Supporting Information for Vegetable Signatures Derived from Human Urinary Metabolomic Data in Controlled Feeding Studies

Ke-Shiuan Lynn[†], Mei-Ling Cheng[‡, §,⊥], Hsin-Chou Yang[∥], Yu-Jen Liang[∥], Mei-Jyh Kang[∇], Fong-Ling Chen[∇], Ming-Shi Shiao[‡], and Wen-Harn Pan*[,∇,⊗]

[†]Department of Mathematics, Fu Jen Catholic University, New Taipei City 24205, Taiwan

[‡]Department of Biomedical Sciences, College of Medicine, Chang Gung University, Taoyuan 33302, Taiwan

[§]Metabolomics Core Laboratory, Healthy Aging Research Center, Chang Gung University, Taoyuan 33305, Taiwan

[⊥]Clinical Metabolomics Core Laboratory, Chang Gung Memorial Hospital, Taoyuan 33305, Taiwan

[∥]Institute of Statistical Science, Academia Sinica, Taipei, 11529, Taiwan

[∇]Institute of Biomedical Sciences, Academia Sinica, Taipei, 11529, Taiwan

[⊗]Institute of Population Health Sciences, National Health Research Institutes, Miaoli 35053, Taiwan

Corresponding Author:

Wen-Harn Pan, Institute of Biomedical Sciences, Academia Sinica, Taipei, 11529, Taiwan.

E-mail: pan@ibms.sinica.edu.tw
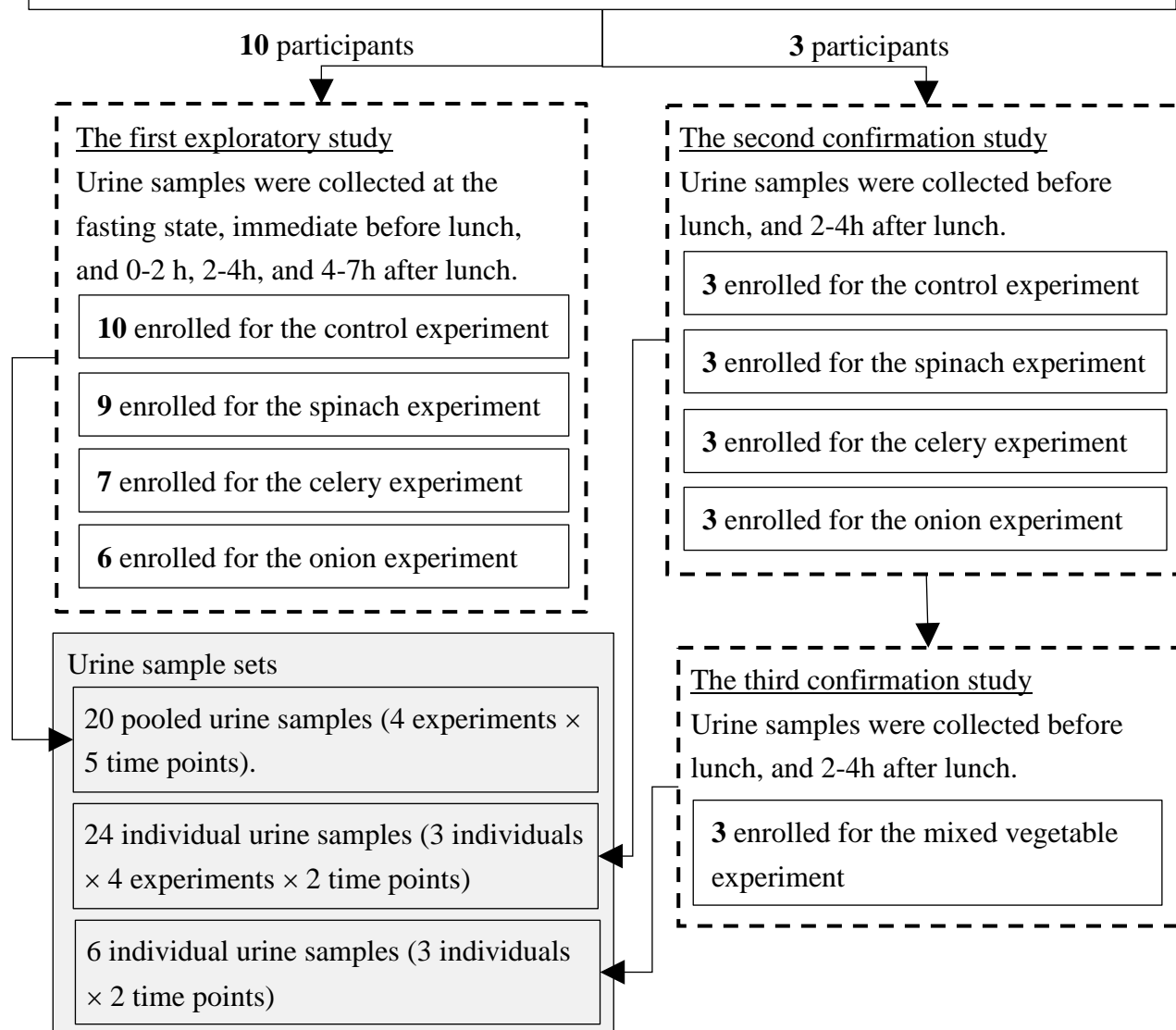
Table of Content

Text S1: Participant Flow Chart

A total of **13** participants recruited at Tri-Service General Hospital, Taipei, Taiwan.

