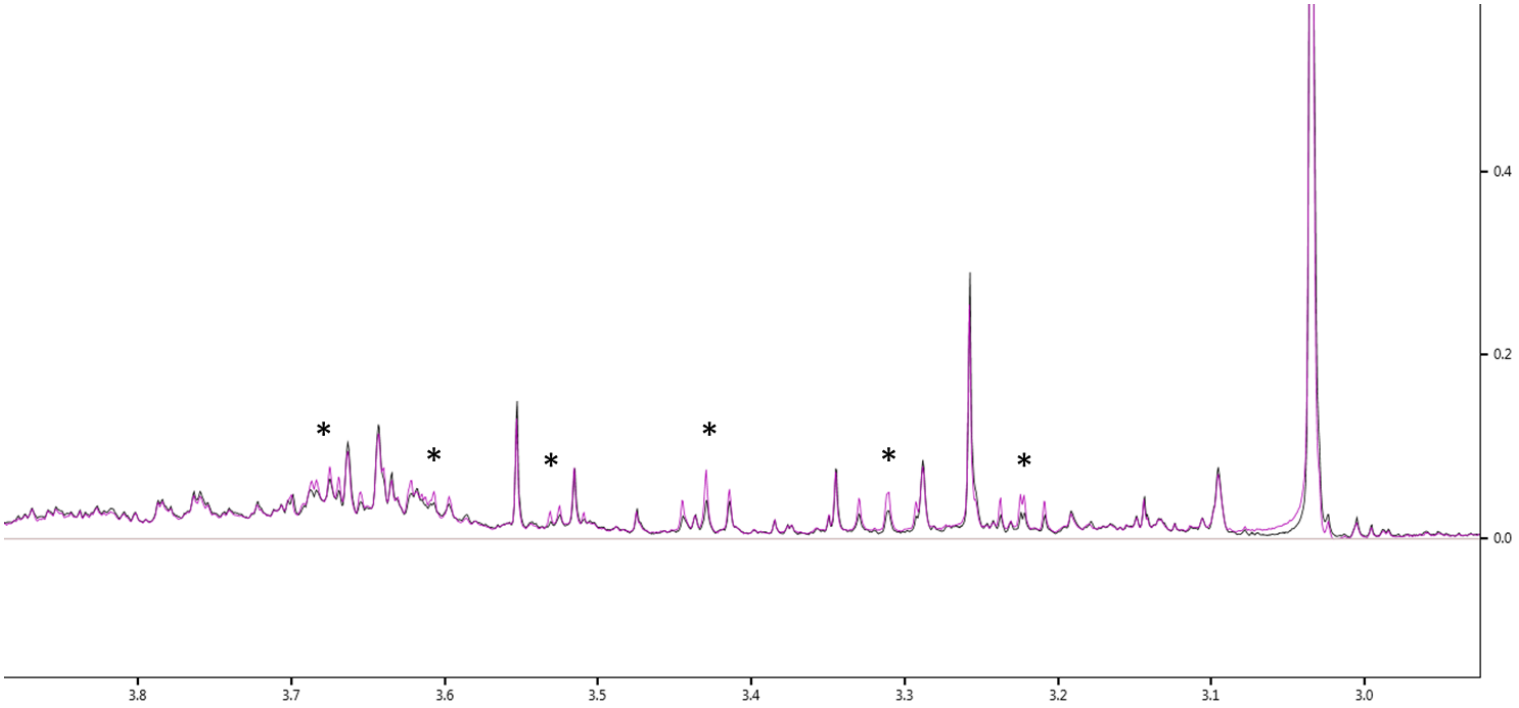
Supplementary Figure 1. Timeline of the A-Diet Discovery study outlining when participants consumed the apples and sampling timepoints.



