Metabolomics (2020) 16:85
https://doi.org/10.1007/s11306-020-01705-y

ORIGINAL ARTICLE

Can we trust biomarkers identified using different non-targeted metabolomics platforms? Multi-platform, inter-laboratory comparative metabolomics profiling of lettuce cultivars via UPLC-QTOF-MS

Carlos J. García[1] · Xiao Yang[2,3] · Danfeng Huang[3] · Francisco A. Tomás-Barberán[1]

Received: 13 April 2020 / Accepted: 20 July 2020 / Published online: 31 July 2020

Abstract

Introduction Data analysis during UPLC-MS non-targeted metabolomics introduces variation as different manufacturers use specific algorithms for data treatment and this makes untargeted metabolomics an application for the discovery of new biomarkers with low confidence in the reproducibility of the results under the use of different metabolomics platforms.
Objectives This study compared the ability of two platforms (Agilent UPLC-ESI-QTOF-MS and Waters UPLC-IMS-QTOF-MS) to identify biomarkers in butterhead and romaine lettuce cultivars.
Methods Two case studies by different metabolomics platforms: (1) Waters and Agilent datasets processed by the same data pre-processing software (Progenesis QI), and (2) Datasets processed by different data pre-processing software.
Results A higher number of candidate biomarkers shared between sample groups in case 2 (101) than in case 1 (26) was found. Thirteen metabolites were common to both cases. Romaine lettuce was characterised by phenolic compounds including flavonoids, hydroxycinnamate derivatives, and 9-undecenal, while Butterhead showed sesquiterpene lactones and xanthosine. This study demonstrates that high percentages of the most discriminatory entities can be obtained by using the manufacturers' embedded pre-processing software and following the recommended processing data guidelines using commercial software to normalise the data matrix.

Keywords Plant metabolomics · Lettuce biomarkers · Metabolic profiling · Multivariate analysis · Multi-platform analysis

Carlos J. García and Xiao Yang have contributed equally to this work.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s11306-020-01705-y) contains supplementary material, which is available to authorized users.

✉ Danfeng Huang
hdf@sjtu.edu.cn

✉ Francisco A. Tomás-Barberán
fatomas@cebas.csic.es

[1] Research Group on Quality, Safety, and Bioactivity of Plant Foods, Center for Applied Soil Science and Biology of the Segura, Spanish National Research Council (CEBAS-CSIC), 30100 Murcia, Spain

[2] Present Address: Institute of Urban Agriculture, Chinese Academy of Agricultural Sciences, Chengdu National Agricultural Science and Technology Center, Chengdu 610213, People's Republic of China

[3] School of Agriculture and Biology, Shanghai Jiao Tong University, Shanghai 200240, People's Republic of China

1 Introduction

Plant metabolomics can reveal altered metabolite expression levels and changes in metabolic pathways in response to disturbances in biotic and abiotic factors within biological systems (Fiehn 2002; Shulaev et al. 2008). Advanced analytical platforms such as ultra-performance liquid chromatography coupled with quadrupole/time-of-flight mass spectrometry (UPLC-QTOF/MS) and ion mobility (IMS) IMS-QTOF have become the preferred technologies for plant metabolomics studies due to their ability to rapidly separate metabolites and enable high-throughput detection of small molecules, including secondary metabolites (Gika et al. 2014; Rochat 2016). UPLC-QTOF/MS is the most widely used platform for plant metabolomics because of its rapid scan rate and high-resolution mass accuracy. The most advanced technique, UPLC-IMS-QTOF/MS, separates ions based on their drift time through the ion mobility chamber,

offering an additional dimension for separation beyond chromatographic and mass spectrometric separation (Lanucara et al. 2014). UPLC-IMS-QTOF/MS generates collision cross-section (CCS, 20–40 eV) values based on the ionic drifting time, which represents a specific characteristic for metabolite identification (Mairinger et al. 2018). However, the reproducibility of the non-targeted plant metabolomics data generated from two of the most sensitive tools, UPLC-ESI-QTOF-MS and UPLC-IMS-QTOF-MS, is not well documented.

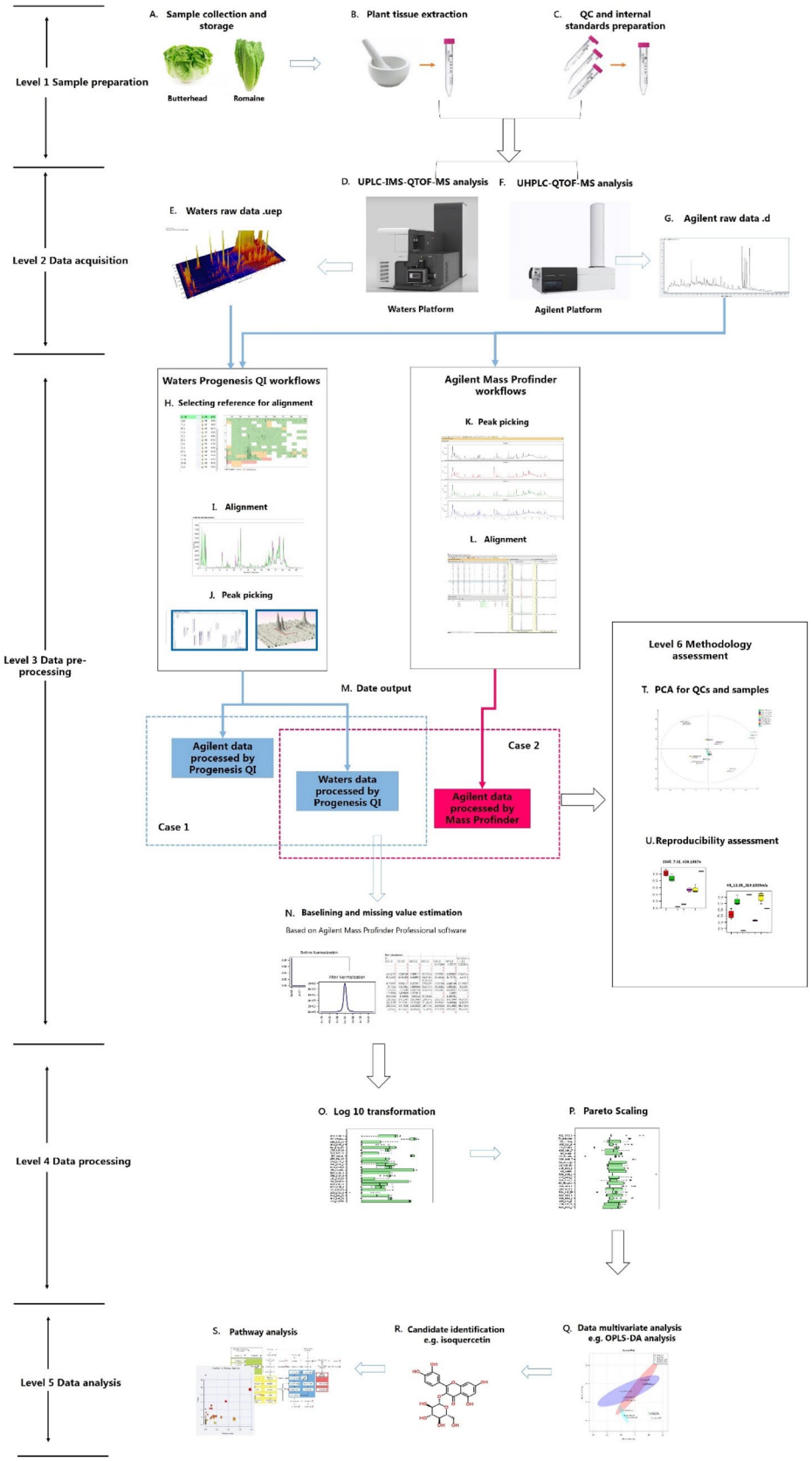
There are two main analytical workflow approaches in metabolomics: 'targeted' and 'non-targeted. The targeted approach focuses on specific analytes, while non-targeted analysis aims to simultaneously detect, identify and quantify the relative levels of maximal numbers of metabolites and generate metabolic fingerprints for biological tissues (Mairinger et al. 2018; Dudzik et al. 2018). However, due to the lack of spectral reference databases based on chemical standards, the precise annotation of all metabolites remains a considerable challenge (Dudzik et al. 2018). The non-targeted metabolomics analysis is most commonly found in differential expression studies: only signals that are significantly different when comparing two metabolic fingerprints (e.g., between control and treatment) are selected and subsequently identified. Thus, the methodological standardisation of differential analysis is urgently required to avoid false discoveries that lead to biologically irrelevant or erroneous hypotheses (Dudzik et al. 2018; Engskog et al. 2016). Methodological standardisation requires the standardisation of each step of the non-targeted metabolomics analysis, including sample preparation, data acquisition, data pre-processing, data processing, data analysis and methodology assessment (Fig. 1) (Benton et al. 2012; Djekic et al. 2016; García et al. 2018).