Inclusion criteria were as follows:

(i)   aged 18–60 years;

(ii) body mass index (BMI) between 18.5 and 30 kg/m2;

(iii) nonsmoker

(iv) no alcohol abuse.

Exclusion criteria were as follows:

(i)   those taking medication for hypertension or diabetes

(ii) patients with immune diseases (allergic or autoimmune), liver disease, metabolic disease (hyperthyroidism or hypothyroidism), and other serious diseases such as cancer

(iii) those unwilling to cease taking dietary supplements

(iv) patients with urinary tract infections or those who had taken antibiotics within 3 weeks of the study.

10 participants   **3** participants

The first exploratory study
Urine samples were collected at the fasting state, immediate before lunch, and 0-2 h, 2-4h, and 4-7h after lunch.

10 enrolled for the control experiment

9 enrolled for the spinach experiment

7 enrolled for the celery experiment

6 enrolled for the onion experiment

The second confirmation study
Urine samples were collected before lunch, and 2-4h after lunch.

3 enrolled for the control experiment

3 enrolled for the spinach experiment

3 enrolled for the celery experiment

3 enrolled for the onion experiment

Urine sample sets

20 pooled urine samples (4 experiments × 5 time points).

24 individual urine samples (3 individuals × 4 experiments × 2 time points)

6 individual urine samples (3 individuals × 2 time points)

The third confirmation study
Urine samples were collected before lunch, and 2-4h after lunch.

3 enrolled for the mixed vegetable experiment

Text S2

We have suggested using the intensity profile of feature urinary metabolites for vegetable signature. Assuming that a metabolomic signature library of m vegetables is available, in which a total of n metabolites is involved. The m vegetable signatures can be regarded as a m basis $\bar{p}_i$, $i = 1, 2, \ldots, m$, spanning an n-dimensional space. Given a measurement $\bar{p}_t$ of the n urinary metabolites, there exists a unique solution $\bar{w} = [w_1, w_2, \ldots, w_m]^T$, $w_i \geq 0$, $i = 1, 2, \ldots, m$, for the weighting (proportions) of the m vegetables such that $\bar{p}_t$ is a linear (or more precisely, conical) combination of the m basis, i.e. $\bar{p}_t = w_1\bar{p}_1 + w_2\bar{p}_2 + \cdots + w_m\bar{p}_m$, if the m basis are linear independent, as illustrated in Supplementary Figure S1.
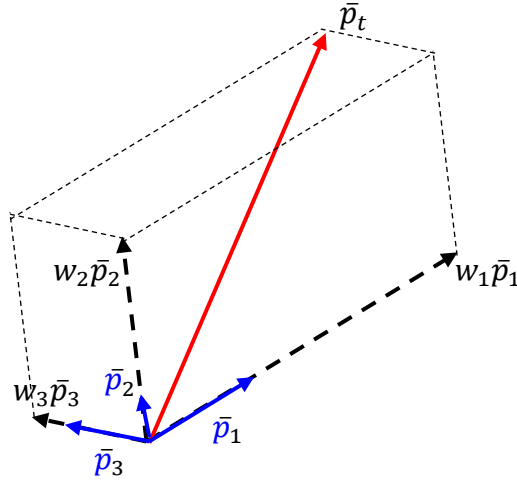
Figure S1. An illustration of any given vector $\bar{p}_t$ in a three-dimensional space can be represented by a unique linear combination of a set of linearly independent basis elements $\bar{p}_1$, $\bar{p}_2$, and $\bar{p}_3$.

We present an example of 3 vegetables in Supplementary Figure S2, where a total of 7 metabolites are involved. As shown in Supplementary Figure S2(a), vegetables 1, 2, and 3 accordingly consist of 3, 4, and 4 feature metabolites. Some of the feature metabolites of the three vegetables are overlapped, resulting in a total of 7 metabolites for intensity profile examination. Therefore, for each vegetable, its vegetable signature can be represented by a vector $\bar{p}_i = [p_{i,1}\ p_{i,2}\ \ldots\ p_{i,7}]^T$, $i = 1, 2,$ and 3, as listed below and plotted in Supplementary Figure S2(b).

$$\bar{p}_1 = [0.3\ 0.0\ 1.0\ 0.0\ 0.0\ 0.4\ 0.0]^T$$

$$\bar{p}_2 = [0.0\ 0.5\ 0.3\ 0.0\ 1.0\ 0.6\ 0.0]^T$$

$$\bar{p}_3 = [1.0\ 0.0\ 0.0\ 0.4\ 0.6\ 0.0\ 0.7]^T$$

On the presence of a measurement $\bar{p}_t$ of the 7 urinary metabolites in a urine sample, it can be represented by $\bar{p}_t = w_1\bar{p}_1 + w_2\bar{p}_2 + w_3\bar{p}_3$ or in the following matrix form.

