


Article

A Metabolomics Workflow for Analyzing Complex Biological Samples Using a Combined Method of Untargeted and Target-List Based Approaches

Thomas Züllig ^{1,†}, Martina Zandl-Lang ^{2,†}, Martin Trötzmüller ¹, Jürgen Hartler ³ , Barbara Plecko ² and Harald C. Köfeler ^{1,*}

¹ Core Facility Mass Spectrometry, Medical University of Graz, 8036 Graz, Austria;

thomas.zuellig@medunigraz.at (T.Z.); martin.troetzmueller@medunigraz.at (M.T.)

² Department of Paediatrics and Adolescent Medicine, Division of General Paediatrics, University Childrens' Hospital Graz, Medical University of Graz, 8036 Graz, Austria; martina.zandl@medunigraz.at (M.Z.-L.); barbara.plecko@medunigraz.at (B.P.)

³ Institute of Pharmaceutical Sciences, University of Graz, 8036 Graz, Austria; juergen.hartler@uni-graz.at

* Correspondence: harald.koefeler@medunigraz.at

† These authors have contributed equally to this work.

Received: 23 July 2020; Accepted: 20 August 2020; Published: 25 August 2020



Abstract: In the highly dynamic field of metabolomics, we have developed a method for the analysis of hydrophilic metabolites in various biological samples. Therefore, we used hydrophilic interaction chromatography (HILIC) for separation, combined with a high-resolution mass spectrometer (MS) with the aim of separating and analyzing a wide range of compounds. We used 41 reference standards with different chemical properties to develop an optimal chromatographic separation. MS analysis was performed with a set of pooled biological samples human cerebrospinal fluid (CSF), and human plasma. The raw data was processed in a first step with Compound Discoverer 3.1 (CD), a software tool for untargeted metabolomics with the aim to create a list of unknown compounds. In a second step, we combined the results obtained with our internally analyzed reference standard list to process the data along with the Lipid Data Analyzer 2.6 (LDA), a software tool for a targeted approach. In order to demonstrate the advantages of this combined target-list based and untargeted approach, we not only compared the relative standard deviation (%RSD) of the technical replicas of pooled plasma samples ($n = 5$) and pooled CSF samples ($n = 3$) with the results from CD, but also with XCMS Online, a well-known software tool for untargeted metabolomics studies. As a result of this study we could demonstrate with our HILIC-MS method that all standards could be either separated by chromatography, including isobaric leucine and isoleucine or with MS by different mass. We also showed that this combined approach benefits from improved precision compared to well-known metabolomics software tools such as CD and XCMS online. Within the pooled plasma samples processed by LDA 68% of the detected compounds had a %RSD of less than 25%, compared to CD and XCMS online (57% and 55%). The improvements of precision in the pooled CSF samples were even more pronounced, 83% had a %RSD of less than 25% compared to CD and XCMS online (28% and 8% compounds detected). Particularly for low concentration samples, this method showed a more precise peak area integration with its 3D algorithm and with the benefits of the LDAs graphical user interface for fast and easy manual curation of peak integration. The here-described method has the advantage that manual curation for larger batch measurements remains minimal due to the target list containing the information obtained by an untargeted approach.

Keywords: LC-MS; metabolomics; LDA; XCMS; plasma; CSF

1. Introduction

Metabolomics is defined as an “omics” approach in which the global set of small molecules (<1.5 kDa) in a distinct biological sample, e.g., tissue, blood, cerebrospinal fluid (CSF), cells or their compartments, are analyzed with the prospect of a better understanding of their biological function [1]. Metabolomic approaches to clinically relevant data include differential analysis between control groups and diseased groups, which can lead to the identification of novel biomarkers reflecting those abnormalities. Various potential biomarker candidates have been introduced in several clinical fields including respiratory [2,3], neurological [4], and kidney diseases [5,6], diabetes [7,8], cancer [9,10], and metabolic disorders [11]. Depending on the underlying hypothesis, different approaches are applied in metabolomics research. On the one hand there are established targeted workflows where samples are screened for a set of specific and pre-defined, metabolites (e.g., amino acids, biogenic amines, nucleobases, organic acids, sterols, organic phosphates, and vitamins) [12]. Untargeted workflows on the other hand, are non-hypothesis-driven, allowing deeper insights into complex biological samples and, therefore, may also lead to the discovery of new metabolic hypotheses as well as novel biomarkers. The challenge of untargeted metabolomics in larger sample cohorts is the time-consuming data processing, the annotation of unknown compounds and their verification [13–15]. In the field of mass spectrometry (MS) metabolomics, MS devices are usually coupled with an additional separation step—except for direct infusion techniques—to reduce sample complexity and to increase coverage and detection sensitivity [16]. The most used coupling technique is liquid chromatography (LC). This technique benefits from a variety of columns with different stationary phases, which allows separating a large number of distinct polar compounds. Frequently used columns are reversed phased (RP) C18 or C8 [17,18] and hydrophilic interaction chromatography (HILIC) [19]. The variety of these columns offers additional information depending on the composition of the stationary phase and the selected pH range [20]. Gas chromatography (GC) is especially used in breath metabolomics, as shown in several studies [21], benefitting from superior separation and less signal-to-noise ratio compared to liquid chromatography (LC). In addition, GC-MS is used beyond volatile organic compounds by applying different derivatization techniques [22,23]. Capillary electrophoresis coupled to MS (CE-MS) is a beneficial method for highly polar compounds of the metabolome, albeit not as popular as the chromatography techniques [24]. For processing of obtained MS data several software tools are available, e.g., XCMS [25], MZmine [26] or OpenMS [27]. As diverse as the software pool is, results can vary widely and data processing tools have to be adapted to individual biomedical applications [28,29]. Several methods for improving annotation and data quality are available using algorithms with false discovery rate [30], frameworks based on machine learning methods which can improve the quality of automated data processing workflows [31], or a sophisticated graphical interface for customizing parameters to reprocess data [26].

The aim of this work was to demonstrate selective separation of metabolites by performing a HILIC approach within a comprehensive metabolomics workflow. Therefore we used two types of software, Compound Discoverer 3.1 for untargeted metabolomics and Lipid Data Analyzer 2.6 for a target-list based approach to benefit from LDA 3D algorithm for more precise peak border confinement. We also compare the performance of this combined approach with typical untargeted metabolomics software tools (CD and XCMS online), for which we have used the precision of repeatedly measured pooled serum and pooled CSF samples. Since annotation of compounds for metabolomics studies can be complex, we have applied a level of trust that was introduced by the Metabolomics Standards Initiative (MSI).