Among the steps above, the principal obstacles that prevent standardisation of non-targeted metabolomics analysis are related to the data acquisition and the data pre-processing steps (Sugimoto et al. 2012). Data acquisition differs between different manufacturers' platforms and laboratories due to variations in the features, performance, and precision of the instruments. Several open-source software and commercial software have been developed to facilitate the data pre-processing procedures and enable the comparison across platforms, such as XCMS online (https://xcmsonline.scripps.edu), MZmine 2 (https://mzmine.github.io/), Progenesis QI(Waters), MetAlign (https://www.wur.nl/en/show/MetAlign-1.htm), MassHunter software suite (Agilent), and MS-DIAL (https://prime.psc.riken.jp/Metabolomics_Software/MS-DIAL/). Each software package supports various raw data formats available from manufacturers and embeds different peak picking and alignment algorithms, which possess unique advantages in the data pre-processing step. For instance, a quantitative comparison of inter-laboratory untargeted lipidomics could be achieved using MS-DIAL that enables that the different raw data formats acquired using various LC–MS platforms can be converted and compared (Cajka et al. 2017). However, several studies have reported significant differences in the discovery of biomarkers in the same samples due to the data matrix generated by different data pre-processing software. Coble and Fraga (2014) observed significant differences in peak detection and quantification of four open-source preprocessing freeware for detecting, matching, and cataloging chromatographic peaks from chemical forensics samples. In this sense, Gürdeniz, Kristensen, Skov, and Dragsted (2012) noted different sets of biomarkers in plasma samples identified by MZmine, XCMS and MarkerLynx due to the considerable variation of algorithms for feature detection by different software. Myers et al. (2017) found several problems in peak detection algorithms of XCMS and MZmine 2, which should be partly responsible for the large numbers of false results in feature identification. More recently, Li et al. (2018) compared five widely used software for data processing in a non-targeted metabolomics study and observed a significant difference in quantification accuracy for true features among the data matrix processed by different software.

Remarkably, the prerequisite for an accurate comparison of non-targeted metabolomics is that the different raw data formats acquired from different LC–MS platforms can be converted, normalised, processed and analysed in the same manner (Gika et al. 2010; Cajka et al. 2017). However, to the best of our knowledge, the data format (.uep) acquired from the Waters UNIFI Scientific Information System (the specific data acquisition software for the Waters UPLC-IMS-QTOF/MS platform) cannot be converted to a format that is compatible with any other commercial or open-source software, except for Progenesis QI, which hampers inter-study comparisons of the reproducibility of non-targeted plant metabolomics data generated from UPLC-IMS-QTOF-MS and other high-resolution mass spectrometry platforms. So far, only a few studies have reported the reproducibility of the data pre-processing by different software in plant non-targeted metabolomics. The samples used in most of the previous studies for the comparison of pre-processing software in non-targeted metabolomics were on human tissues, lipids or other standard mixtures that were far below the complexity of the plant samples of interest in food science that are extremely rich in phytochemicals with approximately 100,000–200,000 metabolites (Dixon 2003; Oksman-Caldentey and Inzé 2004).

To address this issue, we performed a large-scale plant metabolomics reproducibility study using two

Fig. 1 Study design



manufacturers' platforms, the Agilent UPLC-QTOF/MS and Waters UPLC-IMS-QTOF-MS. This inter-laboratory analysis was performed at two independent labs located in CEBAS-CSIC, Spain, and Shanghai Jiao Tong University, China. Lettuce, a leafy vegetable widely consumed by humans, is a rich source of many natural health-promoting compounds, particularly secondary metabolites (Cervera-Mata et al. 2019; Riga et al. 2019). In this study, we aimed to evaluate the effect of the variation introduced by the metabolomics data acquisition platform and during data analysis, particularly the data pre-processing step, on the reproducibility of non-targeted plant metabolomics data. To reach this objective, we assessed two lettuce cultivars using two MS platforms in different laboratories, pre-processed the datasets using the manufacturer's software or Progenesis QI, subsequently processed and analysed the data using the same methods, and then compared the discriminatory metabolite biomarkers identified.

2 Materials and methods

2.1 Chemicals

Ultra-pure water was prepared using a Milli-Q system (Merck Millipore, Burlington, MA, USA). LC–MS-grade acetonitrile was purchased from Fisher Scientific (Thermo Fisher Scientific, Pittsburgh, PA, USA). The internal standards ellagic acid (≥ 95%, HPLC), morin, and ursolic acid were obtained from Sigma-Aldrich (Merck KGaA, Darmstadt, Germany); these internal standards were selected as the biochemical characteristics of ellagic acid, morin and ursolic acid are similar to the traits of phenolic acid and its derivatives, flavonoids, and terpenes in lettuce (Gika et al. 2010).

2.2 Sample preparation (Level 1)

The metabolomic profiles of two widely consumed lettuce varieties were compared: one cultivar, a butterhead type, is commonly cultivated in Shanghai, China (labelled C1,C2…); the other variety, a romaine type is widely grown in Murcia, Spain (labelled S1,S2…). The butterhead- and romaine-type lettuce varieties were cultivated to commercial maturity (60 days after sowing) in a greenhouse at Shanghai Academy of Agriculture Sciences, China (30.89° N, 121.39° E) or an experimental field belonging to Enza-Zaden Company, Spain (37.44°N, 0.59°E), respectively. Lettuce plants were picked randomly, and the outer six leaves were collected, freeze-dried, and finally shattered into fine powders. Twelve biological replicates were collected and labeled Butterhead (C1, C2… and C12) and Romaine (S1, S2… and S12). All samples of buttherhead type from China were sent to CEBAS-CSIC, Spain, for extraction. All samples were processed at CEBAS-CSIC,Spain under the same methodology. (Fig. 1, step A).

Lettuce extracts were prepared as described previously (García et al. 2017). In brief, 300 mg of freeze-dried powder was extracted with 10 mL methanol/water (80:20, *v/v*), sonicated at 30 °C for 30 min, centrifuged at 5000×*g* for 15 min at 4 °C. The samples (500 µL of each supernatant) were thoroughly dried in a SpeedVac coupled to a cold trap (Savant SPD121P, Thermo Fisher Scientific) to facilitate transport to China and to guarantee metabolite stability. Samples in China and Spain were re-dissolved in 500 µL of methanol:water (50:50, *v/v*) before UPLC-QTOF-MS analysis under the same conditions (Fig. 1, step B). Quality control (QC) samples were prepared by mixing 10 µL of each sample. Three internal standards (5 µL of 2 mg/mL morin, 5 µL of 2 mg/mL ellagic acid and 5 µL of 1 mg/mL ursolic acid) were added to each sample (Fig. 1, step C).

2.3 Data acquisition (Level 2)

To conduct an inter-laboratory, multi-platform, non-targeted metabolomics study, two platforms, an Agilent 6550 iFunnel Q-TOF LC/MS (Agilent Technologies, Waldbronn, Germany) located at CEBAS-CSIC, Spain, and a Waters Vion IMS QTOF (Waters Corp., Milford, MA, USA) located at Shanghai Jiao Tong University, China, were employed to profile all of the lettuce samples. The technical settings of both instruments and the chromatographic and MS conditions for sample separation are shown in Supplemental Table S1. Both UPLC-QTOF-MS instruments were adjusted to their respective optimum parameters before injection. A QC sample was injected before and after every six samples (Fig. 1, step D and F). Data acquisition for the Waters and Agilent instruments was performed using Waters UNIFI Scientific Information System software (Waters Corp., Fig. 1, step E) and Agilent MassHunter Workstation Data Acquisition software (Agilent Technologies, Fig. 1, step G), respectively.

2.4 Data pre-processing (Level 3)

The complete raw datasets for the lettuce samples and QCs were separately exported from the Waters (.uep format) and Agilent (.d format) systems into Progenesis QI (Waters Corp.) and pre-processed in the same manner to

enable the comparison between the two platforms. Generic pre-processing parameters regarding feature extraction were performed to normalise the number of entities picked as much as possible, aligned, and extracted by the different pre-processing software. This procedure aims to obtain two data matrices that, although unequal in extension due to the few restrictive filters used, include the most relevant entities. The Progenesis QI parameters included the selection of the alignment reference (Fig. 1, step H), the alignment (Fig. 1, step I) and the peak picking (Fig. 1, step J), using the default parameters, including data filter strength (1.0) and peak picking sensitivity (automatic, strength parameter 3). We exported the data matrices for the Waters data generated by Progenesis QI (Waters_QI) and Agilent data generated by Progenesis QI (Agilent_QI).