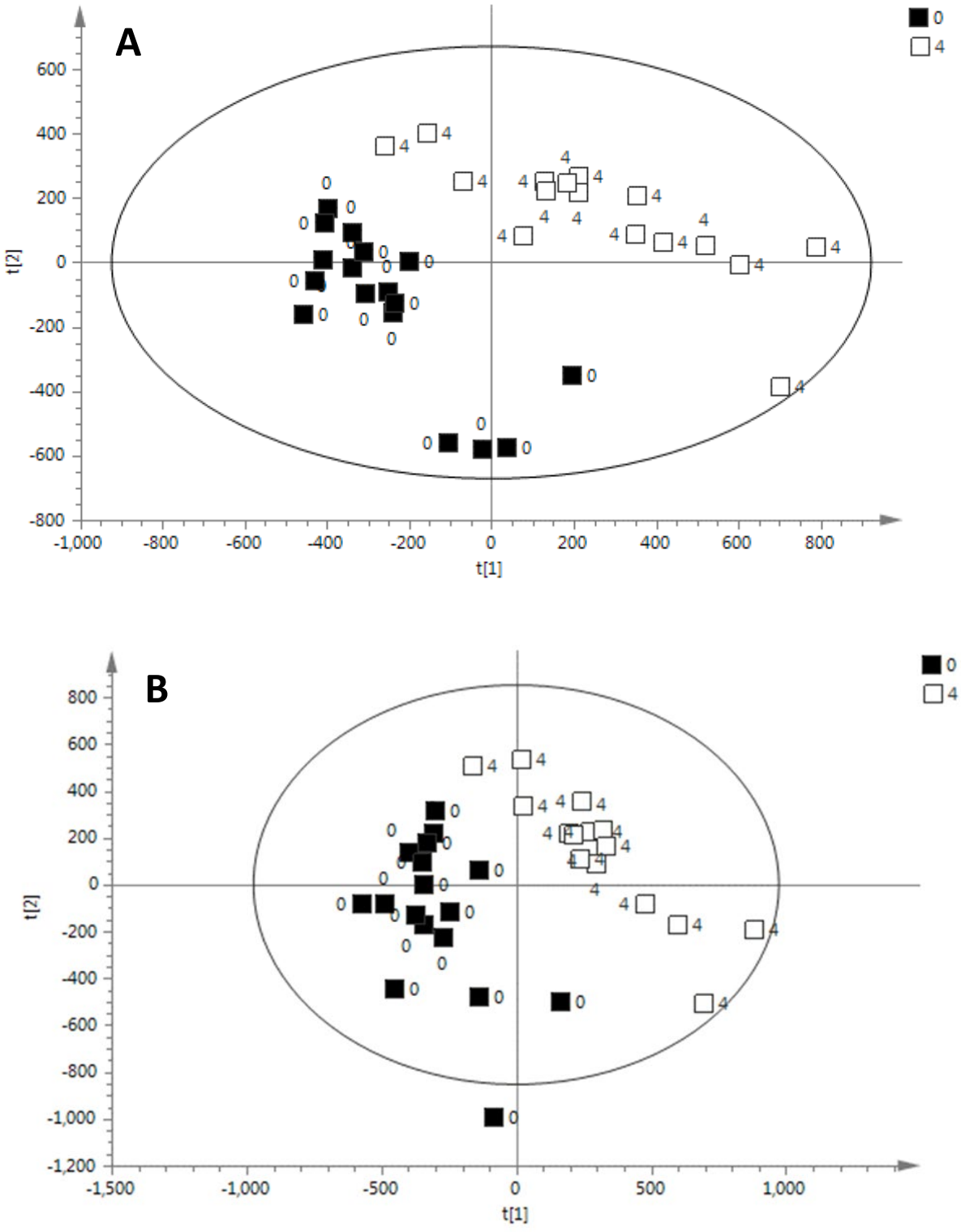
Supplementary Figure 2. Flow diagram of participant recruitment for the Discovery study. A total of 17 participants completed the apple testing session.



