



UNIVERSITÀ  
DEGLI STUDI  
FIRENZE

FLORE

## Repository istituzionale dell'Università degli Studi di Firenze

### **Comparison of chemometric strategies for potential exposure marker discovery and false-positive reduction in untargeted metabolomics:**

Questa è la versione Preprint (Submitted version) della seguente pubblicazione:

*Original Citation:*

Comparison of chemometric strategies for potential exposure marker discovery and false-positive reduction in untargeted metabolomics: application to the serum analysis by LC-HRMS after intake of Vaccinium fruit supplements / Renai L.; Ancillotti C.; Ulaszewska M.; Garcia-Aloy M.; Mattivi F.; Bartoletti R.; Del Bubba M.. - In: ANALYTICAL AND BIOANALYTICAL CHEMISTRY. - ISSN 1618-2642. - STAMPA. - 414:(2022), pp. 1841-1855. [10.1007/s00216-021-03815-5]

*Availability:*

This version is available at: 2158/1256644 since: 2022-02-16T22:54:44Z

*Published version:*

DOI: 10.1007/s00216-021-03815-5

*Terms of use:*

Open Access

La pubblicazione è resa disponibile sotto le norme e i termini della licenza di deposito, secondo quanto stabilito dalla Policy per l'accesso aperto dell'Università degli Studi di Firenze (<https://www.sba.unifi.it/upload/policy-oa-2016-1.pdf>)

*Publisher copyright claim:*

(Article begins on next page)



**Comparison of chemometrics workflows for potential exposure markers discovery and false positive reduction in untargeted metabolomics: application to the serum analysis by LC-HRMS after intake of Vaccinium fruits supplements**

Journal:	<i>Analytical and Bioanalytical Chemistry</i>
Manuscript ID	Draft
Type of Paper:	Research Paper
Date Submitted by the Author:	n/a
Complete List of Authors:	Renai, Lapo; UNIVERSITY OF FLORENCE, DEPARTMENT OF CHEMISTRY Ancillotti, Claudia; UNIVERSITY OF FLORENCE, DEPARTMENT OF CHEMISTRY Ulaszewska, Marynka ; IRCCS Ospedale San Raffaele, Center for Omics Sciences, Proteomics and Metabolomics Facility (ProMeFa) Garcia-Aloy, Mar; Edmund Mach Foundation Research and Innovation Centre, Metabolomics Unit, Department of Food Quality and Nutrition Mattivi, Fulvio; University of Trento Department of Cellular Computational and Integrative Biology Bartoletti, Riccardo; University of Pisa, Department of Translational Research and New Technologies DEL BUBBA, MASSIMO; UNIVERSITY OF FLORENCE, DEPARTMENT OF CHEMISTRY
Keywords:	untargeted metabolomics, multivariate analysis, univariate parametric analysis, univariate non-parametric analysis, exposure markers, nutrikinetic



UNIVERSITÀ  
DEGLI STUDI  
FIRENZE

DIPARTIMENTO  
DI CHIMICA  
"UGO SCHIFF"

EDITORIAL OFFICE  
Analytical and Bioanalytical Chemistry

Dear Editor,

I send you the manuscript entitled "Comparison of chemometrics workflows for potential exposure markers discovery and false positive reduction in untargeted metabolomics: application to the serum analysis by LC-HRMS after intake of *Vaccinium* fruits supplements" for the submission to Analytical and Bioanalytical Chemistry.

Corresponding author and the submitter of the manuscript: Prof. Massimo Del Bubba, Department of Chemistry, University of Florence, Via della Lastruccia n.3, 50019 Sesto Fiorentino (Florence), Italy. E-mail address: [massimo.delbubba@unifi.it](mailto:massimo.delbubba@unifi.it)

In this paper, we performed a non-trivial comparison of different chemometrics workflows for potential exposure markers (PEMs) discovery in nutrimental metabolomics, identifying a novel chemometrics strategy that allows for reducing the false discovery selection of features, which were conversely retained in the highly populated group of "significant features", generated by other statistical methods, commonly adopted in literature.

An untargeted LC-HRMS study of serum samples from interventions with *Vaccinium myrtillus* (VM) and *Vaccinium corymbosum* (VC) supplements was chosen as representative in the field of LC-HRMS nutrimental metabolomics, since it results in very complex outputs of data, which require the implementation of proper chemometrics workflows for data pretreatment and PEMs selection.

This aspect is often disregarded in metabolomics studies, where only one chemometrics protocol is applied without any explanation and in some cases protocols for PEMs discovery are even used not properly, e.g. when univariate

**Prof. Massimo Del Bubba**

Via della Lastruccia, 3 – 50019 Sesto Fiorentino (Firenze)  
phone +39 055 4573326 | e-mail: [massimo.delbubba@unifi.it](mailto:massimo.delbubba@unifi.it)  
P.IVA | Cod. Fis. 01279680480



1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12 parametric tests are used without verifying normality and homoscedasticity of data  
13 distributions or in absence of these prerequisites.  
14

15  
16 Twelve PEMs related to the intake of bilberry and/or blueberry supplements were  
17 annotated in serum samples, five of them (i.e. octahydro-methyl- $\beta$ -carboline-  
18 dicarboxylic acid and tetrahydro-methyl- $\beta$ -carboline-dicarboxylic acid for VC, citric  
19 acid for VM, and caprylic acid and azelaic acid for both VC and VM) reported here  
20 for the first time as serum metabolites of these interventions, thus representing an  
21 additional element of novelty.  
22  
23

24  
25  
26  
27 The comparison of the two interventions through a rigorous chemometrics approach,  
28 based on the analysis of the area under the curve of the longitudinal dataset followed  
29 by principal component analysis and linear discriminant analysis, highlighted  
30 thirteen statistically significant PEMs discriminating the two interventions, including  
31 four intra-intervention relevant metabolites (i.e. abscisic acid glucuronide, catechol  
32 sulphate, methyl-catechol sulphate, and  $\alpha$ -hydroxy-hippuric acid). All these findings  
33 represent a further novelty element.  
34  
35

36  
37  
38  
39 Based on the "Aims and Scope" of Analytical and Bioanalytical Chemistry,  
40 encouraging the publication of papers dealing with advanced chemometrics data  
41 analysis and omics analytical strategies, and the manuscript contents, I think that this  
42 paper is in line with the scope of Analytical and Bioanalytical Chemistry and it has  
43 enough quality to deserve the publication in this high reputable journal.  
44  
45  
46  
47

48  
49 I thank you in advance for your consideration and I send you my best regards.  
50  
51

52 Sesto Fiorentino, October, 5th 2021  
53

54 Massimo Del Bubba

A handwritten signature in black ink that reads 'Massimo Del Bubba'.

55  
56  
57  
58  
59  
60



UNIVERSITÀ  
DEGLI STUDI  
FIRENZE

DIPARTIMENTO  
DI CHIMICA  
"UGO SCHIFF"

- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9
- 10
- 11
- 12
- 13
- 14
- 15
- 16
- 17
- 18
- 19
- 20
- 21
- 22
- 23
- 24
- 25
- 26
- 27
- 28
- 29
- 30
- 31
- 32
- 33
- 34
- 35
- 36
- 37
- 38
- 39
- 40
- 41
- 42
- 43
- 44
- 45
- 46
- 47
- 48
- 49
- 50
- 51
- 52
- 53
- 54
- 55
- 56
- 57
- 58
- 59
- 60

1  
2  
3 **1 Comparison of chemometrics workflows for potential exposure markers**  
4  
5  
6 **2 discovery and false positive reduction in untargeted metabolomics:**  
7  
8 **3 application to the serum analysis by LC-HRMS after intake of *Vaccinium***  
9  
10 **4 fruits supplements**  
11  
12  
13  
14  
15

16 Lapo Renai<sup>a</sup>, Claudia Ancillotti<sup>a</sup>, Marynka Ulaszewska<sup>b,c</sup>, Mar Garcia-Aloy<sup>c</sup>, Fulvio Mattivi<sup>d</sup>,  
17  
18 Riccardo Bartoletti<sup>e</sup>, Massimo Del Bubba<sup>a\*</sup>  
19  
20

21  
22  
23 9 a) Department of Chemistry, University of Florence, Via della Lastruccia 3, 50019 Sesto  
24  
25 10 Fiorentino (Florence, Italy)

26  
27 11 b) IRCCS Ospedale San Raffaele, Center for Omics Sciences, Proteomics and Metabolomics  
28  
29 12 Facility (ProMeFa), Via Olgettina, 60, 20132 Milan (Italy)

30  
31 13 c) Metabolomics Unit, Department of Food Quality and Nutrition, Research and Innovation  
32  
33 14 Centre, Fondazione Edmund Mach (FEM), Via Mach 1, 38098 San Michele all'Adige (Trento,  
34  
35 15 Italy)

36  
37 16 d) Department of Cellular, Computational, and Integrative Biology (CIBIO), University of  
38  
39 17 Trento, Via Sommarive 9, 38123 Povo (Trento, Italy)

40  
41 18 e) Department of Translational Research and New Technologies, University of Pisa, Via  
42  
43 19 Risorgimento 36, 56126 Pisa (Italy)

44  
45  
46 20  
47  
48 21 \* Corresponding author – phone number: +39 055 4573326; e-mail address:  
49  
50 22 massimo.delbubba@unifi.it  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

**Abstract**

Untargeted liquid chromatographic-high resolution mass spectrometric (LC-HRMS) metabolomics for potential exposure markers (PEMs) discovery in nutrkinetic studies generates complex outputs. The correct selection of statistically significant PEMs is a crucial analytical step for understanding nutrition-health interactions. Hence, in this paper different chemometrics selection workflows for PEMs discovery, using multivariate or univariate parametric or non-parametric data analyses were comparatively tested and evaluated. The PEMs selection protocols were applied to a small sample size untargeted LC-HRMS study of a longitudinal set of serum samples from 20 volunteers after a single intake of (poly)phenolic-rich *Vaccinium myrtillus* and *Vaccinium corymbosum* supplements. The non-parametric Games-Howell test identified a restricted group of significant features, thus minimizing the risk of false positive retention. Among the forty-seven PEMs exhibiting a statistically significant postprandial kinetics, twelve were successfully annotated as purine pathway metabolites, benzoic and benzodiol metabolites, indole alkaloids, and organic and fatty acids, and five (i.e. octahydro-methyl- $\beta$ -carboline-dicarboxylic acid, tetrahydro-methyl- $\beta$ -carboline-dicarboxylic acid, citric acid, caprylic acid, and azelaic acid) were associated to *Vaccinium* berry consumption for the first time. The analysis of the area under the curve of the longitudinal dataset highlighted thirteen statistically significant PEMs discriminating the two interventions, including four intra-intervention relevant metabolites (i.e. abscisic acid glucuronide, catechol sulphate, methyl-catechol sulphate, and  $\alpha$ -hydroxy-hippuric acid). Principal component analysis and samples classification through linear discriminant analysis performed on PEMs maximum intensity confirmed the discriminating role of these PEMs.

**Keywords:** untargeted metabolomics; multivariate analysis; univariate parametric analysis; univariate non-parametric analysis; exposure markers; nutrkinetic

## 1 Introduction

Untargeted metabolomics has been extensively recognized as the leading approach for the investigation of potential exposure markers (PEMs) of food (1, 2), also in relation to food-health interaction studies (3) (i.e. nutrimetabolomics), as it allows in principle the comprehensive overview of the human metabolome. Untargeted metabolomics platforms usually consist in liquid chromatography (LC) hyphenated with high-resolution mass spectrometry (HRMS), which guarantees the determination of a much larger number of biological metabolites and/or a more comprehensive structural investigation, compared to other techniques such as gas chromatography coupled to HRMS or LC coupled to nuclear magnetic resonance (4).

The use of LC-HRMS platforms in nutrimetabolomics generally results in very complex outputs of data, which require the implementation of proper workflows for their pretreatment and processing (5). Currently, no standard protocol has been defined to this aim, even though many workflows of data processing and analysis have been proposed (6). Data handling is usually performed by chemometrics tools, specifically selected according to the study design adopted, which may largely vary depending on a number of factors, such as study design (i.e. cross over or parallel), the number of interventions, sample size, and data acquisition strategies (e.g. standalone full scan or coupled with MS<sup>n</sup> acquisitions) (5, 7). In nutrimetabolomics studies, food PEMs discovery commonly involves the application of multivariate analysis, using partial least square-discriminant analysis (PLS-DA) and the variable importance in projection (VIP) classifier as filtering strategy of statistically significant features from the original dataset (8-10). However, multivariate methods such as PLS-DA-VIP, due to their probabilistic nature, tend to select false positive (variables not causally related to groups) and/or false negative (missing of relevant features) (11). On the other hand, the multiple univariate approaches, which have been applied less frequently in nutrimetabolomics, although



1  
2  
3 76 not associated *per se* with false discovery risks, need the verification of some data distribution  
4  
5 77 pre-requisites, i.e. normality or homoscedasticity (12). However, it should be emphasized that  
6  
7 78 multiple univariate approaches have been frequently applied as parametric methods (e.g. *t*-test  
8  
9  
10 79 and ANOVA), although the sample size was so small to suggest the absence of the  
11  
12 80 aforementioned prerequisite (13, 14), thus involving a high risk of generating false positive. In  
13  
14 81 order to reduce the false positive risk, an adjustment of the false positive rate should be  
15  
16 82 performed for the application of multiple inference tests (15). Conversely, the use of non-  
17  
18 83 parametric tests is much less investigated (16), although the application of this type of  
19  
20 84 chemometrics tools would release the data analysis procedure from the constraints of  
21  
22 85 homoscedasticity and normal distribution of data.

23  
24  
25 86 Even though the impact of different chemometrics tools in the PEMs selection is expected to  
26  
27 87 be remarkable, this issue is poorly investigated in the literature, especially in LC-HRMS  
28  
29 88 nutrimentalomics studies that, as mentioned before, are prone by their nature to generate a  
30  
31 89 large number of features to be tested for their statistical significance (17). It is therefore  
32  
33 90 important to compare the effects of the use of different chemometrics approaches for the  
34  
35 91 selection of statistically significant PEMs. This issue is particularly important in the light of  
36  
37 92 the intrinsic complexity of the biological fluids (e.g. serum and urine) investigated in this type  
38  
39 93 of studies. Among these matrices, serum and plasma are particularly relevant as matrices that  
40  
41 94 contain substances responsible for a direct biological activity, the study of which is of  
42  
43 95 remarkable importance to establish a cause-effect role between the intake of a certain food and  
44  
45 96 the human health benefits.

46  
47 97 Based on the aforementioned considerations, this paper aimed at comparing the effectiveness  
48  
49 98 of chemometrics protocols for PEMs discovery and false positive reduction, using different  
50  
51 99 data analysis approaches, which included the widely adopted PLS-DA-VIP, the parametric *t*-  
52  
53 100 test (before and after *P*-value adjustment), and the non-parametric Wilcoxon and Games-  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 101 Howell tests (*P*-value adjusted for both tests), the latter never investigated before in untargeted  
4  
5 102 nutrimentalomics.

