University of Groningen

Dynamic binning peak detection and assessment of various lipidomics liquid chromatography-mass spectrometry pre-processing platforms

Feng, Xiaodong; Zhang, Wenxuan; Kuipers, Folkert; Kema, Ido; Barcaru, Andrei; Horvatovich, Péter

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

[Link to publication in University of Groningen/UMCG research database](#)

Dynamic binning peak detection and assessment of various lipidomics liquid chromatography-mass spectrometry pre-processing platforms

Xiaodong Feng [a], Wenxuan Zhang [b, c], Folkert Kuipers [a, b], Ido Kema [a], Andrei Barcaru [a, 1], Péter Horvatovich [c, *, 1]

[a] Department of Laboratory Medicine, University Medical Center Groningen, Hanzeplein 1, 9713, GZ Groningen, the Netherlands
[b] Department of Pediatrics, University Medical Center Groningen, Hanzeplein 1, 9713, GZ Groningen, the Netherlands
[c] Department of Analytical Biochemistry, University of Groningen, Antonius Deusinglaan 1, 9713, AV Groningen, the Netherlands

HIGHLIGHTS

- The dynamic binning allows accurate reconstruction of extracted ion chromatogram and improve peak picking and quantification in XCMS.
- The peak detection mass tolerance set by dynamic binning method is proportional to $mz2$ for FTICR, to $mz1.5$ for Orbitrap, to mz for Q-TOF and is a constant for Quadrupole mass analyzers.
- The dynamic binning method improves the performance of peak detection and quantification in XCMS.
- XCMS upgraded with dynamic binning and Progenesis perform similarly well to identify compounds with differential concentration levels and shows better performance compared to mzMine.
- Aggregate LC-MS map improve quantification of peaks with low variance but results in artifacts for ones with high variance.

GRAPHICAL ABSTRACT
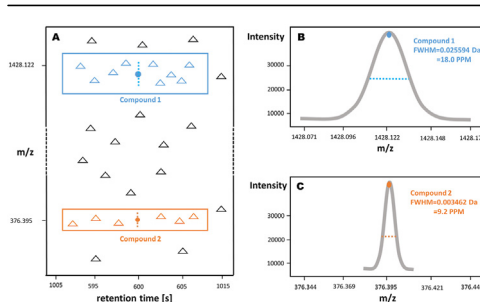
ARTICLE INFO

ABSTRACT

Liquid chromatography-mass spectrometry (LC-MS)-based lipidomics generates large datasets that need to be interpreted using high-performance data pre-processing tools such as XCMS, mzMine, and Progenesis. These pre-processing tools rely heavily on accurate peak detection, which depends on proper setting of the peak detection mass tolerance (PDMT). The PDMT is usually set with a fixed value in either ppm or Da units. However, this fixed value may result in duplicates or missed peak detection and

---

https://doi.org/10.1016/j.aca.2021.338674

inaccurate peak quantification. To improve the accuracy of peak detection, we developed the dynamic binning method, which considers peak broadening described by the physics of ion separation and sets the PDMT dynamically in function of *m/z*. In our method, the PDMT is proportional to $\left(\frac{m}{z}\right)^2$ for Fourier-transform ion cyclotron resonance (FTICR), to $\left(\frac{m}{z}\right)^{1.5}$ for Orbitrap and to *m/z* for Quadrupole time-of-flight (Q-TOF), and is a constant for Quadrupole mass analyzer. The dynamic binning method was implemented in XCMS [1,2], and the adopted source code is available in GitHub at https://github.com/xiaodfeng/DynamicXCMS. We have compared the performance of the XCMS implemented dynamic binning with different popular lipidomics pre-processing tools to find differential compounds. We generated set samples with 43 lipid internal standards that were differentially spiked to aliquots of one human plasma lipid sample using Orbitrap LC-MS/MS. The performance of various pipelines using matched parameter sets was quantified by a quality score system that reflects the ability of a pre-processing pipeline to detect differential peaks spiked at various concentrations. The quality score indicated that our dynamic binning method improves the quantification performance of XCMS (maximum p-value $9.8 \cdot 10^{-3}$ of two-sample Wilcoxon test) over its original implementation. We also showed that the XCMS with dynamic binning found differential spiked-in lipids better or with similar performance as mzMine and Progenesis do.

1. Introduction

In lipidomics, liquid chromatography-mass spectrometry (LC-MS) is commonly used for quantitative profiling because LC has a high separation efficiency and MS has a large measurement range and high specificity and sensitivity [3,4]. A typical LC-MS experiment generates a large amount of complex data that need accurate quantitative processing and powerful identification approaches to identify thousands of lipid species present in complex biological samples. Consequently, accurate data processing has become a major challenge in lipidomics [5].

Significant effort has been made to develop LC-MS(/MS) data processing tools. These include commercial tools such as Progenesis [6] developed by Nonlinear Dynamics and open-source tools such as mzMine [7,8], mzMine ADAP [9], XCMS [1,2], MS-DIAL [10,11], KniMet [12], OpenMS [13], metaX [14], LipidMatch [15], and MetaboAnalystR; [16]. In addition, online tools like XCMS online [17], MetaboAnalyst [18], PiMP my metabolome [19], and Workflow4Metabolomics [20] have also been developed.

These data processing tools typically contain the following processing modules [21]: transformation of data (such as resampling and smoothing), detection of peaks, correction of retention time, grouping of peaks across different samples, filling in missing data, annotation of isotope peak clusters with potential metabolite identity (represented by CAMERA [22]), normalization of quantification values, and differential statistical analysis with univariate and multivariate methods or other types of statistical analysis (Fig. A1). Among them, peak detection is one of the most critical steps in LC-MS(/MS) data pre-processing. Thus, it is crucial to define an algorithm that can distinguish between irrelevant signals from chemical or electronic noise and the compound's actual signal. The compound's signal is usually represented as a 2-dimensional Gaussian peak (Fig. A1b). Several peaks corresponding to different stable isotopes of the compound form an isotope cluster (Fig. A1d).

Algorithms based on extracted ion chromatogram (EIC) peak detection typically consist of EIC construction and chromatographic peak detection in the constructed EICs. *XCMS* [1] was developed in 2006 to facilitate LC-MS data pre-processing and is one of the most widely used tool for peak detection in metabolomics LC-MS datasets [2,23]. It includes three EIC-based peak detection algorithms suitable for LC-MS(/MS) pre-processing: *matchedFilter*, *centWave* [24], and *massifquant* [25]. Among them, *centWave* is the most frequently used peak detection algorithm.