We also compared the pre-processing step of plant non-targeted metabolomics between the two platforms conducted by Waters' or Agilent's algorithms. Waters data file was prepared by Progenesis QI (Waters_QI). The Mass Hunter Profinder (Agilent Technologies) combined the extraction parameters (Fig. 1, step K), the compound binning and alignment, and algorithm filters (Fig. 1, step L). The settings selected in the different procedures of pre-processing software were set to mainly filter by retention time (0.5–25 min) and mass (100.0000–1000.0000 m/*z*). We then imported the Agilent.d files into Profinder to perform the molecular feature extraction algorithm (including peak picking and peak extraction, Fig. 1, step K) and alignment (Fig. 1, step L); the data matrix generated from the Agilent data generated by MassHunter Profinder was named Agilent_PRO.

The three pre-processed data matrices were extracted (Fig. 1, step M), and two comparison cases were defined. In case 1, we compared the Agilent_QI and Waters_QI files (both generated from Progenesis QI software), and in case 2, we compared the Waters_QI and Agilent_PRO data files. The fact the algorithms embedded in each software program, including the parameters used for peak detection, extraction, and alignment as well as ion detection and filtering, are entirely different is a vital issue that needs to be addressed. Therefore, we set the same generic, equally restrictive parameters to pick entities in each software program.

All three data files (Waters_QI, Agilent_QI, Agilent_PRO) were normalised using the same software, Agilent Mass Profiler Professional (version 14.9.1, Agilent Technologies), including two-step data baselining and missing value estimation (Fig. 1, step N). Each data matrix was normalised using its median; features were eliminated from the data matrices if they were only present in the butterhead or romaine cultivar and the corresponding intensity values were missing in > 75% of the replicates for that variety.

2.5 Data processing (Level 4)

Log transformation (Fig. 1, step O) was performed using Simca-P software (version 14.0, Umetrics, Umeå, Sweden) with the formula $X_{new} = log_{10}(C1 * X_{original} + C2)$. The parameters *C1* and *C2* were both 1, and X_{new} and $X_{original}$ represent the transformed value and original abundance of each feature, respectively. The scaling approach (Fig. 1, step P) was Pareto scaling, mean-centred and divided by the square root of the standard deviation of each variable (van den Berg et al. 2006).

2.6 Data analysis (Level 5)

Supervised OPLS-DA was performed to screen for potential biomarkers that differentiate the two lettuce cultivars, Fig. 1, step Q (Bylesjö et al. 2006). The parameters used to select discriminatory variables were *p, p* (*corr*), and variable importance in projection (VIP). The parameter '*p*' represents the importance of the variable in approximating the independent variables (*X*) in component one; '*p* (*corr*)' is the correlation coefficient for the relationship between *X* and linear combinations of *X* (one vector for each model dimension); 'VIP' represents the influence of the variable on this projection, and its value expresses the impact of the terms in the matrix *X* on all group information (*Y*). We defined several criteria to identify potential candidate biomarkers from our datasets: $p > 0.02$ or ≤ 0.02, $p\,(corr) > 0.5$ or ≤ 0.5, and VIP > 1.

Metabolite identification (Fig. 1, step R) was based on the comparison of chromatographic characteristics (e.g., retention time) and spectral information (e.g., *m/z*, MS^2 fragments, CCS values) using the Metlin and NIST databases as well as our in-house lettuce metabolite database (Garcia et al. 2016). Pathway analysis (Fig. 1, step S) was performed based on the reference pathways in KEGG and the literature.

2.7 Methodology assessment (Level 6)

The pre-processed data from step M (Fig. 1) were directly imported into Simca-P to establish principal components analysis (PCA) models. We assessed the PCA score plots and parameters used to develop the models (Fig. 1, step T). Moreover, we evaluated whether the coefficient of variance (CV) distributions for > 70% of the sum of responses was < 30%, and the intensities of the internal standards were stable across different samples (Fig. 1, step U).

3 Results and discussion

3.1 Study design

Numerous factors affect the number and nature of features detected in the non-targeted LC–MS-based metabolomic analysis, e.g., sampling, extraction method, chromatographic characteristics, mobile phases, instrument sensitivity, ionisation mode, MS acquisition mode, data processing procedure, and data normalisation (Cajka and Fiehn 2014). In this study, we investigated whether performing data acquisition and pre-processing using two different UPLC-QTOF-MS platforms (Agilent 6550 iFunnel Q-TOF and Waters Vion IMS QTOF) could affect the biological interpretations of a typical non-targeted plant metabolomic study. We collected raw data using two MS platforms at different laboratories and then performed peak picking and alignment procedures using two software programs, Progenesis QI, and MassHunter Profinder. All data were normalised using Mass Profiler Professional, and data processing was conducted using Simca-P. To accurately assess the effects of variation introduced during data acquisition and pre-processing, all other parameters were kept the same, including sampling, sample preparation, chromatographic separation, data processing, and data analysis.

Plant tissue samples were subjected to an optimised non-targeted metabolomic analysis methodology, including sample extraction, metabolic profiling, and metabolite identification (García et al. 2016; Yang et al. 2017, 2018a, b). The most popular lettuce variety in China (butterhead type) and most widely consumed lettuce cultivar in Spain (romaine type) were collected, extracted, and profiled to identify metabolomic differences between these cultivars. All samples were prepared in the same manner via methanol–water extraction (García et al. 2017). To focus on the effect of variations in data acquisition introduced by the UPLC-QTOF-MS platforms, we performed the analyses using the optimal settings for each instrument. We then analysed the samples using the same chromatographic separation and ionisation mode conditions and MS scan ranges; however, as the MS spectrometers have different structures, the source conditions and data acquisition mode (including collision energy) varied. To generate robust conclusions, we analysed a sequence of 32 injections per mass spectrometer, including 24 lettuce extract samples (12 biological replicates per cultivar), 5 quality controls, 2 blanks, and the phytochemical standard.

3.2 The choice of pre-processing platform

The most current strategy for comparing the data processing level is the use of open-source software (like XCMS online), which supports different data formats. It can perform peak detection and feature alignment in the same way for data acquired from different platforms. However, critical information from the original dataset (e.g., MS^2 fragment abundances, isotope distributions for each ion, drift times for ions and fragments, and time-aligned ions and their fragments from IMS) can inevitably disappear when the raw data is condensed and converted to a specific format in advance of data pre-processing, which may hinder further elucidation of chemical structures. Extraction of features using open source software may result in a large proportion of fragments and artifacts (> 50%) due to various noise sources introduced through the use of non-specific algorithms and parameters for data-pre-processing (Mahieu et al. 2016; Myers et al. 2017).

Furthermore, some raw data formats (such as.uep from Waters UNIFI) cannot be converted to formats that are compatible with other software, making comparisons between experiments conducted on different platforms hard to achieve. For instance, we used $HDMS^E$, an ion mobility based data-independent acquisition mode, to acquire the raw data (file format.uep) from the Waters platform. However, to the best of our knowledge, data format conversion ($HDMS^E$ based. uep to compatible formats for other platforms) cannot be achieved via online open-source software (like XCMS online, MZmine, and MS-DIAl), Waters embedded program (MassLynx and UNIFI Scientific Information System) or state-of-the-art packages (such as msConvert from ProteoWizard), except for Progenesis QI software from Waters platform. Therefore, to assess the effect of variations introduced during peak picking and by the alignment algorithm, the solutions recommended by the manufacturers (Progenesis QI software for Waters raw data and MassHunter Profinder software for Agilent raw data) were employed for the non-targeted metabolomics data pre-processing. We also pre-processed the raw data from the Agilent instrument with Progenesis QI, as Progenesis QI can support both.d (Agilent) and.uep (Waters) files. In contrast, Profinder can only support.d (Agilent) files. Then, the alignment data were baselined, and features were screened via missing value estimation. Data processing procedures (log transformation and Pareto scaling) were carried out using Simca-P, a widely used metabolomics data analysis software. Finally, supervised multivariable analysis was conducted for all features to discriminate potential discriminatory candidate biomarkers that can group the romaine-type and butterhead-type lettuce cultivars.

3.3 Methodology assessment

To ensure the acquisition of high-quality data, we assessed the replicate samples from the two lettuce cultivars in random order and distributed the QC samples throughout the

sample work-flow. Using Progenesis QI to pick and align features, we detected 7534 and 13,815 features in the Waters and Agilent datasets, respectively. Numerous feature/entity variables were different between the Waters and Agilent datasets, even when the same software and algorithm were used to pre-process the data. Thus, these variations were mainly due to the different operational modes of the two platforms and differences between the formats of the original datasets, which could affect ion detection. We also performed an alignment and peak detection for the Agilent dataset using Mass Profinder, a bespoke software package designed by Agilent. Compared to the Agilent_QI dataset (13,815 variables), more than 11,000 variables were missing from the Agilent_PRO dataset (2387 variables). These missing features were mainly false-positives extracted from noise due to the use of non-specific parameters for peak picking. Therefore, these observations suggest that the selection of the proper pre-processing method is vital for inter-laboratory comparisons and cross-instrument studies.