$$\bar{p}_t = \begin{bmatrix} p_{t,1} \\ p_{t,2} \\ \vdots \\ p_{t,7} \end{bmatrix} = [\bar{p}_1\ \bar{p}_2\ \bar{p}_3] \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} = \begin{bmatrix} p_{1,1} & p_{2,1} & p_{3,1} \\ p_{1,2} & p_{2,2} & p_{3,2} \\ \vdots & \vdots & \vdots \\ p_{1,7} & p_{2,7} & p_{3,7} \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} = P\bar{w} \tag{1}$$

The method of ordinary least squares can be used to compute a solution to the equation. For the equation $P\bar{w} = \bar{p}_t$, the least squares formula is obtained from the problem

$$\min_{\bar{w}} \|P\bar{w} - \bar{p}_t\| \tag{2}$$

The weights $[w_1, w_2, w_3]$ of the 3 vegetables can be computed via the following equation.

$$\bar{w} = (P^T P)^{-1} P^T \bar{p}_t \tag{3}$$

We demonstrate an example in Supplementary Figure S2(c) where $\bar{p}_t = 4.1\bar{p}_1 + 7.5\bar{p}_2 + 0.6\bar{p}_3$. If $\bar{p}_1, \bar{p}_2,$ and $\bar{p}_3$ are known, given a measurement $\bar{p}_t$, we can use equation (3) to compute the weights of the three vegetables. In Supplementary Figure S2(c), we can see that the given signature and the computed signature are equivalent.

We also demonstrate that the above procedure is robust to small noise. As mention in our paper, the intensity profile of feature metabolites of a vegetable exhibited small variations among individuals and among sample collection time points. Although such variations may be reduced by adopting a targeted LC/MS approach, measuring error can still occur in the signal. Therefore, we generate a new measurement $\bar{p}_t'$ by adding 15% noise to $\bar{p}_t$ and demonstrate in Supplementary Figure S2(d) that even with such a noise level, we can still compute a weighting vector [1.6 8.5 0.3] close to the designated value of [4.1 7.5 0.6] in the sense of value order.
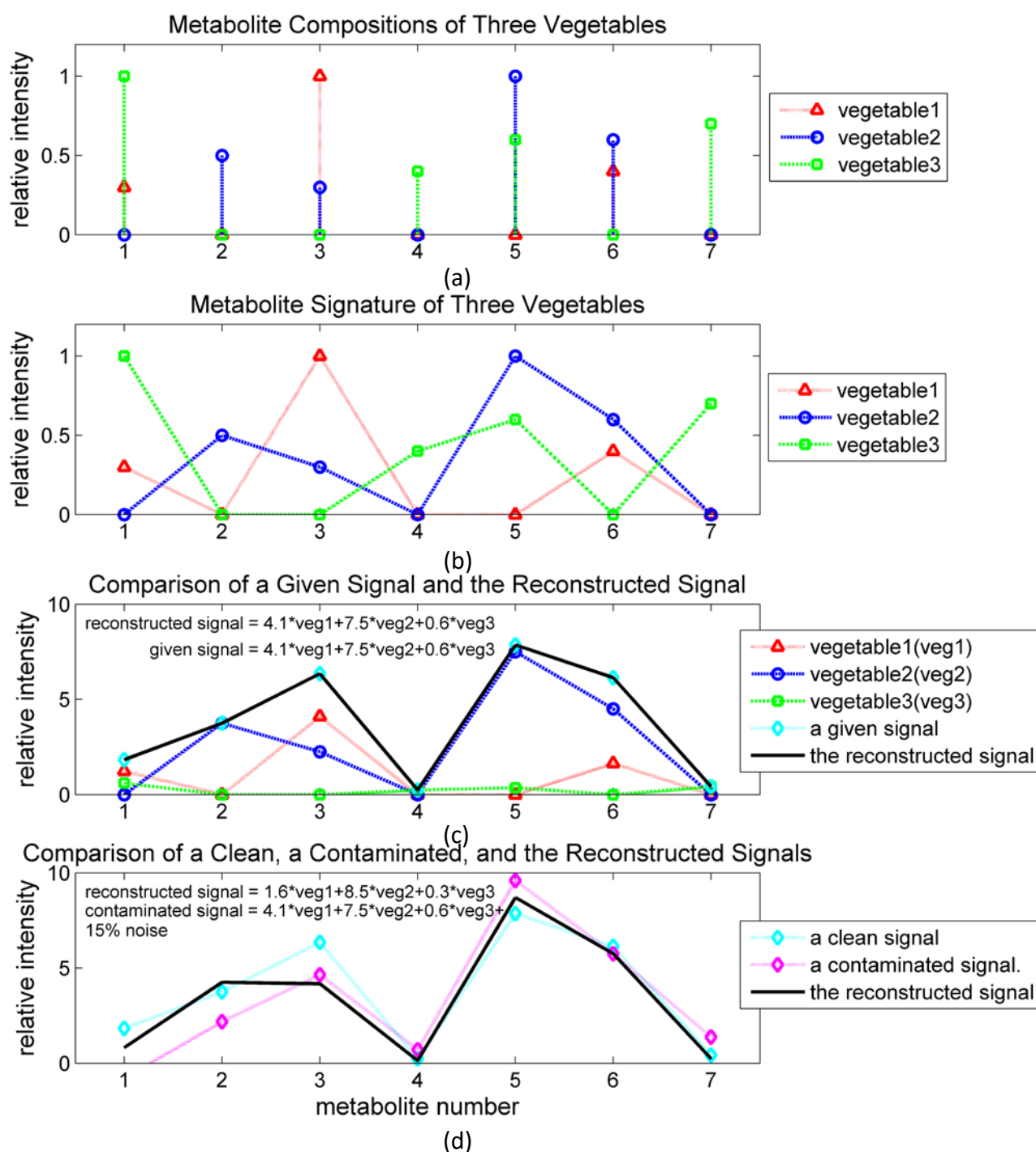
Figure S2. A demonstration that a measurement of a set of designated urinary metabolites can be decomposed into linear combination of vegetables with known metabolite signatures. (a) Three vegetables with known composition of seven urinary metabolites. (b) Metabolite signature of the three vegetables constructed from the intensities of the seven urinary metabolites. (c) Given a clean measurement signal of the seven urinary metabolites from an individual, the signal can be reconstructed by a linear combination of the metabolite signatures of the three vegetables. (d) A contaminated signal with 15% noise can still be approximately reconstructed by the metabolite signatures of the three vegetables.

