Metabolomics (2020) 16:14
https://doi.org/10.1007/s11306-020-1636-9

ORIGINAL ARTICLE

Optimization of XCMS parameters for LC–MS metabolomics: an assessment of automated versus manual tuning and its effect on the final results

Oihane E. Albóniga[1] · Oskar González[1] · Rosa M. Alonso[1] · Yun Xu[2] · Royston Goodacre[2]

Received: 5 September 2019 / Accepted: 4 January 2020 / Published online: 10 January 2020

Abstract

Introduction Several software packages containing diverse algorithms are available for processing Liquid Chromatography-Mass Spectrometry (LC–MS) chromatographic data and within these deconvolution packages different parameters settings can lead to different outcomes. XCMS is the most widely used peak picking and deconvolution software for metabolomics, but the parameter selection can be hard for inexpert users. To solve this issue, the automatic optimization tools such as Isotopologue Parameters Optimization (IPO) can be extremely helpful.

Objectives To evaluate the suitability of IPO as a tool for XCMS parameters optimization and compare the results with those manually obtained by an exhaustive examination of the LC–MS characteristics and performance.

Methods Raw HPLC-TOF–MS data from two types of biological samples (liver and plasma) analysed in both positive and negative electrospray ionization modes from three groups of piglets were processed with XCMS using parameters optimized following two different approaches: IPO and Manual. The outcomes were compared to determine the advantages and disadvantages of using each method.

Results IPO processing produced the higher number of repeatable (%RSD < 20) and significant features for all data sets and allowed the different piglet groups to be distinguished. Nevertheless, on multivariate level, similar clustering results were obtained by Principal Component Analysis (PCA) when applied to IPO and manual matrices.

Conclusion IPO is a useful optimization tool that helps in choosing the appropriate parameters. It works well on data with a good LC–MS performance but the lack of such adequate data can result in unrealistic parameter settings, which might require further investigation and manual tuning. On the contrary, manual selection criteria requires deeper knowledge on LC–MS, programming language and XCMS parameter interpretation, but allows a better fine-tuning of the parameters, and thus more robust deconvolution.

Keywords IPO · XCMS · LC–MS · Metabolomics · Data treatment

1 Introduction

Nowadays, LC–MS is the most widely used technique in metabolomics due to the large number of metabolites that can be measured in a single analysis (Spicer et al. 2017; Tautenhahn et al. 2012). The data obtained from LC–MS system are highly complex since hundreds or thousands of features (e.g., mass-to-charge ratio and retention time, variables) are obtained with relatively a small number of observations (samples). Therefore, handling such complex data requires bioinformatic software for processing the results and reducing the complexity.

Nowadays, there are several software packages with different approaches and options available for LC–MS data

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s11306-020-1636-9) contains supplementary material, which is available to authorized users.

✉ Oihane E. Albóniga
oihaneelena.alboniga@ehu.eus

[1] Department of Analytical Chemistry, Faculty of Science and Technology, University of the Basque Country (UPV/EHU), Barrio Sarriena s/n, Leioa, Spain

[2] Department of Biochemistry, Institute of Integrative Biology, University of Liverpool, Biosciences Building, Crown Street, Liverpool L69 7ZB, UK

processing. Since the algorithms employed within each program are different, the outcomes can also be different and, therefore, the results of a metabolomic study might differ (Myers et al. 2017). In addition, processing methods reply on a variety of adjustable parameters to function. An improper selection can lead to distorted outcomes (Libiseller et al. 2015). As reported Spicer et al., most of the day-to-day activities of metabolomics professionals consist of a combination of wet and dry lab work and only half have dedicated bioinformatics support (Spicer et al. 2017). This certainly means that metabolomics is a multidisciplinary science and that users need a significant knowledge on data processing including parameters selection and optimization. Nevertheless, software choice is not an easy task and some factors should be considered: the ease-of-use, the transparency of the algorithms employed the accessibility, robustness and reproducibility, etc. In this aspect, commercial software packages are generally more user-friendly, but also restrict the parameters that can be tuned and hide the algorithms behind the data processing. Moreover, commercial software can also be extremely expensive (Spicer et al. 2017) and normally, are intended to work with data files that are provided by the same manufacturer. By contrast, freely available software (usually under open-source license) can read files with open format as mzXML, mzML and netCDF, which are independent of instrument manufacturers formats. All these reasons explain the increased tendency of free software usage.

XCMS is a freely available open source package (Smith et al. 2006) implemented in the R software environment. It is the most used software for LC–MS analysis as reported in a recent survey (Weber et al. 2017). In XCMS, data are first filtered and peak detection is performed for each data file. Then, the software matches peaks in common among the various data files and uses them to correct drifts in retention time (Lazar et al. 2015; Smith et al. 2006). Once the data files are aligned, it groups them across samples. Currently, XCMS provides three main peak detection algorithms, two retention time correction methods and three methods for peak grouping (Spicer et al. 2017). These methods have, in turn, an extensive number of parameters that are optimizable and determine the outcome. As a result, an adequate selection of these parameters is critical for a successful metabolomics study. Unfortunately, this is not an easy task, especially for inexpert users, since the choice of some parameters is not intuitive if the process behind the algorithms is not properly understood. Furthermore, the dependence on the instrument used and the quality of the analytical performance also require a deep knowledge on mass spectrometry and liquid chromatography.

In order to solve this issue, Libiseller et al. developed an automatic tool for the optimization of XCMS data processing parameters for metabolomics called Isotopologue Parameter Optimization (IPO) (Libiseller et al. 2015). IPO uses natural stable ^{13}C isotopes to calculate a peak picking score. Then, retention time correction is optimized by minimizing the relative retention time differences within features, and grouping parameters are selected by maximizing the number of features showing exactly one peak from each injection of a pooled sample. The optimization process with IPO package starts with default settings for the XCMS algorithms. Then, an iterative process is followed by design of experiments (DoEs) using response surface models until the optimal processing parameters are achieved. This R package is freely available and could be helpful when a bioinformatician or a homemade optimisation tool is not accessible. IPO has been applied in multiple metabolomic studies, such as metabolite identification in plasma and cerebrospinal fluid for Parkinson's Disease research (Stoessel et al. 2018a), multiple sclerosis metabolomic profiling (Stoessel et al. 2018b), tissue lipidome storage stability studies (Roszkowska et al. 2018), and many others (Grimbs et al. 2017; Harvey Colin et al. 2018; Narath et al. 2016).

The aim of this present study was to evaluate the suitability of IPO as a tool for XCMS parameters optimization and compare the results with the ones obtained by fixing the processing parameters based on an exhaustive examination of the LC–MS performance. XCMS parameters optimization methods were applied to two sample sets (plasma and liver tissue) obtained from piglets of different ages [newborns (A), neonates (B) and infants (C)] analysed both in positive and negative electrospray ionisation (ESI) modes using High Performance Liquid Chromatography coupled to a Time-Of-Flight Mass Spectrometry (HPLC-TOF–MS) system.

