

Article

EASY-FIA: A Readably Usable Standalone Tool for High-Resolution Mass Spectrometry Metabolomics Data Pre-Processing

Aurelia Morabito ^{1,2}, Giulia De Simone ^{1,3}, Manuela Ferrario ² , Francesca Falcetta ⁴, Roberta Pastorelli ¹ 
and Laura Brunelli ^{1,*}

¹ Laboratory of Mass Spectrometry, Istituto di Ricerche Farmacologiche Mario Negri IRCCS, 20156 Milan, Italy

² Department of Electronics, Information and Bioengineering, Politecnico di Milano, 20133 Milan, Italy

³ Department of Biotechnologies and Biosciences, Università degli Studi Milano Bicocca, 20126 Milan, Italy

⁴ Unit of Biophysics, Istituto di Ricerche Farmacologiche Mario Negri IRCCS, 20156 Milan, Italy

* Correspondence: laura.brunelli@marionegri.it; Tel.: +39-0239014742

Abstract: Flow injection analysis coupled with high-resolution mass spectrometry (FIA-HRMS) is a fair trade-off between resolution and speed. However, free software available for data pre-processing is few, web-based, and often requires advanced user specialization. These tools rarely embedded blank and noise evaluation strategies, and direct feature annotation. We developed EASY-FIA, a free standalone application that can be employed for FIA-HRMS metabolomic data pre-processing by users with no bioinformatics/programming skills. We validated the tool's performance and applicability in two clinical metabolomics case studies. The main functions of our application are blank subtraction, alignment of the metabolites, and direct feature annotation by means of the Human Metabolome Database (HMDB) using a minimum number of mass spectrometry parameters. In a scenario where FIA-HRMS is increasingly recognized as a reliable strategy for fast metabolomics analysis, EASY-FIA could become a standardized and feasible tool easily usable by all scientists dealing with MS-based metabolomics. EASY-FIA was implemented in MATLAB with the App Designer tool and it is freely available for download.



Citation: Morabito, A.; De Simone, G.; Ferrario, M.; Falcetta, F.; Pastorelli, R.; Brunelli, L. EASY-FIA: A Readably Usable Standalone Tool for High-Resolution Mass Spectrometry Metabolomics Data Pre-Processing. *Metabolites* **2023**, *13*, 13. <https://doi.org/10.3390/metabo13010013>

Academic Editors: Giuseppe Martano and Giuseppe Paglia

Received: 23 November 2022

Revised: 15 December 2022

Accepted: 16 December 2022

Published: 21 December 2022



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: Flow Injection Analysis-High Resolution Mass Spectrometry; FIA-HRMS data pre-processing; *m/z* alignment and annotation; metabolomics

1. Introduction

Metabolism is at the cornerstone of all cellular functions, and it is deregulated in different and complex diseases. This emphasizes the importance of a comprehensive understanding of metabolic regulation at the whole-organism level [1]. Metabolomics has also been described as the “real-world endpoint” of omics research [2] and is closely linked to the phenotype of an organism [3]. Metabolic phenotypes are the result of the interplay between the genome and the environment, so metabolomics represents the organism's response to perturbations due to either internal or external factors [4,5].

Untargeted metabolomics has been widely used as an unbiased strategy for the whole profiling of small molecules in biological systems, either to point out potential signatures and biomarkers of pathologies or to better understand their underlying mechanisms and progression. Mass spectrometry-based untargeted metabolomics is widely used to identify metabolic alterations associated with human disease [6–14]. Beyond that, untargeted metabolomics has been successfully applied in forensic, environmental, food, and agricultural studies. For instance, metabolomics could possibly be employed to determine putative biomarkers of drug consumption, pollution control, soil improvement and environmental monitoring by studying the metabolism of microorganisms [15]. Other fields which are

exploiting metabolomics are precision nutrition, with the aim of assessing individual responses and profiles of dietary patterns in humans, and food analysis, from risk and safety assessment to quality evaluation [16]. Finally, metabolomics is increasingly employed in combination with other omics for the elucidation of various crop functions and responses to stress [17].

Liquid Chromatography coupled with Mass Spectrometry (LC-MS) is one of the most widely used techniques used in untargeted metabolomics due to its good selectivity and sensitivity allowing the detection of thousands of metabolites in a single analysis [18]. Direct injection of the sample into a carrier stream, without any chromatographic system (Flow Injection Analysis-High Resolution Mass Spectrometry, FIA-HRMS) has found even wider application in untargeted metabolomics as it offers similar advantages as LC-MS in terms of sensitivity, precision and accuracy, but greater method simplicity and fast analysis time [19]. FIA-HRMS has thus become a technique for untargeted metabolomics applications in clinical studies, providing analysis of many samples and fast screening time, as is the case for critically ill patients where patient stratification is time-sensitive [20]. FIA-HRMS has been also used in archaeological samples [21], to discriminate between different kinds of cannabinoids present in cannabis samples [22], or to characterize eventual food adulterations [23].

Metabolomics investigations usually include the collection and pre-processing of the MS raw data, statistical analysis to detect features of interest, and feature identification or annotation to cast light on their chemical and biological functions [24]. Several standardized data pre-processing protocols are available for untargeted LC-MS [25–30] and Nuclear Magnetic Resonance [31]; however, algorithms for FIA-HRMS data pre-processing are few and web-based. This is probably due because each type of mass spectrometer produces data in a proprietary format; therefore, data pre-processing depends on the software distributed by the vendor [32]. Furthermore, only a limited number of free tools merge data pre-processing with feature imputation using databases (e.g., Human Metabolome Database—(HMDB), Kyoto Encyclopedia of Genes and Genomes—(KEGG), and LIPID MAPS), in a single pipeline useful for all type of users, beginners included [33]. For instance, FIA-HRMS pre-processing might be aided by the proFIA software, which is available either as an R package on the Bioconductor repository [34] or as a tool of Workflow4Metabolomics online infrastructure in the Galaxy environment [24]. proFIA implements a strategy to pre-process FIA-HRMS raw data and generates the peak table; however, it does not provide an embedded method for blank subtraction and direct feature annotation. Our present work aims at providing the applicability and performance of easy-to-use standalone software for FIA-HRMS metabolomics data pre-processing.

EASY-FIA Approach

FIA-HRMS experiment is characterized by the direct injection of sequential samples directly into the mass spectrometer source without prior chromatographic separation. FIA-HRMS via modern high-resolution mass spectrometers such as Orbitrap and time-of-flight (ToF) was shown to enable the determination of hundreds to thousands of m/z features in biological samples allowing the efficient discrimination between isobaric compounds and facilitating the determination of molecular formulas, providing key technology for high-throughput metabolomics analysis [35]. Each FIA-HRMS data file consists of successive acquisitions of mass spectra (m/z value) within a defined time window (typically one-two minutes) and has a classical two-dimensional structure (m/z and intensity) as shown in Figure 1.