6  
7  
8 103 The protocols have been tested on the untargeted LC-HRMS analysis of serum samples from  
9  
10 104 an intervention study of *Vaccinium myrtillus* (VM) and *Vaccinium corymbosum* (VC)  
11  
12 105 supplements as a representative application. In this regard, it should be emphasized that this  
13  
14 106 kind of application is of considerable intrinsic importance for nutrimentalomics purposes. In  
15  
16 107 fact, among fruits, small berries represent an abundant source of phenolic compounds in human  
17  
18 108 diet (18). In detail, the fruits belonging to the *Vaccinium* genus, above all *V. myrtillus* berries,  
19  
20 109 have been suggested as functional foods and used for supplement and drug preparation (19,  
21  
22 110 20). Moreover, *in vitro* studies have shown anti-proliferative and pro-apoptotic effects of  
23  
24 111 polyphenol-rich berry extracts against different prostate cancer cell lines (21) and the  
25  
26 112 chemopreventive properties of an anthocyanin-rich *V. myrtillus* extract were suggested in a  
27  
28 113 pilot study with patients affected by colon cancer (22). It should also be noted that few targeted  
29  
30 114 (23, 24) and only one untargeted (1) metabolomics studies have been published so far on human  
31  
32 115 serum and urinary (poly)phenolic metabolites of *Vaccinium* berries. **Figure 1** illustrates a  
33  
34 116 comprehensive and intuitive scheme of the analytical workflow adopted in this study for the  
35  
36 117 PEMs discovery in human serum after the administration of bilberry or blueberry supplement.  
37  
38  
39  
40  
41

## 42 118 **2 Material and Methods**

### 43 119 2.1 Reagents and standards

44  
45  
46 120 Methanol and acetonitrile LC-MS Ultra CHROMASOLV™ and formic acid were purchased  
47  
48 121 from Sigma Aldrich (St. Louis, MO, USA). The ultrapure water was obtained by purifying  
49  
50 122 demineralized water in a Milli-Q system from Millipore (Bedford, MA, USA). O-hydroxy-  
51  
52 123 hippuric acid, abscisic acid, azelaic acid, 1-methyl-1,2,3,4-tetrahydro-β-carboline-3-carboxylic  
53  
54 124 and L-tryptophan-2',4',5',6',7'-d5 (TRI-d5) reference standards were obtained from Sigma  
55  
56  
57  
58  
59  
60

1  
2  
3 125 Aldrich, while trans-cinnamic-d5 acid (CIN-d5), N-Benzoyl-d5-glycine (Hippuric acid-d5,  
4  
5 126 HIP-d5) from CDN ISOTOPES Inc.  
6

## 7 127 2.2 *Vaccinium myrtillus* and *Vaccinium corymbosum* berry supplements

8 128 The supplements were obtained by cryo-milling VM and VC freeze-dried berries, as detailed  
9  
10 129 in the section 1 of the *Supplementary material*. The supplements were characterized for total  
11  
12 130 soluble polyphenols, total monomeric anthocyanins and for some phenolic compounds, using  
13  
14 131 methods elsewhere described (25). Full details of the characterization analysis (**Tables S1 and**  
15  
16 132 **S2**) and results obtained (**Tables S3-S5**) are reported in the section 2 of the *Supplementary*  
17  
18 133 *material*.  
19  
20  
21  
22

## 23 134 2.3 Study design

24 135 The study was a randomized, single-blinded two-arms intervention of single intake of VM and  
25  
26 136 VC, involving twenty healthy voluntary subjects (11 males and 9 females aged between 25 and  
27  
28 137 60 years). The study followed the guidelines set by the Helsinki Declaration  
29  
30 138 (<http://www.fda.gov/ohrms/dockets/dockets/06d0331/06D-0331-EC20-Attach-1.pdf>) and all  
31  
32 139 subjects provided written informed consent prior to the study. Ethical approval for freeze-dried  
33  
34 140 VM and VC powdered supplement administration was obtained for a phase I-II study (approval  
35  
36 141 n. SPE 14.178 AOUC, 30th May 2016).  
37  
38  
39

40 142 All the subjects convened early in the morning after 10 h of fasting and were randomly divided  
41  
42 143 in two groups according to an electronic randomisation key. A single dose of 25 g of VM  
43  
44 144 supplement mixed with 500 mL of water was orally administered to the first group, whereas  
45  
46 145 the same amount of VC berry supplement mixed with 500 mL of water was orally administered  
47  
48 146 to the second group. Serum was collected at baseline and different sampling times (30, 60, 120,  
49  
50 147 240 and 360 minutes) after the supplement assumption, as described in section 3 (**Table S6**) of  
51  
52 148 the *Supplementary material*. After the collection, samples were divided in aliquots of 500  $\mu$ L,  
53  
54 149 frozen at -80°C and stored until LC-HRMS analysis was performed.  
55  
56  
57  
58  
59  
60

## 150 2.4 Serum extraction

151 Serum extraction was performed on Waters Ostro 96-well plates as reported elsewhere (1).

152 Quality of the untargeted analysis was guaranteed by the use of quality control (QC) samples

153 and internal and external standards. See details in the section 4 of the *Supplementary material*.

## 154 2.5 LC-MS and LC-MS/MS analysis

155 LC analysis was performed on a Dionex (Sunnyvale, CA, USA) Ultimate 3000 HPLC system

156 equipped with a Kinetex C18 column (150 mm x 2.1 mm I.D., particle size 2.6  $\mu\text{m}$ ) and a guard

157 column containing the same stationary phase (Phenomenex, Torrance, CA, USA). The LC

158 system was coupled to a hybrid linear ion trap Fourier Transform (LTQ FT) Orbitrap high-

159 resolution mass spectrometer (Thermo Fisher, Waltham, MA, USA) by an electrospray

160 ionization (ESI) probe for MS and MS/MS analysis both in positive (PI) and negative (NI)

161 ionization. Each sample was analysed under PI and NI, using two different mass acquisition

162 methods for each ionization mode. The first method consisted of a Full Scan method (mass

163 range from 100 to 1000 Da) at a mass resolution of 30,000 FWHM ( $m/z$  400) in centroid mode.

164 The second method included data dependent acquisition (DDA) scans for the three most

165 abundant ions per scan at resolution of 7500 FWHM. Product ions were generated in the LTQ

166 trap at collision energy of 35 eV using an isolation width of 2 Da. Further details related to MS

167 acquisition are reported in the section 5 of the *Supplementary material*.

## 168 2.6 Data pre-treatment

169 All raw data were manually inspected using the Qual browser module of Xcalibur version 2.0.7

170 (Thermo Fisher). The LC-MS raw files were converted to mzML format using the MSConvert

171 utility included in ProteoWizard (26). Then, the mzML files were processed with the *XCMS R*

172 package that allows for obtaining detection and retention time alignment of features (i.e. ions

173 characterized by a unique exact mass and retention time) (2, 27). During this processing, the

174 maximum value for mass accuracy and retention time deviation was set equal to 5 ppm and 2

1  
2  
3 175 seconds, respectively. Data pre-processing of Full Scan acquisitions was performed according  
4  
5 176 to Garcia-Aloy et al. 2020 (28). Briefly, each dataset was filtered to discard noisy and irrelevant  
6  
7 177 features. Firstly, those features whose first decimal place took a value between 4 and 8 were  
8  
9 178 removed from each data set. Subsequently, the signals that showed a higher variation  
10  
11 179 coefficient in QC samples than within study samples were also excluded. A random number  
12  
13 180 between 1 and 500 replaced all zero values. These thresholds were fixed empirically according  
14  
15 181 to the intensity level representing the noise of the chromatogram. The principal component  
16  
17 182 analysis (PCA) on log-transformed and Pareto-scaled data was applied to check data quality  
18  
19 183 (i.e. possible batch effect) and to detect the presence of outliers.

## 20 184 2.7 Chemometrics analyses

21 185 PEMs in postprandial responses were selected separately for VM and VC interventions using  
22  
23 186 a selection protocol consisting in a two-step procedure: i) statistical significance among time  
24  
25 187 points within a same intervention and ii) nutrikinetic discrimination between interventions.

### 26 188 2.7.1 *Statistical significance among time points of a same intervention*

27 189 For the identification of PEMs suitable for the discrimination among time points, the following  
28  
29 190 approaches were investigated: univariate post-hoc non-parametric Games-Howell and  
30  
31 191 Wilcoxon signed-rank tests, univariate parametric *t*-test, and multivariate PLS-DA-VIP. A  
32  
33 192 short overview of the characteristics of these tests is reported in section 6 of the *Supplementary*  
34  
35 193 *material (Table S7)*.

36 194 For the univariate approaches, we first selected those features with increasing mean values of  
37  
38 195 chromatographic intensity in at least two consecutive time points, against baseline samples (i.e.  
39  
40 196 time point "0 min"). Normal distribution of data was investigated by the application of Shapiro-  
41  
42 197 Wilk's test for multiple variables, while homogeneity of variance was evaluated adopting a  
43  
44 198 Bartlett's test for multiple independent variables, comparing the variances of VM and VC  
45  
46 199 datasets in both polarities. Then, the *t*-test and the post-hoc non-parametric Games-Howell and  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 200 Wilcoxon tests were applied to evaluate the statistical significance between time point “0 min”  
4  
5 201 and all the others, setting the threshold  $P$ -value at 0.05. For both  $t$ -test and non-parametric tests,  
6  
7 202 the Benjamini-Hochberg procedure was applied to decrease the false discovery rate  
8  
9 203 (significance level 5%).

10  
11  
12 204 The multivariate PLS-DA-VIP approach was performed on (i) the pre-treated VM and VC  
13  
14 205 datasets and (ii) using the *plsda* package implemented with VIP scoring, setting the inclusion  
15  
16 206 cut-off as  $> 1$ , as elsewhere reported (11, 16).

### 17 207 2.7.2 VM and VC discrimination by nutrikinetic data

18  
19 208 For each feature selected as significant in VM and/or VC, the baseline sample was subtracted  
20  
21 209 from the intensities of the other time points, and in case of negative values, the replacement by  
22  
23 210 zero was performed. Nutrikinetic significant features were selected using a two-step procedure  
24  
25 211 based on their nutrikinetic curve behaviour. For that purpose, according to a previous  
26  
27 212 intervention study (28), features that in at least two consecutive points exhibited the 25<sup>th</sup>  
28  
29 213 percentile of one group ( $n = 2$ ) higher than the 75<sup>th</sup> percentile of the other group ( $n = 8$ ), were  
30  
31 214 included in the discrimination process based on comparison of area under the curve (AUC).  
32  
33 215 AUC of each selected feature was calculated between time 0 and 360 min using the *pracma* R  
34  
35 216 package. Differences of AUCs among diets were tested using the Kruskal-Wallis test and the  
36  
37 217 obtained  $P$ -values were adjusted using the Benjamini Hochberg method. Adjusted  $P$ -  
38  
39 218 values  $< 0.05$  were considered statistically significant. Descriptive analysis for evaluating the  
40  
41 219 discrimination between the two interventions was carried out by means of PCA using the  
42  
43 220 *FactoMineR* R package. Quality control of PCA was performed using QCs by visually  
44  
45 221 verifying if PCA object scores obtained by replicated injections of the QC sample were close  
46  
47 222 to the origin of PCA coordinates. Linear discriminant analysis (LDA) was performed on QC,  
48  
49 223 VC, and VM sample groups made up by the annotated PEMs discriminating the two  
50  
51 224 interventions, using the *MASS* R package.  
52  
53  
54  
55  
56  
57  
58  
59  
60

## 225 2.8 Metabolite identification

226 The identification of features of interest, evidenced by data analysis, was performed according  
227 to the criteria previously reported by our research group (19), using the workflow described  
228 below. (i) Comparison of the exact mass of the experimental precursor ion with the pseudo-  
229 molecular ions proposed by the MzCloud ([www.mzcloud.org](http://www.mzcloud.org)), Humane Metabolome Database  
230 (HMDB, [www.hmdb.ca](http://www.hmdb.ca)), MassBank of North America (MoNA,  
231 <http://mona.fiehnlab.ucdavis.edu>), and Kyoto Encyclopedia of Genes and Genomes (KEGG)  
232 libraries, selecting a mass accuracy ( $\Delta$ )  $\leq$  5 ppm as tolerance threshold. (ii) Export of the  
233 isotopic profile of pseudo-molecular ions selected within libraries at step (i) and subsequent  
234 comparison with the isotopic profile of the experimental precursor ions, selecting features  
235 providing an isotope ratio percentage difference of 20% as tolerance threshold. (iii) Structural  
236 elucidation of features of interest through the evaluation of MS/MS spectra obtained with DDA  
237 mass method in comparison with matched library mass spectra. (iv) Feature annotation  
238 performed according to golden standards for metabolomics (29).

## 239 3 Results and discussion

### 240 3.1 Assessment of data quality

241 Internal standards in study samples and QC samples showed a variation of retention time and  
242 mass accuracy below 2 seconds and 5 ppm, respectively, thus highlighting the correct data  
243 acquisition. Moreover, the variation of integrated area of the surrogate standards (added before  
244 the extraction) and internal standard (added before the analysis) in all the QC samples resulted  
245 lower than 20% confirming the repeatability of the extraction and excluding the possibility of  
246 signal drift and carry over phenomena during the LC-MS and LC-MS/MS analysis. Moreover,  
247 by means of Pareto-scaled PCA performed on raw data (data not shown), it was possible to  
248 observe that QC samples are well grouped in component space, showing no batch effect, or

249 trend due to the injection order (i.e. drop in signal intensity). Study samples are distributed  
250 homogeneously with no visible trends according to post-prandial time point.

### 251 3.2 Chemometrics workflow for feature selection

252 After data pre-treatment, complex data sets were obtained for both interventions, being the total  
253 number of features 12091 and 5361 for PI and NI, respectively.

254 The supervised multivariate PLS-DA-VIP analysis was performed on these groups of features,  
255 as commonly done in literature, highlighting a very high number of statistically significant  
256 features (2283-8639, depending on the dataset considered). This result is probably due to the  
257 tendency of the PLS-DA-VIP approach to select false positives. In fact, this criterion is very  
258 reasonable to discard irrelevant variables, but it may have drawbacks if used for assessing the  
259 significance of features (11). Moreover, model classification error rate of maximized distances  
260 was very high, ranging from 0.9 to 0.7, regardless the number of selected components for PLS-  
261 DA modelling.

262 For the application of univariate data analysis, the total number of features was preliminarily  
263 filtered by a selection procedure that preserves only the features showing increasing intensities  
264 in at least two consecutive time points. This feature selection resulted in a number of relevant  
265 postprandial features ranging from 4245 to 4932, depending on the intervention and acquisition  
266 mode (**Table 1**). The obtained datasets were checked for (i) normal distribution and (ii)  
267 homogeneity of variance (i.e. homoscedasticity). Normality was evaluated by the application  
268 of the Shapiro-Wilk's test for multiple variables. Shapiro-Wilk's correlation coefficients and  
269 *P*-values (in bracket) ranged as follows: (i) VC – NI 0.538-0.897 ( $1.0 \cdot 10^{-5}$ - $4.9 \cdot 10^{-2}$ ); (ii) VC –  
270 PI 0.656-0.761 ( $3.0 \cdot 10^{-3}$ - $4.9 \cdot 10^{-2}$ ); (iii) VM – NI 0.366-0.807 ( $1.0 \cdot 10^{-7}$ - $3.4 \cdot 10^{-2}$ ); (iv) VM – PI  
271 0.609-0.734 ( $2.0 \cdot 10^{-4}$ - $5.0 \cdot 10^{-3}$ ), thus highlighting the absence of normality of the investigated  
272 datasets. Homogeneity of variance was evaluated adopting the Bartlett's test for multiple  
273 independent variables, comparing the variances of VC and VM datasets in both polarities. The



274 results obtained were characterized by  $P$ -value  $< 0.05$  (confidence interval = 0.95), thus  
275 evidencing lack of homogeneity among variances of the aforementioned datasets.

276 The aforementioned univariate tests were used for the PEMs selection within the  
277 aforementioned datasets of relevant postprandial features, obtaining more or less populated  
278 groups of statistically significant features in each acquisition polarity (**Table 1**). The adoption  
279 of the  $t$ -test generated the largest group of “statistically significant” PEMs, which surely  
280 includes a high number of false positives, even after the  $P$ -value correction using the  
281 Benjamini-Hochberg procedure (1051-2154 features, depending on the dataset considered),  
282 due to the non-normal and heteroscedastic distribution of data, which should discourage the  
283 use of parametric test for PEMs selection. The Wilcoxon test was found to be much more  
284 restrictive (75-359 features) than the adjusted  $t$ -test, in accordance with its non-parametric  
285 character, which does not require compliance with the conditions of normality and  
286 heteroscedasticity of data distribution, making it therefore more suitable for the treatment of  
287 small sample size, for which the above-mentioned conditions usually do not occur. A  
288 significantly reduced number of statistically significant features (i.e. 80 and 29 for VM and VC  
289 interventions, respectively) was identified using the Games-Howell test, thus highlighting its  
290 strong selectivity, which greatly reduces the probability of including false positives in the  
291 PEMs group.

292 Using some features that can be annotated easily due to their previous identifications in bilberry  
293 and/or blueberry intervention studies as representative cases, it is interesting to evaluate how  
294 these features are treated by the procedures of PEMs selection here tested. For instance, benzoic  
295 acid was annotated here ( $\Delta = -0.4$  ppm) after the intake of VC supplement and retained in the  
296 statistically significant PEMs group using the adjusted  $t$ -test. However, this feature was  
297 excluded when PLS-DA-VIP or the non-parametric Wilcoxon and Games-Howell tests for  
298 PEMs selection were applied. Moreover, hippuric acid was annotated here ( $\Delta = -0.6$  ppm) in

299 both interventions and considered statistically significant by PLS-DA-VIP, adjusted *t*-test, and  
300 Wilcoxon test but excluded by the Games-Howell treatment. It is interesting to note that these  
301 features were considered statistically significant in a study investigating plasma after blueberry  
302 intake using ANOVA and post-hoc Bonferroni correction (24). As further examples, four  
303 metabolites (i.e. hydroxy-methoxy hippuric acid, hydroxy-(hydroxy-methoxyphenyl)-  
304 pentenoic acid glucuronide, dihydroxyphenyl propionic acid glucuronide, and hydroxypheny  
305 propionic acid sulphate) recently annotated as PEMs of bilberry intake (1), resulted here  
306 statistically significant by applying the Wilcoxon signed-rank test, while they were discarded  
307 by the Games-Howell test, due to the different assumptions made with respect to Wilcoxon  
308 test.