2 Experimental

2.1 Reagents and solutions

Acetonitrile (ACN) (LC–MS grade purity) and formic acid used in the mobile phases were purchased from Scharlau (Sentmenat, Spain) and Fisher Scientific (Pittsburgh, PA, USA), respectively. Ultra-high purity water was used in the preparation of mobile phase and reagent solutions and was obtained from tap water pre-treated by Elix reverse osmosis, and subsequent filtration by a Milli-Q system from Millipore (Bedford, MA, USA). Methanol (MeOH) used for standards and sample preparation was also obtained from Scharlau (Sentmenat, Spain).

Standard reagents used to assess the LC–MS system operation were from different manufacturers: paracetamol, cholic acid, (±) verapamil hydrochloride, simvastatin, reserpine and leucine encephalin acetate salt hydrate were provided by Sigma-Aldrich (Steinheim, Germany), caffeine was purchased from Alfa Aesar (Karlsruhe, Germany) and salicylic

acid from Fluka Analytical (Bucharest, Romania). Finally, Novartis (Barcelona, Spain) gratefully donated fluvastatin sodium. A system suitability test (SST) was prepared with these nine compounds at a final concentration of 100 ng/mL in MeOH:H_2O 2:1 (v/v); this solution composition was chosen as this is similar to the supernatant within plasma and liver samples.

2.2 Study design and sample collection

Sample collection was performed by the team of the Experimental Neonatal Physiology Unit of the BioCruces Health Research Institute (Cruces University Hospital, Basque Country, Spain) following the European and Spanish regulations for protection of experimental animals (86/609/EFC and RD 1201/2005), and was approved by the Ethical Committee for Animal Welfare. Samples were obtained from mechanically ventilated newborn piglets or group A (< 5 days, $n = 12$), neonate piglets or group B (2 weeks, $n = 12$) and infant piglets or group C (4 weeks, $n = 12$) of Topig F-1 Large White × Landrave breed. Each group contained the same number of females and males. Whole blood samples were collected in EDTA tubes, and they were immediately centrifuged at 950×*g* for 10 min at room temperature in order to obtain plasma. The supernatant was transferred to a cryovial and stored at − 80 °C until analysis. Liver tissue samples were immediately submerged in liquid nitrogen and stored also at − 80 °C until analysis.

2.3 Plasma, liver and QC samples preparation

Frozen plasma samples were thawed to room temperature and protein precipitation was carried out with 50 µL of plasma and 100 µL of cold MeOH. After vortex mixing for 2 min in a Signature Digital Vortex Mixer 945303 (VWR, Radnor, PA, USA), samples were centrifuged at 16,110×*g* for 15 min at 10 °C in a 5415R Eppendorf centrifuge (Hamburg, Germany). The clean upper layer was transferred to a chromatographic vial to be injected into the HPLC-TOF–MS system.

Liver tissue samples were kept on liquid nitrogen and/or ice during the whole sample manipulation and treatment. 1 mL of MeOH:H_2O 2:1 (v/v) solution was added to 100 mg of tissue weighted in Precellys tubes with six zirconium balls. The samples were extracted in a Precellys 24 Tissue Homogenizer coupled to a Cryolysis cooling system (both from Bertin Instrument, Montigny-le-Bretonneus, France), and this provided a N_2 stream and set to 5 °C. The tissue homogenizer conditions applied were three cycles of 40 s (10 s between cycles) at 4500 rpm. Supernatant was then centrifuged three times at 15,866×*g* during 15 min at 10 °C in order to remove the suspension particles. Finally, the supernatant was transferred to a chromatographic vial to be injected into the HPLC-TOF–MS system.

QC samples were prepared as a pooled of biological samples (Broadhurst et al. 2018; Dunn et al. 2011). In the case of plasma QC sample, 5 µL of each biological sample were taken and thoroughly mixed, reaching a total volume of 180 µL. From this pool, 50 µL were treated as previously described protocol for plasma samples. Liver QC sample was prepared by taking 8 µL of the centrifuged supernatant from each sample. The QC samples were injected at the beginning of the run to set up the system and then every sixth samples, so they were used for signal correction within the analytical sequence. For both plasma and liver analysis samples were run in a randomised order.

2.4 HPLC-TOF–MS analysis

The metabolomic profiling of plasma and liver supernatants were performed using an Agilent 1200 Series HPLC system coupled to an Agilent 6530 hybrid quadrupole-time-of-flight (Q-TOF) mass spectrometer (Agilent Technologies, Santa Clara, CA, USA) accompanied by the MassHunter Workstation software for Q-TOF (version B.05.01 for data acquisition, and version B.07.00 for raw data processing). 5 µL of the extracted samples were injected into a Zorbax SB-C18 (2.1 × 100 mm, 3.5 µm) reverse phase column with a C8 guard column (2.1 × 12 mm, 5 µm), both from Agilent Technologies.

The separation was carried out at a flow rate of 0.4 mL/min and 35 °C. The HPLC mobile phase consisted of 0.1% formic acid in water and 5% of ACN (solution A), and 0.1% formic acid in ACN (solution B). Gradient started from 0 to 100% B in 10 min, and then kept at 100% B for 2.5 min, returned to starting conditions in 1.5 min. Finally, it was kept for re-equilibrations at 0% B for 5 min.

The ESI mass spectra data were acquired both in positive and negative ionization modes using Agilent Jet Stream ESI source (ESI), with capillary voltages of + 3800 V and −2500 V, respectively. The other source parameters were kept constant in all the experiments: drying gas (nitrogen) temperature was 325 °C at a flow rate of 10 L/min, pressure of nebulizer gas (nitrogen) was 30 psi, sheath gas was kept at a flow rate of 11 L/min at 350 °C temperature and the voltages of the skimmer, fragmentor, and octopole RF peak were 65, 125 and 750 V, respectively.

The MS detector operated in low mass range (< 1700 *m/z*) and 2 GHz extended dynamic range. Centroid acquisition mode was used at a full scan range from 50 to 1200 *m/z* with an acquisition rate of 2 spectra/s. In order to guarantee mass accuracy, a reference solution was directly infused into the source enabling continuous internal calibration during analysis and ensuring accuracy and reproducibility. For this purpose, two reference masses at *m/z*

121.0509 (purine, $[C_5H_4N_4+H]^+$) and *m/z* 922.0098 (HP-921, $[C_{18}H_{18}O_6N_3P_3F_{24}+H]^+$) for positive mode, and *m/z* 112.9855 ($TFANH_4$, $[C_2H_4O_2NF_3–NH_4]^-$) and *m/z* 966.0007 (HP-921COOH, $[C_{18}H_{18}O_6N_3P_3F_{24}–COOH]^-$) for negative mode, were used during the HPLC-TOF–MS run. Additionally, in order to control the analytical performance of the MS instrument and the LC system, the SST was injected in the HPLC-TOF–MS system at the beginning, in the middle and at the end of each sequence.