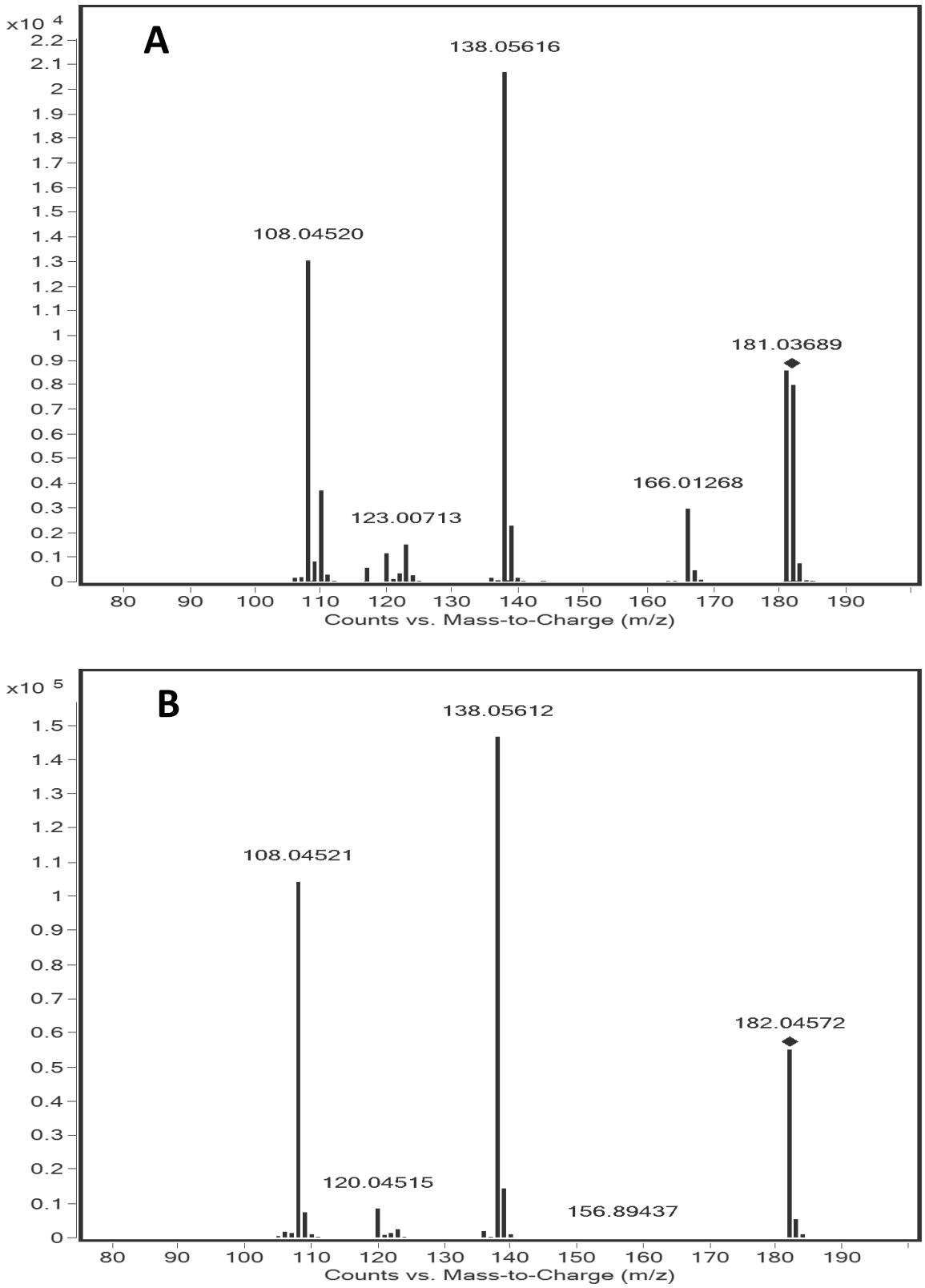
Supplementary Figure 3. Flow diagram of participant recruitment for Validation study.