A fast analysis is pivotal to ensure that large sample cohorts are analyzed within a reasonable time and with optimal allocation of experimental resources. Metabolomics clinical study size can quickly scale to several or tens of thousands of individual samples to fulfill the demand imposed by robust experimental design. Large-scale phenotyping analysis can hardly be performed in a short time when the chromatographic step is included, because of the daily limits of the acquired samples.

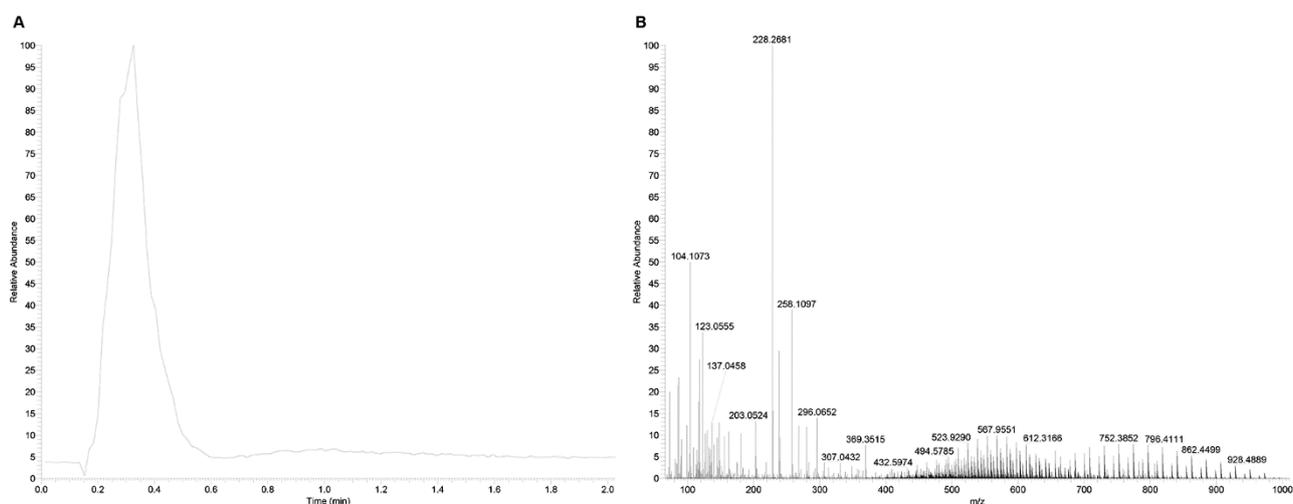


Figure 1. FIA-HRMS spectrum acquisition. (A): FIA-HRMS representative recorded spectrum; (B): m/z feature distribution obtained using the FIA-HRMS approach.

Here, we set out to establish a platform for routine high-throughput and high-mass-accuracy metabolomics data pre-processing, which combines metabolic feature extraction, blank subtraction, feature alignment, and annotation to generate a single table of metabolic features by their intensities. Together with a robust and reliable feature alignment to avoid false positives, first-pass automatic feature annotation is fundamental to speed up metabolomics data elaboration. EASY-FIA implements a non-chromatographic-based m/z annotation using an in-house HMDB-derived database to annotate single or multiple accurate m/z features by a customizable adduct annotation list of $[M+H]^+$ / $[M-H]^-$ candidates. First-round identification by full-scan mass spectra information needs to be completed with tandem (MS/MS) mass spectral data for metabolite identification.

In the following sections, we describe EASY-FIA workflow algorithms (Materials and Methods) and the performance of our tool using two clinical metabolomics datasets (Results). Furthermore, we examined the evaluation of the applicability of blank subtraction and the unbiased mathematical definition of a peak intensity threshold to discard spurious intensities.

We expect EASY-FIA to be highly useful for high-throughput clinical metabolomics applications (e.g., population-level disease screening and omics data integration strategies).

2. Materials and Methods

2.1. EASY-FIA Pre-Processing Workflow

EASY-FIA software for FIA-HRMS metabolomics pre-processing was implemented in MATLAB (version R2021a) environment. If the user does not have a currently available license of MATLAB software, a MATLAB runtime can be downloaded free of charge from the MATLAB website; it is a standalone set of shared libraries that enables the execution of compiled MATLAB components such the one proposed in this work. The required MATLAB Runtime is the release R2021a (Version 9.10). Finally, EASY-FIA can be freely downloaded either from GitHub at <https://github.com/AMrbt20/EASY-FIA/> (accessed on 15 December 2022) under a GNU GPL v3.0 license, or at <https://www.marionegri.it/centro-di-ricerca-spettrometria-di-massa-per-la-salute-e-ambiente> (accessed on 15 December 2022) by following the provided instructions. The user-friendly Graphical Computer Interface (Supplementary Figure S1) aids to set a minimum number of mass spectrometry parameters for data processing (m/z tolerance interval, adducts, and intensity/sample cut-off values). Additional documentation and instructions are available on GitHub.

The EASY-FIA pipeline consists of three steps: blank subtraction, m/z feature alignment and feature annotation using the in-house HMDB database (Figure 2).

2.1.2. Alignment of m/z

Alignment starts from the first m/z of the first sample, and the algorithm searches each m/z in all the other m/z lists within the tolerance interval t using a while loop. When an m/z match is found within the t interval, the matching m/z is progressively averaged, and their mean is saved in an accessory vector. Then, the relative intensities of each matched m/z are reported sequentially in a single matrix, where the vector of averaged masses appears in the first column. The final matrix of intensities contains an identification number for each m/z to aid eventual subsequent analysis (Figure 2B).

Before starting the processing, the user can flag the *Sort matrix by adducts* option, available for the positive ionization mode. This function rearranges the matrix of intensities by sorting and grouping each M+H with its theoretical sodium and potassium adducts (M+Na, M+K). First, the algorithm subtracts the theoretical value of hydrogen from all m/z values in the matrix of intensities. Then, the algorithm adds the theoretical masses of sodium and potassium ions to each of the generated monoisotopic masses. The theoretical adducts are then searched in the experimental m/z in the matrix of intensity. When an m/z match is found within the tolerance interval, the row is moved under the corresponding M+H value.

A matrix cleaning procedure automatically runs at the end of the alignment to replace the zeros in the matrix of intensities with NaN (not a number), to manage the absence of intensity values for specific m/z . The GUI also implements the *Sample cut-off* option, which removes the features whose intensity is not detected in a number of samples defined by the user (Figure 2B). The EASY-FIA default threshold is 1 (it removes m/z detected in just one sample), but the user can change this number in the *Sample cut-off* field of the GUI. In the end, the matrix of intensities is saved in the same folder in both MATLAB (.mat) and Excel file formats with the suffix *_alignment_data_sorted* or *alignment_data* according to the option selected.

2.1.3. Human Metabolome Database (HMDB) Annotation

EASY-FIA automatically identifies the metabolic features (m/z value) by comparing each experimental m/z with the theoretical one present in the Human Metabolome Database (HMDB). This step can be run either independently or immediately after the alignment.