2.5 Data treatment and optimization of the processing parameters

The raw HPLC-TOF–MS data were acquired with the Agilent MassHunter Workstation and were converted into mzXML format using msConvert (proteoWizard) from 0 to 13.5 min RT in order to avoid features coming from cleaning step of the gradient. Feature detection was performed in R 3.4.3 (https://www.r-project.org/) using the XCMS 1.52.0 (Scripps, La Jolla, CA, USA) package (https://www.bioconductor.org/packages/release/bioc/html/xcms.html) (Smith et al. 2006). CentWave algorithm was used for peak picking considering that it fits better with centroid acquisition data and that it can detect partially overlapping features. It also results in higher precision values compared to matchedFilter (Tautenhahn et al. 2008). Retention time correction was directly performed on the profile by Obiwarp algorithm (Prince and Marcotte 2006), which aligns multiple samples by using a center sample. An additional advantage of Obiwarp algorithm is that can be done independently of the peak picking or peak grouping (Prince and Marcotte 2006).

All these algorithms have some critical parameters that should be properly chosen to perform a metabolomics study. In this work, two different approaches to select the XCMS processing parameters were compared: an automated approach using IPO package (https://bioconductor.org/packages/release/bioc/html/IPO.html) developed by Libiseller et al. and a manual approach based on the HPLC-TOF–MS system (Libiseller et al. 2015). In Table 1 the most significant variables optimized for these two approaches are collected.

2.5.1 IPO parameters optimization

The optimization of the processing parameters in IPO was performed using only the QC samples so that the biological or inter-individual variability would not affect the parameter optimization. The pipeline described by Libiseller et al. was followed using the default values included in the package (Libiseller et al. 2015). In short, IPO uses a lower and upper level for each parameter to be optimized and automatically calculates the center point. It generates the first Design of Experiment (DoE) and then, a new DoE is built according to the previous DoE results. An iterative process is performed until the optimized parameter setting was converged. Once the parameters were optimized, the values obtained were checked to confirm that they fitted with the HPLC-TOF–MS system settings and then they were applied to the all samples.

2.5.2 Manual parameters selection

Manual parameter selection was carried out considering the specifications of the HPLC-Q-TOF–MS instrument and the performance of the analysis, which was studied by using the SST. In this way, the *ppm* parameter was established considering the difference between the theoretical and

Table 1 Definition of the optimized XCMS processing parameters according to R documentation

Processing step	XCMS parameter	Definition
Peak picking (centWave)	ppm	Maximal tolerated *m/z* deviation in consecutive scans in ppm. This is usually much larger than the ppm specified by the manufacturer
	peakwidth	Chromatographic peak width, given as a range (min, max) in seconds
	mzdiff	Minimum difference in *m/z* for peaks with overlapping retention times, can be negative to allow overlap
Retention time alignment (Obiwarp)	profStep	Step size (in *m/z*) to use for profile generation from the raw data files
	center	The index of the sample all others will be aligned to. If center ==NULL, the sample with the most peaks is chosen as default
	gapInit	Penalty for gap opening
	gapExtend	Penalty for gap enlargement
Grouping (density)	bw	Bandwidth (standard deviation or half width at half maximum) of Gaussian smoothing kernel to apply to the peak density chromatogram
	mzwid	Width of overlapping *m/z* slices to use for creating peak density chromatograms and grouping peaks across samples
	minfrac	Minimum fraction of samples necessary in at least one of the sample groups for it to be a valid group

experimental mass of the SST compounds and the tolerated mass deviation (μ) (Tautenhahn et al. 2008). The *peakwidth* was determined taking into account the minimum and maximum peak width within the standard compounds. In order to ensure a proper comparison, it was used as *center sample* the one selected by IPO in the optimization process. Then, *mzwid* was determined with the parameter manager tab in the XCMS Online (https://xcmsonline.scripps.edu/landing_page.php?pgcontent=institute). The *bandwith* (*bw*) was retention time deviation dependent and it was obtained from the figure resulted after Obiwarp retention time alignment when only QCs were analyzed.

Once the data were analyzed with the parameters optimized by IPO and manually, and after grouping across samples, a filling step was included to reduce the number of missing peaks that could occurred during the peak identification or because the metabolite was not present in a sample (Smith et al. 2006). It is important to highlight that during the optimization processes *minfrac* parameter was set at 1 as features had to be in all the QC samples to be detected. On the other hand, when the study samples were processed *minfrac* was set at 1/6, so a feature had to be at least in the half of the samples of a group to be included. Finally, CAMERA 1.32.0 package (Bioconductor Open Source Software for Bioinformatics; https://bioconductor.org/packages/release/bioc/html/CAMERA.html) (Kuhl et al. 2012) was used for isotopologues and adducts detections. After the whole data processing, a two-dimensional table (matrix) with a list of features and their intensities was obtained for each sample set and ionization mode. These features include molecules, fragments, small peptides, isotopes, and so on. Then, these matrices were prepared before the statistical analysis in order to remove the isotopes identified by CAMERA ([M + 1], [M + 2] and [M + 3]) and the features before the injection peak (less than 1 min). The percentage of relative standard deviation (%RSD) was calculated for all metabolic features in QC samples and the features with %RSD greater than 20% were removed due to its variability (Dudzik et al. 2018; Dunn et al. 2011; Kirwan et al. 2013). The amount of features that fulfill the %RSD criteria was calculated to compare easily the matrices obtained by both parameters optimization procedure, IPO and manual.

2.6 Statistical analysis

Multivariate analysis was performed in MATLAB (The MathWorks, Naticks, MA, USA) using our homemade MATLAB functions which is freely available online at https://github.com/Biospec/cluster-toolbox-v2.0. The matrices obtained were further processed with QC intensity drop correction (Kirwan et al. 2013) and autoscaling or logarithm ($\log_{10}$) scaling was applied. Unsupervised analysis using principal component analysis (PCA) was performed as an overview for tendencies among piglets with different ages [newborn (A), neonates (B) and infant (C) piglets]. This analysis was applied not only to reduce the dimensionality of the data for grouping the samples, but also to assess the data quality, such as the detection of potential outliers or technical variations (Lazar et al. 2015). The pattern of the data was analyzed visually via PCA scores plots. To evaluate if there were significant differences between the clustering obtained with the two parameters optimization methods a Procrustes distance (Dryden and Mardia 1998) was calculated between each pair of PCA scores (first 3 PCs) obtained by these two methods. Procrustes distance varied from 0 to 1 in which 0 indicates identical two patterns while 1 indicates that the two patterns were completely different.