309 All these examples evidenced that the selection of significant features is strongly dependent on  
310 the statistical tool adopted. Based on these findings, great caution should be used in reporting  
311 metabolites as statistically significant for a given intervention. The application of highly  
312 conservative statistical methods, i.e. prone to minimizing the risk of false positive results is  
313 certainly useful in this sense and should be encouraged, although it involves a certain risk of  
314 excluding false negative from the significant dataset. On the contrary, the use of less restrictive  
315 tests should be used with great caution in PEMs discovery, even though they give a wider  
316 overview of features that could become significant when moving from a pilot study, i.e.  
317 characterized by a limited number of subjects, to one on a larger validation cohort.

### 318 3.3 PEMs annotation

319 Among the features selected by the Games-Howell test, corresponding to 47 PEMs, the  
320 annotation protocol allowed for identifying 5 features in PI and 13 features in NI,  
321 corresponding to a total of 12 PEMs. **Table 2** reports the annotated metabolites, providing their  
322 spectral characteristics, formula, time points for statistically significant maximum intensities  
323 ( $T_{\max}$ ) with related *P*-values, the annotation level, and for annotation levels I and II, the

1  
2  
3 324 HMDB/KEGG identification numbers. Among the twelve PEMs identified, eight were  
4  
5 325 annotated only after VM intervention (peaks 2-6, 9, 11, and 12), three only after VC ingestion  
6  
7 326 (peaks 7, 8, and 10), and only one (peak 1) as common response to the consumption of the two  
8  
9 327 supplements.

10  
11  
12 328 **Peak 1** showed two pseudo-molecular ions in NI Full Scan spectra, i.e. at  $m/z$  335.0504 and  
13  
14 329  $m/z$  167.0214. The second ion was characterized by a fragmentation pattern that matched with  
15  
16 330 uric acid, as reported in mzCloud spectral library and literature findings (30). **Peak 2** was  
17  
18 331 putatively annotated as citric acid, due the occurrence of precursor ions at  $m/z$  193.0335 and  
19  
20 332 191.0202, in PI and NI, respectively. MS/MS fragmentation pattern included as main fragments  
21  
22 333 in NI mode peaks at  $m/z$  173 (loss of a water molecule) and  $m/z$  111.0091 (loss of acetic acid),  
23  
24 334 whilst in PI  $m/z$  132.1019 (loss of acetic acid) was highlighted. This attribution was confirmed  
25  
26 335 by mzCloud spectral library and previous researches (31). **Peak 3** exhibited  $[M-H]^-$  ion at  $m/z$   
27  
28 336 267.0744, characterized by the neutral loss of 132 Da (pentose moiety) that led to the formation  
29  
30 337 of a fragment at  $m/z$  135.0313 associated to hypoxanthine, in accordance with spectral libraries  
31  
32 338 and previous findings (32). Thus, this metabolite was putatively identified as inosine, a  
33  
34 339 nucleoside naturally occurring in the pathway of purine metabolism after vegetable and/or fruit  
35  
36 340 intake (33). **Peak 4** gave rise to a  $[M-H]^-$  pseudo-molecular ion at  $m/z$  194.0454 Da, which  
37  
38 341 fragmented by losing 44 Da ( $CO_2$ ) originating the fragment at  $m/z$  150.0561. Since the pseudo-  
39  
40 342 molecular ion matched with high accuracy ( $\Delta=-2.6$  ppm) the mass of a hydroxy-hippuric acid  
41  
42 343 and no other fragments were detected in the  $MS^2$  spectrum,  $\alpha$ -hydroxy-hippuric acid, p-  
43  
44 344 hydroxy-hippuric acid, and o-hydroxy-hippuric acid reference standards were injected,  
45  
46 345 obtaining  $t_R$  of 3.3, 3.8, and 5.4 minutes, respectively. Accordingly, peak 4 was annotated at  
47  
48 346 level I as  $\alpha$ -hydroxy-hippuric acid. **Peak 5** was putatively identified as benzodiol (catechol)  
49  
50 347 sulphate owing to the presence of the typical loss of sulphate ( $m/z$  79.9576) and the appearance  
51  
52 348 of the hydroxy-phenol ion at  $m/z$  109.0297 in the MS/MS spectrum (1, 34). An analogous  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 349 fragmentation pattern was observed for **peak 6**, putatively identified as methyl-catechol  
4  
5 350 sulphate, due to the presence in the MS/MS spectrum of a peak at  $m/z$  188.9860 (loss of  
6  
7 351 methyl), together with typical losses of sulfonic ( $m/z$  123.0454) and sulphate ( $m/z$  108.0220)  
8  
9 352 groups. **Peak 7** (PI:  $m/z$  279.1329, NI:  $m/z$  277.1185) and **peak 8** (PI:  $m/z$  275.1024, NI:  $m/z$   
10 353 273.0875) were recognized as indole alkaloids derivatives, based on their mass accuracy and  
11  
12 354 similar MS<sup>n</sup> fragmentation patterns (**Figure S1** of the *Supplementary material*). Spectra of both  
13  
14 355 peaks in PI (**Fig. S1, box A** and **box E**) were characterized by the neutral loss of C<sub>2</sub>H<sub>3</sub>NO<sub>2</sub> (73  
15  
16 356 Da, iminoacetic acid), typical of carboxylic acid derivatives of carbolines (35). Moreover,  
17  
18 357 under NI mode, the pseudo-molecular ions (i.e.  $m/z$  277.1185 and 273.0875) underwent the  
19  
20 358 sequential loss of two CO<sub>2</sub> moieties, indicating the presence of two carboxylic groups (**Fig. S1,**  
21  
22 359 **box B-C** and **box F-G**). In order to confirm the attribution of **peaks 7** and **8** to the carbolines  
23  
24 360 class, their fragmentation patterns were compared with the one of 1-methyl-1,2,3,4-tetrahydro-  
25  
26 361 β-carboline-3-carboxylic acid reference standard (**Fig. S1, boxes D-H**). Interestingly, the  
27  
28 362 pseudo-molecular ion of **peak 8** in NI mode ( $m/z$  273.0875) lost a carboxylic group, originating  
29  
30 363 the feature at  $m/z$  229.0982 (**Fig. S1, box F**), the fragmentation pattern of which (**Fig. S1, box**  
31  
32 364 **G**) matched the one of the analytical standard 1-methyl-1,2,3,4-tetrahydro-β-carboline-3-  
33  
34 365 carboxylic acid (**Fig. S1, box D**) with good mass accuracy and similar ions intensities. It should  
35  
36 366 also be noted that **peak 7** showed in both PI (**Fig. S1, boxes A** vs. **E**) and NI (**Fig. S1, boxes**  
37  
38 367 **B** vs. **F** and **C** vs. **G**) MS<sup>n</sup> spectra a constant difference of about 4.03 Da compared to  
39  
40 368 corresponding ions of **peak 8**. Hence, **peak 7** was annotated as octahydro-methyl-β-carboline-  
41  
42 369 dicarboxylic acid. **Peak 9** showed the [M-H]<sup>-</sup> pseudo-molecular ion in NI at  $m/z$  439.1599,  
43  
44 370 which fragmented through the typical neutral loss of glucuronide-conjugated compounds (176  
45  
46 371 Da) giving rise to the product ion at  $m/z$  263.1288. This ion further fragmented originating the  
47  
48 372  $m/z$  219.1391 and 153.0921 ions (**Figure S2-A** of the *Supplementary material*), which can be  
49  
50 373 attributed to abscisic acid, as also highlighted by injecting the abscisic acid reference standard  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 374 **(Figure S2-B). Peaks 10 and 11** were annotated as caprylic acid and azelaic acid, respectively,  
4  
5 375 by comparing their mass spectra with those present in spectral databases (i.e. mzCloud and  
6  
7 376 HMDB, see **Table 2**). In addition, azelaic acid was further confirmed at Level I by injecting  
8  
9 377 the analytical standard. Finally, **peak 12** was putatively annotated as hydroxy-phenyl propionic  
10  
11 378 acid sulphate thanks to the mass accuracy of its  $[M-H]^-$  pseudo-molecular ion ( $m/z$  245.0119,  
12  
13 379  $\Delta=-2.4$  ppm), the comparison with MS<sup>2</sup> spectral libraries, and based on findings of previous  
14  
15 380 clinical studies on VM intake (1).

### 19 381 3.4 Nutrimetabolomics significance

21 382 The twelve annotated metabolites reported in **Table 2** were categorized as (i) purine pathway  
22  
23 383 metabolites, (ii) benzoic and benzodiol metabolites, (iii) indole alkaloids, and (iv) organic and  
24  
25 384 fatty acids, according to their chemical class and metabolism. **Figures 2-5** report the serum  
26  
27 385 postprandial kinetics of the 12 annotated PEMs, expressed as boxplots of the precursor ion  
28  
29 386 intensity as a function of time. Each boxplot represents the interquartile range (75% of the  
30  
31 387 intensity values are less than or equal to the top value of the box and 25% of the intensity values  
32  
33 388 are less than or equal to the bottom value of the box), the upper and lower whiskers refer to the  
34  
35 389 maximum and minimum data point, respectively, and the line within the box represents the  
36  
37 390 median of the data.

#### 42 391 3.4.1 Purine pathway metabolites

44 392 **Figure 2** illustrates the kinetic profile of the pseudo-molecular ions of uric acid (**peak 1**) in  
45  
46 393 VM (**Fig. 2A**) and VC (**Fig. 2B**), and inosine (**peak 3**) in VM (**Fig. 2C**). According to literature,  
47  
48 394 these metabolites are correlated with the ingestion of fructose, rather than phenolic compounds,  
49  
50 395 since fructose induces acute depletion of ATP and inorganic phosphate and causes increased  
51  
52 396 activity of the enzymes involved in the degradation of purine nucleotides to inosine,  
53  
54 397 hypoxanthine, xanthine, and finally uric acid (36). Statistically significant variations ( $P < 0.05$ )  
55  
56 398 of uric acid and inosine were observed at  $T_{max}$  30-60 min, suggesting that the fructose-induced  
57  
58  
59  
60

1  
2  
3 399 response increases uric acid concentration in plasma in the early stages of metabolism. The  
4  
5 400 increase in uric acid is responsible of the raise of plasma antioxidant activity widely observed  
6  
7 401 in literature after ingestion of fructose-rich fruits (33, 37). The fast depletion of these  
8  
9  
10 402 metabolites after two-three hours from food intake is in agreement with the post-prandial trend  
11  
12 403 commonly observed for phase II metabolites of methylxanthines, which are rapidly excreted  
13  
14 404 through the urinary trait within eight hours from the intake (38, 39).

#### 17 405 3.4.2 Benzoic and benzodiol metabolites

19 406 **Figure 3** shows the kinetic profiles of  $\alpha$ -hydroxy-hippuric acid (**peak 4, Fig. 3A**), the sulphate  
20  
21 407 derivatives of catechol (**peak 5, Fig. 3B**) and methyl-catechol (**peak 6, Fig. 3C**), and  
22  
23 408 hydroxyphenyl propionic acid sulphate (**peak 12, Fig. 3D**), which were detected only in  
24  
25  
26 409 response to the VM intervention. Differently from purine derivatives, the trend observed for  
27  
28 410 these compounds reflected a diverse postprandial scenario. In fact, the increase in signals  
29  
30 411 intensities became statistically significant ( $P < 0.05$ ) starting from two hours after the  
31  
32 412 supplement ingestion ( $T_{\max}$  between 120 and 360 min), suggesting that their occurrence in  
33  
34 413 plasma is associated to microbiota activity, which mediates the degradation of polyphenols in  
35  
36 414 smaller and more soluble molecules (40). In detail, the occurrence in human plasma of both  
37  
38 415 hippuric acid and catechol derivatives was already reported in literature (1, 23, 24) as potential  
39  
40 416 markers of acute and/or long-term ingestion of bilberry and blueberry, and their occurrence  
41  
42 417 related to quercetin glycosides, which are dominant flavonols in these fruits (1, 41, 42).  
43  
44 418 Conversely, phenyl-propionic acid derivatives may derive from caffeic acid degradation and/or  
45  
46 419 from anthocyanins after B-ring cleavage, and C-ring opening and oxidation (43).

#### 51 420 3.4.3 Indole alkaloids

53 421 **Figure 4** reports the kinetic profiles of the octahydro-methyl- $\beta$ -carboline-dicarboxylic acid  
54  
55 422 (**peak 7, Fig. 4A-B**) and the tetrahydro-methyl- $\beta$ -carboline-dicarboxylic acid (**peak 8, Fig. 4C-**  
56  
57 423 **D**), which have been annotated here for the first time in human plasma in response to the VC  
58  
59  
60

1  
2  
3 424 intervention. Interestingly, both PEMs showed a significant increase in their intensities ( $P <$   
4  
5 425 0.05) in plasma at  $T_{\max}$  60 min, then decreasing in the following time points, suggesting that  
6  
7 426 they are rapidly absorbed by human organism and likewise excreted through the urinary tract,  
8  
9  
10 427 as highlighted elsewhere (44). This group of chemicals is already known to occur in fruits,  
11  
12 428 including berries, as well as in food processed products, suggesting that diet may directly  
13  
14 429 contribute to their presence in human bio-fluids (35, 45, 46). However, the occurrence of these  
15  
16  
17 430 alkaloids should not be necessarily associated to the intake of fruit, since other foods such as  
18  
19 431 beer, coffee, cereal products, and vegetables contain these alkaloids and may therefore  
20  
21  
22 432 contribute to their presence in human biofluids (35, 44).

#### 23 433 3.4.4 Organic and fatty acids

24  
25  
26 434 **Figure 5** illustrates the kinetic profiles observed in response to the VM intervention for citric  
27  
28 435 acid (**peak 2, Fig. 5A-B**), abscisic acid glucuronide (**peak 9, Fig. 5C-D**), and azelaic acid (**peak**  
29  
30 436 **11, Fig. 5F**). Moreover, the trend of caprylic acid (**peak 10, Fig. 5E**) after VC intervention is  
31  
32  
33 437 also shown.

34  
35 438 Citric acid showed a postprandial maximum intensity in between 30 and 60 min. Even though  
36  
37 439 this compound is already known as one of the main organic acids occurring in various fruits,  
38  
39 440 including blueberry and especially bilberry (47-49), these data represent the first evidence of a  
40  
41  
42 441 correlation between fruit intake and the occurrence of a statistically significant increase of citric  
43  
44 442 acid in plasma.

45  
46  
47 443 Abscisic acid glucuronide reached a  $T_{\max}$  at 120 min after the supplement consumption,  
48  
49 444 suggesting that it is involved in the phase II metabolism, even though the possible endogenous  
50  
51 445 contribution associated with supplement intake cannot be excluded. In fact, abscisic acid has  
52  
53 446 been found at relevant concentrations in bilberry fruits as a well-known plant hormone involved  
54  
55  
56 447 in the regulation of biosynthesis of polyphenols, such as anthocyanins (50).

1  
2  
3 448 Caprylic acid and azelaic acid showed a postprandial maximum intensity at very different  $T_{\max}$ ,  
4  
5 449 i.e. 30 min and 360 min after VC and VM interventions, respectively. In detail, caprylic acid,  
6  
7 450 after the very fast absorption with an early post-prandial maximum intensity, exhibited a rapid  
8  
9 451 decrease in the following time points, ascribable to its excretion by urine. On the other hand,  
10  
11 452 the late intensity increase of the azelaic acid kinetic profile suggests a closer dependence on  
12  
13 453 the phase I metabolism than the phase II one. The presence of fatty acids in serum has been  
14  
15 454 poorly investigated in association with phenolic-rich fruit interventions. However, caprylic  
16  
17 455 acid has been detected in human serum after pomegranate ingestion, probably due to its  
18  
19 456 occurrence in seeds (51). Moreover, fatty acids other than caprylic and azelaic have been  
20  
21 457 detected in human serum after the intake of mulberry and attributed to their occurrence in fruit  
22  
23 458 seeds and peels (52). Hence, to our knowledge, caprylic acid and azelaic acid have been  
24  
25 459 reported here as PEMs of VM and VC interventions for the first time.

### 30 460 3.5 Discrimination between VM and VC interventions

31 461 Comparison of serum metabolome profiles between the two berry interventions (performed  
32  
33 462 according to the protocol described in section 2.7.2) showed ten and fifteen discriminant  
34  
35 463 features in PI and NI, respectively, which correspond to thirteen PEMs. Among these, four  
36  
37 464 metabolites matched the previously annotated PEMs, i.e. abscisic acid glucuronide,  $\alpha$ -hydroxy-  
38  
39 465 hippuric acid, catechol sulphate, and methyl-catechol sulphate, which were found to be  
40  
41 466 statistically significant in VM intervention (**Table 2**).