Two in-house databases were generated for annotation, one for positive and one for negative ionization mode; they are available as MATLAB files in the GitHub repository as HMDB_POS.mat and HMDB_NEG.mat, respectively. The HMDB in-house databases were created from the HMDB repository by downloading the XML file (All Metabolites, version 23 October 2021) and retrieving the identification code, metabolite name and monoisotopic weight for each metabolite. The in-house databases were obtained by adding the theoretical weights of hydrogen to the monoisotopic weight of the retrieved metabolites, sodium and potassium ions, for the positive acquisition mode, and subtracting the theoretical weight of hydrogen for the negative acquisition mode. EASY-FIA requires the user to load first the Excel file containing the matrix of intensities produced in the alignment section, then the HMDB mat file according to the ionization mode of the data under investigation. The algorithm looks for every m/z present in the matrix of intensities in the HMDB in-house database, implementing a cycle and a find function to search each mass within the tolerance interval. For any m/z match, the HMDB identification code, the name and the adduct of the metabolite are inserted in the corresponding row of the matrix of intensities. When an m/z matches more than one metabolic feature, then the number of matches is reported (Figure 2C).

The output of the identification process is saved as an Excel file in the current folder with the suffix HMDB_ID and consists of the cleaned matrix of intensities where each identified mass is associated with the name of the corresponding metabolite or the number of related metabolites.

2.2. Case Studies

EASY-FIA was used for the sample alignment of two published clinical studies, both with the objective of determining metabolites that cast light on different pathological conditions.

Case Study 1 [36]. A longitudinal population-based study that investigated frailty syndrome in older subjects, and aims at identifying metabolic hallmarks of the frailty syndrome. A total of 130 plasma samples were analyzed by the LTQ-OrbitrapXL mass spectrometer (Thermo Fisher Scientific) equipped with an electrospray source operated in negative and positive modes. Briefly, metabolites were extracted by adding cold methanol (4:1, MeOH:plasma) to the plasma samples (20 μ L); samples were incubated at -80 $^{\circ}$ C for 20 min and then centrifuged for 15 min at $14,000\times g$. The supernatant was collected, dried under nitrogen, and suspended in 25 μ L of 0.1% formic acid. Each run of the instrument was carried out by injecting 8 μ L of sample extract at a flow rate of 50 μ L/min of mobile phase consisting of isopropanol/water (60:40, v/v) buffered with 5 nM ammonium at pH 9 for negative mode and methanol/water (60:40, v/v) with 0.1% formic acid at pH 3 for positive mode. The source temperature was set to 240 $^{\circ}$ C with 25 L/in drying gas and a nebulizer pressure of 35 psig. Reference masses for internal calibration were used in continuous infusion during the analysis (m/z 210.1285 for positive and m/z 212.0750 for negative ionization). Mass spectra were recorded from m/z 50 to 1000. The eight quality controls (QC) acquired during the FIA-HRMS analysis were used to test EASY-FIA reproducibility.

Case Study 2 [20]. The longitudinal population-based investigation, aimed at verifying whether different responses to therapy in the acute phase of shock were associated with different plasma metabolic patterns. Forty-two (42) plasma samples were analyzed by the 6550 iFunnel Q-TOF mass spectrometer (Agilent) equipped with an electrospray source operated in negative and positive modes. Metabolites were extracted by adding four volumes of cold methanol to the plasma sample (10 μ L); samples were vortexed and incubated at -20 $^{\circ}$ C for 1 h and then centrifuged 10 min at $14,000\times g$. The supernatant was collected, dried in a SpeedVac and resuspended in 50 μ L of 0.1% formic acid. The flow rate of the instrument was set to 150 μ L of mobile phase consisting of isopropanol/water (60:40, v/v) buffered with 5 nM ammonium at pH 9 for negative mode, and methanol/water (60:40, v/v) with 0.1% formic acid at pH 3 for the positive mode. The source temperature was set to 320 $^{\circ}$ C with 15 L/min drying gas and a nebulizer pressure of 35 psig. Reference masses for internal calibration were used in continuous infusion during the analysis (m/z 121.050873, 922.009798 for positive and m/z 11.9856, 1033.9881 for negative ionization). Mass spectra were recorded from m/z 50 to 1100. Case 2, quantitative data of the metabolic species analyzed both by untargeted metabolomics (FIA-HRMS) and by target metabolomics (AbsoluteIDQ 180 kit, Biocrates, Innsbruck, Austria) [37] were used to demonstrate the correlation between peak intensity and metabolite concentrations (microM).

2.3. Statistical Analysis

Spearman correlation analysis was performed to verify the correlation between peak intensities and concentration (GraphPad Prism 9.2.0) for those metabolites quantified by both metabolomics strategies (untargeted FIA-HRMS and targeted AbsoluteIDQ 180 kit Biocrates).

3. Results and Discussion

3.1. EASY-FIA Performance on FIA-HRMS Clinical Metabolomics Case Studies

The reliability and applicability of the EASY-FIA algorithm were tested on two published case studies of human plasma FIA-HRMS untargeted metabolomics profiling obtained using two different high-resolution mass spectrometers [20,36]. The raw data files acquired from the vendor software were easily converted to .csv files, obtaining the centroided m/z list of the samples and the relative blanks of the two case studies. The tolerance interval was set to 6 ppm for both positive and negative ionization modes for case study 1 (LTQ OrbitrapXL), and to 20 ppm for case study 2 (6550 iFunnel Q-TOF), taking into consideration the different instrumental accuracy; the intensity cut-off was set to 0 and the peaks cut-off to 1. EASY-FIA aligned the samples of both acquisition modes in the two

studies. The intensity matrix in case study 1 contained a number of m/z equal to 248,000 and 251,557 with an average of, respectively, 12,936 and 15,094 m/z per sample for the positive and negative modes, respectively. The matrix of intensities in case study 2 had 38,363 and 74,569 m/z , with averages of, respectively, 18,115 and 37,196 m/z per sample for the positive and negative modes, respectively (Supplementary Tables S1–S5).

After alignment, EASY-FIA automatically annotated the m/z features by using the in-house databases and identified by HMDB 35,017 m/z (25,134 positive, 9883 negative) in case study 1 and 20,084 m/z (12,024 positive, 8060 negative) in case study 2 (Supplementary Tables S1–S5).

Since EASY-FIA provides the first-round identification exclusively based on the m/z , users should confirm identities through MS/MS approaches. For the metabolites that were annotated unambiguously (m/z with a single HMDB metabolite), we observed that FIA-HRMS analysis acquire both nonlipid and lipid metabolites (nonlipid ~93%, lipid ~6.3%) even though with a high prevalence of nonlipid metabolic species.

To evaluate the quality of the EASY-FIA pre-processing workflow, we compared the number of signals detected with peak intensities in the highest quartile (75th) because such features are expected to be less influenced by instrumental background and accurately quantified among the eight QC replicates. Seventy-one percent (71%) of the features were present in all QC replicates, highlighting the reproducibility of EASY-FIA.