Univariate analysis, also performed in MATLAB, was carried out to identify molecular features that discriminate the three groups under study. Non-parametric testing (Kruskal–Wallis) followed by False Discovery Rate (FDR) was applied to detect the significant features. In order to find those significant features that differentiate the three groups of piglets, a post-hoc Tukey HSD test (Honestly Significant Differences) was performed. Then, correlation of those features that show a similar retention time and *m/z* value in both parameters selection methods was studied. A coefficient of determination (R^2) close to 1 indicated that the same feature had been detected by IPO and manual selection method. Finally, a Venn Diagram was constructed with the significant features that fulfill the non-parametric test in order to visualize the impact of the processing parameter optimization method on the detection of features of interest. The diagram was built with the online version Venny 1.0 (https://bioinfogp.cnb.csic.es/tools/venny/).

3 Results and discussion

3.1 IPO and manual parameter selection

The XCMS processing parameters obtained by IPO optimization and the parameters chosen manually for each of the sample sets using the QC samples are collected in Table 2. In addition to the parameters that were optimized (*ppm*, *peakwidth*, *mzdiff*, *center sample*, *bw* and *mzwid*), other parameters that were set by default within the XCMS processing algorithms, but different from the IPOs' values, such as *profStep*, *gapInit* and *gapExtend*, are included.

The values obtained using IPO laid within the expected range for the HPLC-TOF–MS analysis with the exception of the results obtained for the Plasma ESI− data set, which had significantly lower signal-to-noise ratio compare to the other three data sets. For the Plasma ESI− data set, the optimal values estimated for *ppm* and *peakwidth* parameters by IPO did not seem to be realistic. IPO reported an

Table 2 Optimal processing parameters obtained with the QC samples using IPO and manual approaches for the four data sets

Processing step	Parameter	ESI+		ESI−	
		IPO	Manual	IPO	Manual
Liver					
Peak picking (centWave)	ppm	29	30	20.75	30
	peakwidth	12, 39.5	13, 42	9.76, 50	13, 42
	mzdiff	0.0034	0.001	− 0.0048	0.001
Retention time alignment (Obiwarp)	profStep	0.655	1	0.64	1
	center	QC8	QC8	QC4	QC4
	gapInit	0.64	NULL	0.928	NULL
	gapExtend	2.4	NULL	1.668	NULL
Grouping (density)	bw	0.879	6	0.25	6
	mzwid	0.0265	0.025	0.0126	0.025
Plasma					
Peak picking (centWave)	ppm	31.68	30	74.75 [31][a]	30
	peakwidth	22.01, 81.26	17, 84	69.6, 70.6 [20, 80][a]	17, 84
	mzdiff	− 0.0123	0.001	0.0029 [− 0.012][a]	0.001
Retention time alignment (Obiwarp)	profStep	0.7324	1	0.844 [1][a]	1
	center	QC4	QC4	QC7 [QC5][a]	QC 7 [QC5][b]
	gapInit	0.7552	NULL	0.928 [0.928][a]	NULL
	gapExtend	2.4	NULL	2.358 [2.688][a]	NULL
Grouping (density)	bw	0.25	10	0.879 [0.879][a]	55 [50][b]
	mzwid	0.027	0.025	0.0265 [0.0342][a]	0.025

NULL represents the absence of intentional value, that is to say that the variable has an unknown value, for which it is assigned null

[a]In brackets values of the parameters after supervision (i.e., after investigation and manual tuning of *ppm* and *peakwidth* parameters based on the HPLC-TOF–MS performance)

[b]In brackets values of the parameters adapted to allow comparison with IPO results after supervision

optimal 74.75 ppm value that did not fit with the standard accuracy of the MS instrument and a *peakwidth* of (69.6, 70.6) with an apparently overestimated lower range value. To understand the anomalous values for these parameters, results obtained for SST were carefully studied. On the one hand, it was observed that the signal strength was indeed low, but still most significant peaks had peak width ranged between 17 and 84 s. In addition, MS accuracy was also comparable to the one obtained during the analysis of the other sample sets and lower than 5 ppm for all the compounds in the SST. Taking into account this fact, data analysis was performed using the raw values for the parameters obtained by IPO (labelled as "Before Supervision": B.S.) and using the values obtained after *ppm* and *peakwidth* were set to more realistic parameters (labelled as "After Supervision": A.S.). For the latter, *ppm* parameter was set at 31 and *peakwidth* parameter at (20, 80) based on the results obtained for the Plasma ESI+ data set. Since IPO used a different center sample in each case, data analysis using manual parameters was performed using those two center samples optimized by IPO in order to allow a suitable comparison. All this information is gathered in Table 2.

3.2 Influence of the parameter selection on the number and variability of the detected features

Once the processing parameters were optimized using the QC samples, each data set was processed in order to obtain the features matrices. Table 3 shows the total number of features obtained initially for each data set and the number of features after filtering the matrices. For instance, for Liver ESI+ 2472 and 2045 features were eliminated with IPO parameters and manual selection criteria, respectively. In order to study the variability, the %RSD across the QC samples was calculated for each feature and the percentage of features with a %RSD lower than 20% was calculated for each data set. For further data processing, the features with %RSD higher than 20% were removed.

As can be observed, the impact of the processing parameters in the results for ESI+ data sets was minimal in terms of percentage of features with acceptable repeatability, although the total number of repeatable features was higher when IPO parameters were used, especially for liver data set. Regarding ESI− data sets, the percentage of features with a good repeatability was significantly higher with IPO

Table 3 Number of features obtained with XCMS after processing the different data sets and number of features after each filtering step

	Liver				Plasma					
	ESI+		ESI−		ESI+		ESI− (B.S.)		ESI− (A.S.)	
	IPO	Manual	IPO	Manual	IPO	Manual	IPO	Manual	IPO	Manual
Total number of features	9759	8399	4559	4724	4735	5253	2439	3352	4359	3378
After removing isotopes and injection volume features	7287	6354	3307	3782	3750	3909	2005	2636	3536	2602
Number of features with %RSD < 20	5603	4746	2242	2110	2207	2178	2005	2636	1855	1186
Percentage of features with %RSD < 20	76.9	74.7	67.8	49.5	58.9	55.7	37.0	46.9	52.5	45.6

in the case of liver. In plasma data set it is noteworthy the total number of detected features and also the percentage of repeatable features (%RSD < 20%) obtained by using the parameters optimized by IPO B.S., which were both lower than the ones obtained by IPO B.S. (2439 vs. 4359 and 37% vs. 52.5% respectively). This suggested that with such unrealistic parameter setting, many peaks had been wrongly rejected by XCMS.

Considering these results, IPO appears to be a very valuable tool to optimize XCMS parameters. For results obtained by using the optimized XCMS parameters, the percentage of repeatable features was 2.2, 18.3 and 3.2 higher than those using manually picked XCMS parameters for Liver ESI+, Liver ESI−, and Plasma ESI+, respectively. However, extra care is needed when IPO is applied to a data set with poor chromatographic performance, the optimization process can be trapped into an unrealistic region, resulted in impractical parameter settings, and subsequently lead to poorer peak picking and alignment results.