2. Materials and Methods

2.1. Pooled Serum and Pooled CSF Samples

Pooled CSF and pooled plasma samples were obtained from six female patients suffering of Rett Syndrome and 13 (eligible for pooled CSF) and 25 (eligible for pooled plasma) healthy children,

respectively. All individuals were aged between 4 and 16 years. The pooled samples were used to develop and evaluate this metabolomics workflow. The study itself is an ongoing project and not the objective of this publication. The project is ethically approved (EK-number 31-162ex18/19).

2.2. Chemicals

Acetone (34850-2.5L-M; $\geq 99.8\%$) from Sigma-Aldrich (St. Louis, MO, USA), and acetonitrile chromasolv (34851-2.5L; $\geq 99.9\%$), and methanol chromasolv (34885-2.5L; $\geq 99.9\%$) from Honeywell (Charlotte, NC, USA).

2.3. Metabolite Standards

Dopamine HCl (56610-5G; $\geq 98.5\%$), L-ornithine monohydrochloride (75469-25G; $\geq 99.5\%$), L-prolin (81710-10G; $\geq 99\%$), taurine (86330-25G; $\geq 99\%$), creatinine (C4255-10G; $\geq 98\%$), D-(+)-glucose (G8270-100G; $\geq 99.5\%$), L-alanine (A7627-1g; $\geq 98\%$), L-asparagine (A0884-25G; $\geq 98\%$), L-aspartic acid (A8949-25G; $\geq 99\%$), L-citrulline (C7629-10MG; $\geq 98\%$), L-glutamic acid (G1251-100G; $\geq 99\%$), L-glutamine (G8540-25G; $\geq 99\%$), L-histidine (H8000-5G; $\geq 99\%$), L-isoleucine (I2752-1G; $\geq 98\%$), L-leucine (L8000-25g; $\geq 98\%$), L-lysine (L5501-1G; $\geq 98\%$), L-methionine sulfoxide (M1126-1G; NA), L-phenylalanine (P2126-100G; $\geq 98\%$), L-serine (S4500-1G; $\geq 99\%$), L-threonine (T8625-1G; $\geq 98\%$), L-tryptophan (T0254-5G; $\geq 98\%$), L-tyrosine (T3754-50G; $\geq 98\%$), L-valine (V0500-1G; $\geq 98\%$), cholic acid (C1129-25G; $\geq 98\%$), D-carnitine (544361-1G; $\geq 98\%$), decanoyl-L-carnitine (50637-10MG; $\geq 94\%$), Folic acid (F7876-1G; $\geq 97\%$), hippuric acid (112003-5G; $\geq 98\%$), L-carnosine (C9625-10MG; $\geq 99\%$), Palmitoyl-L-carnitine (91503-10MG; $\geq 95\%$), Riboflavin (R4500-5G; $\geq 98\%$), valeryl-L-carnitine (04265-10MG; $\geq 95\%$), adenine (A8626-1G; $\geq 99\%$), cytidine (C122106-1G; $\geq 99\%$), cytosine (C3506-1G; $\geq 99\%$), D-arginine (A2646-250MG; $\geq 98\%$), choline (C7017-10MG; $\geq 99\%$), a-tocopherol (T3251-5G; $\geq 96\%$), adenosine (A9251-1G; $\geq 99\%$), and methionine (M9625-5G; $\geq 98\%$) standards were purchased from Sigma-Aldrich (St. Louis, MO, USA), and Estradiol (E0950-000; NA) from Steraloids (Newport, RI, USA).

2.4. Sample Preparation and Storage

Sample preparation was performed according to Bruce et al. [32]. Briefly, proteins were precipitated by adding a 3:1 volume of ice-cold acetonitrile/methanol/acetone (1/1/1, *v/v/v*) to 50 μL pooled plasma or 100 μL pooled CSF samples and by vortexing these for 15 s. After a precipitation step at 4 $^{\circ}\text{C}$ for 60 min, samples were centrifuged at 12,419 relative centrifugal force (rcf) for 10 min (Hereaus Biofuge pico, Hanau, Germany). The resultant supernatants were aspirated into clean Eppendorf tubes. The samples were concentrated under a stream of dry nitrogen gas at room temperature (Techne, FSC400D, Staffordshire, UK). The resultant plasma and CSF sample pellets were re-suspended in acetonitrile/water (1/1, *v/v*) to 50 μL sample volume, respectively, and immediately stored at -80°C until further analysis. In order to evaluate data processing and annotation a mixture of metabolite standards (shown in Section 2.3) were measured in addition to biological samples. The final injection concentration was 10 μM solved in acetonitrile/water (1/1, *v/v*).

2.5. LC-MS/MS Method

For all samples a full-scan mass-spectrometric interrogation of each sample's small molecule was achieved by Dionex Ultimate XRS UHPLC (ultra-high performance liquid chromatography)-Orbitrap Velos Pro hybrid mass spectrometer (Thermo Fisher Scientific, Waltham, MA, USA). Separation was performed on an Acquity UPLC BEH Amide column (2.1 mm \times 150 mm, 1.7 μm) (Waters Corporation, Milford, USA), thermostated to 40 $^{\circ}\text{C}$. Mobile phase A was 97% ACN + 3% H_2O + 0.1 mM NH_4COOH + 0.16% HCOOH, Mobile phase B was H_2O + 0.1 mM NH_4COOH + 0.16% HCOOH, and autosampler wash solution ACN/ H_2O (1/1, *v/v*). The starting point of gradient elution was 5% mobile phase B and increased up to 30% over 30 min. Mobile phase B was reset to start conditions over a minute and re-equilibrated for 9 min before 2 μL of the next sample was re-injected. Flow rate was 200 $\mu\text{L min}^{-1}$ and samples were thermostated at 8 $^{\circ}\text{C}$ in the autosampler.

The Orbitrap Velos Pro operated in data dependent acquisition mode using a HESI II ion source. Full scan spectra from m/z 60 to 1600 were acquired in the Orbitrap with a resolution of 100,000 (m/z 400) in positive mode and the 10 most abundant ions of the full-scan spectrum were sequentially fragmented with CID (normalized collision energy, 50) and analyzed in the linear ion trap. Isolation width was 1.5, activation Q: 0.2; activation time: 10, and the centroided product spectra at normal scan rate were collected. The exclusion time was set to 11 s and as lock mass a polysiloxane with m/z 536.16537 was chosen.

The following source parameters were used: Source voltage: 4.5 kV, source temperature: 275 °C, sheath gas: 25 arbitrary units, aux gas: eight arbitrary units, sweep gas: zero arbitrary units, capillary temperature: 300 °C.