Despite intensive efforts to develop and benchmark accurate peak detection, the peak detection and quantification performance of many algorithms is not sufficient for accurate quantification and often leads to detection of compounds not present in the sample [23] (i.e., false positives) or to miss (generally low abundant) existing compounds [26] (false negatives) in LC-MS(/MS) data. False-positive and false-negative results and inaccurate quantification may be due to incorrect parameterization of the algorithm. Setting a proper mass tolerance is essential for EIC construction (i.e., binning) used for peak detection. Although there are guidelines for setting the proper mass tolerance [21], the recommended mass tolerance is not always suitable for specific instrumental settings and a specific *m/z* range. For example, a range of 5−15 ppm for Orbitrap data generates a too broad EIC, which may mix signals from multiple compounds. The mass range (and peak width in *m/z* dimension) generally increases in function of *m/z* when using high-resolution mass analyzers such as TOF, FTICR, and Orbitrap instruments. Thus, *Mayers* et al. recommended using Dalton (Da) as unit for mass tolerance instead of parts per million (ppm) as the mass range variation is smaller with Da unit [27]. However, even if the mass tolerance is specified in Da, improperly setting a fixed mass tolerance may still cause duplicates or missed peak detection and inaccurate peak quantification. To address this problem, we dynamically adjusted the mass tolerance to construct EICs that improve the peak detection performance of *centWave* [28]. In the original *centWave* algorithm, mass tolerance is set in ppm and increases proportionally with *m/z*. This relationship is optimal for the TOF mass analyzer but not for the Orbitrap mass analyzer, in which the peak width and *m/z* in mass spectra are proportional to Ref. $\left(\frac{m}{z}\right)^{1.5}$ [29]. Our dynamic binning method of constructing EICs takes into account the correct $\left(\frac{m}{z}\right)^{1.5}$ proportional peak broadening in *m/z* for Orbitrap data in the peak detection step, leading to more precise quantification.

LC-MS(/MS) pre-processing workflows differ in algorithmic design and involve many parameters. Determining the correct parameters for peak detection during data analysis can be difficult. Lipidomic LC-MS(/MS) data pre-processing workflows and parameter sets need to be objectively evaluated to compare the performance of various pipelines [23,30]. In 2012, *Hoekman* et al. [31] introduced a scoring method that compares the performances

of different quantitative LC-MS/MS pre-processing workflows. The score quantifies the capacity of a pre-processing pipeline, with a given set of parameters, to detect differentially spiked compounds in a sample with the same composition (i.e., from the same aliquot of one background sample). In this dataset, the spiked compounds are not present in the background sample. We extended this scoring system to determine the distribution of scores, which allows pipeline performance to be compared using a non-parametric significance test. In this study, we used this modified scoring strategy to compare the performance of *centWave* peak picking with fixed and dynamic EIC construction tolerance in *XCMS*, as well as mzMine and Progenesis. Finally, we compared the performance of these LC-MS pre-processing workflows after optimizing each workflow's parameters to make them as comparable as possible, despite their algorithmic differences.

2. Theory

Fig. A1 shows the most critical aspects of the peak detection process, which consists of the following steps: EIC construction, peak detection, isotope pattern identification, adduct ion, isomer, and feature detection.

2.1. Setting proper mass tolerance for accurate EIC construction

Peak detection provides crucial information required for accurate compound identification and quantification, and accurate EIC construction is fundamental for accurate peak detection. Setting the mass tolerance properly is essential for selecting the ions of one isotope of a compound that forms the EIC. Fig. 1 shows how mass tolerance is related to the width of the selected rectangle. Although there is a detailed guideline for setting the proper mass tolerance in XCMS [21], the mass tolerance is highly dependent on the *m/z* and the type of mass analyzer. As shown in Fig. 1, *compound 1* is located at a high *m/z* of 1428.1220 Da, which has a large mass peak width, around 0.025594 Da (18.0 ppm), while *compound 2* is located at a low *m/z* of 376.3950 Da, which has a smaller mass peak width (and uncertainty) around 0.003462 Da (9.2 ppm). Thus, a low mass tolerance value may not be sufficient to cover the uncertainty of ion *m/z* fluctuation in higher *m/z*. As a result, intensity ions may be

missing from peaks and peak quantity may be underestimated, leading to peak splitting. If a larger mass tolerance value is selected, close mass traces can merge in lower *m/z,* leading to missing peaks and incorrect quantitative values for mixed peaks. Typically, this uncertainty (or variability) may result from two sources: mass fluctuation (MF) and mass dispersion (MD). MF is the fluctuation of the peak maxima in the mass spectrum, which can be observed between different subsequent MS1 mass spectra. Fig. A2 and A.3 show that MF exists in both low and high *m/z*. This is sometimes called mass accuracy by the vendor. For a well-calibrated Orbitrap, the MF should be less than 1 ppm at 200 Da [32]. MD can be estimated according to the relationship between the mass peak width in *m/z* and *m/z*. MD is proportional to $\left(\frac{m}{z}\right)^2$ for FTICR, to $\left(\frac{m}{z}\right)^{1.5}$ for Orbitrap and to *m/z* for Q-TOF, and is a constant value for Quadrupole [29].

This uncertainty is dependent on the *m/z*, so a fixed mass tolerance value for peak detection may result in peak merging and/or failure to detect peaks at a specific *m/z* range. An alternative is to use a dynamic mass tolerance according to the uncertainty of acquired ions. Thus, the peak detection mass tolerance (*PDMT*) should be set as a function of the MF (usually defined in ppm) and MD. The PDMT in Da (PDMT$_{Da}$) can be calculated by

$$PDMT_{Da} = MD_{Da} + MF \cdot \frac{m}{z} \cdot 10^{-6} \tag{1}$$

FTICR:

$$PDMT_{Da} = \left(\frac{A_{fti}}{2.35482}\right) \cdot \left(\frac{m}{z}\right)^2 + MF \cdot \frac{m}{z} \cdot 10^{-6} \tag{1a}$$

Orbitrap:

$$PDMT_{Da} = \left(\frac{A_{orb}}{2.35482}\right) \cdot \left(\frac{m}{z}\right)^{1.5} + MF \cdot \frac{m}{z} \cdot 10^{-6} \tag{1b}$$

Q-TOF:
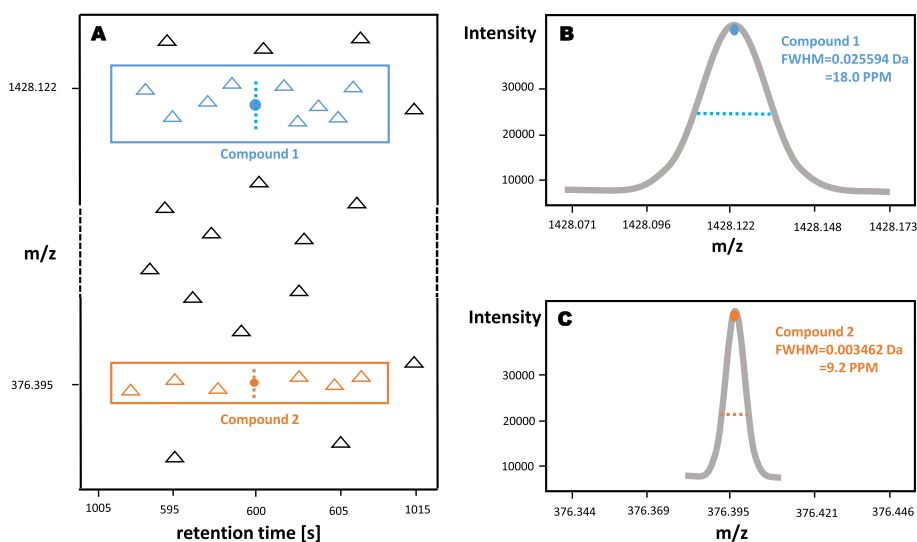
Fig. 1. Scheme showing the main aspects of the dynamic binning method demonstrating the change of peak width defined as *FWHM* (Full Width at Half Maximum) in function of *m/z*. (A) Scatter plot of detected peaks, with retention time in the x-axis and *m/z* in the y-axis. (B) Compound 1 is located at a high *m/z* 1428.1220 Da, which has a large *FWHM*, around 0.025594 Da (18.0 ppm). (C) Compound 2 is situated at a low *m/z* of 376.3950 Da, which has a small *FWHM*, around 0.003462 Da (9.2 ppm).