ID	RT (min)	M/Z	low (Trap/Transfer)	high (Trap/Transfer)
214>146	1.413	214.1527	6/4	17/4
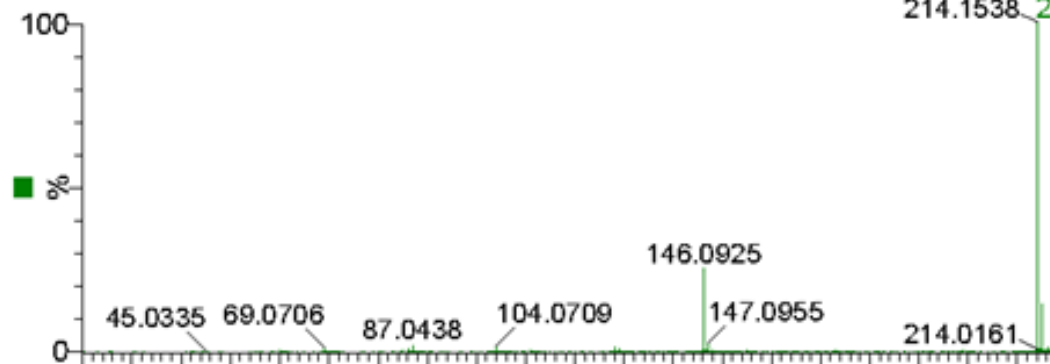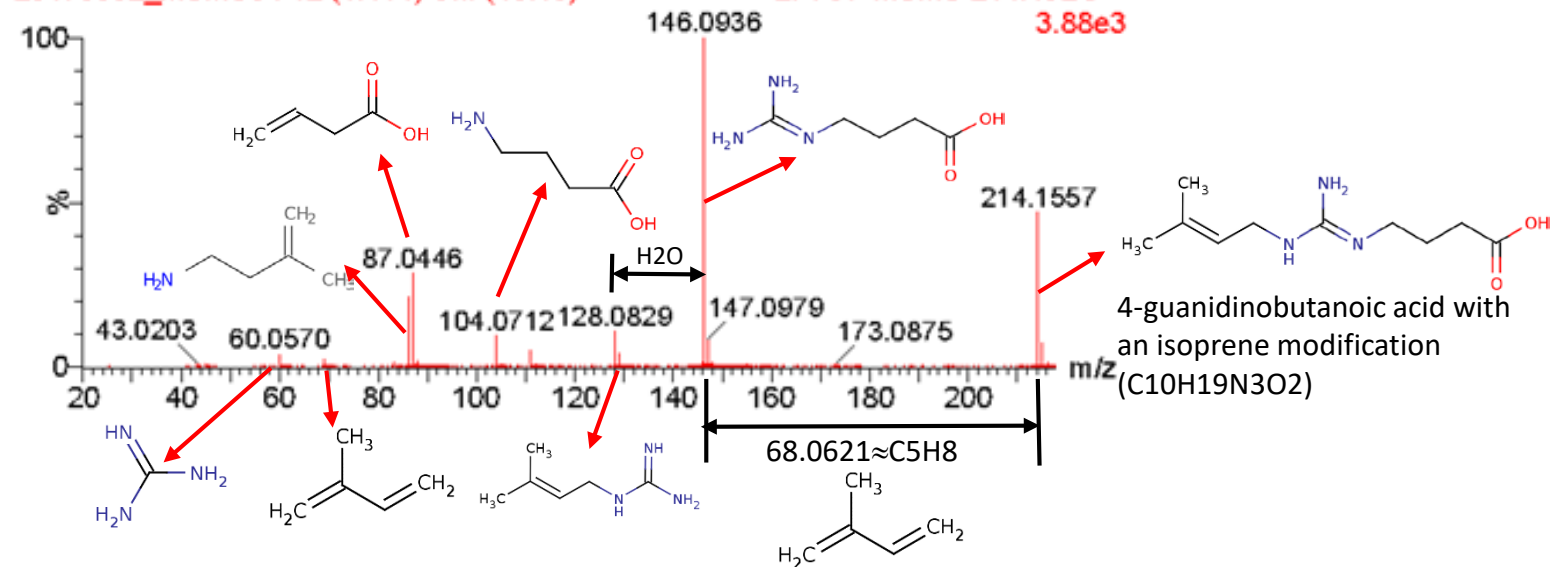
Figure S3. MS/MS spectra of 4-guanidinobutanoic acid with an isoprene modification