To further validate EASY-FIA performance, we verified the correlation between peak intensity and absolute concentration for those features quantified by both approaches (untargeted metabolomics by FIA-HRMS and targeted metabolomics by AbsoluteIDQ 180 kit Biocrates). Spearman correlation between peak intensities and concentrations for metabolic species belonging to amino acids (arginine, glutamine, tyrosine, histidine, proline, lysine, threonine), biogenic amines (taurine) and acylcarnitines (carnitine, acylcarnitine) showed a significant good correlation ($r > 0.7$ and p -value < 0.05) in both positive and negative ionization modes (Figure 3). As such, EASY FIA achieves performances of good linearity in terms of correlation between peak intensity and concentration.

3.2. Unbiased Strategies for Limiting the Matrix Size of the Intensities

FIA-HRMS produces a huge matrix of intensities, with wide scattering due to a large number of missing values (m/z detected in a small number of samples or even in only one) in relation to the number of all the m/z detected (Table 1 and Supplementary Tables S1–S4). Thousands of m/z are generally detected in one single sample, and this poses a problem for statistical and data mining analyses. Moreover, high-dimensional matrices are subject to the so-called curse of dimensionality [38]: the performance of a machine learning model does not necessarily improve with the number of features, so they must be selected by removing noisy features and redundancies.

Table 1. Missing values and Features' intensity values in the two case study matrices. Number of missing values and intensity values detected in the intensity matrix for each case study, in accordance with the acquisition mode.

Case Study	Acquisition Mode	Missing Values	Intensity Values
1	Positive	32,290,973	1,685,027
1	Negative	32,762,205	1,952,661
2	Positive	1,168,071	519,901
2	Negative	2,343,890	937,146

Even though EASY-FIA is a tool for alignment and identification, and it aims at pre-processing data for further data mining applications, we tested some unbiased strategies for data cleaning and for dimensionality reduction by identifying and filtering non-relevant features. We searched for an objective cut-off intensity threshold to discard intensity values, instead of using an arbitrary threshold or hypothesis on noise characteristics.

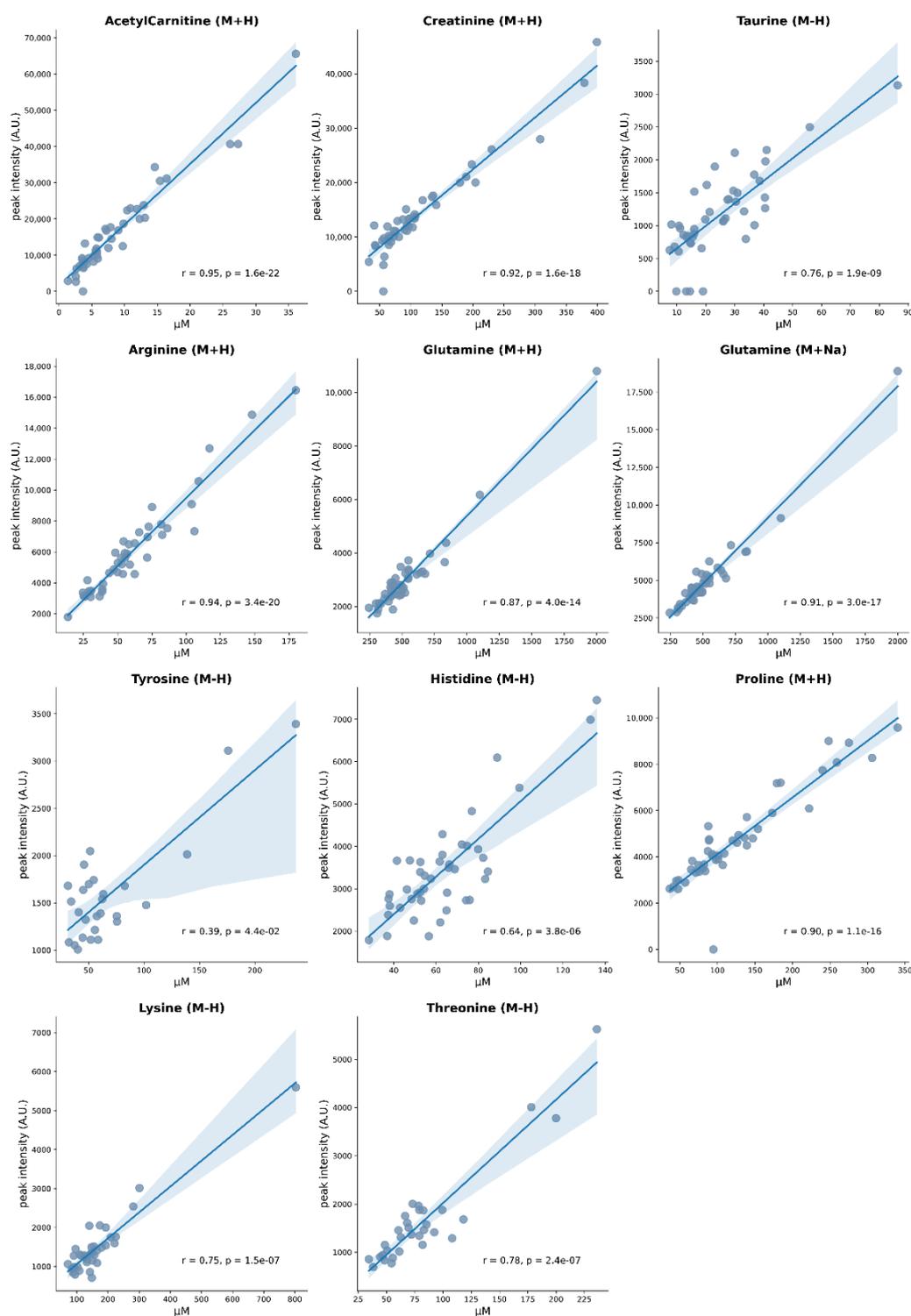


Figure 3. EASY-FIA validation. Spearman correlation between peak intensities and concentrations of metabolites quantified by untargeted FIA-HRMS and targeted AbsoluteIDQ 180 kit approaches. The blue line represents the regression line, blue dots refer to the samples and the shaded area represents the 95% confidence interval for that regression.

3.3. Assessment of an Unbiased Strategy for Intensity Cut-Off to Remove Blank Spectra Noise

In the mass spectrometry-based metabolomics analysis, selecting the intensity cut-off to discard a portion of acquired m/z features is not a trivial issue. Indeed, a wrong value selection may have a detrimental impact on the subsequent data elaboration. To date, the

criterion for selecting a cut-off intensity value is not univocal; in fact, several approaches can be seen in the scientific literature. For instance, Fuhrer et al. suggested filtering peaks of less than 500 ion counts in the summed spectrum [39]. Beuchel et al. implemented an algorithm for the removal of outliers based on the logarithmic transformation of non-zero measurements, and the threshold was set to be less than five times the standard deviation value (SD) [40,41]. Gatto et al. developed an R function called *removePeaks* to remove peaks with intensity below an arbitrary threshold [42]. However, the use of an arbitrary cut-off to discard m/z might be a biased strategy, considering the large value range of FIA-HRMS metabolomics matrices.