$$\text{PDMT}_{\text{Da}} = \left(\frac{A_{qtof}}{2.35482}\right) \cdot \left(\frac{m}{z}\right) + MF \cdot \frac{m}{z} \cdot 10^{-6} \tag{1c}$$

Quadrupole:

$$\text{PDMT}_{\text{Da}} = Constant \tag{1d}$$

While the peak detection mass tolerance in ppm (PDMT_{ppm}) can be calculated by

$$\text{PDMT}_{\text{ppm}} = \text{MD}_{\text{ppm}} + MF \tag{2}$$

In the above equations, the value of 2.35482 corresponds to $2\sqrt{2ln2}$. This constant relates the standard deviations (σ) and the mass peak width using Gaussian distribution. A_{orb}, A_{qtof} and A_{fti} are constants which relates m/z with MD for Orbitrap, Q-TOF, and FTICR data respectively when divided by 2.35482. In the following, we describe the equations for calculating A_{orb} in the Orbitrap mass spectrometer. The equations for A_{qtof} and A_{fti} can be found in Appendix A.

2.2. Estimation of the mass dispersion according to the mass peak width

The variation of the m/z defined by the MD is linked to the mass resolution or mass resolving power of a particular mass analyzer. We first defined the essential concepts necessary to express the MD in terms of m/z and mass resolution based on the theory of Hoffmann and Stroobant. [33]; these terms may be defined differently in other sources. [34]) These definitions are as follows: **mass peak width (Δm)**: full width at half maximum (*FWHM*) of mass spectral peak; **mass resolving power (R)**: the observed mass (m) divided by the mass peak width (Δm) at 50% height for an isolated single mass spectral peak, as illustrated in Eq. (3).

$$R = \frac{m}{\Delta m} \tag{3}$$

Therefore, the mass peak width (Δm) changes according to mass resolving power (R) and the observed mass (m), as illustrated in Eq. (4).

$$\Delta m = \frac{m}{R} \tag{4}$$

It follows from Eq. (2), by replacing mass (m) with the mass-to-charge ratio (m/z),

$$\Delta(m/z) = \frac{m/z}{R} \tag{5}$$

The underlying physical principle used to estimate m/z is different for different mass analyzers. In an Orbitrap mass spectrometer, the frequency (w) is directly linked to the m/z ratio, as illustrated in Eq. (6)

$$w = \sqrt{\frac{zek}{m}} \tag{6}$$

In which e represents the electron charge, and z represents the number of charges of the ions. The letter k represents the field curvature, which is a constant value. Thus,

$$\frac{m}{z} = ekw^{-2} \tag{7}$$

To find the relationship between the mass resolving power R and m/z, we use the derivative of the mass with respect to the angular frequency from Eq. (7) and derive Eq. (8) after rearrangement:

$$\frac{dm}{z} = -2ekw^{-3}dw \tag{8}$$

The ratio of mass-to-mass variation can then be obtained by dividing Eq. (7) by Eq. (8)

$$\frac{m}{dm} = \frac{w^{-2}}{-2w^{-3}dw} = \frac{w}{-2dw} \tag{9}$$

For a slight variation, the derivative operator (d) and the difference used in numerical calculations (Δ) are interchangeable. The mass resolving power of the Orbitrap is further obtained by introducing the expression of the angular frequency from Eq. (6) into Eq. (9) as follows:

$$R_{orb} = \frac{m}{\Delta m} = \frac{w}{-2\Delta w} = \frac{\sqrt{\frac{zek}{m}}}{-2\Delta w} = \frac{\sqrt{ek}\sqrt{\frac{z}{m}}}{-2\Delta w} = \frac{\sqrt{ek}}{-2\Delta w}\left(\frac{m}{z}\right)^{-0.5} \tag{10}$$

Where Δm and Δw are 50% of the peak width measured on the mass scale and frequency scale, respectively. Several publications have indicated that the peak width in the frequency domain is an almost constant value; the seminal work of Makarov pointed out that frequency errors are due to the construction of the instrument rather than the experimental conditions [35]. *Lössl* et al. [36] applied a fitting of Eq. (10) and showed an almost constant behavior of the coefficient of $(m/z)^{-0.5}$. Thus, for fixed acquisition times, Eq. (10) indicates the mass resolving power R_{orb} is inversely proportional to the square root of the m/z ratio [37]. The correlation can be expressed as:

$$R_{orb} \approx Ct \cdot \left(\frac{m}{z}\right)^{-0.5} \tag{11}$$

In which Ct is a constant value. Including Eq. (11) in Eq. (5), the expression of the mass peak width for Orbitrap (i.e., $\Delta(m/z)_{orb}$) becomes:

$$\Delta(m/z)_{orb} = \frac{m/z}{R_{orb}} \approx \frac{(m/z)^{1.5}}{Ct} \tag{12}$$

For a chosen reference m/z (m/z_r):

$$\Delta(mz_r)_{orb} \approx \frac{(m/z_r)^{1.5}}{Ct} \tag{13}$$

From Eq. (12) and Eq. (13) follows the equality:

$$\frac{\Delta(m/z)_{orb}}{\Delta(m/z_r)_{orb}} = \frac{(m/z)^{1.5}}{(m/z_r)^{1.5}} \tag{14}$$

Eq. (14) can be further transformed into:

$$\Delta(m/z)_{orb} = \left(\frac{\Delta(m/z_r)_{orb}}{(m/z_r)^{1.5}}\right)(m/z)^{1.5} = A_{orb}(m/z)^{1.5} \tag{15}$$

In Eq. (15), A_{orb} is considered a constant value that can be calculated from the reference m/z_r value and the reference resolving power R_r. Based on Eq. (5), it becomes:

$$A_{orb} = \frac{\Delta(m/z_r)_{orb}}{(m/z_r)^{1.5}} = \frac{m/z_r}{(m/z_r)^{1.5} \cdot R_r} = \frac{1}{(m/z_r)^{0.5} \cdot R_r} \tag{16}$$

In which R_r indicates the reference resolving power at reference m/z_r. MD is usually quoted in terms of standard deviations (σ)[38],

which can be estimated by the calculated mass peak width in *Eq. (15)*, according to the definition of a Gaussian-shaped function [39].