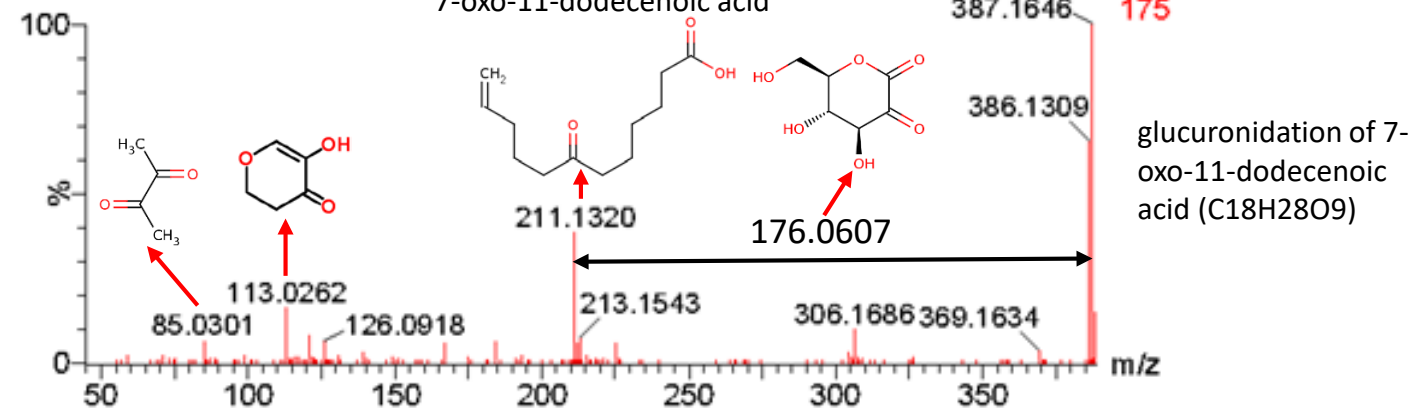| ID	RT (min)	M/Z	low (Trap/Transfer)	high (Trap/Transfer)
403>227 209 163 153	1.64	227.1296	2/4	10/4
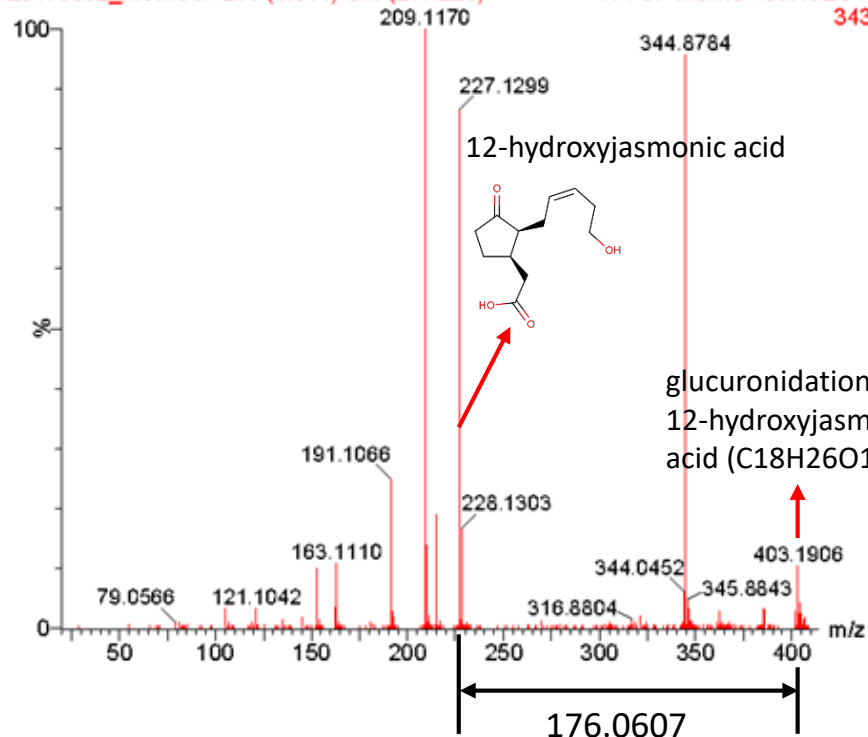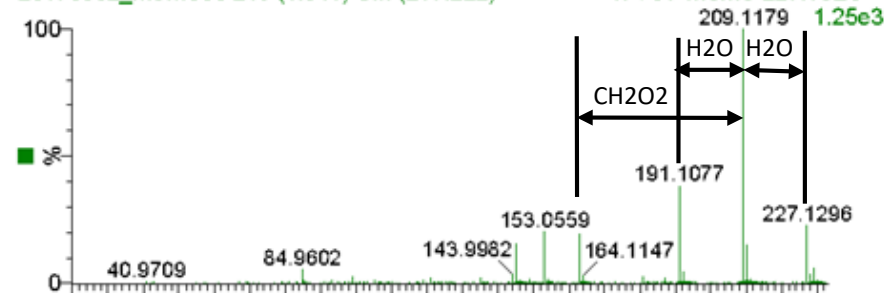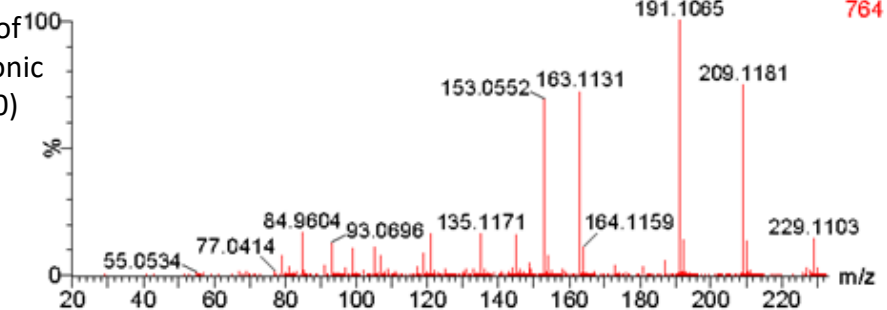
Figure S4. MS/MS spectra of the glucuronidation products of 12-hydroxyjasmonic acid