42  
43 467 In order to provide a summarising description of the discrimination capacity between the two  
44  
45 468 interventions by the thirteen PEMs, it is interesting to perform a multivariate analysis by means  
46  
47 469 of PCA and LDA of the autoscaled intensity data of these PEMs, expressed as the highest  
48  
49 470 intensity value determined for each feature under NI in the kinetic profile of each intervention.  
50  
51 471 Moreover, the intensities of these PEMs in QC samples were included in both PCA and LDA,  
52  
53 472 in order to make more informative the multivariate analyses, also avoiding misclassifications.  
54  
55  
56  
57  
58  
59  
60



1  
2  
3 473 Five principal components (PCs) characterized by eigenvalues  $> 1$  and accounting for  
4  
5 474 percentages of explained variance (EV) of 29.9%, 12.9%, 12.2%, 11.2%, and 8.2% were  
6  
7 475 obtained (total EV=74.6%).  
8  
9

10 476 **Table S8** in section 6 of the *Supplementary material* illustrates the loading values of the five  
11  
12 477 significant PCs. Most features mainly contributed to only one PC with high absolute values of  
13  
14 478 the loadings, thus highlighting a good separation of the original variables among PCs. For  
15  
16 479 example, abscisic acid glucuronide,  $\alpha$ -hydroxy-hippuric acid, catechol sulphate, methyl-  
17  
18 480 catechol sulphate, and the unknown at  $m/z$  287.0224 contributed significantly only to PC1,  
19  
20 481 which alone accounts for about a third of the total variance of the original data. A detailed  
21  
22 482 description of the contribution of the various features to the five significant PCs is reported in  
23  
24 483 section 6 of the *Supplementary material*. **Figure 6** illustrates the score plots of PC1 vs. PC2,  
25  
26 484 PC3, PC4, and PC5, which highlights a quite good separation of QC, VC, and VM samples in  
27  
28 485 all the four PC spaces. Moreover, QCs were very close to the origin of the coordinates,  
29  
30 486 suggesting a high accuracy and precision of the entire analytical procedure. LDA has been also  
31  
32 487 carried out on the dataset of the discriminating PEMs in QC, VC, and VM samples, in order to  
33  
34 488 verify the correct assignment of each sample to its own group (i.e. QC, VM, and VC). The  
35  
36 489 summary of the results obtained for the fitted and cross-validated (leave-one-out method) LDA  
37  
38 490 classification are illustrated in **Table 3**. QC samples were correctly classified in all cases, both  
39  
40 491 in fitting and in cross-validation, thus evidencing the robustness of this classification and  
41  
42 492 confirming the results of PCA regarding the reliability of the analytical procedure here adopted.  
43  
44 493 The LDA classification exhibited a very high robustness also for the VM and VC groups, since  
45  
46 494 all samples were correctly classified in fitting and only two (i.e. one for each intervention) were  
47  
48 495 misclassified in cross-validation. It should however be remarked that the two misclassified  
49  
50 496 samples were attributed to the QC group, without any misleading attribution to the other  
51  
52 497 volunteer group.  
53  
54  
55  
56  
57  
58  
59  
60

#### 498 4 Conclusions

499 The comparative evaluation of different multivariate and univariate methods for PEMs  
500 selection evidenced that the number of significant features is strongly dependent on the  
501 statistical tool adopted, thus highlighting the importance of testing multiple statistical  
502 approaches for PEMs discovery.

503 The protocol of untargeted metabolomics analysis and chemometrics data treatment adopting  
504 the non-parametric Games-Howell test for PEMs selection within interventions, was suitable  
505 for selecting a restricted set of statistically significant features, excluding a number of false  
506 positive, which were conversely retained in the highly populated group of “significant  
507 features”, generated by other statistical methods, commonly adopted in literature. Accordingly,  
508 this chemometrics strategy could be fruitfully extended to the data treatment of intervention  
509 studies characterized by a similar study design (i.e. small sample size, randomized, single-  
510 blinded, parallel studies). In fact, it is more useful to identify a smaller set of PEMs, accepting  
511 a certain risk of excluding from this group some significant metabolites (false negatives), rather  
512 than selecting a larger group of PEMs, which however contains a high number of false positive.  
513 It should also be noted that with the presented approach it is possible to concentrate the  
514 annotation effort on a smaller but highly discriminant and certainly significant set of PEMs.

515 Twelve PEMs related to the intake of bilberry and/or blueberry supplements were annotated in  
516 serum samples, five of them (i.e. octahydro-methyl- $\beta$ -carboline-dicarboxylic acid and  
517 tetrahydro-methyl- $\beta$ -carboline-dicarboxylic acid for VC, citric acid for VM, and caprylic acid  
518 and azelaic acid for both VC and VM) reported here for the first time as serum metabolites of  
519 these interventions.

520 Based on AUC analysis, thirteen PEMs were found statistically significant for the  
521 discrimination of VM and VC interventions. In this regard, it is remarkable that this group of  
522 PEMs included four intra-intervention relevant metabolites (i.e. abscisic acid glucuronide,

523 catechol sulphate, methyl-catechol sulphate, and  $\alpha$ -hydroxy-hippuric acid), thus confirming the  
524 validity of the selection workflow based on the Games-Howell test. PCA data exploration and  
525 LDA samples classification performed on PEMs intensity at  $T_{\max}$  corroborating the  
526 discriminating role of the thirteen PEMs.

## 527 **Acknowledgements**

528 The authors wish to acknowledge the support of the Regione Toscana and the private  
529 companies Il Baggiole S.r.l., Danti Giampiero & C. S.n.c., Azienda Agricola Il Sottobosco,  
530 and Farmaceutica MEV S.r.l., through the PRAF Misura 1.2. e) grant.

## 531 **Conflict of interest**

532 The authors declare that they have no conflict of interest.

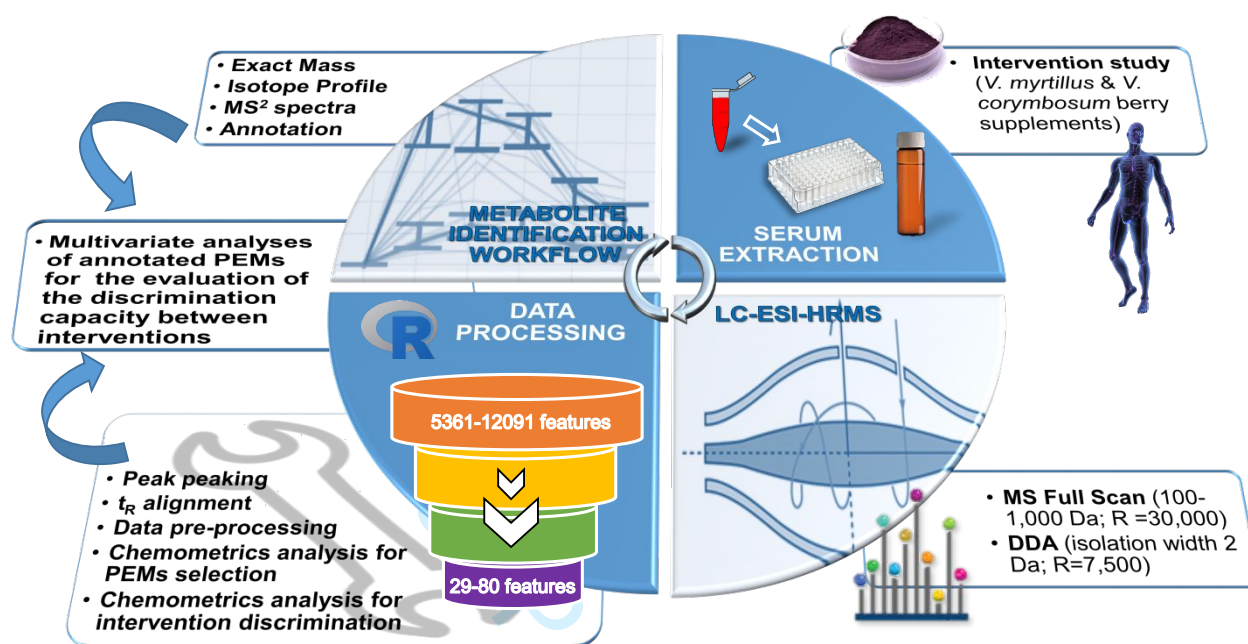
## 533 **References**

- 534 1. Ancillotti C, Ulaszewska M, Mattivi F, Del Bubba M. Untargeted metabolomics  
535 analytical strategy based on liquid chromatography/electrospray ionization linear ion trap  
536 quadrupole/orbitrap mass spectrometry for discovering new polyphenol metabolites in human  
537 biofluids after acute ingestion of vaccinium myrtillus berry supplement. *J Am Soc Mass*  
538 *Spectrom.* 2019;30(3):381-402.
- 539 2. Ulaszewska M, Garcia-Aloy M, Vázquez-Manjarrez N, Soria-Flórido M, Llorach R,  
540 Mattivi F, et al. Food intake biomarkers for berries and grapes. *Genes Nutr.* 2020;15(1):1-35.
- 541 3. Koutsos A, Riccadonna S, Ulaszewska MM, Franceschi P, Trošt K, Galvin A, et al.  
542 Two apples a day lower serum cholesterol and improve cardiometabolic biomarkers in mildly  
543 hypercholesterolemic adults: a randomized, controlled, crossover trial. *Am J Clin Nutr.*  
544 2020;111(2):307-18.
- 545 4. Lacalle-Bergeron L, Izquierdo-Sandoval D, Sancho JV, López FJ, Hernández F,  
546 Portolés T. Chromatography hyphenated to high resolution mass spectrometry in untargeted  
547 metabolomics for investigation of food (bio) markers. *TrAC, Trends Anal Chem.* 2020;116:161.
- 548 5. Ulaszewska MM, Weinert CH, Trimigno A, Portmann R, Andres Lacueva C,  
549 Badertscher R, et al. Nutrimentalomics: an integrative action for metabolomic analyses in  
550 human nutritional studies. *Mol Nutr Food Res.* 2019;63(1):1800384.
- 551 6. Stanstrup J, Broeckling CD, Helmus R, Hoffmann N, Mathé E, Naake T, et al. The  
552 metaRbolomics Toolbox in Bioconductor and beyond. *Metabolites.* 2019;9(10):200.
- 553 7. Hendriks MM, van Eeuwijk FA, Jellema RH, Westerhuis JA, Reijmers TH, Hoefsloot  
554 HC, et al. Data-processing strategies for metabolomics studies. *TrAC, Trends Anal Chem.*  
555 2011;30(10):1685-98.
- 556 8. Cuparencu CS, Andersen M-BS, Gürdeniz G, Schou SS, Mortensen MW, Raben A, et  
557 al. Identification of urinary biomarkers after consumption of sea buckthorn and strawberry, by  
558 untargeted LC-MS metabolomics: a meal study in adult men. *Metabolomics.* 2016;12(2):31.
- 559 9. Xu T, Feng G, Zhao B, Zhao J, Pi Z, Liu S, et al. A non-target urinary and serum  
560 metabolomics strategy reveals therapeutical mechanism of Radix Astragali on adjuvant-  
561 induced arthritis rats. *J Chromatogr B.* 2017;1048:94-101.

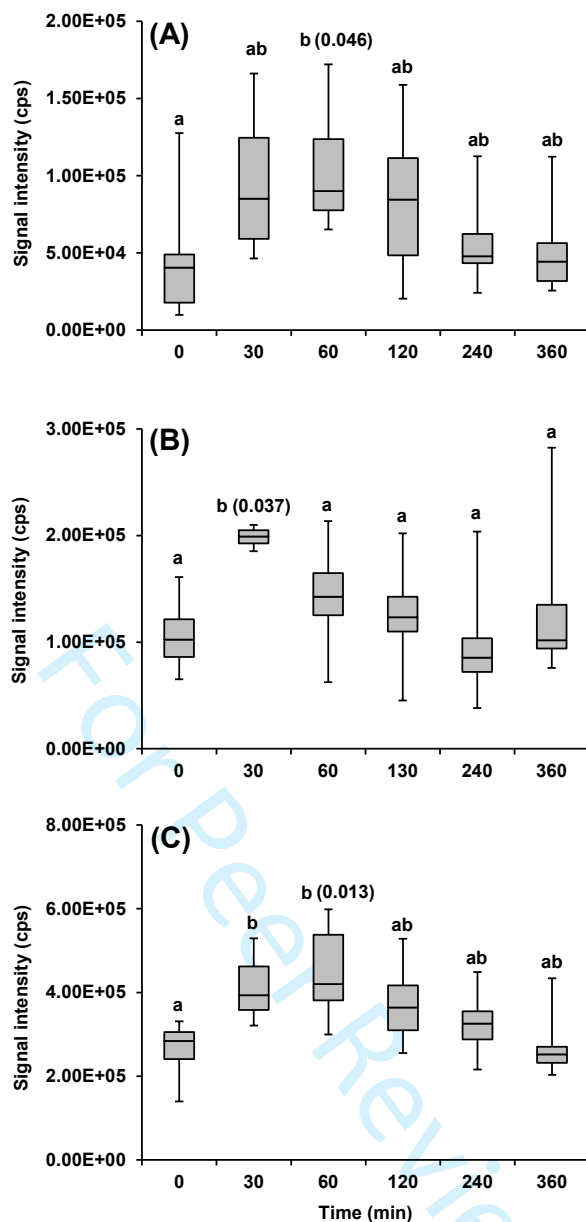
10. Dickson L, Tenon M, Svilar L, Faça-Berthon P, Martin J-C, Rogez H, et al. Genipap (*Genipa americana* L.) juice intake biomarkers after medium-term consumption. *Food Res Int.* 2020;137:109375.
11. Cocchi M, Biancolillo A, Marini F. Chemometric methods for classification and feature selection. *Comprehensive Analytical Chemistry.* 82: Elsevier; 2018. p. 265-99.
12. Vinaixa M, Samino S, Saez I, Duran J, Guinovart JJ, Yanes O. A guideline to univariate statistical analysis for LC/MS-based untargeted metabolomics-derived data. *Metabolites.* 2012;2(4):775-95.
13. Banaszewski K, Park E, Edirisinghe I, Cappozzo JC, Burton-Freeman BM. A pilot study to investigate bioavailability of strawberry anthocyanins and characterize postprandial plasma polyphenols absorption patterns by Q-TOF LC/MS in humans. *J Berry Res.* 2013;3(2):113-26.
14. Medina S, Ferreres F, Garcia-Viguera C, Horcajada M-N, Orduna J, Saviron M, et al. Non-targeted metabolomic approach reveals urinary metabolites linked to steroid biosynthesis pathway after ingestion of citrus juice. *Food Chem.* 2013;136(2):938-46.
15. Franceschi P, Giordan M, Wehrens R. Multiple comparisons in mass-spectrometry-based-omics technologies. *TrAC, Trends Anal Chem.* 2013;50:11-21.
16. Lamichhane S, Sen P, Dickens AM, Hyötyläinen T, Orešič M. An overview of metabolomics data analysis: current tools and future perspectives. *Comprehensive analytical chemistry.* 2018;82:387-413.
17. Godzien J, Ciborowski M, Angulo S, Barbas C. From numbers to a biological sense: How the strategy chosen for metabolomics data treatment may affect final results. A practical example based on urine fingerprints obtained by LC-MS. *Electrophoresis.* 2013;34(19):2812-26.
18. Manach C, Scalbert A, Morand C, Rémésy C, Jiménez L. Polyphenols: food sources and bioavailability. *Am J Clin Nutr.* 2004;79(5):727-47.
19. Ancillotti C, Ciofi L, Rossini D, Chiuminatto U, Stahl-Zeng J, Orlandini S, et al. Liquid chromatographic/electrospray ionization quadrupole/time of flight tandem mass spectrometric study of polyphenolic composition of different *Vaccinium* berry species and their comparative evaluation. *Anal Bioanal Chem.* 2017;409(5):1347-68.
20. Patel S. Blueberry as functional food and dietary supplement: The natural way to ensure holistic health. *Mediterr J Nutr Metab.* 2014;7(2):133-43.
21. Del Bubba M, Di Serio C, Renai L, Scordo CVA, Checchini L, Ungar A, et al. *Vaccinium myrtillus* L. extract and its native polyphenol-recombined mixture have anti-proliferative and pro-apoptotic effects on human prostate cancer cell lines. *Phytother Res.* 2020;35(2):1089-98.
22. Thomasset S, Berry DP, Cai H, West K, Marczylo TH, Marsden D, et al. Pilot study of oral anthocyanins for colorectal cancer chemoprevention. *Cancer Prev Res.* 2009;2(7):625-33.
23. Feliciano RP, Istaş G, Heiss C, Rodriguez-Mateos A. Plasma and urinary phenolic profiles after acute and repetitive intake of wild blueberry. *Molecules.* 2016;21(9):1120.
24. Rodriguez-Mateos A, Feliciano RP, Cifuentes-Gomez T, Spencer JP. Bioavailability of wild blueberry (poly) phenols at different levels of intake. *J Berry Res.* 2016;6(2):137-48.
25. Ancillotti C, Ciofi L, Pucci D, Sagona E, Giordani E, Biricolti S, et al. Polyphenolic profiles and antioxidant and antiradical activity of Italian berries from *Vaccinium myrtillus* L. and *Vaccinium uliginosum* L. subsp. *gaultherioides* (Bigelow) SB Young. *Food Chem.* 2016;204:176-84.
26. Chambers MC, Maclean B, Burke R, Amodei D, Ruderman DL, Neumann S, et al. A cross-platform toolkit for mass spectrometry and proteomics. *Nat Biotechnol.* 2012;30(10):918-20.