3.3 Influence of the parameters selection in multivariate analysis

PCA scores plot was firstly constructed with the pre-processing methods that best fits grouping (see Supplementary Fig. 1), and when the intensity drop and/or the lack of grouping was observed, QC correction function (available at https://github.com/Biospec/cluster-toolbox-v2.0) was applied. After that, different transformations were tested again. In this way, for Liver ESI+, Liver ESI− and Plasma ESI+, QC correction and autoscaling transformation were used, whereas for Plasma ESI− (B.S. and A.S.), QC correction and logarithmic transformation were employed (Fig. 1). In order to allow a proper comparison, the scaling transformation applied in each data set was the same for both IPO and manual selection procedure.

A visual inspection of the PCA did not show significant differences between the groupings obtained with the different parameter setting optimization methods for any of the data sets, except for Plasma ESI−, which is clearly better with manual selection criteria (Fig. 1), and IPO (B.S) had resulted in artificial splitting in group B and C. Such split largely disappeared after parameter supervision. This clearly demonstrated the effect of an improper parameter setting in the final results. The Procrustes distances confirmed such observation. For Liver ESI+ and Liver ESI− data sets, Procrustes distances were 0.0096 and 0.0082 respectively, suggesting the two patterns were virtually identical. For Plasma ESI+, the Procrustes distance was 0.0313, also suggested that the two patterns were highly similar. For Plasma ESI− before IPO supervision, the Procrustes distance was 0.5707, which indicates that there were substantial differences between the two patterns while after supervision this distance decreased to 0.0821, indicating that the artefacts introduced by improper parameter settings were mostly removed.

4 Influence of the parameters selection in the detection of features of interest

The ultimate goal of the data processing in metabolomics is the detection of metabolites that can explain the differences between different groups of samples. Therefore, it was our aim to compare the features that better explain the differences among the three groups of the study and to see if these features match regardless of the methodology used for processing parameters selection. For this purpose, univariate analyses were used in order to find features that significantly distinguish newborns (A), neonates (B) and infants (C),

In Table 4 the number of significant features after the univariate statistical analysis are gathered. Even though there were remarkable differences between the number of features initially detected, especially in Liver ESI+ and Plasma ESI− data sets, slightly differences were observed between the total significant features after the univariate statistical analysis. The number of significant features obtained with IPO (B.S.) was not calculated considering the poor repeatability of the features extracted.

For each data set the retention time and *m/z* values of the significant features obtained with IPO were compared with those ones obtained with manual parameter selection and

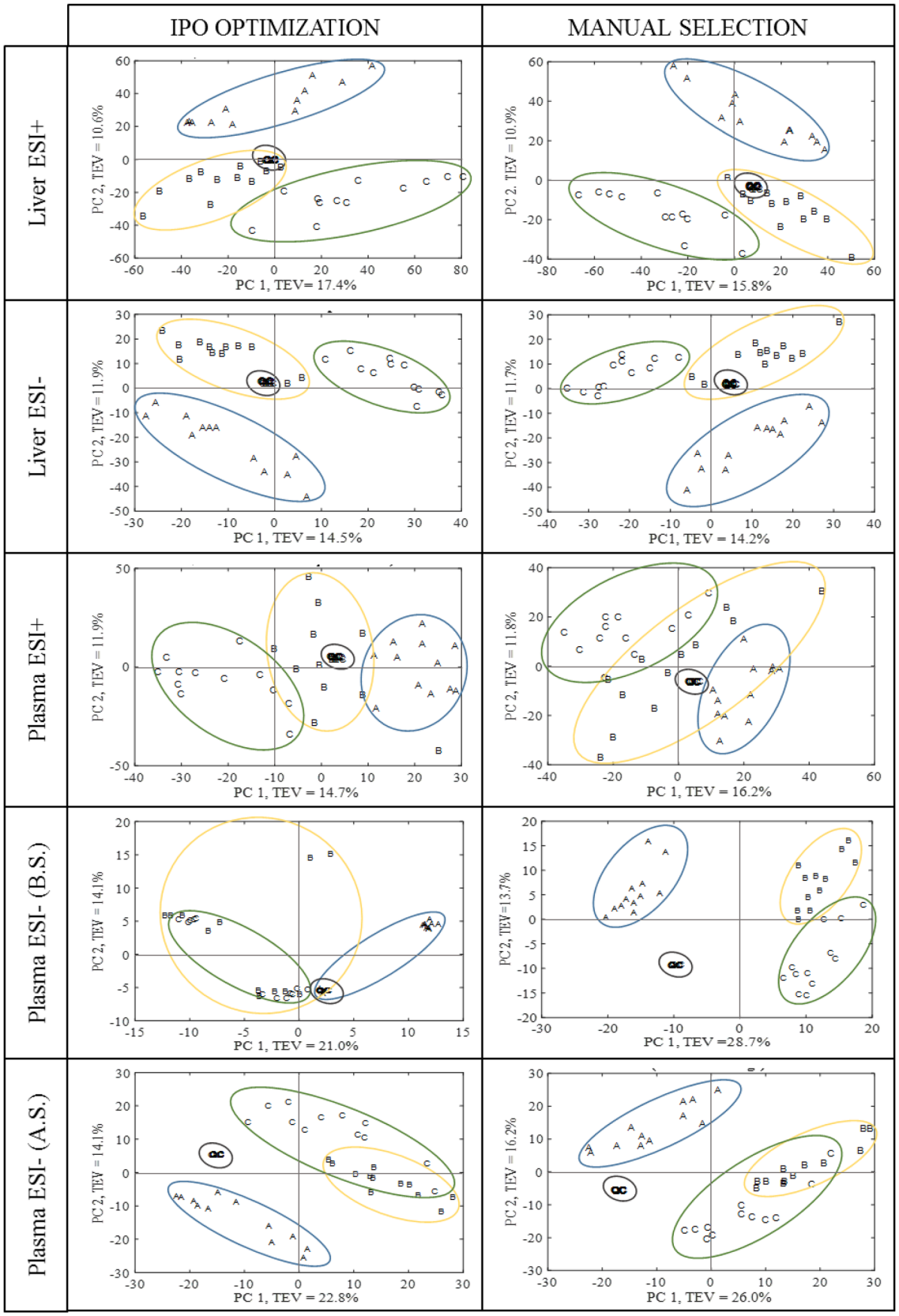

Fig. 1 PCA scores plot obtained with IPO and manual parameters selection criteria for the all data sets (Liver ESI+, Liver ESI−, Plasma ESI+, Plasma ESI− B.S, and Plasma ESI− A.S.). Piglets groups: newborns (A), neonates (B) and infants (C)

Table 4 Number of features of interest in the fourth data sets after applying univariate tests

	Liver				Plasma			
	ESI+		ESI−		ESI+		ESI− (A.S.)	
	IPO	Manual	IPO	Manual	IPO	Manual	IPO	Manual
Total number after matrix filtering	5603	4746	2242	2110	2207	2178	1855	1186
Kruskal–Wallis and FDR ($p < 0.001$)	623	587	318	302	109	102	462	329
Total significant features A ≠ B ≠ C (Tukey HSD)	59	57	27	25	6	8	7	5

potential common features were identified. Those features with R^2 higher than 0.9 were considered to be the same. In Figure S.2 an example of a two features that matched on retention time and *m/z* showed a good correlation and hence, were considered to be the same, can be observed. Similarly, another feature with $R^2 < 0.9$ is also shown. In this case, features were considered to be different, although the retention time and the *m/z* values were very similar.