ID	RT (min)	M/Z	low (Trap/Transfer)	high (Trap/Transfer)
383>207	1.707	383.1360	6/4	11/4
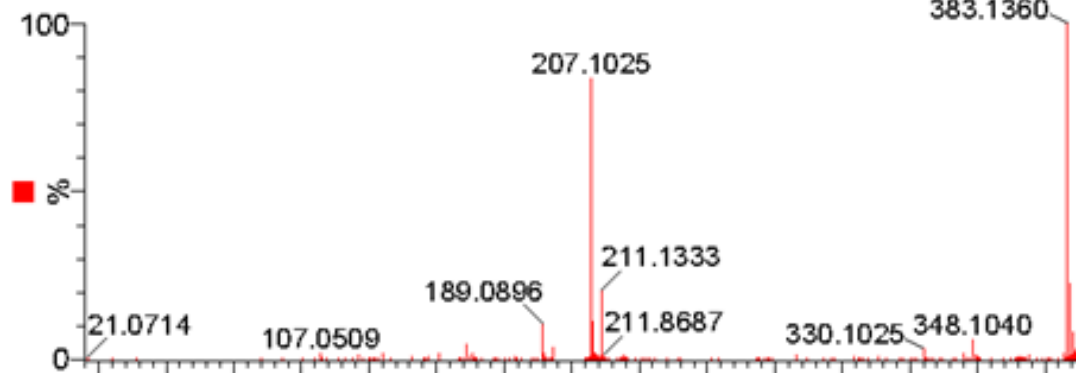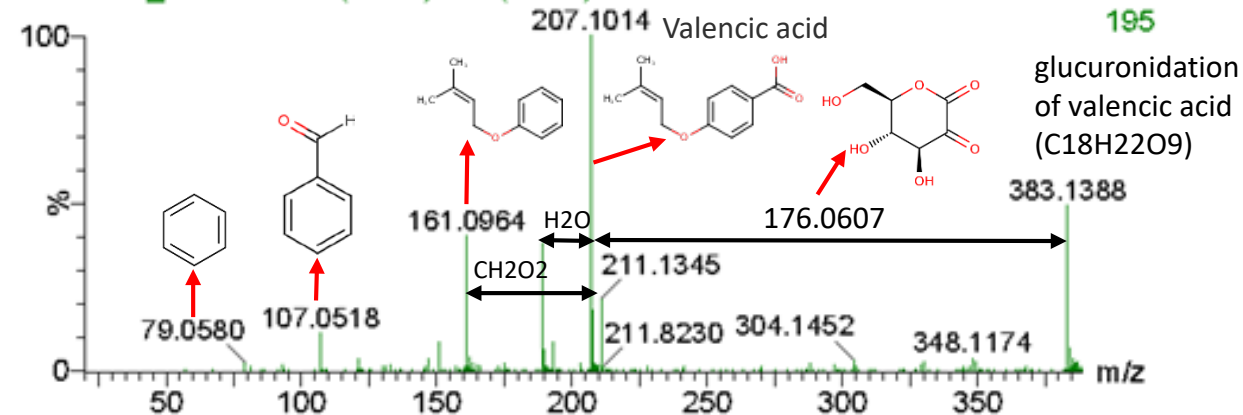
Figure S5. MS/MS spectra of the glucuronidation products of valencic acid