$$\sigma \approx \frac{\Delta (m/z)_{orb}}{2.35482} = \left(\frac{A_{orb}}{2.35482} \right) (m/z)^{1.5} \tag{17}$$

Finally, MD in Da (MD_{da}) can be expressed as,

$$MD_{da} \approx \sigma \approx \left(\frac{A_{orb}}{2.35482} \right) (m/z)^{1.5} \tag{18}$$

While the MD in ppm (MD_{ppm}) can be expressed as:

$$MD_{ppm} = \left(\frac{MD_{Da}}{m/z} \right) \cdot 10^6 \approx \left(\frac{A_{orb}}{2.35482} \right) \cdot (m/z)^{0.5} \cdot 10^6 \tag{19}$$

Eq (18) and (*19*) indicate that the mass dispersion changes as a function of *m/z*. According to *Eq. (1)* and *Eq. (2)*, the PDMT should also change as a function of *m/z*. The reference m/z (m/z_r) is set to 200 Da, and the reference mass resolving power (R_r) is set to 70,000. According to *Eq. (16)*, A_{orb} equals $1.0102 \cdot 10^{-6}$. The mass fluctuation (MF) is less than one ppm for a calibrated instrument at 200 Da [32]. We use MF = 1 ppm in the following equations. $PDMT_{Da}$ is defined according to *Eq. (1)* and *Eq. (18)*

$$PDMT_{Da} = 4.2897 \cdot 10^{-7} \cdot (m/z)^{1.5} + 1 \cdot (m/z) \cdot 10^{-6} \tag{20}$$

and $PDMT_{ppm}$ is defined according to *Eq. (2)* and *Eq. (19)*

$$PDMT_{ppm} = 0.42897 \cdot (m/z)^{0.5} + 1 \tag{21}$$

3. Materials and methods

3.1. Materials and methods

Materials. LC-MS grade acetonitrile (ACN), methanol (MeOH), isopropanol (IPA), and chloroform were purchased from Biosolve BV (Valkenswaard, The Netherlands). Ammonium formate (AmF), formic acid (FA), and *tert*-Butyl methyl ether (MTBE) were purchased from Sigma Aldrich (St. Louis, MO). Lipid standards were purchased from Avanti Polar Lipids, Inc. (Alabaster, AL). Heparin-anticoagulated plasma samples were obtained anonymously from an adult patients at the University Medical Center Groningen (UMCG) and were combined to generate a standard plasma sample as described previously [40]. The study design was in accordance with the current revision of the Helsinki Declaration (2013).

Deuterium lipid internal standard (IS) mixture preparation. 20 different deuterium-labeled lipid internal standard (IS), and 4 deuterium-labeled lipid IS premixes were selected to cover the major lipid classes and distributed evenly in *m/z* and retention time range. All lipid standard stock solutions were diluted with chloroform: MeOH (1:1, v/v) and mixed to generate a lipid IS mixture with optimized concentrations for each standard to acquire adequate signal intensity (as listed in Table A.1). The lipid IS mixture was used to create a dilution series where concentration ratios were set to a factor of two starting from concentration 1 up to concentration 1/16.

Lipid extraction. Plasma lipid extraction was performed following the protocol of *Matyash* et al. [41] with slight modifications. In brief, 60 µl of plasma was mixed with 300 µl of MeOH and sonicated for 10 min. Subsequently, 1000 µl MTBE was added, and the mixture was kept at 25 °C on a shaker (900 rpm) for 30 min. Phase separation was induced by adding 190 µl ultrapure water. Then the mixture was centrifuged at 3000 RCF for 10 min, and the 850 µl upper phase was transferred to a new tube. The re-extraction

was performed by adding 600 µl MTBE/MeOH/ultrapure water (10:3:2.5, v/v/v) into the lower phase, and 500 µl were collected after centrifugation to combine with the previous organic phase. The combined lipid extract solution was aliquoted into 6 tubes (190 µl per tube) to generate a plasma lipid matrix. Different concentrations of lipid standard mixture were added to the plasma lipid extract aliquots and dried in a vacuum centrifuge at 45 °C. The dried lipid extracts were resuspended with 30 µl chloroform:MeOH:MQ (60:30:4.5, v/v/v) and further diluted with 90 µl of IPA:ACN:MQ (2:1:1 v/v/v) for LC-MS analysis.

LC-MS analysis. LC-MS lipid analysis was performed on an Ultimate 3000 High-Performance UPLC coupled with a QExactive Orbitrap instrument (Thermo Fischer Scientific, Darmstadt, Germany). Chromatography separation was achieved with an Acquity UPLC CSH column [1.7 µm, 100 × 2.1 mm, (Waters Corporation, Milford, MA)] at 55 °C with a flow rate of 0.4 ml/min. Mobile phase A was composed of ultrapure water/acetonitrile 40:60 (v/v), 10 mM ammonium formate, and 0.1% formic acid. Mobile phase B contained ACN/IPA 10:90 (v/v) with 10 mM ammonium formate and 0.1% formic acid. The LC gradient was modified from *Damen* et al. [42] It started with 40% mobile phase B and increased to 43% mobile phase B in 2 min. The percentage of mobile phase B increased to 50% in the next 0.1 min and increased to 54% in the next 9.9 min. Mobile phase B increased to 70% in 0.1 min and to 99% in 5.9 min and was maintained at 99% for 1 min. The percentage of mobile phase B went back to 40% in 0.1 min, and the system was equilibrated for 3.9 min before the next run started. The MS was set for positive mode and data-dependent acquisition. A full MS scan ranging from 250 to 1750 Da was acquired at resolution 70,000 *FWHM* at 200 Da. The AGC target was set to $1 \cdot 10^6$. The maximum injection time was 50 ms and the MS1 scan was followed by up to 8 MS/MS events with a collision energy of 25 eV at resolution 17,500 *FWHM* at 200 Da. The precursor isolation window was set to 1.5 Da with a dynamic exclusion time of 6 s. The ionization settings were as follows: capillary voltage: +3.2 kV; capillary temperature: 320 °C; sheath gas/auxiliary gas: 60/20.

3.2. Computational methods

Quality score calculation. Fig. A.4 shows the experimental design of spiked-in lipidomics dataset preparation as well as the steps of data processing using different LC-MS preprocessing workflows. Different concentrations of lipid IS mixture spiked in one aliquoted human plasma lipid extract. Four-fold changes were generated by comparing IS 1 with IS 1/16 (fold change 16, FC16), IS 1/8 (fold change 8, FC8), IS 1/4 (fold change 4, FC4), IS 1/2 (fold change 2, FC2). Two replicates of all ISs without plasma lipid extract were used to identify the spiked-in ISs in the plasma background manually. t-statistics were calculated according to Hoekman et al., using the t-statistics of standard two independent samples *t*-test between pairs of these four-fold changes. [31] Based on these t-statistics, the features were ranked and quality scores were calculated. To get the distribution of these quality scores, we randomly selected from 9 to 2 out of the 10 replicates to calculate the t-statistics and the quality score. This selection was repeated 20 times resulting in 20 quality scores for each comparison.