- 1  
2  
3 611 27. Ulaszewska MM, Trost K, Stanstrup J, Tuohy KM, Franceschi P, Chong MF-F, et al.  
4 612 Urinary metabolomic profiling to identify biomarkers of a flavonoid-rich and flavonoid-poor  
5 613 fruits and vegetables diet in adults: the FLAVURS trial. *Metabolomics*. 2016;12(2):32.  
6 614 28. Garcia-Aloy M, Ulaszewska M, Franceschi P, Estruel-Amades S, Weinert CH,  
7 615 Tor-Roca A, et al. Discovery of Intake Biomarkers of Lentils, Chickpeas and White Beans by  
8 616 Untargeted LC-MS Metabolomics in Serum and Urine. *Mol Nutr Food Res*. 2020:1901137.  
9 617 29. Sumner LW, Amberg A, Barrett D, Beale MH, Beger R, Daykin CA, et al. Proposed  
10 618 minimum reporting standards for chemical analysis. *Metabolomics*. 2007;3(3):211-21.  
11 619 30. Kim KM, Henderson GN, Ouyang X, Frye RF, Sautin YY, Feig DI, et al. A sensitive  
12 620 and specific liquid chromatography–tandem mass spectrometry method for the determination  
13 621 of intracellular and extracellular uric acid. *J Chromatogr B*. 2009;877(22):2032-8.  
14 622 31. Li C-R, Hou X-H, Xu Y-Y, Gao W, Li P, Yang H. Manual annotation combined with  
15 623 untargeted metabolomics for chemical characterization and discrimination of two major  
16 624 crataegus species based on liquid chromatography quadrupole time-of-flight mass  
17 625 spectrometry. *J Chromatogr A*. 2020;1612:460628.  
18 626 32. Lv Y, Liu X, Yan S, Liang X, Yang Y, Dai W, et al. Metabolomic study of myocardial  
19 627 ischemia and intervention effects of Compound Danshen Tablets in rats using ultra-  
20 628 performance liquid chromatography/quadrupole time-of-flight mass spectrometry. *J Pharm*  
21 629 *Biomed Anal*. 2010;52(1):129-35.  
22 630 33. Jakše B, Jakše B, Pajek M, Pajek J. Uric acid and plant-based nutrition. *Nutrients*.  
23 631 2019;11(8):1736.  
24 632 34. Liu H, Garrett TJ, Su Z, Khoo C, Gu L. UHPLC-Q-Orbitrap-HRMS-based global  
25 633 metabolomics reveal metabolome modifications in plasma of young women after cranberry  
26 634 juice consumption. *J Nutr Biochem*. 2017;45:67-76.  
27 635 35. Vázquez-Manjarrez N, Ulaszewska M, Garcia-Aloy M, Mattivi F, Praticò G, Dragsted  
28 636 LO, et al. Biomarkers of intake for tropical fruits. *Genes Nutr*. 2020;15(1):1-21.  
29 637 36. Lotito SB, Frei B. Consumption of flavonoid-rich foods and increased plasma  
30 638 antioxidant capacity in humans: cause, consequence, or epiphenomenon? *Free Radical Biol*  
31 639 *Med*. 2006;41(12):1727-46.  
32 640 37. Godycki-Cwirko M, Krol M, Krol B, Zwolinska A, Kolodziejczyk K, Kasielski M, et  
33 641 al. Uric acid but not apple polyphenols is responsible for the rise of plasma antioxidant activity  
34 642 after apple juice consumption in healthy subjects. *J Am Coll Nutr*. 2010;29(4):397-406.  
35 643 38. Martínez-López S, Sarriá B, Gómez-Juaristi M, Goya L, Mateos R, Bravo-Clemente L.  
36 644 Theobromine, caffeine, and theophylline metabolites in human plasma and urine after  
37 645 consumption of soluble cocoa products with different methylxanthine contents. *Food Res Int*.  
38 646 2014;63:446-55.  
39 647 39. Oracz J, Nebesny E, Zyzelewicz D, Budryn G, Luzak B. Bioavailability and  
40 648 metabolism of selected cocoa bioactive compounds: A comprehensive review. *Crit Rev Food*  
41 649 *Sci Nutr*. 2020;60(12):1947-85.  
42 650 40. Pimpao RC, Ventura MR, Ferreira RB, Williamson G, Santos CN. Phenolic sulfates as  
43 651 new and highly abundant metabolites in human plasma after ingestion of a mixed berry fruit  
44 652 puree. *Br J Nutr*. 2015;113(3):454-63.  
45 653 41. Bresciani L, Martini D, Mena P, Tassotti M, Calani L, Brigati G, et al. Absorption  
46 654 profile of (poly) phenolic compounds after consumption of three food supplements containing  
47 655 36 different fruits, vegetables, and berries. *Nutrients*. 2017;9(3):194.  
48 656 42. Jaganath IB, Jaganath IB, Mullen W, Edwards CA, Crozier A. The relative contribution  
49 657 of the small and large intestine to the absorption and metabolism of rutin in man. *Free Radical*  
50 658 *Res*. 2006;40(10):1035-46.  
51 659 43. Nurmi T, Mursu J, Heinonen M, Nurmi A, Hiltunen R, Voutilainen S. Metabolism of  
52 660 berry anthocyanins to phenolic acids in humans. *J Agric Food Chem*. 2009;57(6):2274-81.

- 1  
2  
3 661 44. Hövelmann Y, Lewin L, Steinert K, Hübner F, Humpf HU. Mass Spectrometry-Based  
4 662 Analysis of Urinary Biomarkers for Dietary Tomato Intake. *Mol Nutr Food Res.*  
5 663 2020;64(12):2000011.  
6 664 45. Herraiz T. Analysis of the bioactive alkaloids tetrahydro- $\beta$ -carboline and  $\beta$ -carboline  
7 665 in food. *J Chromatogr A.* 2000;881(1-2):483-99.  
8 666 46. Koleva II, van Beek TA, Soffers AE, Dusemund B, Rietjens IM. Alkaloids in the human  
9 667 food chain—natural occurrence and possible adverse effects. *Mol Nutr Food Res.*  
10 668 2012;56(1):30-52.  
11 669 47. Akagić A, Oras AV, Oručević Žuljević S, Spaho N, Drkenda P, Bijedić A, et al.  
12 670 Geographic Variability of Sugars and Organic Acids in Selected Wild Fruit Species. *Foods.*  
13 671 2020;9(4):462.  
14 672 48. Mikulic-Petkovsek M, Schmitzer V, Slatnar A, Stampar F, Veberic R. Composition of  
15 673 sugars, organic acids, and total phenolics in 25 wild or cultivated berry species. *J Food Sci.*  
16 674 2012;77(10):C1064-C70.  
17 675 49. Phillips MM, Case RJ, Rimmer CA, Sander LC, Sharpless KE, Wise SA, et al.  
18 676 Determination of organic acids in Vaccinium berry standard reference materials. *Anal Bioanal*  
19 677 *Chem.* 2010;398(1):425-34.  
20 678 50. Karppinen K, Hirvelä E, Nevala T, Sipari N, Suokas M, Jaakola L. Changes in the  
21 679 abscisic acid levels and related gene expression during fruit development and ripening in  
22 680 bilberry (*Vaccinium myrtillus* L.). *Phytochemistry.* 2013;95:127-34.  
23 681 51. Medjakovic S, Jungbauer A. Pomegranate: a fruit that ameliorates metabolic syndrome.  
24 682 *Food Funct.* 2013;4(1):19-39.  
25 683 52. Zhang H, Ma ZF, Luo X, Li X. Effects of mulberry fruit (*Morus alba* L.) consumption  
26 684 on health outcomes: A mini-review. *Antioxidants.* 2018;7(5):69.  
27  
28  
29  
30  
31 685  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

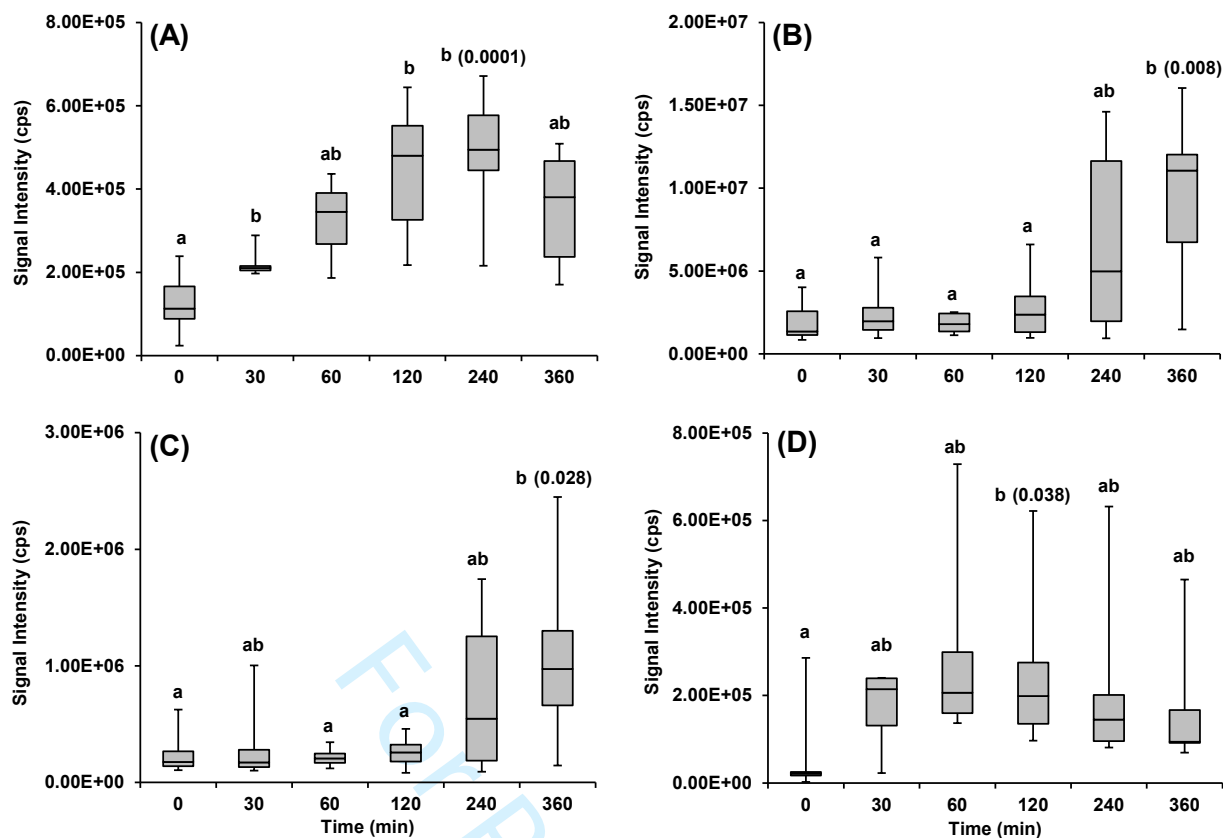


**Figure 1** – Graphical workflow of the experimental steps followed in this study.

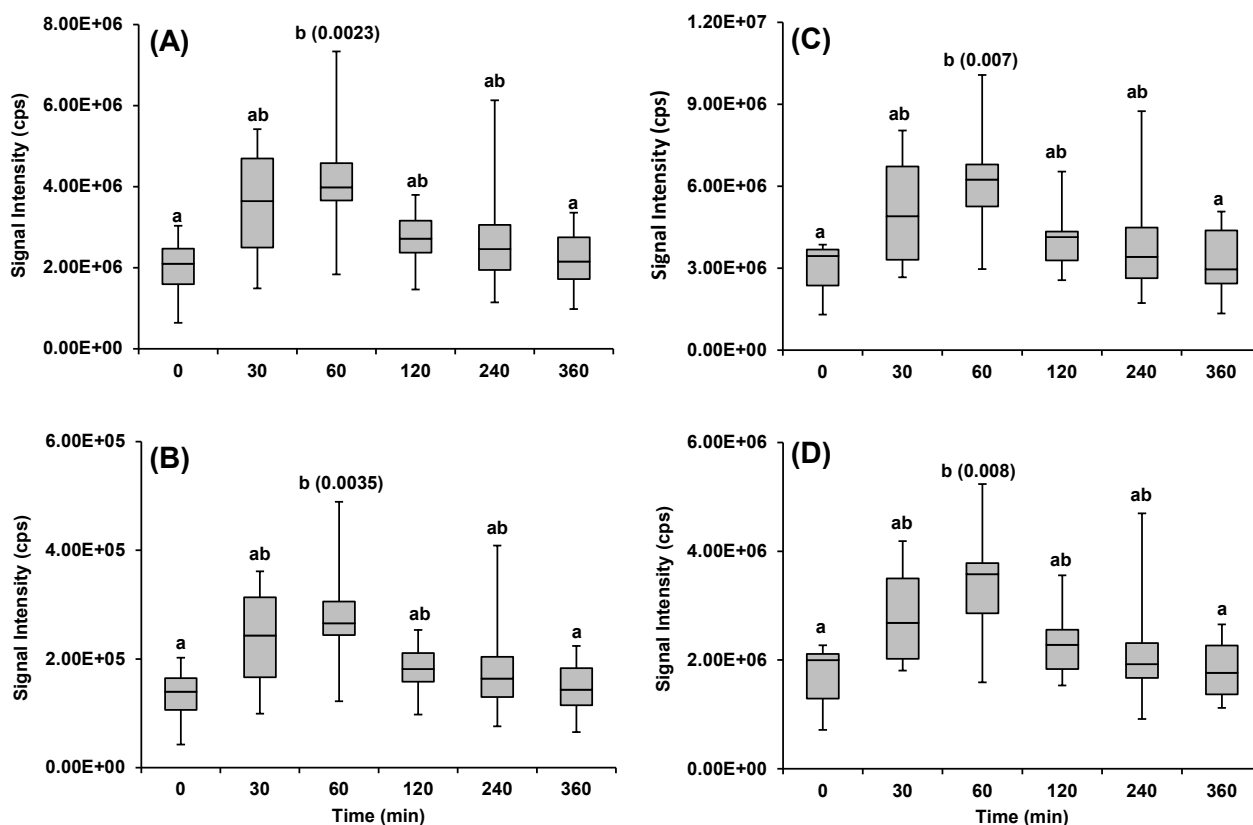


**Figure 2** – Kinetic profiles of the pseudo-molecular ions (negative ionization mode) at  $m/z$  (A) 335.0504 (uric acid in VM), (B) 167.0214 (uric acid in VC), and (C) 267.0744 (inosine in VM). Boxplots exhibiting different letters are significantly different according to the Games-Howell comparison test ( $P < 0.05$ ).  $T_{\max}$  of each pseudo-molecular ion is labelled in bracket with its corresponding  $P$ -value.