ID	RT (min)	M/Z	low (Trap/Transfer)	high (Trap/Transfer)
387	1.854	387.1649	6/4	21/4
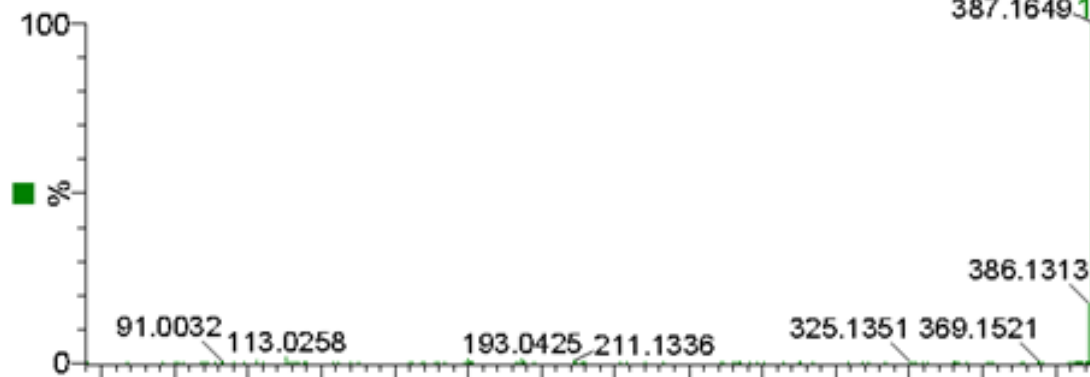
Figure S6. MS/MS spectra of the glucuronidation products of 7-oxo-11-dodecenoic acid

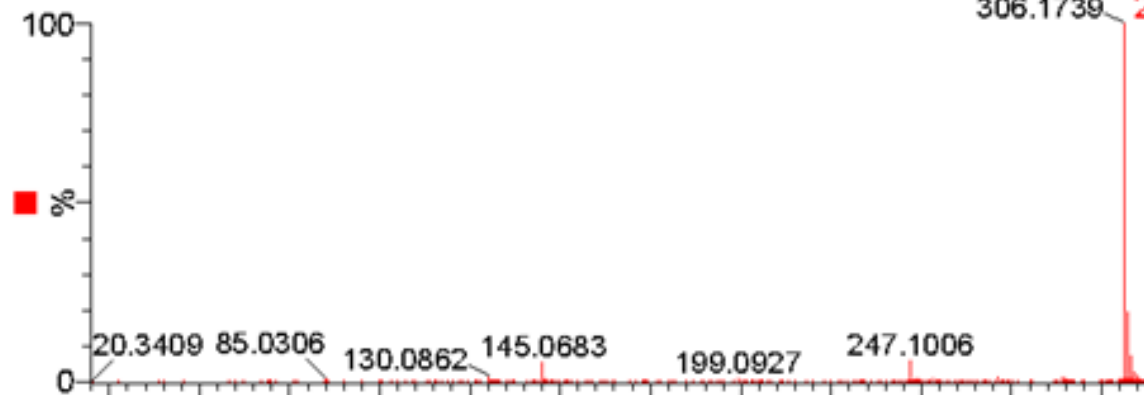| ID	RT (min)	M/Z	low (Trap/Transfer)	high (Trap/Transfer)
220>130 88 84	0.763	130.0508	6/4	14/4
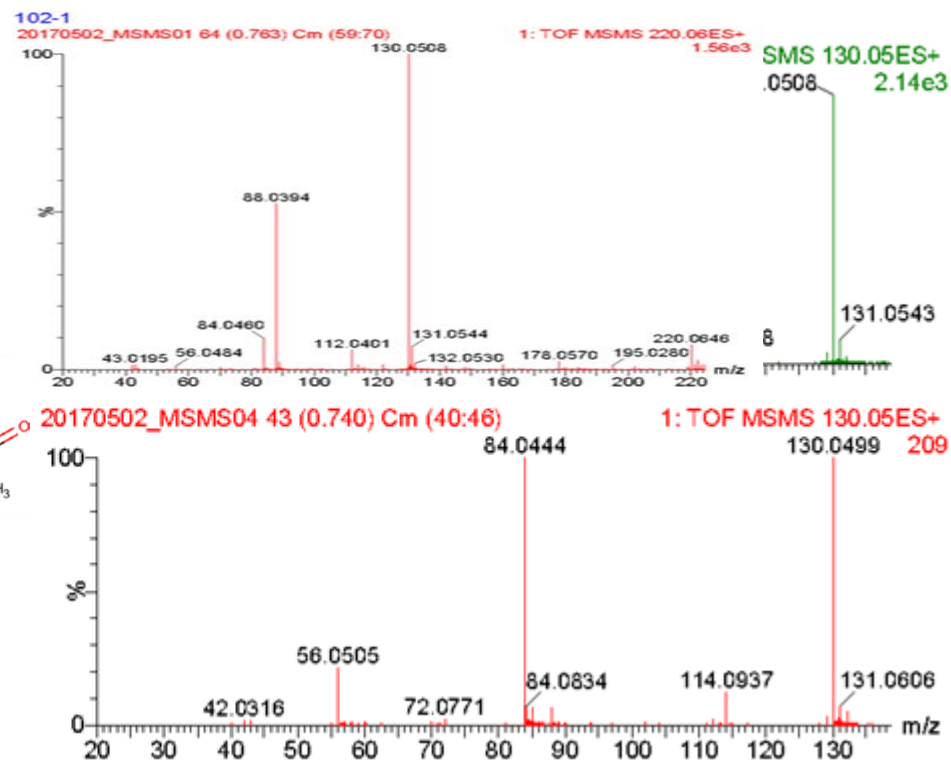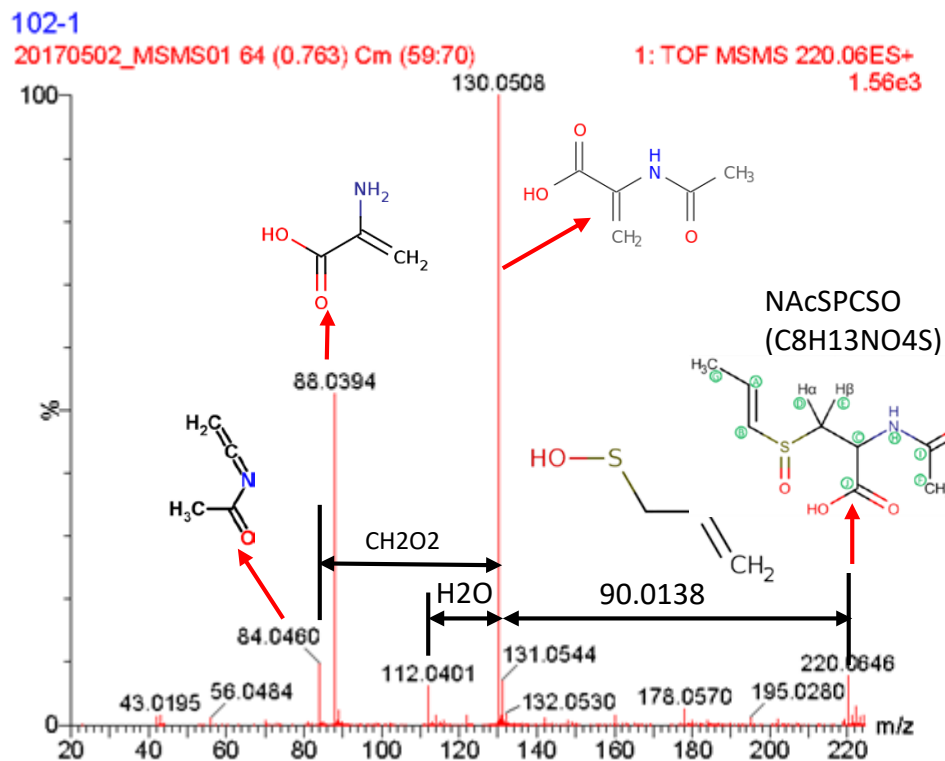
Figure S7. MS/MS spectra of N-acetyl-S-(1Z)-propenyl-cysteine-sulfoxide (NAcSPCSO)