P-value calculation. P-values were calculated to measure the significance level between the quality scores of the dynamic binning and fixed binning with a threshold of 7 ppm. This was calculated using a two-samples Wilcoxon non-parametric test.

IS adducts confirmation. Each IS described in materials and methods was annotated as 10 adduct types: M + H - H_2O, M + H, M + NH_4, M + Na, M + K, M + 2Na-H, 2 M + H, 2 M + NH_4, 2 M + Na, and 2 M + K, and the *m/z* value were calculated accordingly. According to the calculated *m/z* value, the EICs were

visualized. Based on these visualizations, we confirmed the presence of 159 IS adducts. The criteria for confirming that an EIC corresponded to an IS were: (1) the corresponding EIC should not show a bell-shaped peak signal in the blank samples; (2) the corresponding EIC should form a bell-shaped peak in both IS mixture with and without human plasma lipid extract background replicates (3) the apex of the bell-shaped signal (i.e., peak height); should pass the noise level of 10^5 ion counts in the IS mixture and in the highest spiked-in concentration in human plasma samples; (4) the EIC should be identified in both pure IS mixture and IS mixture spiked plasma samples in at least three spiked-in concentrations; (5) the EIC should follow the concentration fold changes in IS mixture spiked plasma samples in at least in three spiked-in concentrations. The EICs for the IS adducts are available in Appendix B. An example of EIC of confirmed IS adducts is shown in Fig. A.5.

IS fragment confirmation. After the collected LC-MS data were carefully inspected, we identified 416 IS fragments (in addition to the IS adducts described in the previous paragraph) and included these in the IS list. The criteria for selecting these IS fragments were the same as for IS adducts. The EICs for the IS fragments are available in Appendix B. An example of EIC of an IS fragment included in the analysis is shown in Fig. A.6.

Lipidomics data pre-processing and filtering. *Thermo Xcalibur® software* (version 3.2.63, Thermo Scientific, Waltham, MA) was used for data acquisition (Fig. A.4). The acquired .raw thermo files were used directly by *Progenesis QI* (version 2.1) for peak picking, grouping, and isotope filtration. The .raw data were further converted into .mzML format using *MSConvert* (Version: 3.0.18234) tool of ProteoWizard [43] package, and the *Vendor Peak Picking* filter was selected to export centroid data. The .mzML files were sent to *mzMine* (version 2.40.1) for peak picking, grouping, and isotope filtration. *XCMS* (standalone version 3.8.2, developed by Steffen Neumann et al.) also used .mzML files as input for peak picking and grouping for the dynamic and fixed binning methods. The results were further isotope filtered by CAMERA (version 1.43.2). The exported .csv files were used to assess the performance of the various LC-MS pre-processing tools. The fixed binning term reflected one parameter set in ppm that was used to linearly expand the EIC construction m/z range in function of m/z.

Dynamic binning. In *XCMS*, the *centWave* [28] algorithm is one of the most used peak detection algorithms and applies a mass tolerance for EIC construction in *ppm*. In this study, the *centWave* algorithm in XCMS source code was modified to implement the dynamic binning approach, which is available in GitHub at https://github.com/xiaodfeng/DynamicXCMS. The m/z of the dataset ranges from 250 Da to 1750 Da; thus, the modified XCMS source code's mass tolerance was changed according to the presented theory.

4. Results and discussion

4.1. Model validation

To validate our model to set the EIC construction threshold, we visually inspected the sampling interval in the profile mode dataset. Sampling intervals are the distances in m/z between two adjacent acquisition points. They are determined by an algorithm implemented on the electronic card of the Orbitrap instrument and ensure a constant number of sampling points for each peak in the mass spectrum. Since the number of sampling points in the Orbitrap mass spectra is constant, the increase in sampling interval should be in accordance with our theory on increasing peak width. Fig. 2a shows that the manually checked sampling intervals (red dots in the figure) increase proportionally to $(m/z)^{1.5}$. The exact values of the sampling intervals are shown in Appendix C. This

increase in the sampling intervals set on the acquisition instrument and found in the profile mass spectrum follows the increase in peak width in the function of m/z described by Eq. (15).

The exact number of sampling intervals defining the m/z interval (which contains all ion intensity information for a peak in profile mode) may vary depending on instrumental setting such as resolution. Empirically, this number can be estimated as 4. *Myers* et al. [27] used centroided data to show that most peaks in a mass spectra span between 0 and 3 sampling intervals. They chose a fixed m/z tolerance value of 0.01 Da for Orbitrap and 0.02 Da for Q-TOF. This setting ensures that most of the centroid maxima of the analyzed isotopic ions in consecutive mass spectra (Fig. 1a) are included in the EIC construction. The m/z tolerance value is estimated according to 1 sampling interval at the highest m/z. In our model, the $PDMT_{Da}$ calculated according to Eq. (20) is between 1 and 2 sampling intervals (Fig. 2a). This dynamic setting for peak detection mass tolerance could avoid merging EICs at low m/z values and avoid splitting EICs of one isotope peak at high m/z values.

In the implemented dynamic binning method, the m/z tolerance value in *centWave* was set to the dynamic *PDMT* value, while for the fixed binning method, 7, 13, or 19 ppm (Fig. 2b) were used for peak detection in the entire m/z range.

4.2. Comparison of dynamic and fixed binning in XCMS

Figs. 1 and 2 show that proper binning is important for the EIC construction. Because of the less accurate binning of EIC, the quantification of detected peaks may be less accurate when fixed binning is used. The peaks in Fig. 3 can be detected using the fixed binning and dynamic binning methods, but the two methods yield different quantification (Table 1). The parameter was set following
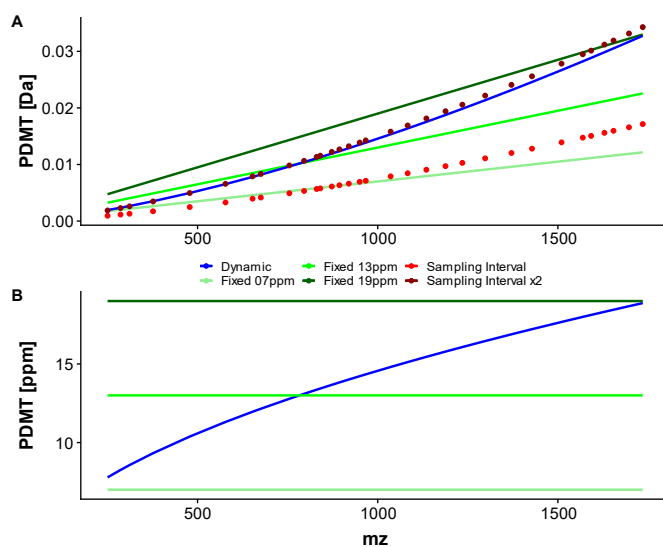