We questioned the possibility of using the m/z spectrum of the blanks to establish an unbiased cut-off intensity value, which would be tailored to the intensity matrix of the investigated data. In case study 1, the intensity range goes from 13.6 arbitrary units (AU) to 1.08×10^7 AU in positive mode, and from 14.6 to 4.11×10^6 AU in negative mode. Case study 2 has an intensity range from 2.53 to 9.88×10^5 AU for the positive mode and from 2.47 to 1.11×10^6 AU for the negative mode. In case study 1, the blank had an intensity range between 14.60 and 7.53×10^5 AU for positive mode and between 13.20 and 8.93×10^5 AU for negative; in case study 2 the blank intensity ranged between 2.50 and 1.09×10^6 AU for positive and 2.47 and 9.48×10^5 for negative mode.

We compared the distributions of peak intensities values between the blank and the samples in order to see if there were a putative cutoff value to be used as a threshold to separate them. As Figure 4 clearly shows the peak intensity distributions are completely overlapped and the threshold-based approach cannot be applied.

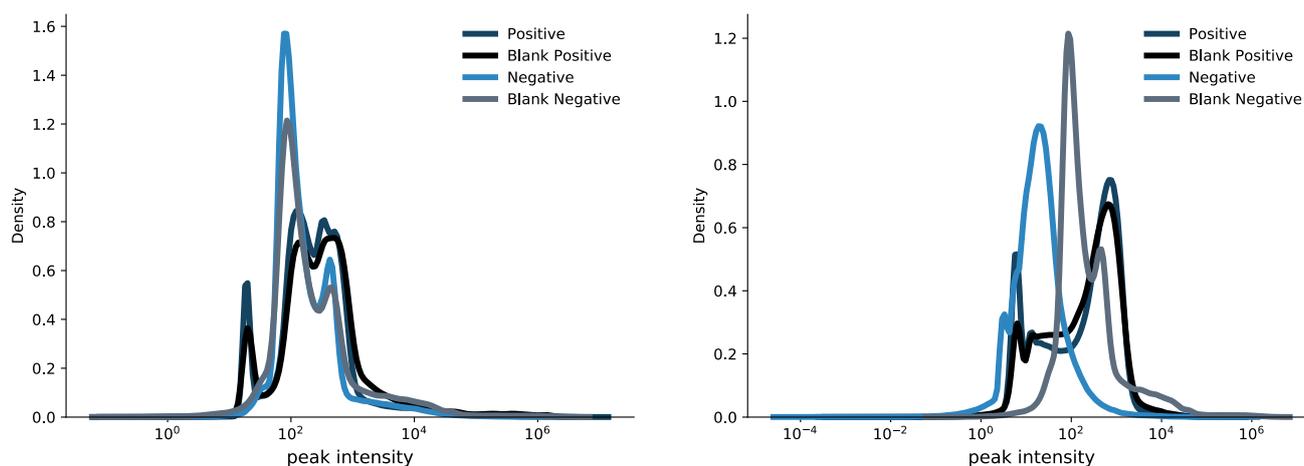


Figure 4. Distribution of peak intensity in the blank and sample spectra. The graphs show the peak intensity distribution density and thus the range of peak intensity in the blank and in the sample for case study 1 (left panel) and for case study 2 (right panel). m/z intensities are reported as arbitrary units and are displayed in logarithm scale.

Note, that there was a low percentage of common m/z values (Supplementary Table S6) between blanks and samples, thus leading us to overlook the use of blank spectra for determining a cut-off threshold on peak intensities. Although we implemented the *intensity cut-off* functionality in EASY-FIA to allow the user to set a threshold under which intensities are discarded, we would like to emphasize that this biased strategy may affect the subsequent data elaboration, considering the high dynamic range of the acquired features.

3.4. Assessment of an Unbiased Strategy for m/z Reduction

We evaluated the possibility of creating a mathematical model of the acquisition noise, in order to discard the m/z values associated with it. Several approaches have been proposed for the characterization of the acquisition noise [43], such as modeling baseline noise by using baseline functions (e.g., linear, logarithmic, exponential or piecewise [44]), the top-hat

operator [45], or applying specific filters that remove the electronic noise based on the physics of the acquisition system [44]. We assume the acquisition noise to be contained in the spectrum of the blanks, so we evaluated the blank spectrum trend over the runs for positive (Figure 5A,C) and negative (Figure 5B,D) ionization modes in both case studies, seeking any temporal pattern. The blank spectra were characterized neither by a specific shape or trend that could be modeled by a mathematical function nor by any linear offset superimposed on the intensities due to the carry-over effect (Figure 5A–D).