Therefore, we assessed the reproducibility of the datasets generated using the two UPLC-QTOF-MS platforms and different pre-processing steps based on several analytical characteristics of metabolic profiling. Unsupervised PCA analysis was performed on all lettuce and QC samples to discriminate group differences in highly complex datasets. Model parameters, including R^2 *(cum)* and Q^2 *(cum),* were calculated to evaluate the predictive ability of the models. As shown in the PCA score plots (Supplemental Figure S1), although the data matrix generated using Progenesis QI (case 1) contained a higher number of features than the data extracted using the manufacturers' software packages (case 2), greater similarity in data trend of the samples and QCs was observed in case 2. A PCA model containing three PCs was obtained from the 7534 features derived from the Waters_QI dataset (Supplemental Fig. S1A). The first two PCs described 57.5% of the total variation in *X* ($R^2 = 0.575$) and predicted 50.7% ($Q^2 = 0.507$) of the total variation of the model. A PCA model containing three PCs was established from the 13,815 features in the Agilent_QI dataset; the first two PCs explained 52.3% ($R^2 = 0.523$) of the total variation and predicted 47.2% ($Q^2 = 0.472$) of total variation (Supplemental Fig. S1B). Three PCs were generated from the 2387 features extracted from the Agilent_PRO dataset; this PCA model described 51.6% of the total variation ($R^2 = 0.516$ and predicted 40% ($Q^2 = 0.4$) of total variation (Supplemental Figure S1 C). When the three PCA scores were plotted, an apparent separation was observed between the two lettuce varieties and all QC samples clustered together in each PCA model (Supplemental Fig. S1A–C).

Moreover, 79.7%, 70.1%, and 75.8% of the features in the QC samples for the Waters_QI, Agilent_QI, and Agilent_PRO datasets had CV distributions < 30% (Supplemental Table S2). Furthermore, stable intensities were observed for the internal standards in the samples, with retention time shifts of < 0.02 min for the specific internal standards on each platform (Supplemental Fig. S3); the relative standard CV of the three internal standards between sample replicates across all datasets (Waters_QI, Agilent_QI and Agilent_PRO) were less than 16.1% (ellagic acid), 29.2% (ursolic acid) and 5.3% (morin; Supplemental Table S2). Collectively, these results indicate the robustness of the data generated from the two platforms using different pre-processing steps was acceptable.

In spite of using identical chromatographic conditions to analyse the same samples, we observed differences in chromatographic separation between the two UPLC-QTOF-MS platforms (Supplemental Figure S3). Supplemental Figure S1 shows examples of the separation of the three internal standards in the lettuce extracts. The post-alignment retention time (RT) drift values for ellagic acid, morin, and ursolic acid were within 1.02, 1.26, and 1.22 min, respectively. We also expected to observe varied peak intensities in the data acquired using two different instruments located in different laboratories. The ion intensities of the Agilent data were dozens of times higher than the Waters data. For instance, we compared the data extracted using Mass Profinder from the Agilent and Waters data generated using Progenesis QI software. We found the average absolute abundance of the feature with an *m/z* 455.3529 (RT 22.51, ursolic acid) was 799 in the Waters dataset and 99,542 in the Agilent run (*m/z* 455.3530, RT 23.73), 124 times higher than its abundance in the Waters data. Furthermore, the raw average intensity of ellagic acid (*m/z* 300.9996) was 4250 (Waters, RT 11.03) and 19,461,257 (Agilent, RT 12.06), representing a 4000-fold difference between the datasets from different platforms. The optimal analytical runs for each platform had many differences, including the chromatographic column conditions (changes in column temperature, fluctuations in column pressure, column size), the design of the electrospray source and QTOF (ionisation efficiency), and the ion mobility separation technology.

3.4 Metabolic biomarker screening

Before multivariate analysis, the raw data were filtered, and features that were only present in one variety and were missing from > 75% of the replicate samples for that cultivar were removed. A total of 1193, 1761 and 671 entities were removed from Waters_QI dataset, Agilent_QI dataset and Agilent_PRO respectively. Then the data was normalised and processed by log transformation and Pareto scaling. To identify potential variables that discriminate between the butterhead and romaine lettuce varieties, we performed multivariate pattern recognition based on Orthogonal Projections to Latent Structures Discriminant Analysis (OPLS-DA), as shown in Fig. 2. An OPLS-DA model based on a

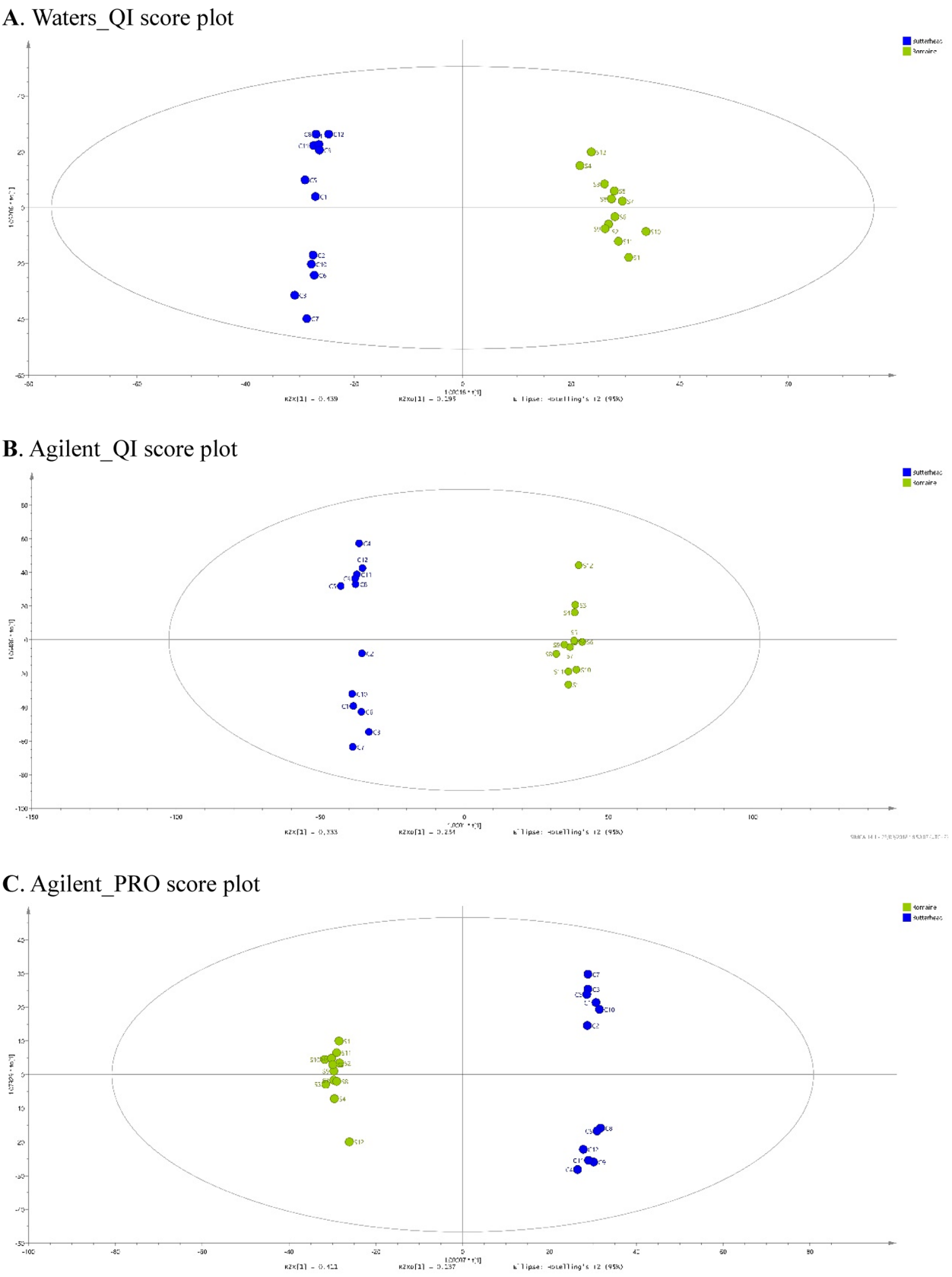
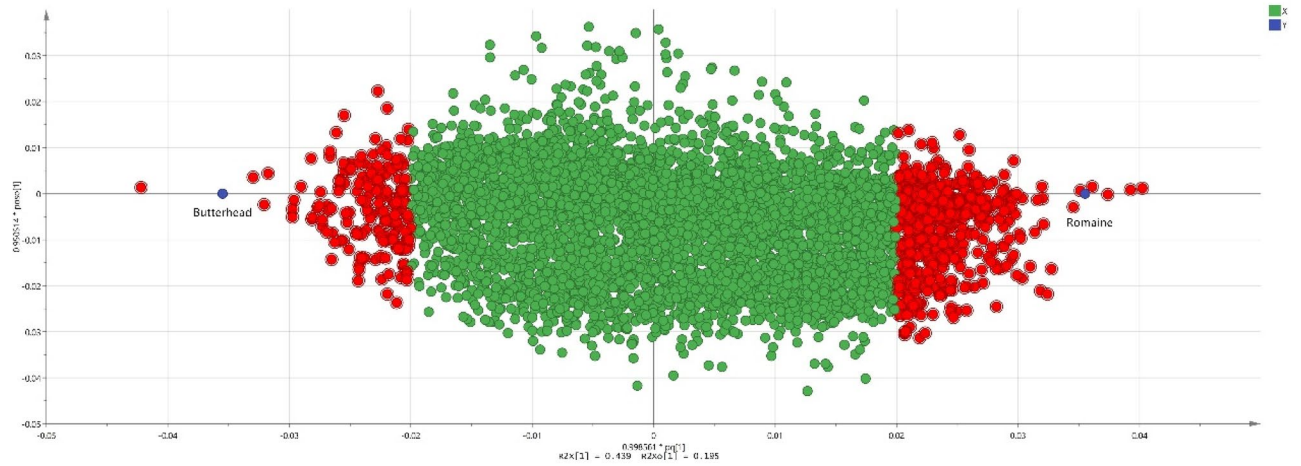
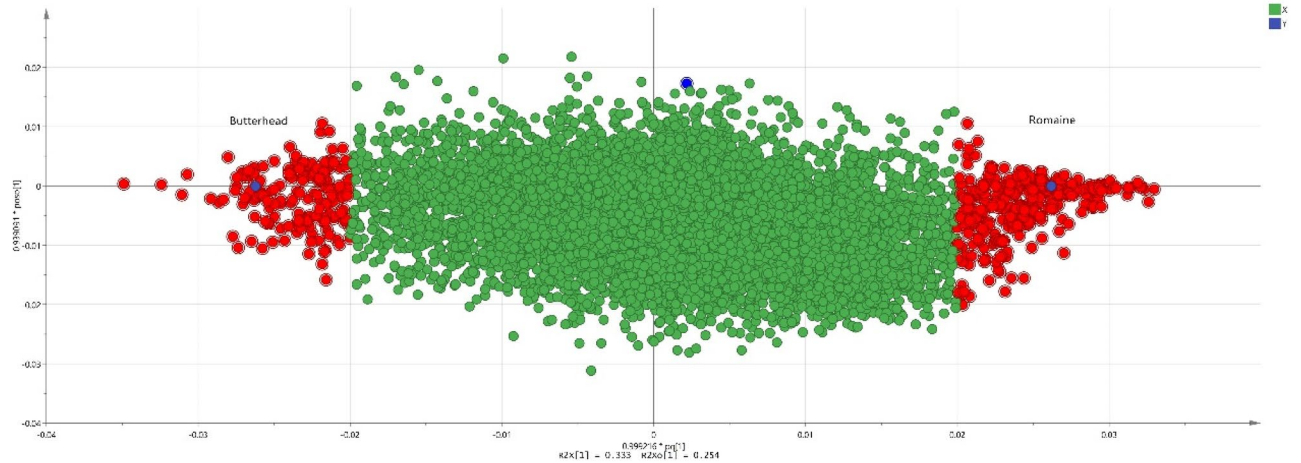