Fig. 2. Validation of the theoretical model calculating PDMT in the function of m/z with sampling intervals of mass spectra from Orbitrap mass analyzer. The red dots (sampling interval, see definition in the text) represent the manually determined sampling intervals and the dark red dots (sampling interval × 2) represent double the manually determined sampling intervals. (a) The manually checked sampling intervals (red dots) increase proportionally to $(m/z)^{1.5}$.This increase in sampling intervals follows equation (15) of our theory. (b) The PDMT m/z tolerance in the function of m/z as implemented in the dynamic binning method using the *centWave* XCMS peak detection algorithm. The dynamic binning threshold is shown with the blue line, while the fixed binning method initially implemented in XCMS with a 7, 13, or 19 ppm thresholds are shown as light green, green, and dark green lines, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

the guideline for setting XCMS parameters [21]. The lower part of the figure shows the scatter plot of *m/z* and retention time, while the upper part shows the peak shape as a scatter plot of intensity and retention time. For example, the highlighted point with a black square in Fig. 3a is included only by dynamic binning, while from the LC profile (upper part of Fig. 3a), it is evident that the point is part of the bell-shaped curve. As a result, the area of the peak detected by dynamic binning is higher than the one detected by fixed binning (i.e., $1.278 \cdot 10^7$ compared with $1.241 \cdot 10^7$). Fig. 3b shows an example of peak splitting using fixed binning peak detection. The green and red rectangles indicate that the fixed binning method detected this peak twice at retention times of 415 and 424 s, while the dynamic binning method correctly detected only one peak at retention time of 415 s, as indicated by the blue rectangle. To illustrate the peak splitting clearly, we zoomed on the green and red rectangles in Fig. 3b, as shown in Fig. 3c and d, respectively.

The difference between the number of detected peaks was slight. Based on the presented data of a few examples, it is not apparent how dynamic binning performs at the whole dataset level. To observe differences in quantification between dynamic binning and fixed binning for all the peaks, we used the Bland−Atman plot of quantification with dynamic and static EIC binning, as shown in Fig. 4. Fig. 4 shows the comparison between

concentrations 1/16, 1/8, 1/4, 1/2, and 1. In each concentration, the x-axis shows the mean $\log_2$ of peak intensities obtained with dynamic and fixed binning (7 ppm), and the y-axis represents the differences of the $\log_2$ of the peak intensities obtained with both methods (i.e., $\log_2(I_{Dynamic})_i - \log_2(I_{Fixed})_i$ for the ith peak). For each concentration, bias (dashed blue line) was higher for dynamic binning (i.e., dashed blue line above 0). This was confirmed by a significantly more dots above the upper limit of agreement (dashed green line) compared with the lower limit of agreement (red line), both expressed as two standard deviations from the overall center (dashed blue line).

Dynamic binning improved XCMS performance primarily by more accurate quantification. Identification was also slightly improved. The examples only existed at higher *m/z* ranges, indicating high uncertainty. At these ranges, the fixed binning method may fail to detect several ISs because limited ions are included for EIC construction, so the constructed EICs may not pass the XCMS's filtration parameters such as *firstBaselineCheck, prefilter,* or *snthresh*. Fig. A.7 shows two ISs that cannot be detected using a fixed 7 ppm *mzTolerance* for the EIC building. For example, the peak in Fig. A.7a can be detected by using 13 ppm, 19 ppm with a fixed value of *mzTolerance*, and by applying the dynamic binning method. This peak should be detected in XCMS because 1) it matches M + Na adduct of IS 24:1(3)-14:1 CA; 2) it contains intensive signal in the
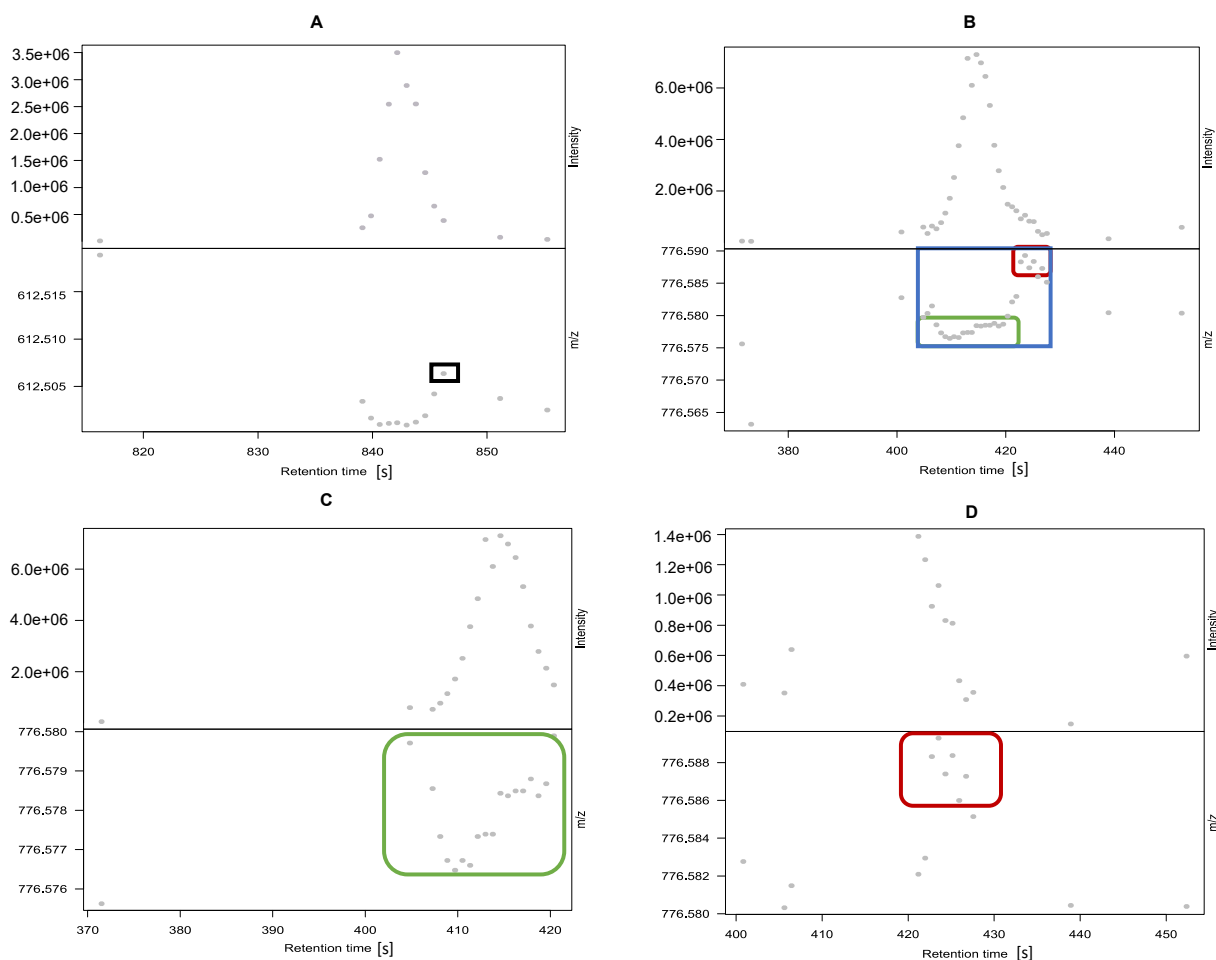