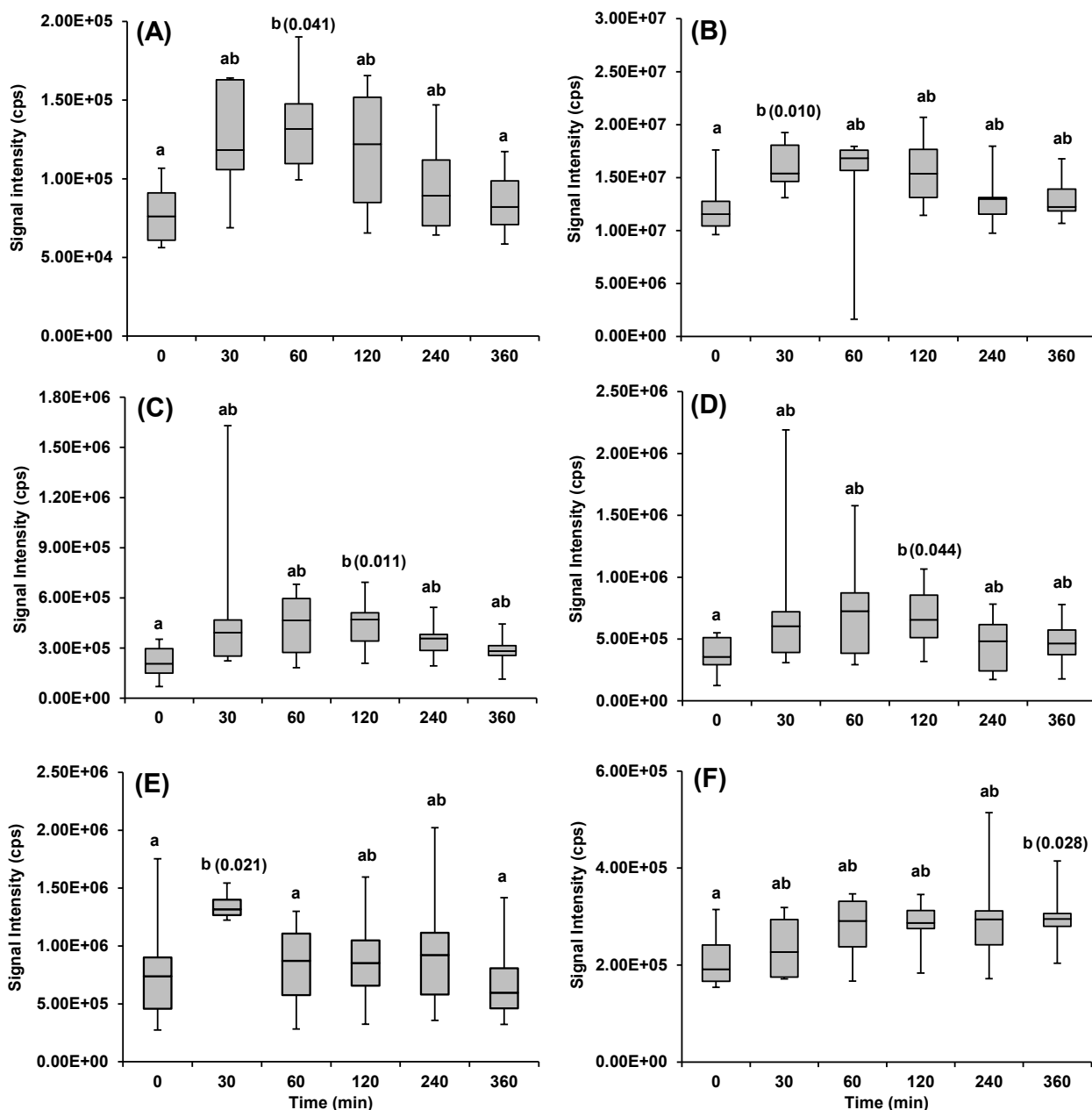




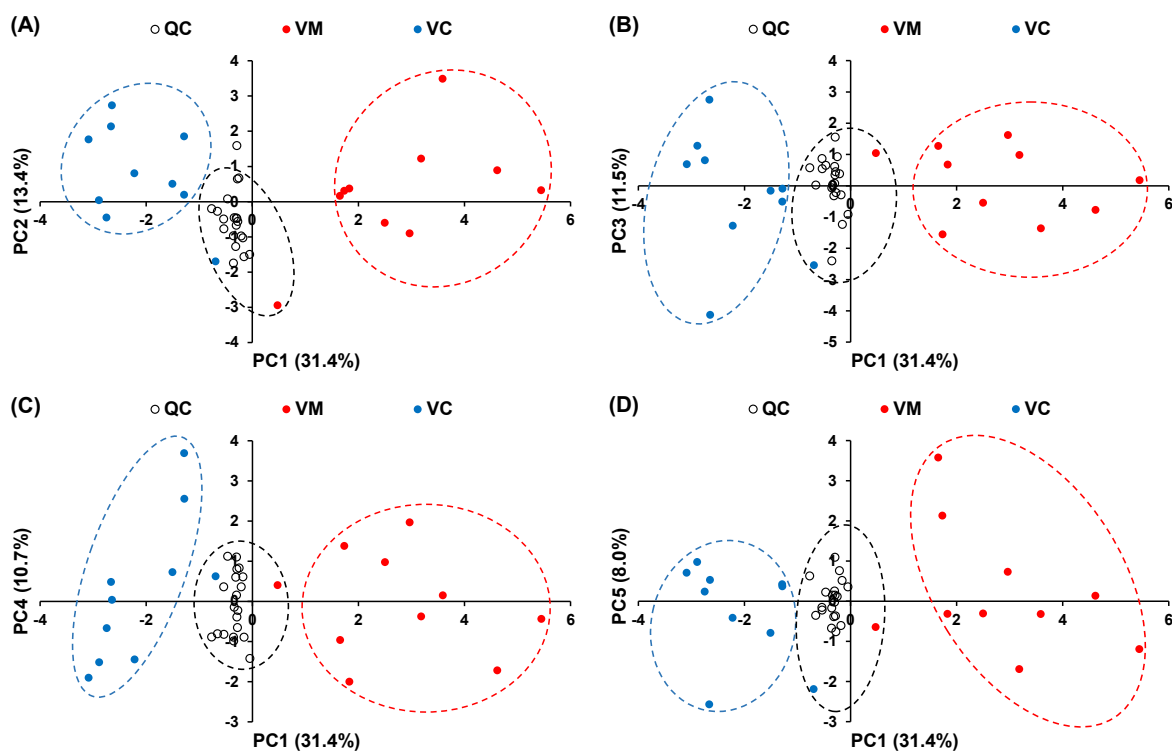
**Figure 3** – Kinetic profiles of the pseudo-molecular ions (negative ionization mode) at  $m/z$  (A) 194.0454 ( $\alpha$ -hydroxy-hippuric acid in VM), (B) 188.9856 (catechol sulphate in VM), (C) 203.0014 (methyl-catechol sulphate in VM), and (D) 245.0119 (hydroxyphenyl propionic acid sulphate in VM). Boxplots exhibiting different letters are significantly different according to the Games-Howell comparison test ( $P < 0.05$ ).  $T_{\max}$  of each pseudo-molecular ion is labelled in bracket with its corresponding  $P$ -value.



**Figure 4** – Kinetic profiles of the pseudo-molecular ions (positive and negative ionization, respectively) at  $m/z$  (A) 279.1329 and (B) 277.1185 (octahydro-methyl- $\beta$ -carboline-dicarboxylic acid in VC), (C) 275.1024 and (D) 273.0875 (tetrahydro-methyl- $\beta$ -carboline-dicarboxylic acid in VC). Boxplots exhibiting different letters are significantly different according to the Games-Howell comparison test ( $P < 0.05$ ).  $T_{\max}$  of each pseudo-molecular ion is labelled in bracket with its corresponding  $P$ -value.



**Figure 5** – Kinetic profiles of the pseudo-molecular ions at  $m/z$  (A) 191.0202 and (B) 193.0335 (citric acid in VM under negative and positive ionization, respectively), (C) 265.1431 and (D) 439.1599 (abscisic acid glucuronide in VM under positive and negative ionization, respectively), (E) 145.1220 (caprylic acid in VC under positive ionization), and (F) 187.0980 (azelaic acid in VM under negative ionization). Boxplots exhibiting different letters are significantly different according to the Games-Howell comparison test ( $P < 0.05$ ).  $T_{\max}$  of each pseudo-molecular ion is labelled in bracket with its corresponding  $P$ -value.



**Figure 6** – Score plots of (A) PC1 vs. PC2, (B) PC1 vs. PC3, (C) PC1 vs. PC4, and (D) PC1 vs. PC5. PCA plots were obtained using the autoscaled intensity data of the PEMs discriminating the two interventions in serum samples from each volunteer, expressed as their features measured under NI at the maximum of their kinetic profile in each intervention. Ellipses identify samples assigned to the QC, VM, and VC groups by applying linear discriminant analysis of maximum intensity of each annotated PEMs after cross validation with the leave-one-out method.

**Table 1** – Number of relevant postprandial features and potential exposure markers (PEMs) in each dataset (intervention and acquisition polarity) identified using the selection strategies based on univariate non-parametric adjusted post-hoc Games-Howell test, pairwise Wilcoxon signed-rank test and the univariate parametric *t*-Test, before and after the Benjamini-Hochberg adjustment procedure (BH-adj). VM = *Vaccinium myrtillus* supplement intervention; VC = *Vaccinium corymbosum* supplement intervention.

Dataset	Relevant postprandial features	<i>t</i> -test	BH-adj	Wilcoxon	Games-Howell
VM ESI (+)	4731	4682	1957	359	29
VC ESI (+)	4932	4395	2154	146	18
VM ESI (–)	4655	3286	1075	224	51
VC ESI (–)	4245	2582	1051	75	11

For Peer Review

**Table 2** — List of metabolites found in serum samples after acute ingestion of *V. myrtillus* (VM) and *V. corymbosum* (VC) supplements. Retention time ( $t_R$ , min), experimental mass in Full Scan spectrum (Da), main MS/MS fragments (base peak in bold), proposed formula, mass accuracy ( $\Delta$ , ppm), time point of max intensity ( $T_{max}$ , min) and  $P$ -value within interventions (in brackets), and HMDB/KEGG identification numbers. Roman numbers refer to the level of annotation: (I) spectra and  $t_R$  matched with reference standard; (II) spectra matched with reference libraries; (III) spectra matched with literature information. n.a.= not available.

Peak	$t_R$	Full Scan	MS/MS	Formula	$T_{max}$ ( $P$ -value)	Identification	HMDB/KEGG
1	1.2	$\ominus$ 335.0504 [2M-H] <sup>-</sup> $\ominus$ 167.0214	<b>167.0214</b> 124.0156, 96.0207	C <sub>5</sub> H <sub>4</sub> N <sub>4</sub> O <sub>3</sub>	VM: 60 min (0.046) VC: 30 min (0.037)	Uric acid <sup>(II)</sup>	HMDB0000289/C00366
2	1.2	$\oplus$ 193.0335 [M+H] <sup>+</sup> $\ominus$ 191.0202 [M-H] <sup>-</sup>	<b>132.1019</b> 173.0093, <b>111.0091</b>	C <sub>6</sub> H <sub>8</sub> O <sub>7</sub>	VM: 30 min (0.041) VM: 60 min (0.010)	Citric acid <sup>(II)</sup>	HMDB0000094/C00158
3	1.5	$\ominus$ 267.0744 [M-H] <sup>-</sup>	<b>135.0313</b>	C <sub>10</sub> H <sub>12</sub> N <sub>4</sub> O <sub>5</sub>	VM: 60 min (0.013)	Inosine <sup>(II)</sup>	HMDB0000195/C00294
4	3.1	$\ominus$ 194.0454 [M-H] <sup>-</sup>	<b>150.0561</b>	C <sub>9</sub> H <sub>9</sub> NO <sub>4</sub>	VM: 240 min (0.0001)	$\alpha$ -Hydroxy-hippuric acid <sup>(I)</sup>	HMDB0013678 HMDB0006116
5	3.9	$\ominus$ 188.9856 [M-H] <sup>-</sup>	<b>109.0297</b>	C <sub>6</sub> H <sub>6</sub> O <sub>5</sub> S	VM: 360 min (0.008)	Benzendiol (catechol) sulphate <sup>(II)</sup>	HMDB0059724
6	4.7	$\ominus$ 203.0014 [M-H] <sup>-</sup>	188.9860, <b>123.0454</b> , 108.0220	C <sub>7</sub> H <sub>8</sub> O <sub>5</sub> S	VM: 360 min (0.028)	Methyl-catechol sulphate <sup>(II)</sup>	HMDB0240663
7	5.23	$\oplus$ 279.1329 [M+H] <sup>+</sup> $\ominus$ 277.1185 [M-H] <sup>-</sup> $\ominus$ 233.1232	235.1376, 206.1110, <b>163.1165</b> <b>233.1232</b> <b>189.1336</b> , 187.1179, 146.0914, 120.0759, 96.0736	C <sub>14</sub> H <sub>18</sub> N <sub>2</sub> O <sub>4</sub>	VC: 60 min (0.023 <sup>(+)</sup> , 0.035 <sup>(-)</sup> )	Octahydro-methyl- $\beta$ -carboline-dicarboxylic acid <sup>(III)</sup>	n.a.
8	5.30	$\oplus$ 275.1024 [M+H] <sup>+</sup> $\oplus$ 202.0861 $\ominus$ 273.0875 [M-H] <sup>-</sup> $\oplus$ 229.0982	231.1128, 202.0861, <b>159.0916</b> <b>184.0755</b> , 156.0802 <b>229.0982</b> <b>185.1084</b> , 183.0928, 142.0663, 116.0508, 92.0508,	C <sub>14</sub> H <sub>14</sub> N <sub>2</sub> O <sub>4</sub>	VC: 60 min (0.007 <sup>(+)</sup> , 0.008 <sup>(-)</sup> )	Tetrahydro-methyl- $\beta$ -carboline-dicarboxylic acid <sup>(II)</sup>	HMDB0035115
9	5.9	$\oplus$ 265.1431 $\oplus$ 247.1326 $\ominus$ 439.1599 [M-H] <sup>-</sup>	<b>247.1329</b> , 199.1023 229.1431, 211.1327 <b>263.1288</b> , 219.1395, 175.0248, 153.0922	C <sub>21</sub> H <sub>28</sub> O <sub>10</sub>	VM: 120 min (0.044)	Abscisic acid glucuronide <sup>(II)</sup>	HMDB0036093 HMDB0035140
10	6.3	$\oplus$ 145.1220 [M+H] <sup>+</sup>	<b>127.1130</b> , 109.1008	C <sub>8</sub> H <sub>16</sub> O <sub>2</sub>	VC: 30 min (0.021)	Caprylic acid (hydroxy-octanone) <sup>(II)</sup>	HMDB00482/ HMDB0062511 HMDB62511/C06423
11	6.4	$\ominus$ 187.0980 [M-H] <sup>-</sup>	<b>125.0973</b>	C <sub>9</sub> H <sub>16</sub> O <sub>4</sub>	VM: 360 min (0.028)	Azelaic acid <sup>(I)</sup>	HMDB00784/C08261
12	6.5	$\ominus$ 245.0119 [M-H] <sup>-</sup>	<b>165.0556</b>	C <sub>9</sub> H <sub>10</sub> O <sub>6</sub> S	VM: 120 min (0.038)	Hydroxyphenyl propionic acid sulphate <sup>(II)</sup>	HMDB0094710

**Table 3** – Fitting (FIT) and cross-validation (CV) results of the Linear Discriminant Analysis based on the values of intensity data of the thirteen discriminating PEMs in serum samples from each volunteer, expressed as their features measured at the maximum of their kinetic profile in each intervention. Groups: quality control (QC), intervention with *Vaccinium corymbosum* supplement (VC), and intervention with *Vaccinium myrtillus* supplement (VM).