2.6. Data Processing

The data processing workflow was split into two parts (Figure 1). The compound discoverer 3.1 (CD) [28] was used for untargeted and novel feature detection and annotation with library scoring (Figure 1, left side). Previously generated target lists were merged and used to process data with the Lipid Data Analyzer 2.6 (LDA) software tool (Figure 1, right side).

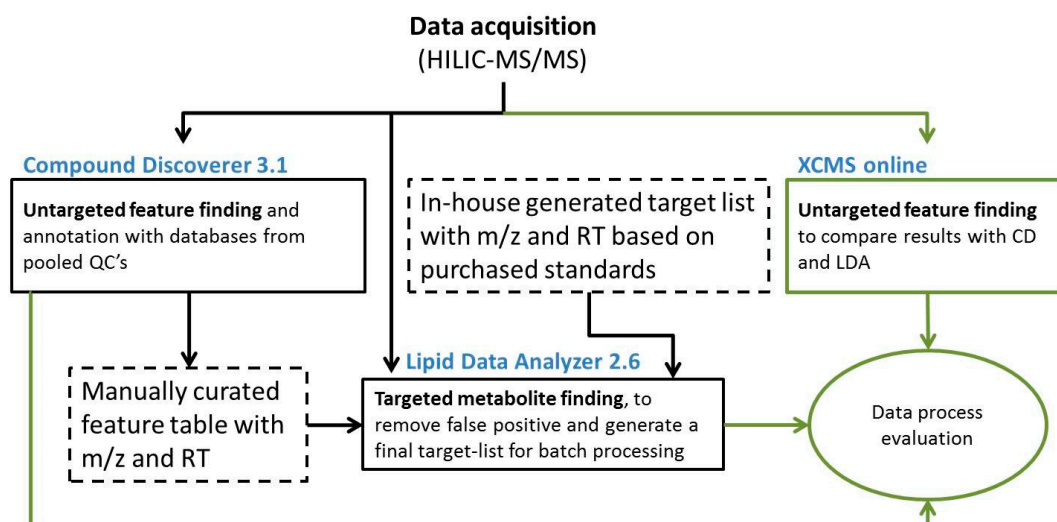


Figure 1. Flowchart of the combined untargeted and target-list based metabolomics workflow used in this study (black) with the extension of data process evaluation of the three software tools (green). Blue titles show software tools used, dashed boxes are target lists based on reference material and an untargeted approach, and boxes with solid lines are the processing steps. (RT: Retention Time; m/z : mass to charge ratio).

Raw files were processed with CD in the following steps: First, spectra were selected from raw data followed by retention time (RT) alignment with RT tolerance of 1 min and 5 ppm mass precision. Then unknown compounds were detected and grouped with RT tolerance of 0.7 min and 5 ppm mass deviation. In a second step missing values were filled and the compounds were annotated with different types of databases. mzCloud was used to annotate compounds on MS/MS level with a mass tolerance of 10 ppm. Chempidder, which integrates BioCyc [33], Human Metabolome Database [34], and KEGG database [35], was used to annotate features based on exact mass with a mass tolerance of 5 ppm as well as the CD internal database (an endogenous metabolites database of 4400 compounds). The data set obtained was used to generate a list of potential targets taking into account the level of confidence in annotations implemented by the Metabolomics Standards Initiative (MSI) of the Metabolomics Society (Table 1) [14,36]. In detail, the confidence level of annotation was created as follows: In the first step, all hits from the mzCloud—MS/MS scoring database—greater than 80% score were exported. Due to known RT and better fragmentation spectra, the samples found in the list of

purchased standards belong to confidence level 1, compounds detected in addition to the purchased standards to confidence level 2. Compounds detected on the basis of the monoisotopic mass in Chemspider or the internal databases were defined as confidence level 3. Unknown compounds of the analyzed CSF or plasma samples were classified as confidence level 4. The list of detected compounds included exact molecular mass, RT, compound names, if known, and were completed with purchased standards not found by the untargeted approach.

Table 1. Requirements as promoted by the Metabolomics Society [14,36] and their implementation into this study.

Level of Confidence	Description	Data Requirements in This Study	Certainty
4	Unknown feature	Recognized feature with CD and detected with LDA	-
3	Possible structure matched on a single information e.g., database	Matched to the exact mass (5 ppm) in an MS 1 database (Chemspider)	BAD
2	At least two sources of information to match 2D structure. e.g., exact mass and MS/MS score	Matching exact mass (10 ppm) with a fragmentation score over 80 of the mzcloud database	GOOD
1	Confidently matched 2D structure with two technics e.g., MS/MS and RT	Exact mass (5 ppm) and experimental MS/MS information and RT information of pure standard	VERY GOOD
0	Compound with determination of the 3D structure	Not possible with this method	BEST

The list was manually curated before it was used as a target list for processing with LDA software tool [37,38]. For data processing we disabled isotopic quantitation as CD already used isotopic pattern matching. After processing, extracted ion chromatograms (EICs) for all compounds were exported and corrected manually if necessary. False positive, noise, or chromatographically not resolved compounds were removed from the final target list. To compare results data were independently processed with the XCMS Online platform. Therefore, raw data sets were uploaded to XCMS Online and a single group job was created to analyze the same pooled plasma and pooled CSF samples as already described with CD and LDA. We selected the predefined workflow settings for Orbitrap (Table S2). The CentWave algorithm was used for features detection, the OBI-warp algorithm for RT correction, CAMERA for annotation of isotopes and adducts, and the METLIN database was used for MS/MS score matching to identify compounds.

3. Results

3.1. Chromatographic Separation of Metabolites

For development of chromatographic separation we used a set of 41 purchased reference standards, representing a selection of different compound classes found in the human metabolome [39] i.e., amino acids ($n = 21$), biogenic amines ($n = 5$), sugars ($n = 1$), acyl carnitines ($n = 4$), nucleobases ($n = 4$), organic acids ($n = 1$), sterols ($n = 2$), and vitamins ($n = 3$). Since ionization and chromatographic separation strongly depend on functional groups, we observed that these compounds are a good representative set for testing the LC separation capability of the used HILIC BEH amide column in combination with eluents and gradient elution. Due to the different chemical properties of different metabolites, separation of all compounds is challenging. Total ion chromatogram (TIC) (Figure 2, top) and extracted ion chromatograms (EIC) of each reference standard (Figure 2, bottom) demonstrated good separation of the purchased reference material by applying the setup as explained above (Section 2). All standards were detectable and separated either by chromatography, including isobaric compounds such as leucine and isoleucine, or by their mass-to-charge (m/z) ratio in the mass spectrometric dimension.