Fig. 3. Examples of improved quantification by the dynamic binning algorithm. All peaks in the figure can be detected using both fixed binning and dynamic binning, but the quantified values are different. (A) The highlighted point is included only with dynamic binning, while it is missing with fixed binning using the 7 ppm threshold. This results in a higher peak area with dynamic binning than with fixed binning. (B) An example of peak splitting in fixed binning with a 7 ppm threshold. The green and red rectangles indicate the two peaks detected by fixed binning and the blue rectangle indicates one peak correctly identified by dynamic binning. (C) Enlarged EIC related to the green rectangle in Fig. 3b. (D) Enlarged EIC related to the red rectangle in Fig. 3b. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)
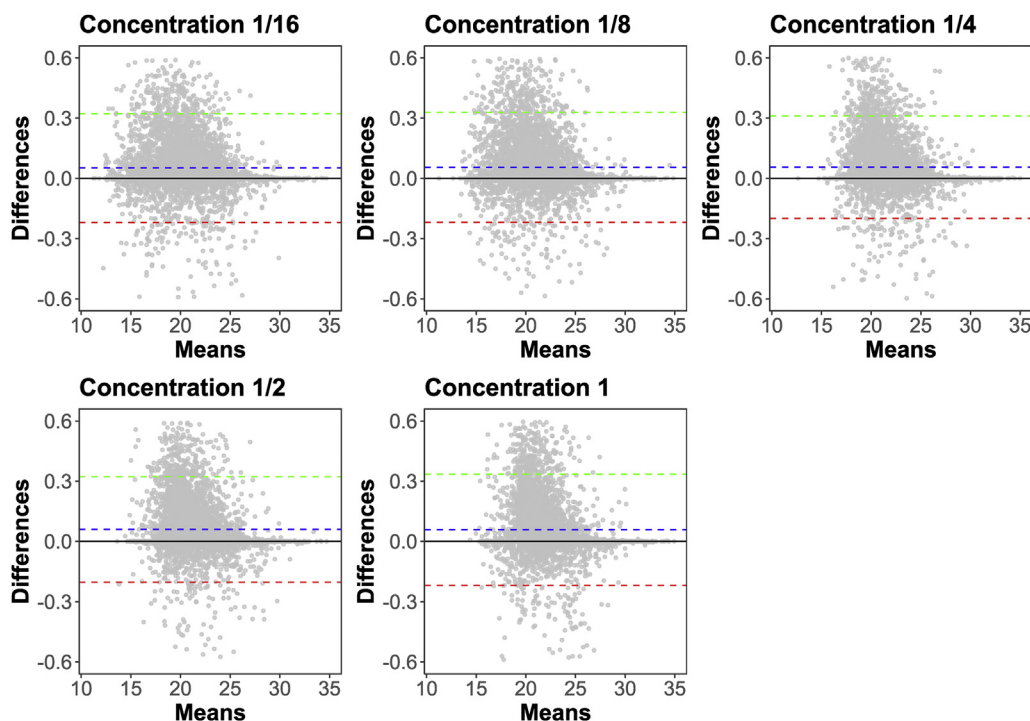
Fig. 4. Bland-Altman plot showing differences in quantification between dynamic and fixed binning with a 7 ppm EIC construction threshold in human plasma lipidomics datasets spiked with IS at 1/16, 1/8, 1/4, 1/2 and 1 dilution level. The x-axis shows the mean $\log_2$ peak intensities obtained with dynamic and fixed binning. The y-axis represents the differences of $\log_2$ peak intensities obtained with both methods. The bias of the two methods is shown with a blue line. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

m/z range between 1676.236 and 1676.239 in Da; 3) it follows a bell-shaped curve with tailing in the retention time (between 1099 and 1134 s); and 4) it has high-intensity values, namely the area under the chromatographic peak is $1.648 \cdot 10^8$, which becomes $1.646 \cdot 10^8$ after baseline correction and the maximum intensity measured in the m/z and retention time rectangle containing the peak is $6.441 \cdot 10^6$ (Table 1). The peak in Fig. A.7b shows similar detection behavior with respect of peak detection by both methods as the peak in Fig. A.7a.

To further assess if the higher values reflect more accurate quantification, we used a quality score introduced by *Hoekman* et al. [31] that assesses the ability to identify differential spiked-in compounds in samples with the same complex lipid background. Fig. 5. shows the distribution of quality scores for the studied pipelines for four fold-change concentrations (16, 8, 4, and 2). As indicated by the asterisks, with all four fold change concentrations, dynamic binning achieved a significantly higher quality score than fixed binning did with 7 ppm (p-values $2.4 \cdot 10^{-4}, 1.9 \cdot 10^{-5}, 2.7 \cdot 10^{-4}$ and $9.8 \cdot 10^{-3}$ of two-samples Wilcoxon tests). Dynamic binning improved the ability of XCMS to identify the biomarker from the background, even when compared with other fixed PDMT thresholds such as 13 and 19 ppm. These results showed that dynamic binning overperformed original XCMS implementation. Apart from sampling 9 out of 10 replicates in Fig. 5, we also sampled from 8 to 2 out of 10 replicates, which gave similar results (Fig. A.8). The use of a lower number of replicates results in lower quality scores. This can be explained by a lower number of replicates leading to lower t-statistics and earlier dilution of spiked-in related features in the

Table 1
Examples of missing peaks and improved quantification.

Method	m/z	m/z min	m/z max	rt	rt min	rt max	int o	int b	max o
				Details of the missing peaks in Fig. A.7a					
Fixed 7 ppm	–	–	–	–	–	–	–	–	–
Dynamic	1676.238	1676.236	1676.239	1118	1099	1134	$1.648 \cdot 10^8$	$1.646 \cdot 10^8$	$6.441 \cdot 10^6$
				Details of the missing peaks in Fig. A.7b					
Fixed 7 ppm	–	–	–	–	–	–	–	–	–
Dynamic	1668.240	1668.239	1668.242	1050	1045	1057	$1.528 \cdot 10^7$	$1.480 \cdot 10^7$	$1.771 \cdot 10^6$
				Details of the peaks with improved quantification in Fig. 3a					
Fixed 7 ppm	612.5013	612.5009	612.5042	843	838	846	$1.241 \cdot 10^7$	$1.241 \cdot 10^7$	$3.498 \cdot 10^6$
Dynamic	612.5014	612.5009	612.5063	843	838	847	$1.278 \cdot 10^7$	$1.276 \cdot 10^7$	$3.498 \cdot 10^6$
				Details of the peaks with improved quantification in Fig. 3b					
Fixed 7 ppm	776.5780	776.5774	776.5785	415	404	423	$5.628 \cdot 10^7$	$5.613 \cdot 10^7$	$7.296 \cdot 10^6$
Fixed 7 ppm	776.5882	776.5860	776.5893	424	422	428	$3.798 \cdot 10^6$	$3.798 \cdot 10^6$	$1.063 \cdot 10^6$
Dynamic	776.5783	776.5765	776.5883	415	404	425	$5.758 \cdot 10^7$	$5.716 \cdot 10^7$	$7.296 \cdot 10^6$