Put into Group		True Group		
		QC	VC	VM
QC	FIT	20	0	0
	CV	20	1	1
VC	FIT	0	10	0
	CV	0	9	0
VM	FIT	0	0	10
	CV	0	0	9
Total Number of Samples		20	10	10
Correct Attributions	FIT	20	10	10
	CV	20	9	9
Percentage of Correct Attribution	FIT	100%	100%	100%
	CV	100%	90%	90%

1  
2  
3  
4 **Manuscript: “Comparison of chemometrics workflows for potential exposure**  
5 **markers discovery and false positive reduction in untargeted metabolomics:**  
6 **application to the serum analysis by LC-HRMS after intake of *Vaccinium* fruits**  
7 **supplements” by L. Renai et al.**  
8  
9

## 10 SUPPLEMENTARY MATERIAL

### 11 1 Supplement preparation

12  
13  
14  
15  
16 The supplements were obtained by cryo-milling separately 5-kg aliquots of bilberries and blueberries  
17 (Laboratorio Terapeutico M.R., Florence, Italy), which were freeze-dried not later than one week  
18 after fruit harvesting, so as to maintain as much as possible the berry native composition. In the period  
19 between harvest and freeze-drying the fruits were kept at -20°C. Both supplements appeared as very  
20 homogeneous powders, characterized by a significant degree of hygroscopicity and were therefore  
21 kept at -20°C in polyethylene food storage containers, until their use.  
22  
23  
24  
25  
26

### 27 2 Supplement characterization

28  
29 In order to assess the amount of total soluble polyphenols (TSP), total monomeric anthocyanins  
30 (TMA) and the most important phenolic compounds commonly found in bilberries and blueberries,  
31 the following protocols were adopted.  
32  
33  
34

#### 35 2.1 Supplement extraction

36  
37 About 500 mg aliquots of supplement were homogenized in an ice bath under magnetic stirring with  
38 15 mL of a methanol/water solution 8/2 (v/v) containing 10 mM NaF to inactivate polyphenol oxidase;  
39 the mixture was centrifuged at 1800xg for 5 min and the supernatant recovered. This procedure was  
40 repeated twice and the extracts combined and analysed for TSP, TMA, and selected phenolic  
41 compounds.  
42  
43  
44  
45

#### 46 2.2 Analysis of TSP

47  
48 TSP were spectrophotometrically determined as following described. Extract aliquots (100-200 µL,  
49 depending on the TSP concentration in the extract) were mixed with 200 µL of Folin-Ciocalteu  
50 reagent. After 3 min, 400 µL of an aqueous solution saturated with sodium carbonate were added and  
51 the mixture obtained was made up to 10 mL with ultra-pure water. The solution was dark incubated  
52 for 1 h. Afterwards, the absorbance was measured at 740 nm and TSP concentration calculated on the  
53 basis of a catechin calibration curve; accordingly, the results were expressed as milligrams of  
54 catechin/g of supplement.  
55  
56  
57  
58  
59  
60



### 2.3 Analysis of TMA

TMA were determined with the pH differential method using cyanidin-3-glucoside as reference standard. Aliquots of 100-200  $\mu\text{L}$  of extract were diluted in buffer solutions at  $\text{pH}=1$  and  $\text{pH}=4.5$ , so as to obtain a final volume of 10 mL. The absorbance (Abs) of both solutions were measured at 520 and 700 nm and the quantity “ $\Delta\text{Abs}$ ” was calculated according to equation 1.

$$\Delta\text{Abs} = (\text{Abs}_{\text{pH}=1}^{520\text{ nm}} - \text{Abs}_{\text{pH}=1}^{700\text{ nm}}) - (\text{Abs}_{\text{pH}=4.5}^{520\text{ nm}} - \text{Abs}_{\text{pH}=4.5}^{700\text{ nm}}) \quad (1)$$

Similarly, “ $\Delta\text{Abs}$ ” values were also calculated for different concentrations of cyanidin-3-glucoside reference standard and plotted as a function of corresponding cyanidin-3-glucoside concentrations. The equation best fitting the experimental data was calculated by the least square regression method, thus obtaining a linear calibration curve, which was used for measuring TMA in the extracts. The results were expressed as milligrams of cyanidin-3-glucoside per g of supplement.

### 2.4 HPLC-MS/MS analysis of selected phenolic compounds

The following polyphenolic compounds were investigated: cyanidin-3-galactoside (CYA-3-GAL), cyanidin-3-glucoside (CYA-3-GLU), cyanidin-3-arabinoside (CYA-3-ARA), delphinidin-3-glucoside (DEL-3-GLU), delphinidin-3-galactoside (DEL-3-GAL), delphinidin-3-rabinoside (DEL-3-ARA), malvidin-3-glucoside (MAL-3-GLU), malvidin-3-galactoside (MAL-3-GAL), malvidin-3-arabinoside (MAL-3-ARA), peonidin-3-glucoside (PEO-3-GLU), peonidin-3-galactoside (PEO-3-GAL), peonidin-3-arabinoside (PEO-3-ARA), petunidin-3-glucoside (PET-3-GLU), petunidin-3-galactoside (PET-3-GAL), petunidin-3-arabinoside (PET-3-ARA), gallic acid (GAL), caffeic acid (CAF), p-coumaric acid (COU), ferulic acid (FER), salicylic acid (SAL), chlorogenic acid (CHL), neochlorogenic acid (NEO), cryptochlorogenic acid (CRY), (+)-catechin (CAT), epicatechin (EPI), procyanidin A2 (PA2), procyanidin B1 (PB1), procyanidin B2 (PB2) procyanidin C1 (PC1), kaempferol-3-glucoside (KAE-3-GLU), myricetin (MYR), quercetin (QUE), quercetin-3-galactoside (QUE-3-GAL), quercetin-3-glucoside (QUE-3-GLU), quercetin-3-rhamnoside (QUE-3-RHA), quercetin-3-rutinoside (QUE-3-RUT).

The analysis of phenolic compounds was carried out on the extracts after removal of organic solvent by vacuum evaporation, followed by acidification with formic acid up to  $\text{pH} = 2.0 \pm 0.1$  and filtration at 0.2  $\mu\text{m}$  on nylon membranes.

HPLC–MS/MS analysis was performed on an Acquity BEH C18 column (150 x 2.1 mm i.d.; particle size 1.7  $\mu\text{m}$ ) equipped with a guard column containing the same stationary phase (Waters, Milford, MA, USA), using a Shimadzu (Kyoto, Japan) chromatographic system consisting of a low pressure gradient quaternary pump Nexera X2 LC-30AD, a CTO/20AC thermostatted column compartment, a SIL-30AC autosampler, a DGU-20A 5R degassing unit and a CBM-20A module controller. The

1  
2  
3  
4 UHPLC system was coupled with a 5500 QTrap<sup>TM</sup> mass spectrometer (ABSciex, Ontario, Canada)  
5 by a Turbo VTM interface equipped with an electrospray probe. Chromatograms were elaborated by  
6 the 1.6.1 release of the software Analyst (AB Sciex).  
7

8  
9 Chromatographic analysis of anthocyanins was performed at 50°C with water/formic acid 95:5 v/v  
10 (eluent A) and methanol/formic acid 95:5 v/v (eluent B) according to the following gradient: 0.1-10  
11 min linear gradient 7-17% B, 10-20 min linear gradient 17-22% B, 20-22 min linear gradient 22-95%  
12 B, 22-27 min isocratic 95% B.  
13  
14

15  
16 Chromatographic analysis of phenolic acids, flavanols and flavonols was performed at 55°C, by  
17 eluting with water/formic acid 99.9:0.1 v/v (eluent A) and methanol/formic acid 99.9:0.1 v/v (eluent  
18 B) according to the following gradient: 0.1-1 min isocratic 4% B, 1-20 min linear gradient 4-70.5%  
19 B, 20-20.5 min linear gradient 70.5-95% B, 20.5-25 min isocratic 95% B.  
20  
21

22  
23 The flow rate was 0.3 mL/min and the injection volume was 5 µL for both chromatographic analyses.  
24 Polyphenol MS/MS analysis was carried out using the Multiple Reaction Monitoring mode by ESI  
25 both in negative (for phenolic acids, flavanols and flavonols) and positive mode (for anthocyanins).  
26 For each investigated compound, the most intense transition was used for quantification and the  
27 second most intense for confirming identification. Compound dependent parameters were optimized  
28 by direct infusion of properly diluted target analyte standard solutions (**Tables S1** and **S2**). The  
29 quantification of DEL-3-ARA, PET-3-GAL, PET-3-ARA and MAL-3-ARA was performed in  
30 respect to the corresponding glucoside because of the absence of commercial standards.  
31  
32

33  
34 Source dependent parameters were optimized in flow injection analysis at optimal LC flow and  
35 mobile phase composition. For negative ion determination optimal source dependent parameters were  
36 as follows: Curtain Gas 40, CAD Gas Medium, Temperature 550°C, Gas 1 60, Gas 2 40, Interface  
37 Heater ON and IonSpray Voltage -4500 V. For positive ion determination optimal source dependent  
38 parameters were as follows: Curtain Gas 40, CAD Gas Medium, Temperature 500°C, Gas 1 40, Gas  
39 2 30, Interface Heater ON and IonSpray Voltage 5500 V.  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

**Table S1** - Optimized MS parameters for the investigated anthocyanins. Letters A and B after each compound name refer to the quantifier and qualifier transitions, respectively. (DP) declustering potential; (EP) entrance potential; (CE) collision energy; (CXP) collision cell exit potential.

Compound	Precursor Ion (m/z)	Product Ion (m/z)	DP (V)	EP (V)	CE (V)	CXP (V)
SAL-A	137	93	-45	-7	-25	-10
SAL-B	137	65	-45	-7	-43	-20
GAL-A	169	125	-50	-9	-20	-11
GAL-B	169	137	-50	-9	-28	-15
COU-A	163	119	-50	-8	-20	-11
COU-B	163	93	-50	-8	-45	-10
CAF-A	179	135	-65	-10	-25	-12
CAF-B	179	134	-65	-10	-35	-10
FER-A	193	134	-50	-10	-20	-12
FER-B	193	149	-50	-10	-15	-14
CHL-A	353	191	-170	-13	-25	-20
CHL-B	353	85	-170	-13	-60	-10
NEO-A	353	191	-70	-8	-25	-25
NEO-B	353	179	-70	-8	-25	-20
CRY-A	353	173	-60	-8	-22	-25
CRY-B	353	179	-60	-8	-23	-20
CAT-A	289	245	-15	-9	-20	-20
CAT-B	289	109	-15	-9	-35	-10
EPI-A	289	245	-15	-9	-20	-20
EPI-B	289	109	-15	-9	-35	-10
PA2-A	575	285	-130	-8	-40	-25
PA2-B	575	289	-130	-8	-35	-20
PB1-A	577	289	-140	-7	-35	-30
PB1-B	577	407	-140	-7	-35	-20
PB2-A	577	289	-80	-7	-35	-30
PB2-B	577	407	-80	-7	-35	-25
PC1-A	865	407	-160	-7	-55	-30
PC1-B	865	289	-160	-7	-52	-25
KAE-3-GLU-A	447	284	-35	-8	-37	-12
KAE-3-GLU-B	447	255	-35	-8	-53	-15
MYR-A	317	151	-120	-8	-32	-15
MYR-B	317	179	-120	-8	-30	-15
QUE-A	301	151	-20	-7	-30	-14
QUE-B	301	179	-20	-7	-27	-17
QUE-3-GAL-A	463	300	-120	-8	-37	-15
QUE-3-GAL-B	463	271	-120	-8	-58	-15
QUE-3-GLU-A	463	300	-120	-8	-37	-15
QUE-3-GLU-B	463	271	-120	-8	-57	-15
QUE-3-RUT-A	609	300	-20	-5	-52	-15
QUE-3-RUT-B	609	271	-20	-5	-75	-15
QUE-3-RHA-A	447	300	-120	-8	-35	-15
QUE-3-RHA-B	447	271	-120	-8	-60	-15

**Table S2** - Optimized MS parameters for the investigated polyphenols. Letters A and B after each compound name refer to the quantifier and qualifier transitions respectively. (DP) declustering potential; (EP) entrance potential; (CE) collision energy; (CXP) collision cell exit potential.

Compound	Precursor Ion (m/z)	Product Ion (m/z)	DP (V)	EP (V)	CE (V)	CXP (V)
DEL-3-GLU-A	465	303	110	10	40	25
DEL-3-GLU-B	465	229	110	10	75	25
DEL-3-GAL-A	465	303	110	10	40	25
DEL-3-GAL-B	465	229	110	10	75	25
CYA-3-GAL-A	449	287	100	10	40	25
CYA-3-GAL-B	449	137	100	10	75	15
CYA-3-GLU-A	449	287	120	10	40	20
CYA-3-GLU-B	449	137	120	10	75	12
CYA-3-ARA-A	419	287	90	12	35	25
CYA-3-ARA-B	419	137	90	12	70	15
PET-3-GLU-A	479	317	100	10	35	25
PET-3-GLU-B	479	302	100	10	60	25
PEO-3-GAL-A	463	301	100	10	35	20
PEO-3-GAL-B	463	286	100	10	55	25
PEO-3-GLU-A	463	301	100	10	35	20
PEO-3-GLU-B	463	286	100	10	55	25
PEO-3-ARA-A	433	301	100	10	35	20
PEO-3-ARA-B	433	286	100	10	55	25
MAL-3-GAL-A	493	331	120	8	35	15
MAL-3-GAL-B	493	315	120	8	65	15
MAL-3-GLU-A	493	331	110	9	35	20
MAL-3-GLU-B	493	315	110	9	65	20

## 2.5 Phenolic compounds concentrations in the supplements

Concentrations of TSP and TMA (**Table S3**), anthocyanins (**Table S4**) and other phenolic compounds (**Table S5**) determined in *V. myrtillus* and *V. corymbosum* supplements are reported below.

**Table S3** – Total soluble polyphenol (TSP) and total monomeric anthocyanin (TMA) concentrations of *V. myrtillus* and *V. corymbosum* supplements administered to volunteers during the study.

	<i>V. myrtillus</i> supplement	<i>V. corymbosum</i> supplement
TSP (mg CAT/g supplement dry weight)	38.8 ± 2.1	9.3 ± 1.0
TMA (mg CYA-3-GLU/g supplement dry weight)	30.1 ± 2.8	6.4 ± 0.9

**Table S4** – Targeted anthocyanin concentrations expressed as mg/g dry weight of *V. myrtillus* and *V. corymbosum* supplements.

Anthocyanins	<i>V. myrtillus</i> supplement	<i>V. corymbosum</i> supplement
DEL-3-GAL	4.07 ± 0.05	0.57 ± 0.04
DEL-3-GLU	4.24 ± 0.08	0.14 ± 0.05
DEL-3-ARA	2.90 ± 0.04	0.37 ± 0.02
CYA-3-GAL	3.69 ± 0.08	0.12 ± 0.01
CYA-3-GLU	3.04 ± 0.09	0.04 ± 0.01

CYA-3-ARA	2.27 ± 0.08	0.06 ± 0.01
PET-3-GAL	0.92 ± 0.04	0.42 ± 0.02
PET-3-GLU	2.29 ± 0.03	0.17 ± 0.02
PET-3-ARA	0.74 ± 0.04	0.26 ± 0.03
PEO-3-GAL	0.27 ± 0.02	0.07 ± 0.01
PEO-3-GLU	1.10 ± 0.07	0.03 ± 0.01
PEO-3-ARA	0.16 ± 0.01	0.03 ± 0.01
MAL-3-GAL	0.94 ± 0.04	1.69 ± 0.06
MAL-3-GLU	2.36 ± 0.05	0.63 ± 0.02
MAL-3-ARA	0.56 ± 0.02	1.28 ± 0.05

**Table S5** – Targeted phenolic acid, flavanol and flavonol concentrations expressed as µg/g dry weight of *V. myrtillus* and *V. corymbosum* supplements.

Phenolic acids	<i>V. myrtillus</i> supplement	<i>V. corymbosum</i> supplement
GAL	41.6 ± 1.5	7.72 ± 0.8
p-COU	4.12 ± 0.05	0.58 ± 0.06
CAF	9.02 ± 0.2	3.96 ± 0.3
FER	0.93 ± 0.01	6.06 ± 0.07
SAL	0.61 ± 0.02	0.44 ± 0.03
CHL	2793 ± 35	796 ± 5
NEO	0.75 ± 0.08	3.38 ± 0.09
CRY	1.58 ± 0.05	1.50 ± 0.01
Flavanols	<i>V. myrtillus</i> supplement	<i>V. corymbosum</i> supplement
CAT	23.3 ± 0.9	17.0 ± 0.2
EPI	224 ± 16	235 ± 15
PA2	3.7 ± 0.4	0.18 ± 0.05
PB1	5.2 ± 0.4	10.1 ± 0.7
PB2	198 ± 21	5.3 ± 0.6
PC1	17.2 ± 0.8	3.2 ± 0.4
Flavonols	<i>V. myrtillus</i> supplement	<i>V. corymbosum</i> supplement
MYR	57.9 ± 1.3	3.88 ± 1.2
QUE	11.5 ± 0.1	5.93 ± 0.3
QUE-3-GAL	640 ± 38	268 ± 18
QUE-3-GLU	77.8 ± 1.4	69.2 ± 1.2
QUE-3-RHA	33.9 ± 1.0	43.4 ± 1.2
QUE-3-RUT	1.64 ± 0.21	58.8 ± 0.5
KAE-3-GLU	1.92 ± 0.08	4.64 ± 0.07

### 3 Study design

Volunteers participating to the study were divided in two groups depending on the supplement to be administered, which were indexed as VM and VC group, for *V. myrtillus* and *V. corymbosum*, respectively. Serum of each volunteer was collected at baseline and at different sampling times (30, 60, 120, 240 and 360 minutes) after the supplement taking and the name of each sample group was reported in **Table S6**.