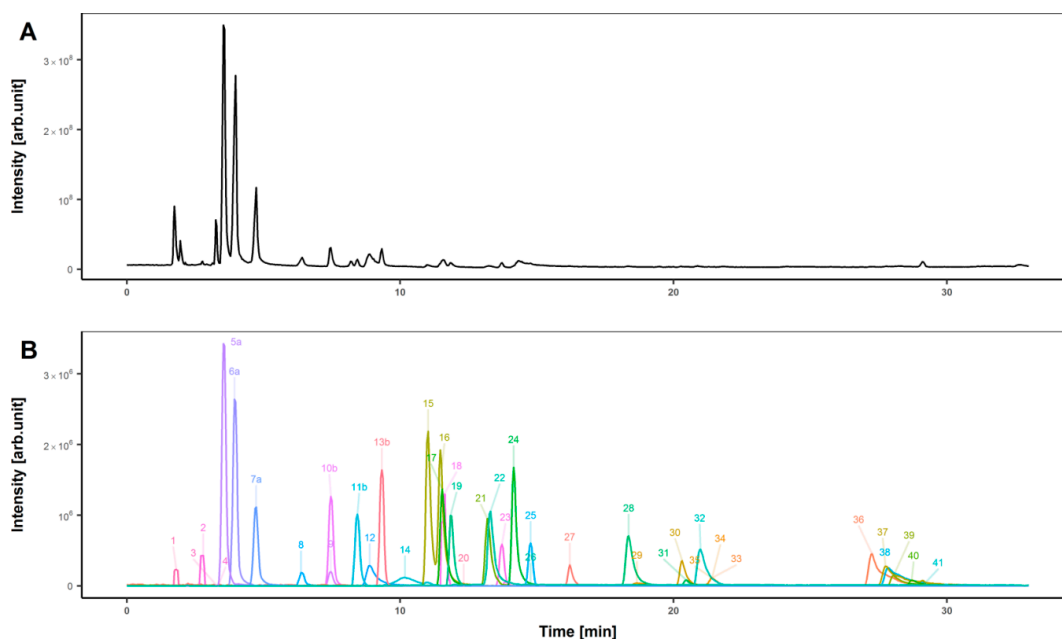


Figure 2. Total ion chromatogram (A) of measured reference material with Orbitrap Velos Pro (Thermo). Separation was performed with BEH amide (2.1 mm × 150 mm particle size 1.7 μm) at 40 °C. (B) Extracted ion chromatograms (EICs) of identified purchased standards. All standards were injected at a concentration of 10 μM. Due to the different signal abundances, the intensities of six compounds were manually reduced for better illustration (a: with factor 50, b: with factor 5). 1, a-tocopherol; 2, hippuric acid; 3, estradiol; 4, cholic acid; 5a, palmitoylcarnitine; 6a, decanoylcarnitine; 7a, valerylcarnitine; 8, choline; 9, adenine; 10b, adenosine; 11b, creatinine; 12, carnitine; 13b, riboflavin; 14, dopamine; 15, leucine; 16, isoleucine; 17, phenylalanine; 18, cytosine; 19, tryptophan; 20, folic acid; 21, methionine; 22, valine; 23, cytidine; 24, proline; 25, taurine; 26, tyrosine; 27, alanine; 28, threonine; 29, glutamate; 30, glutamine; 31, serine; 32, methionine sulfoxide; 33, asparagine; 34, aspartate; 35, citrulline; 36, arginine; 37, histidine; 38, carnosine; 39, lysine; 40, ornithine; 41, glucose.

3.2. Precision Evaluation of LDA Compared to CD and XCMS Online

A key element in finding differences in biological samples is efficient and reliable data processing. By measuring pooled samples, which are technical replicas, we would expect the relative standard deviation to be ideally zero. Due to sample stability, instrument drift, or ionization variability, and other batch effects, this is generally not the case. To demonstrate the analytical concept of an untargeted metabolomics approach using CD and transforming it into a target list-based metabolomics approach using LDA we analyzed pooled plasma ($n = 5$) and CSF ($n = 3$) samples. These pooled samples were repeatedly measured—QC n , extracted blank n , 10 biological samples, QC $n + 1$, extracted blank $n + 1$, followed by 10 biological samples, and so on—between a longer sequence to simulate the acquisition of a typical larger metabolomics study with a QC-based normalization strategy for regression of systematic errors [40,41]. This takes into account sample stability, instrument drifts, and other errors. For evaluation of the difference between the combined approach of the LDA and CD, and XCMS online, we used the relative standard deviation in percent of the detected compounds. Features that were only detectable in XCMS online or with CD and could not be confirmed with the LDA were not evaluated (Figure 3, Table S1). Within the pooled plasma samples XCMS online identified 2510 and CD identified 1880 features. Within the pooled CSF samples XCMS online identified 1446 features and CD detected 1005 features. After adduct filtering and isotope pattern recognition, CD detected 1119 compounds in pooled plasma samples and 495 in pooled CSF samples. These results were used in combination with the list of measured purchased standards to process the raw data with LDA. After reprocessing, we identified 580 compounds in pooled plasma samples and 94 compounds in pooled CSF samples. The lower number of compounds in the CSF samples is partly due to the limitation of the sample

volume. Concentration of the CSF sample in less injection solvent may result in higher signal intensity. Variation of the %RSD between the combined approach using both CD and LDA compared to CD and XCMS strongly depends on the raw data quality. The difference between pooled CSF samples with signals of lower intensity shows a stronger effect than pooled plasma samples with overall better signal intensity (Figure 3). The results of the pooled plasma samples with a %RSD of less than 25% showed that 68% of the compounds processed with LDA, 57% of the compounds processed with XCMS Online and 55% of the compounds processed with CD fall into this category. The differences between CSF samples with less than 25% %RSD were even greater. In the processed data with LDA it was 83%, with XCMS Online, 28%, and with CD, 8% of the compounds with a %RSD of less than 25%.