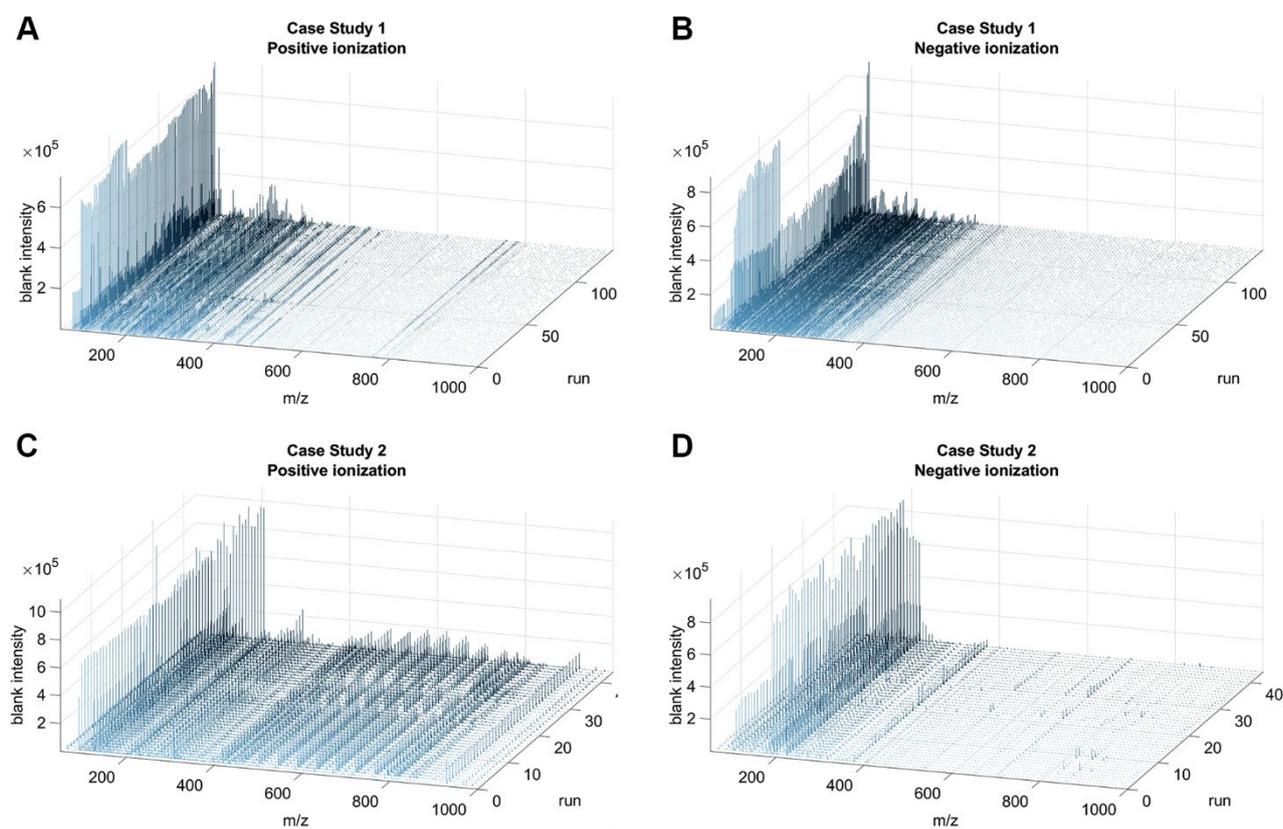


Figure 5. Blank spectra over runs. Three-dimensional plot of the m/z spectra of the blanks over the runs (y-axis) in the positive (A,C) and negative ionization (B,D) modes for each case study.

To further confirm these observations, we repeatedly picked 100 m/z per ionization mode in a random way, and we plotted their intensities through the runs in both case studies (Supplementary Figures S2 and S3). Since we had no evidence of a clear trend in the blank or a particular behavior depending on the run, EASY-FIA only implements the simple blank subtraction in order to painlessly remove the minimal contribution of the blank.

Overall, the impossibility of identifying an unbiased cut-off or a modellable noise for the FIA-HRMS analysis suggested setting the cut-off intensity to 0 (i.e., we took all the acquired m/z) and the sample cut-off to 1 (i.e., we discarded the m/z for which only one intensity was detected) for our analyses.

Our strategy still enabled us to make a first features reduction: in case study 1, the m/z were reduced from 263,203 to 248,000 and from 263,061 to 251,557 in positive and negative mode, respectively, and in case study 2 from 53,314 to 38,363 and from 75,706 to 74,569, obtaining a matrix of intensities reduced by, respectively, the 7% and 17% for case study 1 and 2. If users wish to further reduce the number of m/z for subsequent data analysis, they may consider only the annotated features (i.e., case study 1: 35,017 annotated m/z , case study 2: 20,084 annotated m/z).

4. Conclusions

We develop EASY-FIA as a free, reliable, standalone tool for FIA-HRMS fast metabolomics data pre-processing; it is usable by users with no informatics/programming skills, and it only requires mass spectrometry parameters to be set up. Moreover, EASY-FIA automatically allows a non-chromatographic-based feature annotation according to the widely recognized Human Metabolome Database, thus providing rapid first-pass information about metabolic elements.

Performance on both data reproducibility (QC) and linearity (correlation with absolute metabolite concentrations) was validated, showing that EASY-FIA pre-processing achieves robust feature alignment. Furthermore, we demonstrated that EASY-FIA is intended for any kind of mass spectrometer data by applying our approach to two case studies of metabolomics clinical data obtained using two different high-resolution mass spectrometers. EASY-FIA implements an automatic blank subtraction to remove the blank's minimal contribution since we assessed that neither an intensity cut-off value nor mathematical modeling of the noise worked as reliable methods for the efficient size reduction of the m/z intensity matrix.

EASY-FIA code is fully embedded in an intuitive GUI and requires no programming skills. The GUI allows the user to customize a minimum number of alignment parameters related to the mass spectrometry tolerance interval, adducts, and intensity/sample cut-off values. The EASY-FIA metabolomics intensity matrix with huge numbers of metabolic features is then suitable for in-depth data analysis, including data mining, feature reduction, and importance ranking methods such as PLS-DA, SVM, RF, and mRMR [37] to select the features according to their importance in identifying the target class in the classification problem.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/metabo13010013/s1>, Figure S1: Graphic User Interface of EASY-FIA; Figure S2: Case study 1-Multiple plots of 100 random m/z intensities (y-axis) through the runs (x-axis) for positive (A) and negative (B) ionization modes; Figure S3: Case study 2-Multiple plots of 100 random m/z intensities (y-axis) through the runs (x-axis) for positive (A) and negative (B) ionization modes.; Table S1: Matrix of intensities of case study 1 in positive ionization; Table S2: Matrix of intensities of case study 1 in negative ionization; Table S3: Matrix of intensities of case study 2 in positive ionization; Table S4: Matrix of intensities of case study 2 in negative ionization; Table S5: Summary of the number of m/z identified, not identified, and correctly aligned by EASY-FIA after discarding the m/z detected in a single sample; Table S6: Number of m/z acquired in the samples and blanks, in accordance with the acquisition mode. The table also reports the percentage of shared m/z between samples and blanks in both case studies.