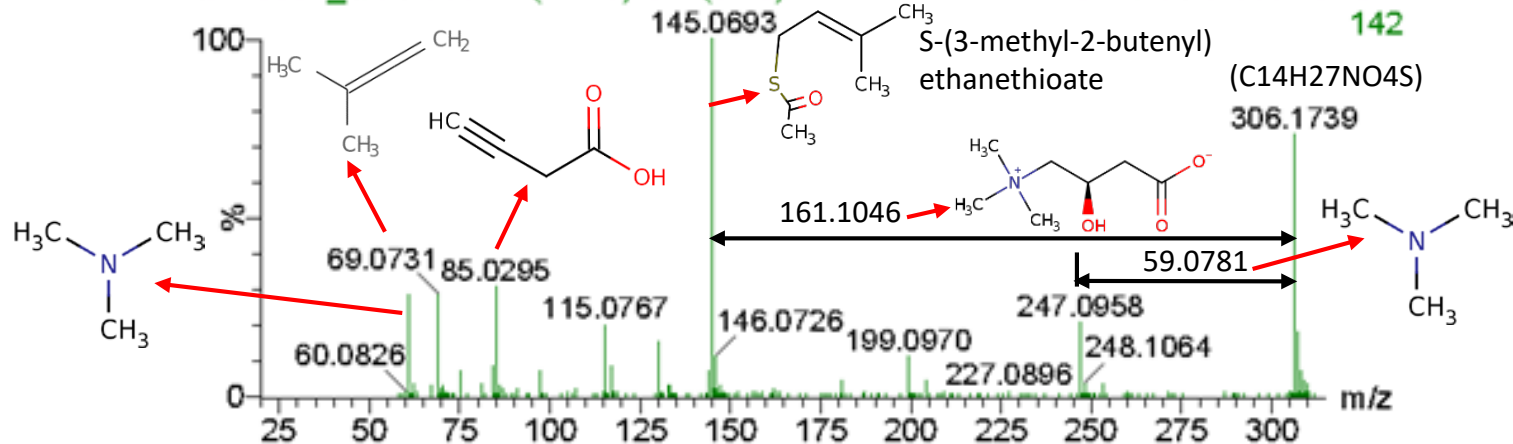| ID	RT (min)	M/Z	low (Trap/Transfer)	high (Trap/Transfer)
306	1.779	306.1739	6/4	19/4

Figure S8. MS/MS spectra of S-(3-Methyl-2-butenyl) ethanethioate with a carnitine modification