Fig. 2 OPLS-DA analysis of UPLC-QTOF-MS analysis for different datasets. **a** Waters_QI score plot; **b** Agilent_QI score plot; **c** Agilent_PRO score plot; **d** Waters_QI loading plot; **e** Agilent_QI loading plot; **f** Agilent_PRO loading plot

D. Waters_ QI loading plot



E. Agilent_QI loading plot



F. Agilent_PRO loading plot

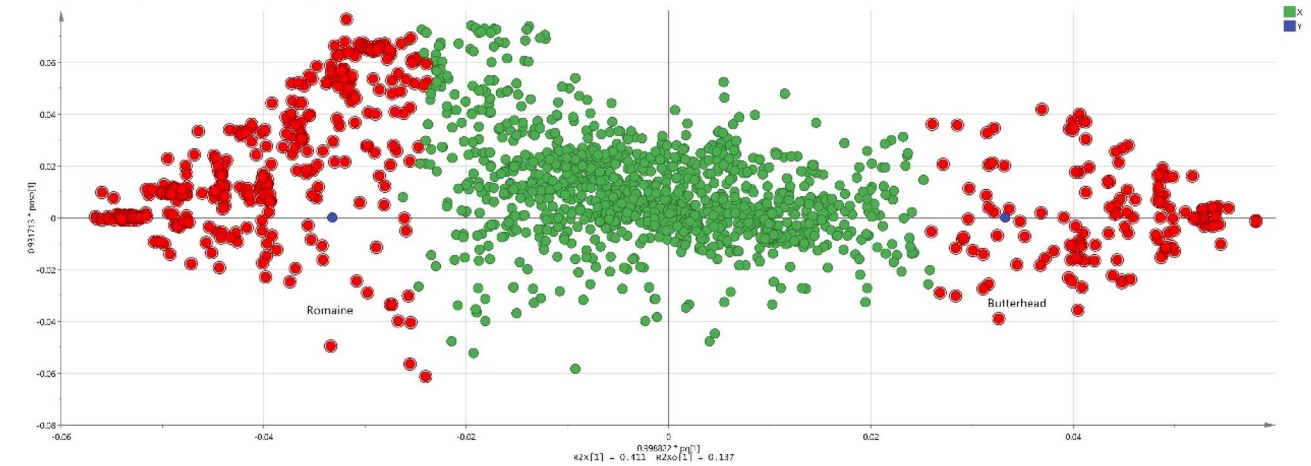


Fig. 2 (continued)

single predictive and orthogonal component (R2X = 0.635, R2Y = 0.993, and Q2 = 0.986) was generated based on the 6341 features acquired in the Waters_QI dataset (Fig. 2A, D). The OPLS-DA model based on the 12,054 features from the Agilent_QI dataset contained one predictive and one orthogonal component, explaining 58.7% and 99.6% of the variation in X and Y, respectively, and predicted 98.7% of the total variation in the model (Fig. 2B, E). As shown in Fig. 2C, F, the OPLS-DA model based on the 1716 variables in the Agilent_PRO dataset contained two components, describing 54.9% of the variation in X and 99.7% of the variation in Y (cross-validation; $Q^2 = 0.997$). The parameters generated by the OPLS-DA models had a high predictive capability. As shown in Fig. 2, the romaine lettuce samples clustered tightly together and apart from the butterhead lettuce samples, and there was a clear discrimination trend along the direction of the predictive component. A distinct separation between the butterhead and romaine varieties was observed in the OPLS-DA score plots generated from the Waters_QI, Agilent_PRO, and Agilent_QI datasets.

The use of screening parameters reduced the risk of identifying false positives and enabled the identification of the features that contributed most significantly to the separation of the lettuce cultivars. Candidate metabolites with a VIP > 1, $p > 0.02$ or ≤ 0.02, and p (*corr*) > 0.5 or ≤ 0.5 were selected for further analysis. In this respect (Supplemental Table S3), 682 of 7543 total entities (9%), 423 of 13,815 total entities (3%), and 473 of 2387 total entities (20%) were plotted as candidates on the V-plots for the Waters_QI (Fig. 3A), Agilent_QI (Fig. 3B), and Agilent_PRO (Fig. 3C) datasets, respectively. Although the Agilent_QI dataset provided approximately 80% more information than the Agilent_PRO dataset, the total numbers of potential candidates from the OPLS-DA models were similar, suggesting that the additional data in the Agilent_QI dataset did not contribute to discrimination of the lettuce cultivars. This result suggests that the parameters of the Progenesis QI software may be inadequate for pre-processing data from the Agilent platform and that most of the entities extracted from the Agilent dataset by Progenesis QI were unsatisfactory, as they were not matching and discriminant between the two study cases.