Once the common significant features were found, Venn Diagrams were generated in order to visualize the coincidence between IPO and manual selection criteria for each data set (Fig. 2). The number of common significant features was 44 for Liver ESI+, 15 for Liver ESI−, 1 for Plasma ESI+ and 3 for Plasma ESI− According to these results, the percentage of significant features obtained with IPO that were also detected with manual parameters were 74.6% for Liver ESI+, 55.6% for Liver ESI−, 16.7% for Plasma ESI+ and 42.9% for Plasma ESI−. Similarly, the percentage of significant features detected by manual criteria that were also detected with IPO parameters were 77.2% for Liver ESI+, 60% for Liver ESI−, 12.5% for Plasma ESI+ and 60% for Plasma ESI−. Thus, the two methods employed to choose the processing parameters offered significantly different results, and many features were only detected using one of the methods.

In order to find an explanation for this lack of congruence in features selected, further study of these features was performed. On one hand, it was observed that, in some cases, isotopologues had not been properly identified by CAMERA (Kuhl et al. 2012) during the XCMS processing for one of the approaches, overestimating the number of non-matching features. On the other hand, it was observed that some of the features had a non-Gaussian shape and could not be considered proper chromatographic peaks after extracting the *m/z* value of each significant feature from the raw data (extracted ion chromatogram). It is remarkable that the amount of these

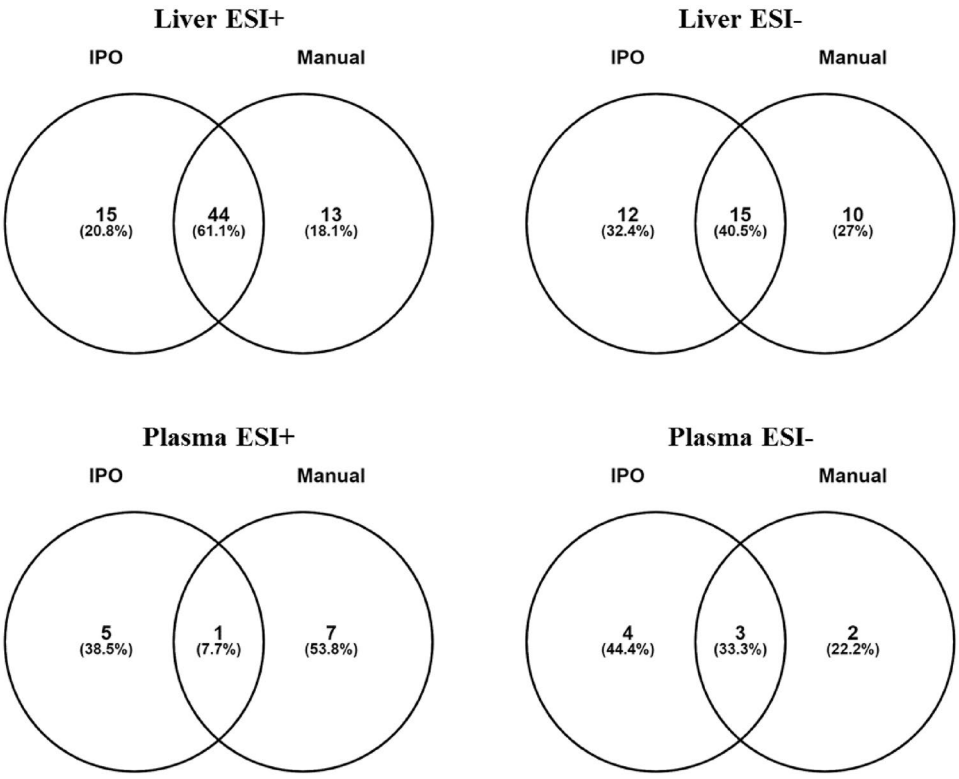

Fig. 2 Venn Diagrams highlighting the common and uncommon features of interest for liver and plasma matrices at both positive and negative ionization modes

artifacts was much higher among those features that had been only considered to be significant with one of the methods. For instance, for Liver ESI+ data set 8 out of the 15 (53.3%) features that were detected only with IPO parameters showed a proper chromatographic signal, and among them, 3 were isotopes of other features. This means that out of the 15 features that could be considered as potential biomarkers only 5 (33.3%) would be taken into consideration for further identification studies. Similarly, 6 out of the 13 (46.1%) features detected only using manual parameters criteria were considered to be suitable peaks. This contrasts with the high number of significant features (79.6%) that show an adequate response among the common features. Therefore, many of the significant features that were not in common could lead to misleading results and they would be rejected in the first steps of identification. Obviously, the fact that the features obtained by both methods differ might have an impact in the outcome of the metabolomic study, but lower than the expected from the results observed initially in the Venn Diagrams.

A deeper study was next performed in the filtered matrices in order to know if the dissimilarities are due to a lack of detection during the preprocessing or to the subsequent statistical analysis. Therefore, the features that were found to be significant only with one approach were searched in the filtered matrix obtained with the other approach. For instance, the 15 significant features detected only with IPO in Liver ESI+ were searched in the matrix obtained with the manual approach. In this way, 10 of these features were observed in the manual matrix, which means that they were detected during the preprocessing but did not fulfill Kruskal–Wallis and/or post-hoc Tukey's HSD test. Thus, a total of 54 features out of the 59 (91.5%) significant features detected by IPO were also in the manual matrix. Similarly, 55 features out of the 57 (96.5%) significant features detected by manual approach were commonly detected in IPO for Liver ESI+. In the same way, 23 out of 27 (85.2%) and 20 out of 25 (80%) were commonly detected in Liver ESI−, 3 out of 6 (50%) and 5 out of 8 (62.5%) in Plasma ESI+ and 6 out of 7 (85.7%) and 4 out of 5 (80%) in Plasma ESI− as can be observed in Fig. 3.