**Table S6** – Groups and sampling time used for statistical comparison serum samples.

Group Name	Sampling Time	Berry type
VM 0min	Baseline	<i>V. myrtillus</i>
VM 30min	30 minutes after supplement taking	<i>V. myrtillus</i>
VM 60min	60 minutes after supplement taking	<i>V. myrtillus</i>
VM 120min	120 minutes after supplement taking	<i>V. myrtillus</i>
VM 240min	240 minutes after supplement taking	<i>V. myrtillus</i>
VM 360min	360 minutes after supplement taking	<i>V. myrtillus</i>
VC 0min	Baseline	<i>V. corymbosum</i>
VC 30min	30 minutes after supplement taking	<i>V. corymbosum</i>
VC 60min	60 minutes after supplement taking	<i>V. corymbosum</i>
VC 120min	120 minutes after supplement taking	<i>V. corymbosum</i>
VC 240min	240 minutes after supplement taking	<i>V. corymbosum</i>
VC 360min	360 minutes after supplement taking	<i>V. corymbosum</i>

#### 4 Serum extraction

As regards serum samples, the precipitation of proteins and the removal of phospholipids were performed using the Ostro 96-well plate (Waters, Milford, MA, USA). In detail, for each sample, 100  $\mu\text{L}$  of serum were placed into the well and 100  $\mu\text{L}$  of surrogate standards in methanol (12.5  $\mu\text{g}/\text{mL}$  TRI-d5 and HIP-d5) were added. Then, 300  $\mu\text{L}$  of the first cold solvent (acetonitrile/formic acid 99:1  $v/v$ ) were added to each well, the system was covered and the well plate/collection plate were firstly mixed for 5 minutes and then placed on the positive pressure processor for the filtration (60 psi for 10 minutes). This procedure was repeated after the addition of 400  $\mu\text{L}$  of the second cold solvent (acetonitrile/water 3:1  $v/v$  with the 1% of formic acid). The resulting solutions were dried under nitrogen flow and re-diluted in 250  $\mu\text{L}$  of the internal standard solution (0.42  $\mu\text{g}/\text{mL}$  CIN-d5). The resulting serum extracts were transferred in labelled brown vials for LC-MS analysis. For testing the serum extraction procedure, quality control (QC) samples were prepared, by mixing same volumes of all serum samples. Deuterated analytical standards were added to all QC samples and study samples to check for mass accuracy and drift of retention time. The same extraction protocol was performed on these quality control samples randomly placed in the well plate. The QC extracts together with the extraction blanks were injected before the analysis of study samples in both ionization modes. The QC samples were acquired each ten samples during LC-MS and LC-MS/MS analysis in both ionization modes.

## 5 LC-MS and LC-MS/MS analysis

The chromatographic analysis was performed on a Kinetex C18 column (150 mm x 2.1 mm I.D., particle size 2.6  $\mu\text{m}$ ) and a guard column containing the same stationary phase (Phenomenex, Torrance, CA, USA). Column temperature was set at 40°C. Water (eluent A) and acetonitrile (eluent B), both with 0.1 % formic acid, were used as mobile phases according to the following gradient: 0-1 min isocratic 5% B, 1-7 min linear gradient 5-45% B, 7-8.5 min linear gradient 45-80% B, 8.5-10.5 min isocratic 80% B. The flow rate was 350  $\mu\text{L}/\text{min}$  and the injection volume was 5  $\mu\text{L}$ .

The ESI conditions in positive (and negative) mode were: spray voltage 5.0 kV (-3.5 kV), heated capillary temperature 320°C, capillary voltage 35 V (-35 V) and tube lens 110 V (-110 V).

In the LTQ component of the instrument, nitrogen was used as both the sheath gas (35 U) and auxiliary gas (5 U), and helium was used as the damping gas. All measurements were done using the automatic gain control of the LTQ to adjust the number of ions entering the trap. Mass calibration was performed with every sequence run just prior to starting the batch by using flow injection of the manufacturer's calibration standards mixture allowing for mass accuracies lower than 5 ppm in external calibration mode.

## 6 Chemometrics

### 6.1 Statistical methods for inter-intervention selection of potential exposure markers (PEMs)

**Table S7** – Brief overview of the statistical approaches and methods used in this study to test inter-intervention PEMs selection.

Methods for PEMs selection	Description
PLS-DA VIP	Combination of the multivariate linear regression technique PLS-R with discriminant classification. Variable importance in projection (VIP) scores are defined for each <b>X</b> variable as the sum of the PLS-weight value, weighted by the percentage of $\hat{y}$ (or <b>Y</b> ) variance.
(paired) <i>t</i> -test	Parametric test used to compare the means of two paired groups.
Wilcoxon Signed-rank test	Non-parametric alternative of paired <i>t</i> -test, used to determine whether the mean ranks of two dependent samples were selected from populations having the same distribution.
Post-hoc Games-Howell	Non-parametric post-hoc test to compare combinations of groups of data ( $n > 2$ ), which does not assume equal variances and sample sizes, as well as normal distribution of data, since it works on ranked variables.

## 6.2 Principal component analysis of features discriminating the two interventions

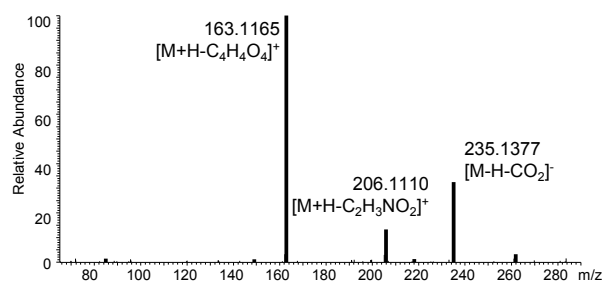
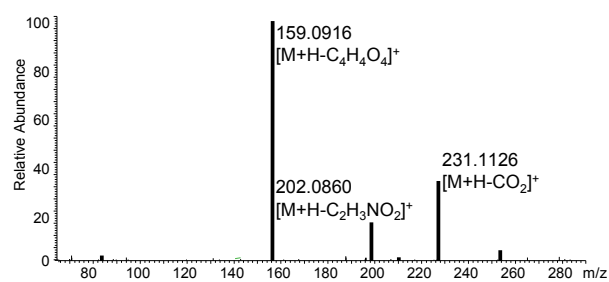
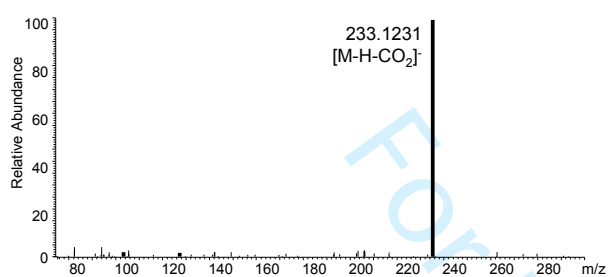
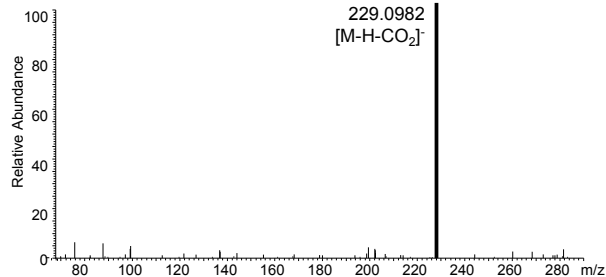
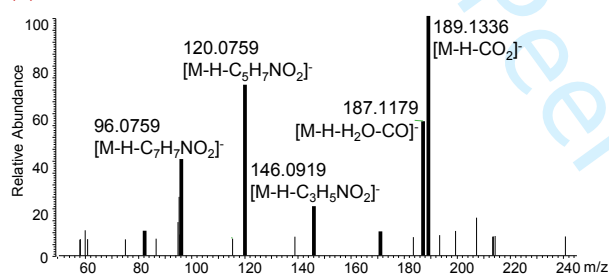
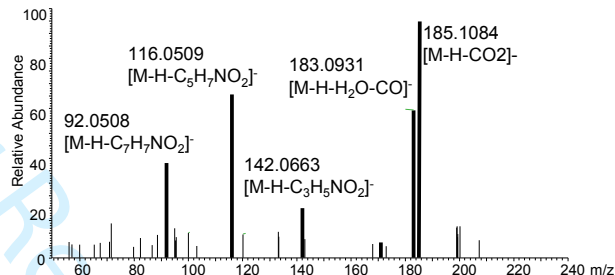
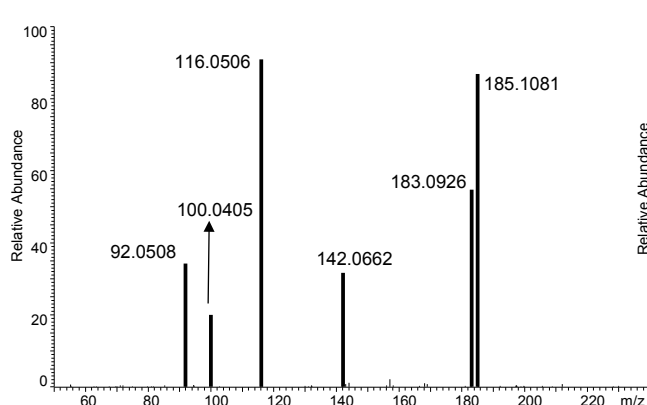
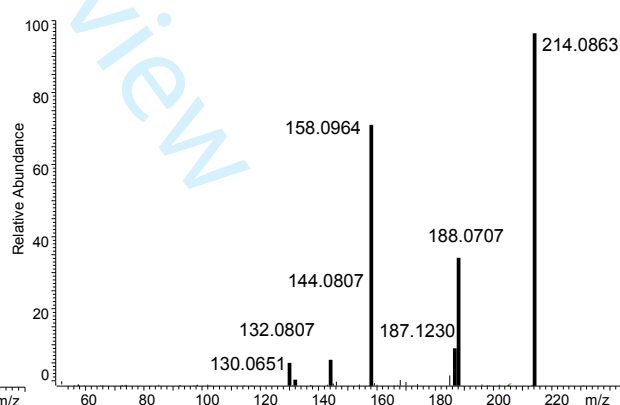
The loadings obtained with principal component analysis of the thirteen features found to be significant for the discrimination of interventions are shown in **Table S8**. High and positive loadings were found on PC1 for  $\alpha$ -hydroxy-hippuric acid, catechol sulphate, methyl-catechol sulphate, and the unknown at  $m/z$  287.0224, which were well differentiated from abscisic acid glucuronide, the latter exhibiting a high but negative projection on this PC. The unknown at  $m/z$  427.2535 and 643.3681 significantly contributed only to PC2 with negative and positive loadings, respectively. PC3 was mainly represented by the positive loading of the unknown at  $m/z$  158.9390 and 420.0832, and the negative loadings of the unknowns at  $m/z$  385.2375 and 465.2845. The unknown at  $m/z$  273.9592 (positive loading) and 381.2306 (negative loading) were those exhibiting a much higher contribution to PC4. Both these features significantly contributed with negative loadings also to PC5, which was however represented by  $m/z$  158.9390 (negative loading) and 420.0832 (positive loading).

**Table S8** – Loading values of the thirteen features found to be statistically significant for the discrimination of the VM and VC interventions. Loading more representative of each PC are in bold.

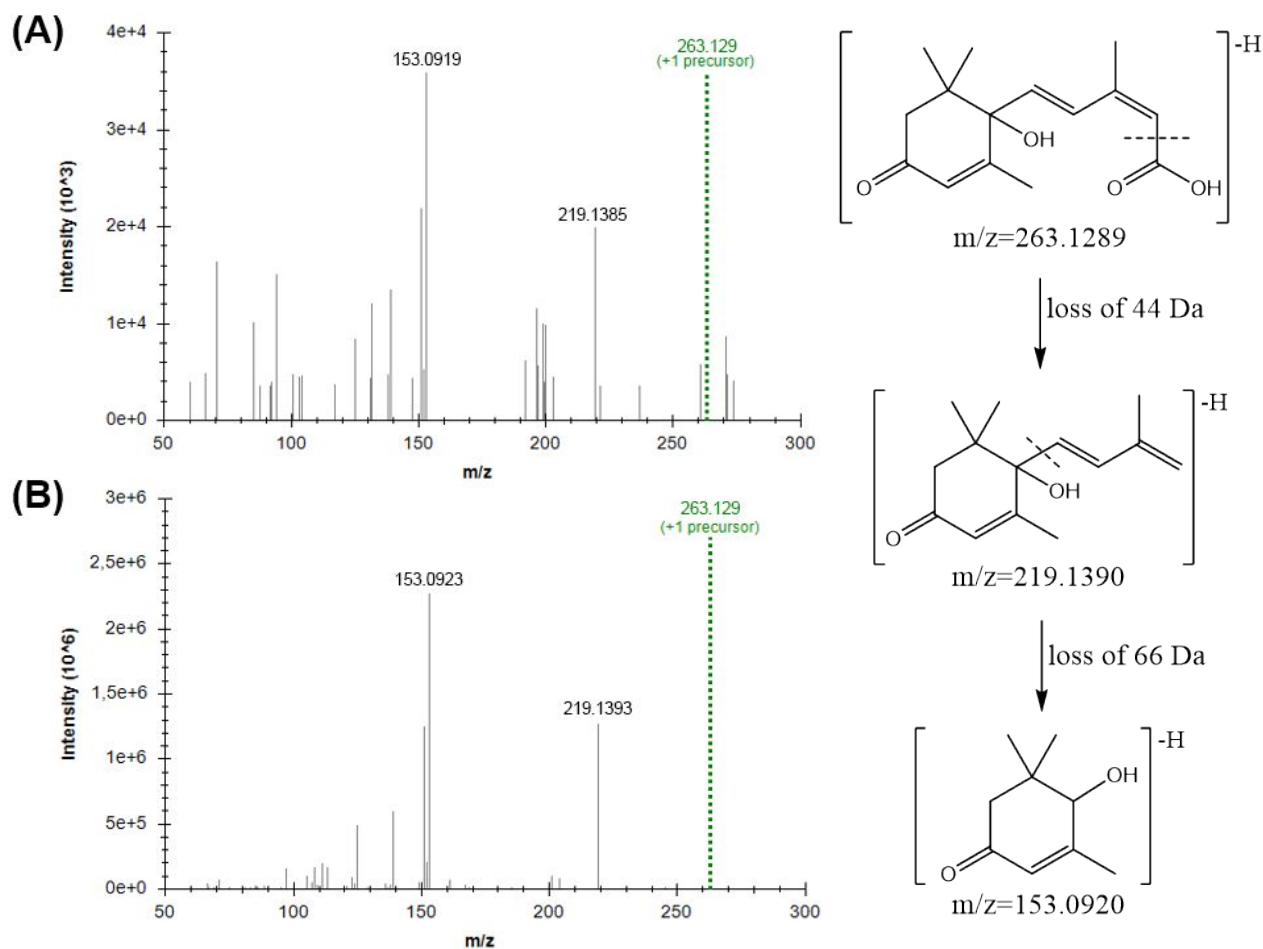
Features	PC1	PC2	PC3	PC4	PC5
Abscisic acid glucuronide	<b>-0.328</b>	0.277	-0.154	0.174	-0.165
$\alpha$ -Hydroxy hippuric acid	<b>0.423</b>	-0.084	0.054	0.079	0.317
Methyl-cathecol sulphate	<b>0.427</b>	0.097	-0.025	-0.127	-0.268
Cathecol sulphate	<b>0.448</b>	0.133	0.050	-0.108	-0.227
420.0832	-0.072	0.229	<b>0.342</b>	-0.216	<b>0.510</b>
158.9390	-0.138	0.224	<b>0.471</b>	0.123	<b>-0.411</b>
465.2845	-0.140	0.266	<b>-0.457</b>	0.289	0.157
427.2535	0.051	<b>-0.534</b>	-0.182	-0.355	-0.030
287.0224	<b>0.423</b>	0.134	-0.063	0.086	0.061
385.2375	0.052	0.275	<b>-0.587</b>	-0.220	0.080
643.3681	0.219	<b>0.537</b>	0.075	-0.081	0.005
273.9592	0.176	-0.183	-0.151	<b>0.504</b>	<b>-0.348</b>
381.2306	-0.153	0.126	-0.113	<b>-0.590</b>	<b>-0.404</b>



## 7 Annotation process

(A) FTMS POS MS<sup>2</sup> 279.1329(E) FTMS POS MS<sup>2</sup> 275.1024(B) FTMS NEG MS<sup>2</sup> 277.1185(F) FTMS NEG MS<sup>2</sup> 273.0875(C) FTMS NEG MS<sup>2</sup> 233.1232(G) FTMS NEG MS<sup>2</sup> 229.0982(D) NEG MS<sup>2</sup> 229.0978(H) POS MS<sup>2</sup> 331.1128

**Figure S1** – Fragmentation patterns of peak 7 (boxes A, B, and C), annotated as octahydro-methyl- $\beta$ -carboline-dicarboxylic acid, peak 8 (boxes E, F, and G), annotated as tetrahydro-methyl- $\beta$ -carboline-dicarboxylic acid, and analytical standard of the carbolines reference compound 1-methyl-1,2,3,4-tetrahydro- $\beta$ -carboline-3-carboxylic acid (boxes D and H).



**Figure S2** – Comparison of the MS<sup>2</sup> spectra acquired in negative ionization mode of (A) the product ion (m/z 263.1288) of peak 9 annotated as abscisic acid glucuronide in serum samples and (B) abscisic acid reference standard. The hypothesized scheme of abscisic acid fragmentation explaining the experimental spectra is also shown.

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41