Author Contributions: Conceptualization, R.P., L.B. and M.F.; methodology, L.B. and G.D.S.; software, A.M. and F.F.; validation, L.B. and G.D.S.; formal analysis, L.B.; investigation, L.B. and G.D.S.; resources, L.B. and G.D.S.; data curation, A.M.; writing—original draft preparation, A.M.; writing—review and editing, R.P., M.F. and L.B.; visualization, A.M.; supervision, R.P. and M.F.; project administration, R.P. and M.F. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The ethical aspects of the Case study 1 (NCT01345110) and Case study2 (NCT02141607) have been already evaluated and approved by the Ethical Committee of the University of Pavia and by Geneva regional research ethics committee (Commission cantonale d'éthique de la recherche, President Prof. Bernhard Hirschel, study number 14-041) and by ethical committee of Hôpital Erasme-Université Libre De Bruxelles (study number P2014/171) respectively.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data underlying this article are available in the Supplementary Materials and on GitHub at <https://github.com/AMrbt20/EASY-FIA/>.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Patti, G.J.; Yanes, O.; Siuzdak, G. Metabolomics: The apogee of the omics trilogy. *Nat. Rev. Mol. Cell Biol.* **2012**, *13*, 263–269. [[CrossRef](#)]
2. Nicholson, J.K.; Connelly, J.; Lindon, J.; Holmes, E. Metabonomics: A platform for studying drug toxicity and gene function. *Nat. Rev. Drug Discov.* **2002**, *1*, 153–161. [[CrossRef](#)]
3. Misra, B.B. New tools and resources in metabolomics: 2016–2017. *Electrophoresis* **2018**, *39*, 909–923. [[CrossRef](#)]
4. Goodacre, R.; Vaidyanathan, S.; Dunn, W.B.; Harrigan, G.G.; Kell, D.B. Metabolomics by numbers: Acquiring and understanding global metabolite data. *Trends Biotechnol.* **2004**, *22*, 245–252. [[CrossRef](#)]
5. Oliver, S.G.; Winson, M.K.; Kell, D.B.; Baganz, F. Systematic functional analysis of the yeast genome. *Trends Biotechnol.* **1998**, *16*, 373–378. [[CrossRef](#)]
6. Jin, Q.; Ma, R.C.W. Metabolomics in Diabetes and Diabetic Complications: Insights from Epidemiological Studies. *Cells* **2021**, *10*, 2832. [[CrossRef](#)]
7. Schmidt, D.R.; Patel, R.; Kirsch, D.G.; Lewis, C.A.; Vander Heiden, M.G.; Locasale, J.W. Metabolomics in cancer research and emerging applications in clinical oncology. *CA Cancer J. Clin.* **2021**, *71*, 333–358. [[CrossRef](#)]
8. Dubin, R.F.; Rhee, E.P. Proteomics and Metabolomics in Kidney Disease, including Insights into Etiology, Treatment, and Prevention. *Clin. J. Am. Soc. Nephrol.* **2020**, *15*, 404–411. [[CrossRef](#)]
9. Taylor, A.L.; Davis, D.E.; Codreanu, S.G.; Harrison, F.E.; Sherrod, S.D.; McLean, J.A. Targeted and Untargeted Mass Spectrometry Reveals the Impact of High-Fat Diet on Peripheral Amino Acid Regulation in a Mouse Model of Alzheimer’s Disease. *J. Proteome Res.* **2021**, *20*, 4405–4414. [[CrossRef](#)]
10. He, Z.; Liu, Z.; Gong, L. Biomarker identification and pathway analysis of rheumatoid arthritis based on metabolomics in combination with ingenuity pathway analysis. *Proteomics* **2021**, *21*, e2100037. [[CrossRef](#)]
11. Wei, Y.; Jasbi, P.; Shi, X.; Turner, C.; Hrovat, J.; Liu, L.; Rabena, Y.; Porter, P.; Gu, H. Early Breast Cancer Detection Using Untargeted and Targeted Metabolomics. *J. Proteome Res.* **2021**, *20*, 3124–3133. [[CrossRef](#)] [[PubMed](#)]
12. Moreau, R.; Clària, J.; Aguilar, F.; Fenaille, F.; Lozano, J.J.; Junot, C.; Colsch, B.; Caraceni, P.; Trebicka, J.; Pavesi, M.; et al. Blood metabolomics uncovers inflammation-associated mitochondrial dysfunction as a potential mechanism underlying ACLF. *J. Hepatol.* **2020**, *72*, 688–701. [[CrossRef](#)] [[PubMed](#)]
13. Cui, H.; Chen, Y.; Li, K.; Zhan, R.; Zhao, M.; Xu, Y.; Lin, Z.; Fu, Y.; He, Q.; Tang, P.C.; et al. Untargeted metabolomics identifies succinate as a biomarker and therapeutic target in aortic aneurysm and dissection. *Eur. Heart J.* **2021**, *42*, 4373–4385. [[CrossRef](#)] [[PubMed](#)]
14. Buentzel, J.; Klemp, H.G.; Kraetzner, R.; Schulz, M.; Dihazi, G.H.; Streit, F.; Bleckmann, A.; Menck, K.; Wlochowitz, D.; Binder, C. Metabolomic Profiling of Blood-Derived Microvesicles in Breast Cancer Patients. *Int. J. Mol. Sci.* **2021**, *22*, 13540. [[CrossRef](#)] [[PubMed](#)]
15. Li, C.; Zhang, X.; Wei, L.; Wei, D.; Chen, Z.; Cao, Z.; Zhao, Q.; Chang, C. Molecular biological methods in environmental engineering. *Water Environ. Res.* **2020**, *92*, 1786–1793. [[CrossRef](#)] [[PubMed](#)]
16. Shepherd, L.V.; Fraser, P.; Stewart, D. Metabolomics: A second-generation platform for crop and food analysis. *Bioanalysis* **2011**, *3*, 1143–1159. [[CrossRef](#)] [[PubMed](#)]
17. Yang, Y.; Saand, M.A.; Huang, L.; Abdelaal, W.B.; Zhang, J.; Wu, Y.; Li, J.; Sirohi, M.H.; Wang, F. Applications of Multi-Omics Technologies for Crop Improvement. *Front. Plant Sci.* **2021**, *12*, 563953. [[CrossRef](#)]
18. Tautenhahn, R.; Patti, G.J.; Rinehart, D.; Siuzdak, G. XCMS Online: A Web-Based Platform to Process Untargeted Metabolomic Data. *Anal. Chem.* **2012**, *84*, 5035–5039. [[CrossRef](#)]
19. Nanita, S.C.; Kaldon, L.G. Emerging flow injection mass spectrometry methods for high-throughput quantitative analysis. *Anal. Bioanal. Chem.* **2016**, *408*, 23–33. [[CrossRef](#)]
20. Cambiaghi, A.; Pinto, B.B.; Brunelli, L.; Falcetta, F.; Aletti, F.; Bendjelid, K.; Pastorelli, R.; Ferrario, M. Characterization of a metabolomic profile associated with responsiveness to therapy in the acute phase of septic shock. *Sci. Rep.* **2017**, *7*, 9748. [[CrossRef](#)]
21. La Nasa, J.; Nardella, F.; Andrei, L.; Giani, M.; Degano, I.; Colombini, M.P.; Ribechini, E. Profiling of high molecular weight esters by flow injection analysis-high resolution mass spectrometry for the characterization of raw and archaeological beeswax and resinous substances. *Talanta* **2020**, *212*, 120800. [[CrossRef](#)]
22. Büttenbender, S.L.; Carvalho, R.; Barbosa, F.D.S.; Ortiz, R.S.; Limberger, R.P.; Mendez, A.S.L. Fragmentation of Cannabinoids by Flow Injection Analysis Tandem Mass Spectrometry (FIA–MS/MS). *J. AOAC Int.* **2022**, *105*, 915–927. [[CrossRef](#)]
23. Vilà, M.; Bedmar, À.; Saurina, J.; Núñez, O.; Sentellas, S. High-Throughput Flow Injection Analysis–Mass Spectrometry (FIA–MS) Fingerprinting for the Authentication of Tea Application to the Detection of Teas Adulterated with Chicory. *Foods* **2022**, *11*, 2153. [[CrossRef](#)]
24. Guitton, Y.; Tremblay-Franco, M.; Le Corguillé, G.; Martin, J.-F.; Pétéra, M.; Roger-Mele, P.; Delabrière, A.; Goulitquer, S.; Monsoor, M.; Duperier, C.; et al. Create, run, share, publish, and reference your LC–MS, FIA–MS, GC–MS, and NMR data analysis workflows with the Workflow4Metabolomics 3.0 Galaxy online infrastructure for metabolomics. *Int. J. Biochem. Cell Biol.* **2017**, *93*, 89–101. [[CrossRef](#)]