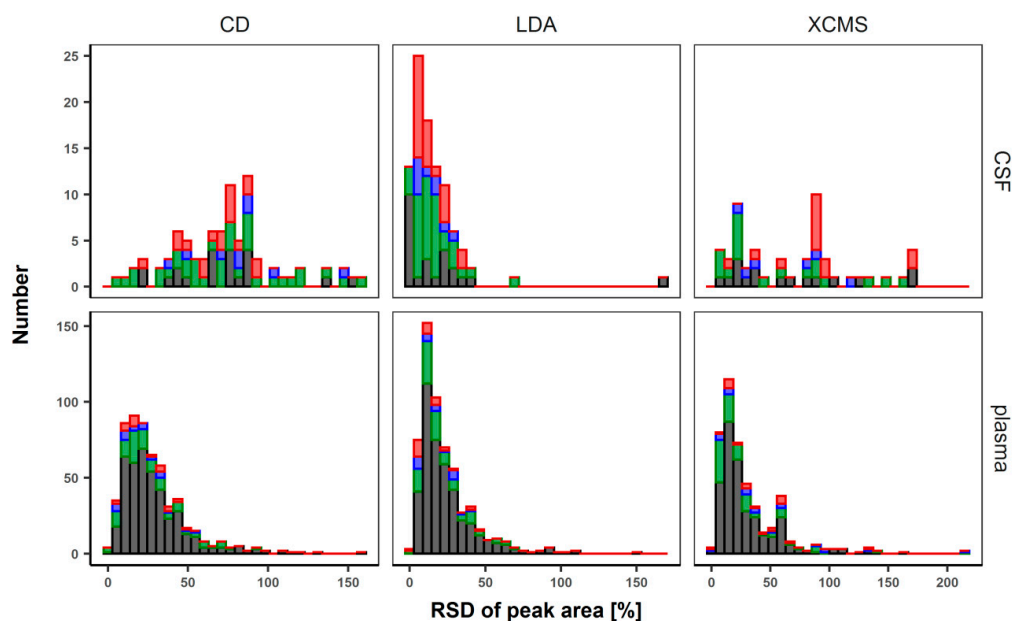


Figure 3. Evaluation of selected and processed data with XCMS online, Compound Discoverer 3.1 (CD) and Lipid Data Analyzer 2.6 (LDA). Relative standard deviations (%RSD) of the peak areas of pooled plasma samples ($n = 5$) and pooled CSF samples ($n = 3$) are shown as a histogram. The color code indicates the confidence level to which the detected compound belongs: Red: group 1, blue: group 2, green: group 3, and black: group 4.

3.3. Metabolite Validation

In many MS based metabolomics studies, the annotation of features can be misleading because not all information is equally valid. For this reason, the 2005 founded Metabolomics Standards Initiative (MSI) has the aim of establishing reporting standards which are agreed upon by the community. One point was the definition of the quality level for the identification of metabolites [42]. Blaženovic et al. published an extended version with one additional category with confidence levels from 0 to 4 [14]. These definitions clarify how trustworthy the annotation is. The following example shows how these criteria affect the annotation of the compound pipecolic acid (m/z , 130.0863, M+H) at different confidence levels. Two-hundred nineteen compounds were revealed with a METLIN database search of pipecolic acids precursor mass with a mass tolerance of 5 ppm, which corresponds to a confidence level of 3. With the additional MS/MS spectral information, four compounds were found, two were L- and D-pipecolic acid and the other two were cycloleucine and nipecotinic acid, which represents confidence level 2. Additional reference material would reduce the potential compounds to L- and D-pipecolic acid as the isomer structure could not be elucidated with this method and would be in the second-best category of confidence level 1. This example is limited by the METLIN-database search and the numbers beyond METLIN annotated metabolites can be even greater.

In the pooled plasma 30 compounds were detected with confidence level 1 and, in the pooled CSF samples, there were 23 compounds with level 1 confidence. For these, standards were purchased and RT and fragmentation patterns were known. At confidence level 2, 26 compounds were detected in the pooled plasma and nine in the pooled CSF samples. For them, we used the fragmentation patterns annotated in mzcloud. The rest of the detectable compounds was within confidence level 3 or 4 (plasma 522, CSF 62 compounds). Monoisotopic masses found in the databases explained differences between levels 3 and 4. Despite the indistinct annotation quality of compounds with confidence levels 3 or 4, differential studies are further needed for analysis and comparison, e.g., if these compounds are up- or down-regulated.

4. Discussion

In this study, we developed a comprehensive method to separate and detect metabolites with high-resolution mass spectrometry in a data-dependent full-scan acquisition method. We compared and evaluated three different data processing workflows, two MS-based metabolomics software tools (CD, XCMS online) and one combination of a MS-based metabolomics software with a target list-based software tool (CD/LDA). A variety of software tools exist in the field of untargeted metabolomics. Misra et al. have compiled a comprehensive list [43]. One of the most common software tools is MS-Dial [44], which, in addition to the DDA mode in combination with liquid chromatography, also enables data-independent acquisition (DIA) or sequential window acquisition of all theoretical fragment ions (SWATH) and provides an MS/MS scoring database to match experimental spectra against spectral libraries, such as MassBank [45] or LipidBlast [42]. MZmine 2 [26] is another frequently used open source software tool for untargeted metabolomics. It offers a graphical user interface (GUI) and has built-in multiple previews to control and modify setting for peak integration. The compound identification is based on a user-defined database with m/z values with retention times and can be connected directly to several online resources, e.g., PubChem [33], HMDB [34], or METLIN [35,36]. In our study, we used Compound Discoverer 3.0 (Thermo), a commercial software tool, XCMS online, and Lipid Data Analyzer 2.6, both open-source tools. Compound Discoverer 3.0 offers a user-friendly interface and has built-in statistical analysis, e.g., PCA, Volcano plot, S-plot, hierarchical cluster analysis, and also univariate statistics. The disadvantage is that commercial software is expensive and limited to processing vendor-specific files. XCMS online offers a platform for uploading and processing your datasets online, is manufacturer-independent, and has advanced statistical tests and path analyses. Compared to the other two non-targeted software tools, LDA is a target-list based tool. Peak integration is based on a unique 3D algorithm which showed in the lipidomics field superior accuracy and sensitivity compared to mzMine2 and LIMSA [37]. Li et al. compared Compound Discoverer 2.0 [38], MS-Dial, MZmine 2 and XCMS with a standard mix of 970 compounds analyzed with a Q exactive HF [40]. They used Tracefinder 4.0 (Thermo) with a targeted approach as a benchmark to compare the results against the non-targeted workflows. They showed that all untargeted software tools had similar overall performance in recognizing features. XCMS had the highest identification rate, but MZmine 2 had slightly better performance in terms of accuracy and quantification. Compared to the benchmark method with the targeted approach, all untargeted analyzes lack accuracy and quantification. They summarized that impaired quantification of features is mainly due to incorrect EIC peak detection. In order to reduce false positives and increase the number of features, a combined approach with XCMS and MZmine2 was successfully tested [46]. This could also be an option to use more than one non-targeted metabolomics software tool for increasing the number of features to be used as a target-list for processing the data with LDA to increase precision. The question is whether it is worth spending the additional time to find a few additional features. In our opinion, there is no best metabolomics software solution. In this study we benefit from the combination of the strengths of complementary types of software. Another general problem of LC-MS instruments are batch and time drift effects, especially when data acquisition takes a long time, as instruments tend to change in response over time [47,48]. Nowadays, various methods for correcting these effects are known, e.g.,