These results show that the preprocessing step has a significant impact not only in the number of features that are actually detected but also in the measured response of those features. This can lead to different statistical results and thus jeopardize the reliability of the results, even if the differences in responses are minimal. A remarkable case of this situation can be observed in Supplementary Fig. 3. A feature that was found to be significant to distinguish the three studied groups with manual approach in Plasma ESI+ was not observed among the significant features obtained with IPO. Nevertheless, this feature was present in the IPO matrix, and the responses were very similar to the ones obtained with the manual approach (maximum difference of 2.2% and $R^2 = 0.99995$, Supplementary Fig. 3A). Surprisingly this feature did not show significant differences among the three groups when the post-hoc Tukey HSD test was applied to the IPO data (see Supplementary Fig. 3B).

5 Concluding remarks

Metabolomics study using XCMS requires careful processing parameters selection due to the importance of a reliable data matrix for the subsequent steps of the data treatment. Although there is not such a thing as the optimal processing solution, the values for these parameters must guarantee a reliable data processing. In this sense, the parameters optimization tools such as IPO can be of great interest. As reported in this study, the resulting matrices after all data filtration steps show that IPO offers a higher number of repeatable features compared with manual parameter selection and, therefore, more robust results. As an exception, the results obtained with IPO for a data set with a poorer chromatographic performance (Plasma ESI−) was less repeatable. Results were greatly improved after modifying some critical parameters (*peakwidth* and *ppm*), but this fact clearly indicates that IPO optimization should be supervised (i.e., have manual steps during the processing) or unrealistic parameters might be chosen, especially when dealing with challenging data sets. Even if similar results were obtained for multivariate analysis, significant differences were observed when applying univariate analysis to find features of interest that distinguish the study groups. Although the number of significant features is similar, there are remarkable differences when comparing IPO and manual results that can be explained to some extend by the appearance of artefacts during XCMS data processing. Also, even a feature had been detected by different parameter settings, small variations in the integrated peak areas can lead to different statistical conclusions. This demonstrated that *p*-values and FDRs also cannot be taken as granted and the threshold for statistical significance assessment may also need to be scrutinized before a conclusion can be drawn. Furthermore, the assessment of the IPO and manual matrices before the application of univariate analysis shows that most of the significant features in IPO data were also found in the manual data and vice versa. Nevertheless, due to differences in the responses of these features the statistical results were affected. These facts clearly show the high impact of the processing parameters in the outcome of a metabolomic study.

The IPO tool can be useful to optimize the parameter selection in a less time-consuming manner and can be especially helpful for untrained users as a starting point for data treatment. Despite the ease of use of IPO package, the obtained parameters require supervision in order to find

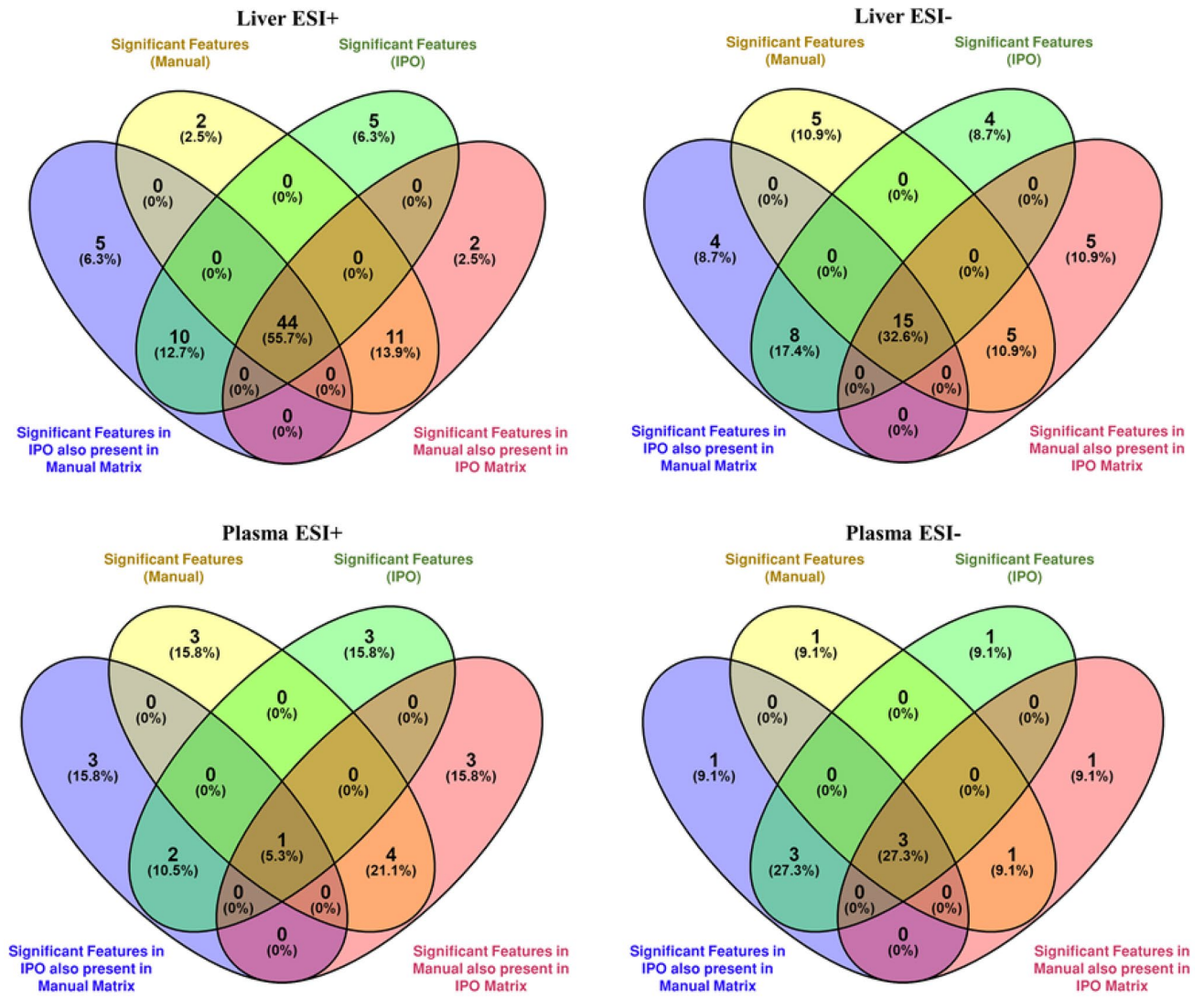

Fig. 3 Venn Diagrams highlighting the significant features detected with IPO or Manual approaches that are coincident in the other approach matrix (but did not meet the statistical tests to be considered as significant features)

unreliable results as we observed for Plasma ESI− data set. Therefore, users need certain liquid chromatography and mass spectrometry knowledge. On the other hand, manual selection criteria require not only a deeper knowledge on these topics, but also on programming language and XCMS parameters interpretation. In return, this approach allows a better fine-tuning of the optimization parameters. In short, IPO optimization and manual selection methods have advantages and disadvantages to consider when untargeted metabolomic data has to be analysed.