Twenty-six discriminatory metabolites were common to the Waters and Agilent datasets generated from the data matrices pre-processed using Progenesis QI, representing 4% (Waters_QI) and 6% (Agilent_QI) of total candidate biomarkers screened using the predefined parameters (case 1, Fig. 4A and Supplemental Table S4). However, the Waters dataset generated using the data matrix and pre-processed by Progenesis QI, and the Agilent dataset made using the data matrix and pre-processed by Profinder shared 101 candidates. Therefore, 3.9-fold more candidates were identified in case 2 than case 1; case 1 only detected 15% and 21% of the total candidate biomarkers detected in Waters_QI and Agilent_PRO (case 2, Fig. 4B and Supplemental Table S5). A higher percentage of the most discriminatory entities (i.e., entities with high VIP values and high absolute p and p (*corr*) values were detected in case 2 than case 1, although the biomarker with the highest VIP value was also detected in case 1 (Fig. 5). These results suggest that the manufacturers' specific software packages enable a better screening of candidate biomarkers. Thirteen discriminatory candidate metabolites were common to case 1 and case 2, accounting for 50% of the total discriminatory metabolites from case 1 and 13% from case 2 (Fig. 4C and Supplemental Table S6).

Identification of potential biomarkers from the differential candidates is the most time-consuming step of non-targeted plant metabolomics and remains a significant challenge (Creek et al. 2014). By comparison with m/z, MS^2 fragments and CCS value described in our previous studies and available online databases (García et al. 2016; Yang et al. 2018b), ten metabolites were tentatively annotated related to lettuce metabolism. These differential metabolites consisted of one aliphatic aldehyde, one nucleotide, six phenolic compounds (five flavonoids and one hydroxycinnamic acid derivative), and two sesquiterpene lactone derivatives (Table 1). Comparison with reference pathways available in the KEGG database and the literature suggested these metabolites are mainly involved in the biosynthesis of alkaloids derived from histidine and purine, flavone and flavonol biosynthesis, and the mevalonate pathway and are related to plant resistance to herbivores, microbial infection, and UV radiation (Brunetti et al. 2013; Chadwick et al. 2013). These differences between the two lettuce varieties may be the result of many factors, such as genetic diversity, environmental factors (e.g., temperature and light conditions), fertilization and irrigation management, wounding, and agricultural practices (Tavarini et al. 2015; García et al. 2018).

4 Conclusions

We analysed the metabolic differences between butterhead and romaine lettuce types using UPLC-QTOF-MS platforms from two manufacturers in different laboratories in parallel to assess how two crucial factors, namely data acquisition and data pre-processing, affect the reproducibility of non-targeted plant metabolomic studies. The PCA scores of the lettuce and QC samples, and the relative standard desviation values of the internal standards spiked into the samples demonstrated the robustness and reproducibility of the separate runs performed on the two

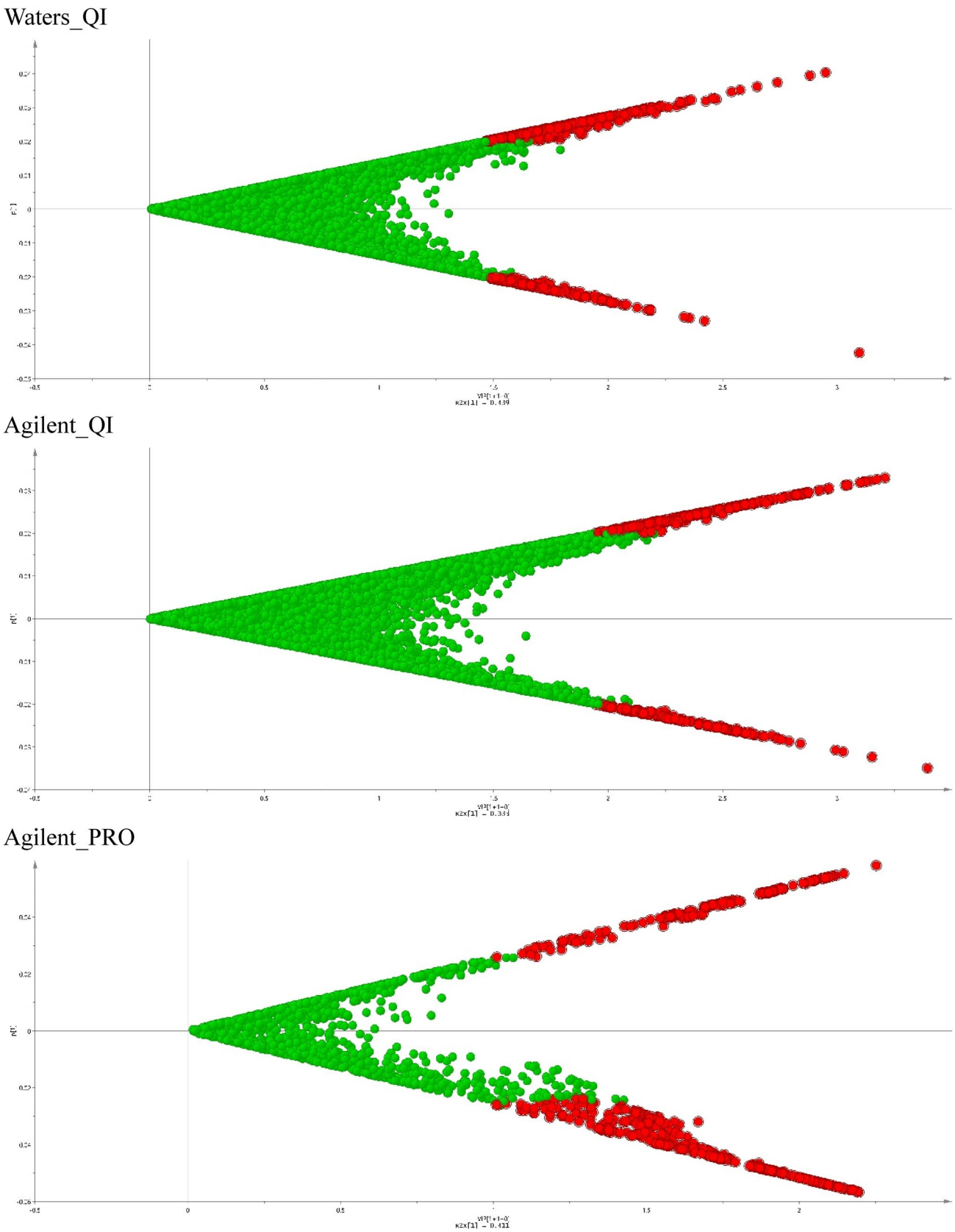


Fig. 3 Candidate biomarkers identified using UPLC-QTOF-MS analysis in the different datasets

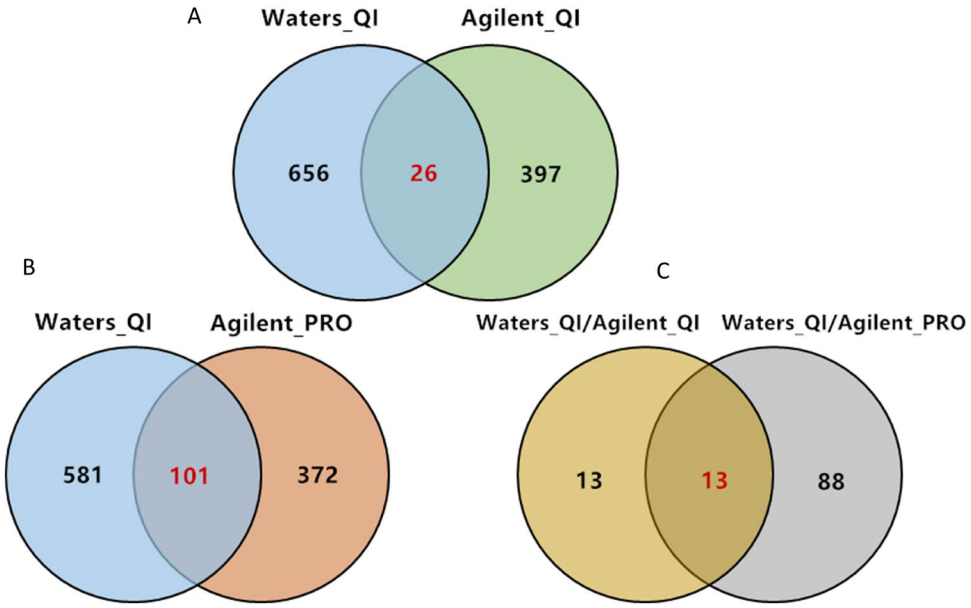


Fig. 4 Venn diagram of biomarkers. **a** Case 1: Waters_QI *vs*. Agilent_QI; **b** Case 2: Waters_QI *vs*. Agilent_PRO; **c** Case 1 *vs*. Case 2

platforms were acceptable. The OPLS-DA models discriminated 423 of 13,815 entities, 682 of 7534 entities, and 473 of 2387 entities from the Agilent_QI, Waters_QI, and Agilent_PRO data matrices, respectively. These results suggest that although the Progenesis QI pre-processing algorithm can extract data from the Agilent platform and results in a high number of entities, the manufacturers' pre-processing software identified the highest final numbers of discriminatory entities and possible biomarker candidates. The Waters and Agilent datasets generated by the data matrix pre-processed using Progenesis QI contained 26 discriminatory metabolites shared, while the Waters dataset generated by the data matrix pre-processed using Progenesis QI and the Agilent dataset generated by the data matrix pre-processed using MassHunter Profinder shared 101 candidates, indicating that the manufacturers' specific pre-processing software packages enable better identification of discriminatory metabolites than Progenesis QI. This inter-laboratory comparison further shows that untargeted UPLC-QTOF MS-based plant metabolomics analysis can achieve similar results, even when using different instruments and the specific manufacturers' pre-processing algorithms, as long as standardised handling of the data matrix is applied during data processing.