the use of multiple internal standards associated with higher costs, or use of pooled QCs, which are measured repeatedly during the acquisition sequence [41,49,50]. However, if algorithms are optimized in respect to find as many features as possible and thereby, the integration of the peak area is difficult to adapt, it is not possible to curate larger data sets manually. This leads to an increased variance, as shown in Figure 3 and affects the identification and removal of outliers, which is needed for proper batch corrections. The here-presented combined approach demonstrates many advantages in peak area integration and data processing, thereby leading to less variance. This reduction offers benefits for further data processing as well as for statistical tests, but has slight disadvantages in the detection and identification of novel compounds, which are caused by the dilution of the pooled QCs in comparison to individual phenotype samples.

This combined approach will be further applied on plasma and CSF samples from a larger ongoing study with control groups of healthy subjects and study groups with neurological pediatric disorders.

In conclusion, this two-step software approach combines the strength of a broad range of compounds detectable by untargeted metabolomics with the highly reliable peak area integration inherent to LDA, which, at the end of the day, results in more accurate up- and down-regulation patterns of expected but also unexpected individual metabolites.

Supplementary Materials: The following are available online at <http://www.mdpi.com/2218-1989/10/9/342/s1>, Table S1. List of detected compounds with the combined approach of Lipid Data Analyzer and Compound Discoverer 3.1; Table S2. List of parameters used for data processing with XCMS online (version of XCMS online 2.7.2, XCMS 1.47.3, CAMERA 1.34.0); Figure S1. Increase of statistical power with the LDA approach. Compounds with confident level 1 found in CSF and Plasma samples with all three data processing workflows: Compound Discoverer 3.0 (CD), XCMS, and the combined approach with CD and Lipid Data analyzer 2.6 (LDA). ns: $p > 0.05$, *: $p \leq 0.05$, **: $p \leq 0.01$. ***: $p \leq 0.001$, ****: $p \leq 0.0001$.

Author Contributions: T.Z. and M.Z.-L. are shared first authors. They developed the theory and performed experiments as well as data processing with consultation of M.T. J.H. helped with all LDA software specific questions. T.Z. wrote the manuscript in consultation with M.Z.-L. H.C.K. and B.P. supervised the project and edited the manuscript. All authors provided critical feedback and contributed to the final shape of the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Austrian Federal Ministry of Education, Science, and Research grant number BMWFW-10.420/0005-WF/V/3c/2017.

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

References

1. Johnson, C.H.; Ivanisevic, J.; Siuzdak, G. Metabolomics: Beyond biomarkers and towards mechanisms. *Nat. Rev. Mol. Cell Biol.* **2016**, *17*, 451–459. [[CrossRef](#)] [[PubMed](#)]
2. Reinke, S.N.; Gallart-Ayala, H.; Gómez, C.; Checa, A.; Fauland, A.; Naz, S.; Kamleh, M.A.; Djukanovic, R.; Hinks, T.S.; Wheelock, C.E. Metabolomics analysis identifies different metabolotypes of asthma severity. *Eur. Respir. J.* **2017**, *49*, 1601740. [[CrossRef](#)] [[PubMed](#)]
3. Martín-Vicente, M.; González-Riaño, C.; Barbas, C.; Jiménez-Sousa, M.Á.; Brochado-Kith, O.; Resino, S.; Martínez, I. Metabolic changes during respiratory syncytial virus infection of epithelial cells. *PLoS ONE* **2020**, *15*, e0230844. [[CrossRef](#)]
4. Zhang, A.-H.; Sun, H.; Wang, X.-J. Recent advances in metabolomics in neurological disease, and future perspectives. *Anal. Bioanal. Chem.* **2013**, *405*, 8143–8150. [[CrossRef](#)] [[PubMed](#)]
5. Hocher, B.; Adamski, J. Metabolomics for clinical use and research in chronic kidney disease. *Nat. Rev. Nephrol.* **2017**, *13*, 269–284. [[CrossRef](#)] [[PubMed](#)]
6. Chen, Z.; Coy, S.L.; Pannkuk, E.L.; Laiakis, E.C.; Fornace, A.J.; Vouros, P. Differential Mobility Spectrometry-Mass Spectrometry (DMS-MS) in Radiation Biodosimetry: Rapid and High-Throughput Quantitation of Multiple Radiation Biomarkers in Nonhuman Primate Urine. *J. Am. Soc. Mass Spectrom.* **2018**, *29*, 1650–1664. [[CrossRef](#)] [[PubMed](#)]
7. Wang, T.J.; Larson, M.G.; Vasan, R.S.; Cheng, S.; Rhee, E.P.; McCabe, E.; Lewis, G.D.; Fox, C.S.; Jacques, P.F.; Fernandez, C.; et al. Metabolite profiles and the risk of developing diabetes. *Obes. Metab.* **2011**, *8*, 72. [[CrossRef](#)]