Acknowledgements Authors thank the Experimental Neonatal Physiology Unit of the BioCruces Health Research Institute (Cruces University Hospital, Basque Country, Spain) for collecting and providing the samples. O.E.A. thanks the Ministry of Economy and Competitiveness for her predoctoral contract. Authors thank for technical and human support provided by SGIker of UPV/EHU and European funding (ERDF and ESF).

Author contributions The following authors contributed to conception and design of the study: O.E.A., O.G., R.M.A., Y.X. and R.G.; funding acquisition: R.M.A; experimental performance and acquisition of data: O.E.A. and O.G.; data analysis: O.E.A. and Y.X.; writing – original draft preparation; O.E.A. and O.G., and supervision: O.G., R.M.A., Y.X. and R.G. All authors revised the article critically for important intellectual content and approved the final version.

Funding This research was funded by UPV/EHU (Project GIU16/04), the Spanish Ministry of Economy and Competitiveness (Project CTQ2013-46179-R) and the UK Wellcome Trust for MetaboFlow (Grant 202952/Z/16/Z).

Data availability The datasets generated and analysed during the current study are available from the corresponding author on request.

Compliance with ethical standards

Conflict of interest The authors declare that there is no conflict of interest.

Ethical approval The study was approved by the Ethical Committee for Animal Welfare following the European and Spanish regulations for protection of experimental animals (86/609/EFC and RD 1201/2005).

References

Broadhurst, D., Goodacre, R., Reinke, S. N., Kuligowski, J., Wilson, I. D., Lewis, M. R., et al. (2018). Guidelines and considerations for the use of system suitability and quality control samples in mass spectrometry assays applied in untargeted clinical metabolomic studies. *Metabolomics, 14*(6), 72.

Dryden, I. L., & Mardia, K. V. (1998). *Statistical shape analysis*. Wiley series in probability and statistics. Chichester: Wiley.

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

Dunn, W. B., Broadhurst, D., Begley, P., Zelena, E., Francis-McIntyre, S., Anderson, N., et al. (2011). Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. *Nature Protocols, 6*(7), 1060–1083.

Grimbs, A., Shrestha, A., Rezk Ahmed, S. D., Grimbs, S., Hakeem, S. I., Hutt, M., et al. (2017). Bioactivity in rhododendron: A systemic analysis of antimicrobial and cytotoxic activities and their phylogenetic and phytochemical origins. *Frontiers in Plant Science, 8*, 551.

Harvey Colin, J. B., Schlecht, U., Horecka, J., Fischer, C. R., Li, J., Naughton, B., et al. (2018). HEx: A heterologous expression platform for the discovery of fungal natural products. *Science Advances, 4*(4),eaar5459.

Kirwan, J., Broadhurst, D., Davidson, R., & Viant, M. (2013). Characterising and correcting batch variation in an automated direct infusion mass spectrometry (DIMS) metabolomics workflow. *Analytical and Bioanalytical Chemistry, 405*(15), 5147–5157.

Kuhl, C., Tautenhahn, R., Boettcher, C., Larson, T. R., & Neumann, S. (2012). CAMERA: An integrated strategy for compound spectra extraction and annotation of liquid chromatography/mass spectrometry data sets. *Analytical Chemistry (Washington, DC), 84*(1), 283–289.

Lazar, G., Florina, R., Socaciu, M., & Socaciu, C. (2015). Bioinformatics tools for metabolomic data processing and analysis using untargeted liquid chromatography coupled with mass spectrometry. *Bulletin of University of Agricultural Sciences and Veterinary Medicine Cluj-Napoca Animal Science and Biotechnologies*. https://doi.org/10.15835/buasvmcn-asb:11536.

Libiseller, G., Dvorzak, M., Kleb, U., Gander, E., Eisenberg, T., Madeo, F., et al. (2015). IPO: A tool for automated optimization of XCMS parameters. *BMC Bioinformatics, 16*, 1–10.

Myers, O. D., Sumner, S. J., Li, S., Barnes, S., & Du, X. (2017). Detailed investigation and comparison of the XCMS and MZmine 2 chromatogram construction and chromatographic peak detection methods for preprocessing mass spectrometry metabolomics data. *Analytical Chemistry (Washington, DC), 89*(17), 8689–8695.

Narath, S. H., Mautner, S. I., Svehlikova, E., Schultes, B., Pieber, T. R., Sinner, F. M., et al. (2016). An untargeted metabolomics approach to characterize short-term and long-term metabolic changes after bariatric surgery. *PLoS ONE, 11*(9), e0161425/1–e0161425/18.

Prince, J. T., & Marcotte, E. M. (2006). Chromatographic alignment of ESI-LC-MS proteomics data sets by ordered bijective interpolated warping. *Analytical Chemistry, 78*(17), 6140–6152.

Roszkowska, A., Yu, M., Bessonneau, V., Bragg, L., Servos, M., & Pawliszyn, J. (2018). Tissue storage affects lipidome profiling in comparison to in vivo microsampling approach. *Science Reports, 8*(1), 1–10.

Smith, C. A., Want, E. J., O'Maille, G., Abagyan, R., & Siuzdak, G. (2006). XCMS: Processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Analytical Chemistry, 78*(3), 779–787.

Spicer, R., Salek, R. M., Moreno, P., Canueto, D., & Steinbeck, C. (2017). Navigating freely-available software tools for metabolomics analysis. *Metabolomics, 13*(9), 1–16.

Stoessel, D., Schauer, N., Stoessel, D., Walther, D., Stoessel, D., Walther, D., et al. (2018a). Promising metabolite profiles in the plasma and CSF of early clinical Parkinson's disease. *Frontiers in Aging Neuroscience, 10*, 51.

Stoessel, D., Schauer, N., Stoessel, D., Walther, D., Stoessel, D., Walther, D., et al. (2018b). Metabolomic profiles for primary progressive multiple sclerosis stratification and disease course monitoring. *Frontiers in Human Neuroscience, 12*, 226.

Tautenhahn, R., Boettcher, C., & Neumann, S. (2008). Highly sensitive feature detection for high resolution LC/MS. *BMC Bioinformatics, 9*, 504.

Tautenhahn, R., Patti, G. J., Rinehart, D., & Siuzdak, G. (2012). XCMS online: A web-based platform to process untargeted metabolomic data. *Analytical Chemistry (Washington, DC), 84*(11), 5035–5039.

Weber, R. J. M., Lawson, T. N., Salek, R. M., Ebbels, T. M. D., Glen, R. C., Goodacre, R., et al. (2017). Computational tools and workflows in metabolomics: An international survey highlights the opportunity for harmonisation through galaxy. *Metabolomics, 13*(2), 1–5.