In summary, this study suggests that we can achieve consistent biomarkers through different metabolomics data acquisition platforms when the raw data is pre-processed by the manufacturer's software (which guarantees the use of all the specific tool's potential on the raw data), and data matrices are normalised by the application of the same criteria from a commercial multivariate software (e.g., SIMCA).

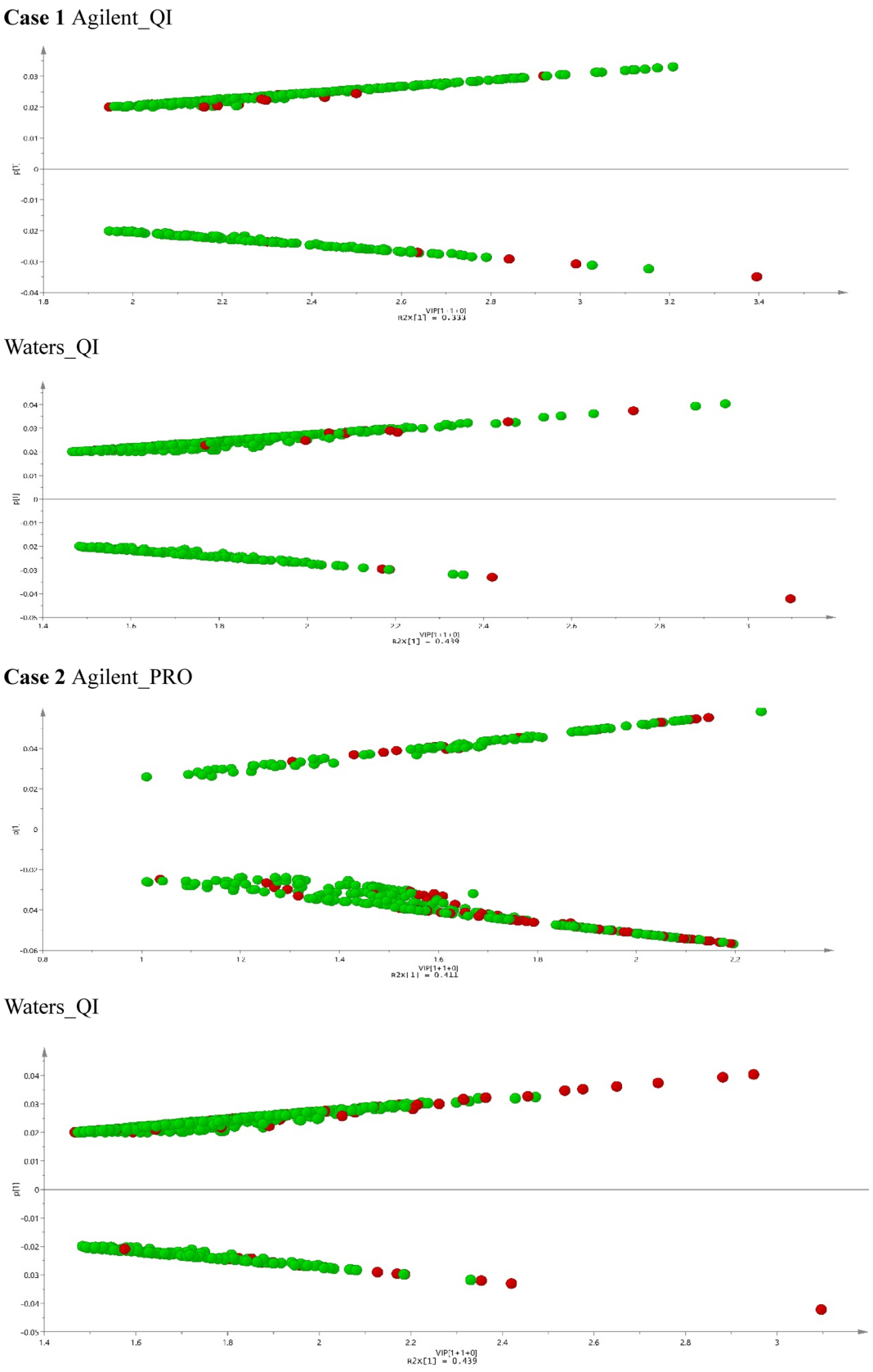


Fig. 5 Biomarkers common to the total candidate biomarkers identified in Case 1 and Case 2

Table 1 Metabolites identified that discriminate between the Butterhead and Romaine lettuce cultivars

Compound	Agilent_QI (*m/z*_min)	Waters_QI (*m/z*_min)	Agilent_PRO (*m/z*_min)	Molecular formula	Error[a] (mDa)	Adduct	Fold Change (BR)[b]	Higher concentration detected in Butterhead or Romaine
Sesquiterpene lactones derivatives								
8-Deoxylactucin	259.0981_13.30	259.0977_12.26		$C_{15}H_{16}O_4$	0.1	M−H	159.10	Butterhead
Santamarin-sulfate	327.0911_15.01	327.0911_14.08	327.0912_15.01	$C_{15}H_{20}O_6S$	0.3	M−H	Only present in B	Butterhead
Phenolic compounds								
Caffeoyltartaric acid		311.0410_4.99	311.0408_5.70	$C_{13}H_{12}O_9$	− 0.1	M−H	0.23	Romaine
Quercetin 3- glucoside isomer		463.0884_10.55	463.0887_11.59	$C_{21}H_{20}O_{12}$	0.5	M−H	0.06	Romaine
Quercetin 3-glucoside-6″-acetate		505.0989_7.88	505.0986_8.69	$C_{23}H_{22}O_{13}$	− 0.2	M−H	0.08	Romaine
Quercetin-3-O-arabinoside		433.0776_11.74	433.0776_12.79	$C_{20}H_{18}O_{11}$	0	M−H	0.17	Romaine
Luteolin 7-neohesperidoside		593.1515_11.18	593.1511_12.15	$C_{27}H_{30}O_{15}$	− 0.1	M−H	0.11	Romaine
Quercetin 3-*O*-(6″-*O*-malonyl)-glucoside 7-*O*-glucuronide		725.1214_7.88	725.1201_9.03	$C_{30}H_{30}O_{21}$	− 0.6	M−H	0.20	Romaine
Others								
9-Undecenal		213.1495_14.40	213.1496_15.40	$C_{11}H_{20}O$	0	M+FA−H	0.13	Romaine
Xanthosine		283.0688_3.14	283.0684_3.99	$C_{10}H_{12}N_4O_6$	0	M−H	14.43	Butterhead

[a]Error calculated based on m/z for Waters QI

[b]Fold change calculated according to raw abundance in the Waters_QI dataset

Acknowledgements This research was supported by CSIC 201870E014, Fundación Séneca de la Región de Murcia, Ayudas a Grupos de Excelencia 19900/GERM/15, Central Public-interest Scientific Institution Basal Research Fund (Y2020XK01), Sichuan Science and Technology Program (2020JDRC0043), and Shanghai Municipal Agricultural Commission (T2017-3-4). We acknowledge our colleagues Miss Lina Zhao and Mr Kai Dou, and Dr Shiwei Wei (Shanghai Agrobiological Gene Center) for help with sample preparation, Dr Hongyan Liu (Chinese Academy of Agricultural Sciences) kindly provides some useful suggestions for the manuscript, and Dr Lei Feng (Instrumental Analysis Center, Shanghai Jiao Tong University) helped with the UPLC-QTOF-MS analysis.

Author contributions CJG and XY designed the metabolomics study, prepared the samples, performed the untargeted metabolomics analysis and run the analyses in the metabolomics platform. CJG created the ISAtab file from the study protocols and rawdata to update the study data in the MetaboLights repository (MTBLS1593). DH and FTB conceived and drafted the manuscript. All the authors contributed with critical intellectual input, read and revised the final draft.