8. O'Connor, S.; Greffard, K.; Leclercq, M.; Julien, P.; Weisnagel, S.J.; Gagnon, C.; Droit, A.; Bilodeau, J.F.; Rudkowska, I. Increased Dairy Product Intake Alters Serum Metabolite Profiles in Subjects at Risk of Developing Type 2 Diabetes. *Mol. Nutr. Food Res.* **2019**, *63*, e1900126. [[CrossRef](#)] [[PubMed](#)]
9. Spratlin, J.L.; Serkova, N.J.; Eckhardt, S.G. Clinical Applications of Metabolomics in Oncology: A Review. *Clin. Cancer Res.* **2009**, *15*, 431–440. [[CrossRef](#)] [[PubMed](#)]
10. Bobrovnikova-Marjon, E.; Hurov, J.B. Targeting Metabolic Changes in Cancer: Novel Therapeutic Approaches. *Annu. Rev. Med.* **2014**, *65*, 157–170. [[CrossRef](#)]
11. Rhee, E.P.; Gerszten, R.E. Metabolomics and Cardiovascular Biomarker Discovery. *Clin. Chem.* **2012**, *58*, 139–147. [[CrossRef](#)] [[PubMed](#)]
12. Yoon, H.-R. Screening newborns for metabolic disorders based on targeted metabolomics using tandem mass spectrometry. *Ann. Pediatr. Endocrinol. Metab.* **2015**, *20*, 119–124. [[CrossRef](#)] [[PubMed](#)]
13. Blaženović, I.; Kind, T.; Sa, M.R.; Ji, J.; Vaniya, A.; Wancewicz, B.; Roberts, B.S.; Torbašinović, H.; Lee, T.; Mehta, S.S.; et al. Structure Annotation of All Mass Spectra in Untargeted Metabolomics. *Anal. Chem.* **2019**, *91*, 2155–2162. [[CrossRef](#)]
14. Blaženović, I.; Kind, T.; Ji, J.; Fiehn, O. Software Tools and Approaches for Compound Identification of LC-MS/MS Data in Metabolomics. *Metabolites* **2018**, *8*, 31. [[CrossRef](#)] [[PubMed](#)]
15. Schrimpe-Rutledge, A.C.; Codreanu, S.G.; Sherrod, S.D.; McLean, J.A. Untargeted Metabolomics Strategies—Challenges and Emerging Directions. *J. Am. Soc. Mass Spectrom.* **2016**, *27*, 1897–1905. [[CrossRef](#)] [[PubMed](#)]
16. Kuehnbaum, N.L.; Britz-McKibbin, P. New Advances in Separation Science for Metabolomics: Resolving Chemical Diversity in a Post-Genomic Era. *Chem. Rev.* **2013**, *113*, 2437–2468. [[CrossRef](#)] [[PubMed](#)]
17. Zhang, J.; Bowers, J.; Liu, L.; Wei, S.; Gowda, G.A.N.; Hammoud, Z.; Raftery, D. Esophageal Cancer Metabolite Biomarkers Detected by LC-MS and NMR Methods. *PLoS ONE* **2012**, *7*, e30181. [[CrossRef](#)]
18. Willmann, L.; Schlimpert, M.; Hirschfeld, M.; Erbes, T.; Neubauer, H.; Stickeler, E.; Kammerer, B. Alterations of the exo- and endometabolite profiles in breast cancer cell lines: A mass spectrometry-based metabolomics approach. *Anal. Chim. Acta* **2016**, *925*, 34–42. [[CrossRef](#)]
19. Elmsjö, A.; Haglöf, J.; Engskog, M.K.R.; Erngren, I.; Nestor, M.; Arvidsson, T.; Pettersson, C. Method selectivity evaluation using the co-feature ratio in LC/MS metabolomics: Comparison of HILIC stationary phase performance for the analysis of plasma, urine and cell extracts. *J. Chromatogr. A* **2018**, *1568*, 49–56. [[CrossRef](#)]
20. Contrepois, K.; Jiang, L.; Snyder, M. Optimized Analytical Procedures for the Untargeted Metabolomic Profiling of Human Urine and Plasma by Combining Hydrophilic Interaction (HILIC) and Reverse-Phase Liquid Chromatography (RPLC)-Mass Spectrometry. *Mol. Cell. Proteom.* **2015**, *14*, 1684–1695. [[CrossRef](#)]
21. Fiehn, O. Extending the breadth of metabolite profiling by gas chromatography coupled to mass spectrometry. *TrAC Trends Anal. Chem.* **2008**, *27*, 261–269. [[CrossRef](#)] [[PubMed](#)]
22. Migné, C.; Durand, S.; Pujos-Guillot, E. Exploratory GC/MS-Based Metabolomics of Body Fluids. In *Clinical Metabolomics*; Humana Press: New York, NY, USA, 2018; pp. 239–246. [[CrossRef](#)]
23. Beale, D.; Pinu, F.; Kouremenos, K.A.; Poojary, M.M.; Narayana, V.K.; Boughton, B.A.; Kanojia, K.; Dayalan, S.; Jones, O.A.H.; Dias, D.A. Review of recent developments in GC-MS approaches to metabolomics-based research. *Metabolomics* **2018**, *14*, 152. [[CrossRef](#)] [[PubMed](#)]
24. Barbas, C.; Moraes, E.P.; Villaseñor, A. Capillary electrophoresis as a metabolomics tool for non-targeted fingerprinting of biological samples. *J. Pharm. Biomed. Anal.* **2011**, *55*, 823–831. [[CrossRef](#)]
25. Mahieu, N.G.; Genenbacher, J.L.; Patti, G.J. A roadmap for the XCMS family of software solutions in metabolomics. *Curr. Opin. Chem. Biol.* **2016**, *30*, 87–93. [[CrossRef](#)]
26. Pluskal, T.; Castillo, S.; Villar-Briones, A.; Orešič, M. MZmine2: Modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinform.* **2010**, *11*, 395. [[CrossRef](#)]
27. Pfeuffer, J.; Sachsenberg, T.; Alka, O.; Walzer, M.; Fillbrunn, A.; Nilse, L.; Schilling, O.; Reinert, K.; Kohlbacher, O. Openms—A Platform for Reproducible Analysis of Mass Spectrometry Data. *J. Biotechnol.* **2017**, *261*, 142–148. [[CrossRef](#)]
28. Hao, L.; Wang, J.; Page, D.; Asthana, S.; Zetterberg, H.; Carlsson, C.; Okonkwo, O.C.; Li, L. Comparative Evaluation of Ms-Based Metabolomics Software and Its Application to Preclinical Alzheimer's Disease. *Sci. Rep.* **2018**, *8*, 1–10.