25. Lewis, M.R.; Pearce, J.T.M.; Spagou, K.; Green, M.; Dona, A.C.; Yuen, A.H.Y.; David, M.; Berry, D.J.; Chappell, K.; der Sluis, V.H.-V.; et al. Development and Application of Ultra-Performance Liquid Chromatography-TOF MS for Precision Large Scale Urinary Metabolic Phenotyping. *Anal. Chem.* **2016**, *88*, 9004–9013. [[CrossRef](#)]
26. Naz, S.; dos Santos, D.C.M.; García, A.; Barbas, C. Analytical protocols based on LC–MS, GC–MS and CE–MS for nontargeted metabolomics of biological tissues. *Bioanalysis* **2014**, *6*, 1657–1677. [[CrossRef](#)]
27. Riquelme, G.; Zabalegui, N.; Marchi, P.; Jones, C.M.; Monge, M.E. A Python-Based Pipeline for Preprocessing LC–MS Data for Untargeted Metabolomics Workflows. *Metabolites* **2020**, *10*, 416. [[CrossRef](#)]
28. Sussulini, A. (Ed.) *Metabolomics: From Fundamentals to Clinical Applications*; Advances in Experimental Medicine and Biology; Springer International Publishing: Cham, Switzerland, 2017; Volume 965. [[CrossRef](#)]
29. The Human Serum Metabolome (HUSERMET) Consortium; Dunn, W.B.; Broadhurst, D.; Begley, P.; Zelena, E.; Francis-McIntyre, S.; Anderson, N.; Brown, M.; Knowles, J.D.; Halsall, A.; et al. Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. *Nat. Protoc.* **2011**, *6*, 1060–1083. [[CrossRef](#)]
30. Tsai, T.-H.; Wang, M.; Resson, H.W. Preprocessing and Analysis of LC-MS-Based Proteomic Data. In *Statistical Analysis in Proteomics*; Jung, K., Ed.; Methods in Molecular Biology; Springer: New York, NY, USA, 2016; Volume 1362, pp. 63–76. [[CrossRef](#)]
31. Beckonert, O.; Keun, H.C.; Ebbels, T.M.D.; Bundy, J.; Holmes, E.; Lindon, J.C.; Nicholson, J.K. Metabolic profiling, metabolomic and metabonomic procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts. *Nat. Protoc.* **2007**, *2*, 2692–2703. [[CrossRef](#)]
32. Pedrioli, P.G.A.; Eng, J.; Hubley, R.; Vogelzang, M.; Deutsch, E.; Raught, B.; Pratt, B.; Nilsson, E.; Angeletti, R.H.; Apweiler, R.; et al. A common open representation of mass spectrometry data and its application to proteomics research. *Nat. Biotechnol.* **2004**, *22*, 1459–1466. [[CrossRef](#)]
33. Giacomoni, F.; Le Corguillé, G.; Monsoor, M.; Landi, M.; Pericard, P.; Pétéra, M.; Duperier, C.; Tremblay-Franco, M.; Martin, J.-F.; Jacob, D.; et al. Workflow4Metabolomics: A collaborative research infrastructure for computational metabolomics. *Bioinformatics* **2015**, *31*, 1493–1495. [[CrossRef](#)] [[PubMed](#)]
34. Delabrière, A.; Hohenester, U.M.; Colsch, B.; Junot, C.; Fenaille, F.; Thévenot, E.A. proFIA: A data preprocessing workflow for flow injection analysis coupled to high-resolution mass spectrometry. *Bioinformatics* **2017**, *33*, 3767–3775. [[CrossRef](#)] [[PubMed](#)]
35. Fuhrer, T.; Zamboni, N. High-throughput discovery metabolomics. *Curr. Opin. Biotechnol.* **2015**, *31*, 73–78. [[CrossRef](#)] [[PubMed](#)]
36. Brunelli, L.; Davin, A.; Sestito, G.; Mimmi, M.C.; De Simone, G.; Balducci, C.; Pansarasa, O.; Forloni, G.; Cereda, C.; Pastorelli, R.; et al. Plasmatic Hippuric Acid as a Hallmark of Frailty in an Italian Cohort: The Mediation Effect of Fruit–Vegetable Intake. *J. Gerontol. Ser. A* **2021**, *76*, 2081–2089. [[CrossRef](#)] [[PubMed](#)]
37. Ferrario, M.; Cambiaghi, A.; Brunelli, L.; Giordano, S.; Caironi, P.; Guatteri, L.; Raimondi, F.; Gattinoni, L.; Latini, R.; Masson, S.; et al. Mortality prediction in patients with severe septic shock: A pilot study using a target metabolomics approach. *Sci. Rep.* **2016**, *6*, 20391. [[CrossRef](#)]
38. Boccard, J.; Rudaz, S. Harnessing the complexity of metabolomic data with chemometrics: Metabolomic Data Analysis with Chemometrics. *J. Chemom.* **2014**, *28*, 1–9. [[CrossRef](#)]
39. Fuhrer, T.; Heer, D.; Begemann, B.; Zamboni, N. High-Throughput, Accurate Mass Metabolome Profiling of Cellular Extracts by Flow Injection–Time-of-Flight Mass Spectrometry. *Anal. Chem.* **2011**, *83*, 7074–7080. [[CrossRef](#)]
40. Beuchel, C.; Kirsten, H.; Ceglarek, U.; Scholz, M. Metabolite-Investigator: An integrated user-friendly workflow for metabolomics multi-study analysis. *Bioinformatics* **2021**, *37*, 2218–2220. [[CrossRef](#)]
41. Beuchel, C.; Becker, S.; Dittrich, J.; Kirsten, H.; Toenjes, A.; Stumvoll, M.; Loeffler, M.; Thiele, H.; Beutner, F.; Thiery, J.; et al. Clinical and lifestyle related factors influencing whole blood metabolite levels—A comparative analysis of three large cohorts. *Mol. Metab.* **2019**, *29*, 76–85. [[CrossRef](#)]
42. Gatto, L.; Gibb, S.; Rainer, J. MSnbase, Efficient and Elegant R-Based Processing and Visualization of Raw Mass Spectrometry Data. *J. Proteome Res.* **2021**, *20*, 1063–1069. [[CrossRef](#)]
43. Mayer, B. (Ed.) *Bioinformatics for Omics Data: Methods and Protocols*; Methods in Molecular Biology; Humana Press: Totowa, NJ, USA, 2011. [[CrossRef](#)]
44. Gullo, F.; Ponti, G.; Tagarelli, A.; Tradigo, G.; Veltri, P. MSPtool: A Versatile Tool for Mass Spectrometry Data Preprocessing. In Proceedings of the 2008 21st IEEE International Symposium on Computer-Based Medical Systems, Jyväskylä, Finland, 17–19 June 2008; pp. 209–214. [[CrossRef](#)]
45. Sauve, A.C.; Speed, T.P. Normalization, baseline correction and alignment of high-throughput mass spectrometry data. *Proc. Gensips* **2004**, *4*, 1–4.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.