Compliance with ethical standards

Conflict of interest The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

Benton, H. P., Want, E., Keun, H. C., Amberg, A., Plumb, R. S., Goldfain-Blanc, F., et al. (2012). Intra- and interlaboratory reproducibility of ultra performance liquid chromatography-time-of-flight mass spectrometry for urinary metabolic profiling. *Analytical Chemistry, 84*, 2424–2432.

Brunetti, C., Di Ferdinando, M., Fini, A., Pollastri, S., & Tattini, M. (2013). Flavonoids as antioxidants and developmental regulators: relative significance in plants and humans. *International Journal of Molecular Sciences, 14*, 3540–3555.

Bylesjö, M., Rantalainen, M., Cloarec, O., Nicholson, J. K., Holmes, E., & Trygg, J. (2006). OPLS discriminant analysis: combining the strengths of PLS-DA and SIMCA classification. *Journal of Chemometrics, 20*, 341–351.

Cajka, T., & Fiehn, O. (2014). Comprehensive analysis of lipids in biological systems by liquid chromatography-mass spectrometry. *TrAC-Trends in Analytical Chemistry, 61*, 192–206.

Cajka, T., Smilowitz, J. T., & Fiehn, O. (2017). Validating quantitative untargeted lipidomics across nine liquid chromatography-high-resolution mass spectrometry platforms. *Analytical Chemistry, 89*, 12360–12368.

Cervera-Mata, A., Navarro-Alarcón, M., Delgado, G., Pastoriza, S., Montilla-Gómez, J., Llopis, J., et al. (2019). Spent coffee grounds improve the nutritional value in elements of lettuce (*Lactuca sativa* L.) and are an ecological alternative to inorganic fertilizers. *Food Chemistry, 282*, 1–8.

Chadwick, M., Trewin, H., Gawthrop, F., & Wagstaff, C. (2013). Sesquiterpenoids lactones: benefits to plants and people. *International Journal of Molecular Sciences, 14*, 12780–12805.

Coble, J. B., & Fraga, C. G. (2014). Comparative evaluation of preprocessing freeware on chromatography/mass spectrometry data for signature discovery. *Journal of Chromatography A, 1358*, 155–164.

Creek, D. J., Dunn, W. B., Fiehn, O., Griffin, J. L., Hall, R. D., Lei, Z., et al. (2014). Metabolite identification: are you sure? And how do your peers gauge your confidence? *Metabolomics, 10*, 350–353.

Dixon, R. A. (2003). Phytochemistry meets genome analysis, and beyond. *Phytochemistry, 62*, 815–816.

Djekic, D., Pinto, R., Vorkas, P. A., & Henein, M. Y. (2016). Replication of LC-MS untargeted lipidomics results in patients with calcific coronary disease: An interlaboratory reproducibility study. *International Journal of Cardiology, 222*, 1042–1048.

Dudzik, D., Barbas-Bernardos, C., García, A., & Barbas, C. (2018). Quality assurance procedures for mass spectrometry untargeted metabolomics. A review. *Journal of Pharmaceutical and Biomedical Analysis, 147*, 149–173.

Engskog, M. K. R., Haglöf, J., Arvidsson, T., & Pettersson, C. (2016). LC-MS based global metabolite profiling: the necessity of high data quality. *Metabolomics, 12*, 1–19.

Fiehn, O. (2002). Metabolomics-the link between genotypes and phenotypes. *Plant Molecular Biology, 48*, 155–171.

García, C. J., García-Villalba, R., Garrido, Y., Gil, M. I., & Tomás-Barberán, F. A. (2016). Untargeted metabolomics approach using UPLC-ESI-QTOF-MS to explore the metabolome of fresh-cut iceberg lettuce. *Metabolomics, 12*, 138.

García, C. J., Garcia-Villalba, R., Gil, M. I., & Tomás-Barberán, F. A. (2017). LC-MS untargeted metabolomics to explain the signal metabolites inducing browning in fresh-cut lettuce. *Journal of Agricultural and Food Chemistry, 65*, 4526–4535.

García, C. J., Gil, M. I., & Tomás-Barberán, F. A. (2018). LC-MS untargeted metabolomics reveals early biomarkers to predict browning of fresh-cut lettuce. *Postharvest Biology and Technology, 146*, 9–17.

Gika, H. G., Theodoridis, G. A., Earll, M., Snyder, R. W., Sumner, S. J., & Wilson, I. D. (2010). Does the mass spectrometer define the marker? A comparison of global metabolite profiling data generated simultaneously via UPLC-MS on two different mass spectrometers. *Analytical Chemistry, 82*, 8226.

Gika, H. G., Theodoridis, G. A., Plumb, R. S., & Wilson, I. D. (2014). Current practice of liquid chromatography-mass spectrometry in metabolomics and metabonomics. *Journal of Pharmaceutical and Biomedical Analysis, 87*, 12–25.

Gürdeniz, G., Kristensen, M., Skov, T., & Dragsted, L. O. (2012). The effect of LC-MS data preprocessing methods on the selection of plasma biomarkers in fed *vs.* fasted rats. *Metabolites, 2*, 77–99.

Lanucara, F., Holman, S. W., Gray, C. J., & Eyers, C. E. (2014). The power of ion mobility-mass spectrometry for structural characterization and the study of conformational dynamics. *Nature Chemistry, 6*, 281–294.

Li, Z., Lu, Y., Guo, Y., Cao, H., Wang, Q., & Shui, W. (2018). Comprehensive evaluation of untargeted metabolomics data processing software in feature detection, quantification and discriminating marker selection. *Analytica Chimica Acta, 1029*, 50–57.

Mahieu, N. G., Genenbacher, J. L., & Patti, G. J. (2016). A roadmap for the XCMS family of software solutions in metabolomics. *Current Opinion in Chemical Biology, 30*, 87–93.

Mairinger, T., Causon, T. J., & Hann, S. (2018). The potential of ion mobility-mass spectrometry for non-targeted metabolomics. *Current Opinion in Chemical Biology, 42*, 9–15.

Myers, O. D., Sumner, S. J., Li, S., Barnes, S., & Du, X. (2017). Detailed investigation and comparison of the XCMS and MZmine^2 chromatogram construction and chromatographic peak detection methods for preprocessing mass spectrometry metabolomics data. *Analytical Chemistry, 89*, 8689–8695.

Oksman-Caldentey, K., & Inzé, D. (2004). Plant cell factories in the post-genomic era: New ways to produce designer secondary metabolites. *Trends in Plant Science, 9*, 433–440.

Riga, P., Benedicto, L., Gil-Izquierdo, Á., Collado-González, J., Ferreres, F., & Medina, S. (2019). Diffuse light affects the contents of vitamin C, phenolic compounds and free amino acids in lettuce plants. *Food Chemistry, 272*, 227–234.

Rochat, B. (2016). From targeted quantification to untargeted metabolomics: Why LC-high-resolution-MS will become a key instrument in clinical labs. *TrAC-Trends in Analytical Chemistry, 84*, 151–164.

Shulaev, V., Cortes, D., Miller, G., & Mittler, R. (2008). Metabolomics for plant stress response. *Physiologia Plantarum, 132*, 199–208.

Sugimoto, M., Kawakami, M., Robert, M., Soga, T., & Tomita, M. (2012). Bioinformatics tools for mass spectroscopy-based metabolomic data processing and analysis. *Current Bioinformatics, 7*, 96–108.

Tavarini, S., Sgherri, C., Ranieri, A. M., & Angelini, L. G. (2015). Effect of nitrogen fertilization and harvest time on steviol glycosides, flavonoid composition, and antioxidant properties in *Stevia rebaudiana* Bertoni. *Journal of Agricultural and Food Chemistry, 63*, 7041–7050.

van den Berg, R. A., Hoefsloot, H. C. J., Westerhuis, J. A., Smilde, A. K., & van der Werf, M. J. (2006). Centering, scaling, and transformations: Improving the biological information content of metabolomics data. *BMC Genomics, 7*, 142.

Yang, X., Cui, X., Zhao, L., Guo, D., Feng, L., Wei, S., et al. (2017). Exogenous glycine nitrogen enhances accumulation of glycosylated flavonoids and antioxidant activity in lettuce (*Lactuca sativa* L.). *Frontiers in Plant Science, 8*, 2098.

Yang, X., Feng, L., Zhao, L., Liu, X., Hassani, D., & Huang, D. (2018a). Effect of glycine nitrogen on lettuce growth under soilless culture: A metabolomics approach to identify the main changes occurred in plant primary and secondary metabolism. *Journal of the Science and Food Agriculture, 98*, 467–477.

Yang, X., Wei, S., Liu, B., Guo, D., Zheng, B., Feng, L., et al. (2018b). A novel integrated non-targeted metabolomic analysis reveals significant metabolite variations between different lettuce (*Lactuca sativa* L.) varieties. *Horticulture Research, 5*, 33.