29. Myers, O.D.; Sumner, S.J.; Li, S.; Barnes, S.; Du, X. Detailed Investigation and Comparison of the XCMS and MZmine2 Chromatogram Construction and Chromatographic Peak Detection Methods for Preprocessing Mass Spectrometry Metabolomics Data. *Anal. Chem.* **2017**, *89*, 8689–8695. [[CrossRef](#)]
30. Matsuda, F.; Shinbo, Y.; Oikawa, A.; Hirai, M.Y.; Fiehn, O.; Kanaya, S.; Saito, K. Assessment of Metabolome Annotation Quality: A Method for Evaluating the False Discovery Rate of Elemental Composition Searches. *PLoS ONE* **2009**, *4*, e7490. [[CrossRef](#)]
31. Zhou, Z.; Tu, J.; Zhu, Z. Advancing the large-scale CCS database for metabolomics and lipidomics at the machine-learning era. *Curr. Opin. Chem. Biol.* **2017**, *42*, 34–41. [[CrossRef](#)]
32. Bruce, S.J.; Jönsson, P.; Antti, H.; Cloarec, O.; Trygg, J.; Marklund, S.L.; Moritz, T. Evaluation of a protocol for metabolic profiling studies on human blood plasma by combined ultra-performance liquid chromatography/mass spectrometry: From extraction to data analysis. *Anal. Biochem.* **2008**, *372*, 237–249. [[CrossRef](#)] [[PubMed](#)]
33. Wang, Y.; Xiao, J.; Suzek, T.O.; Zhang, J.; Wang, J.; Bryant, S.H. PubChem: A public information system for analyzing bioactivities of small molecules. *Nucleic Acids Res.* **2009**, *37*, W623–W633. [[CrossRef](#)] [[PubMed](#)]
34. Wishart, D.S.; Feunang, Y.D.; Marcu, A.; Guo, A.C.; Liang, K.; Vázquez-Fresno, R.; Sajed, T.; Johnson, D.; Li, C.; Karu, N.; et al. HMDB 4.0: The human metabolome database for 2018. *Nucleic Acids Res.* **2018**, *46*, D608–D617. [[CrossRef](#)] [[PubMed](#)]
35. Smith, C.A.; O'Maille, G.; Want, E.J.; Qin, C.; Trauger, S.A.; Brandon, T.R.; Custodio, D.E.; Abagyan, R.; Siuzdak, G. METLIN: A metabolite mass spectral database. *Ther. Drug Monit.* **2005**, *27*, 747–751. [[CrossRef](#)]
36. Xue, J.; Domingo-Almenara, X.; Guijas, C.; Palermo, A.; Rinschen, M.M.; Isbell, J.; Benton, H.P.; Siuzdak, G. Enhanced in-Source Fragmentation Annotation Enables Novel Data Independent Acquisition and Autonomous METLIN Molecular Identification. *Anal. Chem.* **2020**, *92*, 6051–6059. [[CrossRef](#)]
37. Hartler, J.; Trötz Müller, M.; Chittraju, C.; Spener, F.; Köfeler, H.; Thallinger, G.G. Lipid Data Analyzer: Unattended identification and quantitation of lipids in LC-MS data. *Bioinformatics* **2010**, *27*, 572–577. [[CrossRef](#)]
38. Najdekr, L.; Friedecký, D.; Tautenhahn, R.; Pluskal, T.; Wang, J.; Huang, Y.; Adam, T. Influence of Mass Resolving Power in Orbital Ion-Trap Mass Spectrometry-Based Metabolomics. *Anal. Chem.* **2016**, *88*, 11429–11435. [[CrossRef](#)]
39. Psychogios, N.; Hau, D.D.; Peng, J.; Guo, A.C.; Mandal, R.; Bouatra, S.; Sinelnikov, I.; Krishnamurthy, R.; Eisner, R.; Gautam, B.; et al. The Human Serum Metabolome. *PLoS ONE* **2011**, *6*, e16957. [[CrossRef](#)]
40. Li, B.; Tang, J.; Yang, Q.; Cui, X.; Li, S.; Chen, S.; Cao, Q.; Xue, W.; Chen, N.; Zhu, F. Performance Evaluation and Online Realization of Data-driven Normalization Methods Used in LC/MS based Untargeted Metabolomics Analysis. *Sci. Rep.* **2016**, *6*, 38881. [[CrossRef](#)]
41. Fan, S.; Kind, T.; Cajka, T.; Hazen, S.L.; Tang, W.H.W.; Kaddurah-Daouk, R.; Irvin, M.R.; Arnett, D.K.; Barupal, D.K.; Fiehn, O. Systematic Error Removal Using Random Forest for Normalizing Large-Scale Untargeted Lipidomics Data. *Anal. Chem.* **2019**, *91*, 3590–3596. [[CrossRef](#)]
42. Cajka, T.; Fiehn, O. LC–MS–Based Lipidomics and Automated Identification of Lipids Using the LipidBlast In-Silico MS/MS Library. In *Lipidomics*; Humana Press: New York, NY, USA, 2017; pp. 149–170. [[CrossRef](#)]
43. Misra, B.B.; Mohapatra, S. Tools and resources for metabolomics research community: A 2017–2018 update. *Electrophoresis* **2018**, *40*, 227–246. [[CrossRef](#)] [[PubMed](#)]
44. Tsugawa, H.; Cajka, T.; Kind, T.; Ma, Y.; Higgins, B.T.; Ikeda, K.; Kanazawa, M.; VanderGheynst, J.; Fiehn, O.; Arita, M. MS-DIAL: Data-independent MS/MS deconvolution for comprehensive metabolome analysis. *Nat. Methods* **2015**, *12*, 523–526. [[CrossRef](#)] [[PubMed](#)]
45. Horai, H.; Arita, M.; Kanaya, S.; Nihei, Y.; Ikeda, T.; Suwa, K.; Ojima, Y.; Tanaka, K.; Tanaka, S.; Aoshima, K.; et al. MassBank: A public repository for sharing mass spectral data for life sciences. *J. Mass Spectrom.* **2010**, *45*, 703–714. [[CrossRef](#)]
46. Li, Z.; Lu, Y.; Guo, Y.; Cao, H.; Wang, Q.; Shui, W. Comprehensive evaluation of untargeted metabolomics data processing software in feature detection, quantification and discriminating marker selection. *Anal. Chim. Acta* **2018**, *1029*, 50–57. [[CrossRef](#)]
47. Watrous, J.D.; Henglin, M.; Claggett, B.; Lehmann, K.A.; Larson, M.G.; Cheng, S.; Jain, M. Visualization, Quantification, and Alignment of Spectral Drift in Population Scale Untargeted Metabolomics Data. *Anal. Chem.* **2017**, *89*, 1399–1404. [[CrossRef](#)] [[PubMed](#)]

48. Sánchez-Illana, Á.; Piñeiro-Ramos, J.D.; Sanjuan-Herráez, J.D.; Vento, M.; Garcia-Manero, G.; Kuligowski, J. Evaluation of batch effect elimination using quality control replicates in LC-MS metabolite profiling. *Anal. Chim. Acta* **2018**, *1019*, 38–48. [[CrossRef](#)]
49. Luan, H.; Ji, F.; Chen, Y.; Cai, Z. statTarget: A streamlined tool for signal drift correction and interpretations of quantitative mass spectrometry-based omics data. *Anal. Chim. Acta* **2018**, *1036*, 66–72. [[CrossRef](#)]
50. Sysi-Aho, M.; Katajamaa, M.; Yetukuri, L.; Orešič, M. Normalization method for metabolomics data using optimal selection of multiple internal standards. *BMC Bioinform.* **2007**, *8*, 93. [[CrossRef](#)]



© 2020 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 (<http://creativecommons.org/licenses/by/4.0/>).