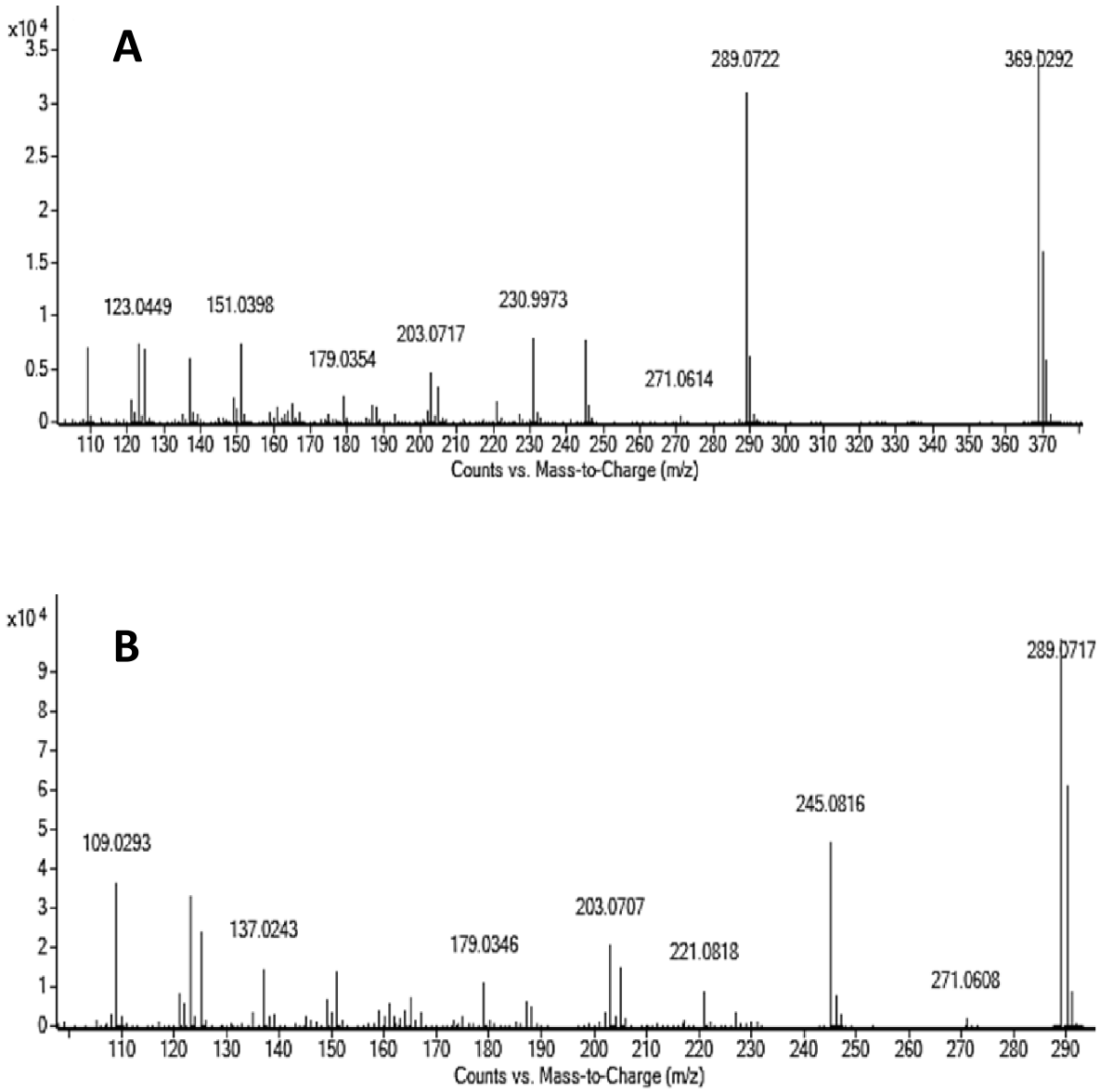
Supplementary Figure 4. 1HNMR spectra of urine sample following apple consumption (black) overlaid with spectra of a urine sample spiked with a xylose standard (pink). Identified peaks of xylose are marked by an asterisk (*). X-axis represents spectral region (parts per million (ppm)) and Y-axis represents peak intensity.



Supplementary Figure 5.(A) PLS-DA model of LC-MS urine data in positive mode. Comparison of baseline (T0) and four hour (T4) time points in positive mode R^2X, 0.37; Q^2 0.71; $R^2X(1)$, 0.21, $R^2X(2)$ 0.10; t[1]. PLS component 1; t[2], PLS component 2. (B) PLS-DA model of LC-MS urine data in negative mode. Comparison of baseline (T0) and four hour (T4) time points in negative mode R^2X, 0.42; Q^2 0.77; t[1]. PLS component 1; t[2], PLS component 2;.



Supplementary Figure 6. MS/MS spectra of 4-pyridoxic acid at average collision energies in the urine sample (A) and of a pure analytical standard (B)



Supplementary Figure 7. MS/MS spectra of Epicatechin at average collision energies in the urine sample (A) and of a pure analytical standard (B). The 79.957 Da difference between the precursor of the urine spectra and of the standard can be attributed to a sulfate group.