m/z = mass to charge value; m/z min, m/z max = minimal and maximum m/z.
rt = measured retention time in seconds; rt min, rt max = minimal and maximum rt.
int o, int b = intensity calculated based on area before and after baseline correction; max o = intensity calculated based on height.
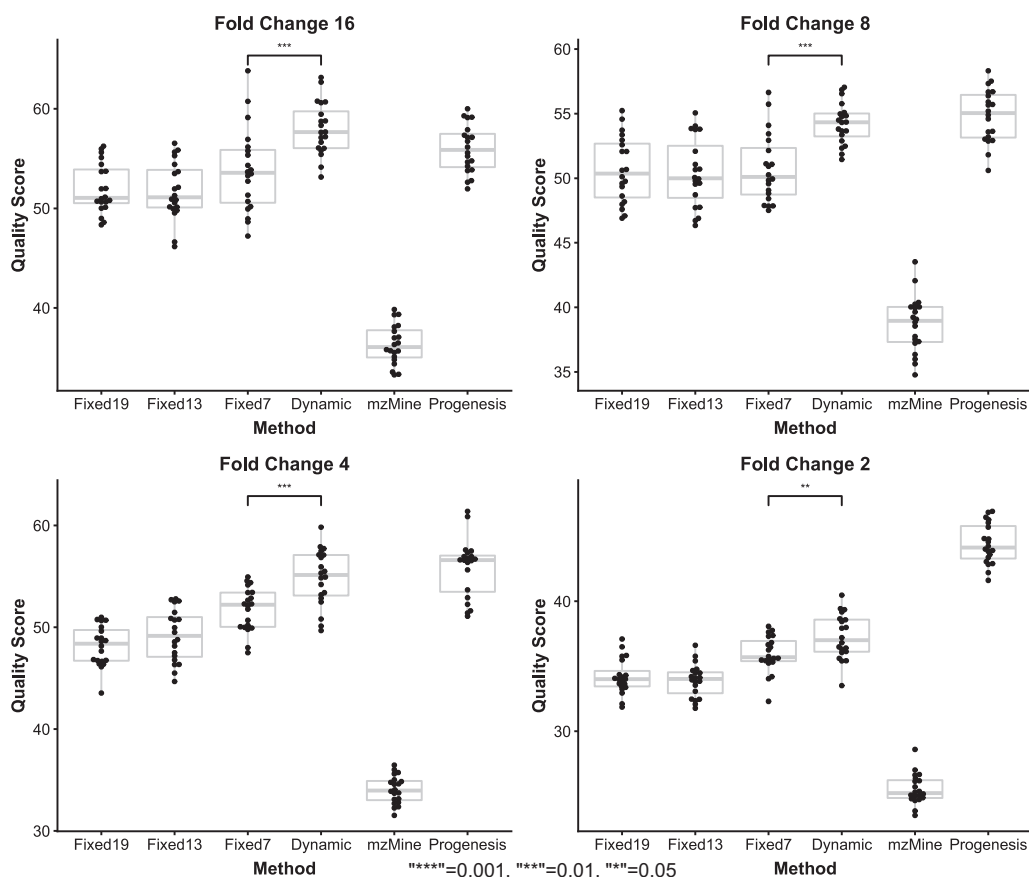
Fig. 5. Box plot showing quality score, which measures the pipeline's ability to identify differential spiked-in lipid related peaks. The figure shows the quality scores distribution of XCMS with dynamic binning and fixed binning (19, 13 and 7 ppm), mzMine and Progenesis for four different fold changes (16, 8, 4, and 2), where a dilution series of 1/16, 1/8, 1/4, and 1/2 is compared with a dilution series of 1. The asterisks indicate the p-values ($2.4 \cdot 10^{-4}$, $1.9 \cdot 10^{-5}$, $2.7 \cdot 10^{-4}$ and $9.8 \cdot 10^{-3}$) of two-sample Wilcoxon non-parametric tests performed using dynamic binning and fixed binning with a threshold of 7 ppm. Dynamic binning performs better than fixed binning and dynamic binning improves XCMS's quantification performance. The performance of dynamic binning is much better than that of mzMine and similar to that of Progenesis.

ranked discriminating feature list, leading to lower cumulative scores. Furthermore, using lower number of replicates increase the score differences between Progenesis and the other workflows. This can be explained by the replicate sampling procedure, which takes peak quantities from a quantitative table obtained using an average LC-MS map constructed with all replicates by Progenesis. Progenesis is a GUI based program and cannot be included in a sampling loop where peak picking is performed in each loop iteration using an average LC-MS map constructed only from the replicates selected in a particular iteration. In contrast, XCMS performs peak picking individually in each chromatogram, which is not influenced by other chromatograms. Peak matching performance could be affected by the number of replicates, but this has much weaker effect than peak picking on an average chromatogram constructed with all available replicates.

The quality score assessing the identification of spiked-in features in a stable molecular background is a cumulative score; therefore, it is worth exploring the ranking of spiked-compound-related features in identified features that are differential between spiking levels. The cumulative quality scores in the function of a ranked list of discriminating features are shown in Fig. 6. In this figure, a binary heat map is included above the x-axis. Each row in this heat map indicates a differential feature ranked (using t-statistics) from the most differential on the left to the least differential on the right. The colored bars indicate features related to the spiked-in compounds, which contribute to the cumulative quality scores as indicated by the y-axis of the line plots. In contrast, the

white bars indicate features that correspond to compounds from the background sample, which are non-discriminatory between the different spiked-in concentration levels. These non-discriminatory features will lower the increase in quality scores for subsequent lower-ranked IS features. For example, the dynamic binning contains more colored bars than white bars among the most discriminant features on the left side of the plots. This indicates that dynamic binning detected more differentially spiked-in IS-related features as reflected in the quality score. This indicates that dynamic binning detects differential features more accurately than fixed binning does. The higher number of colored cells on the left side of the plot results in the higher cumulative score of dynamic binning than fixed binning at 7, 13, and 19 ppm.

4.3. Comparison between XCMS, mzMine, and progenesis

To consistently compare between XCMS, mzMine, and Progenesis, we used the quantitative feature tables containing matched features across all chromatograms in the dataset. The parameters for peak detection, grouping, and isotope filtration are shown in Table A.2, A.3, and A.4. Progenesis and dynamic binning XCMS showed the highest quality scores for spiked-in levels FC16, FC8, FC4, and FC2, followed by fixed binning XCMS and mzMine (Fig. 5). These results indicate that Progenesis and dynamic binning XCMS have a strong ability to find spiked-in compound-related discriminating features and identify fewer background peaks with the same level across all spiked-in levels.
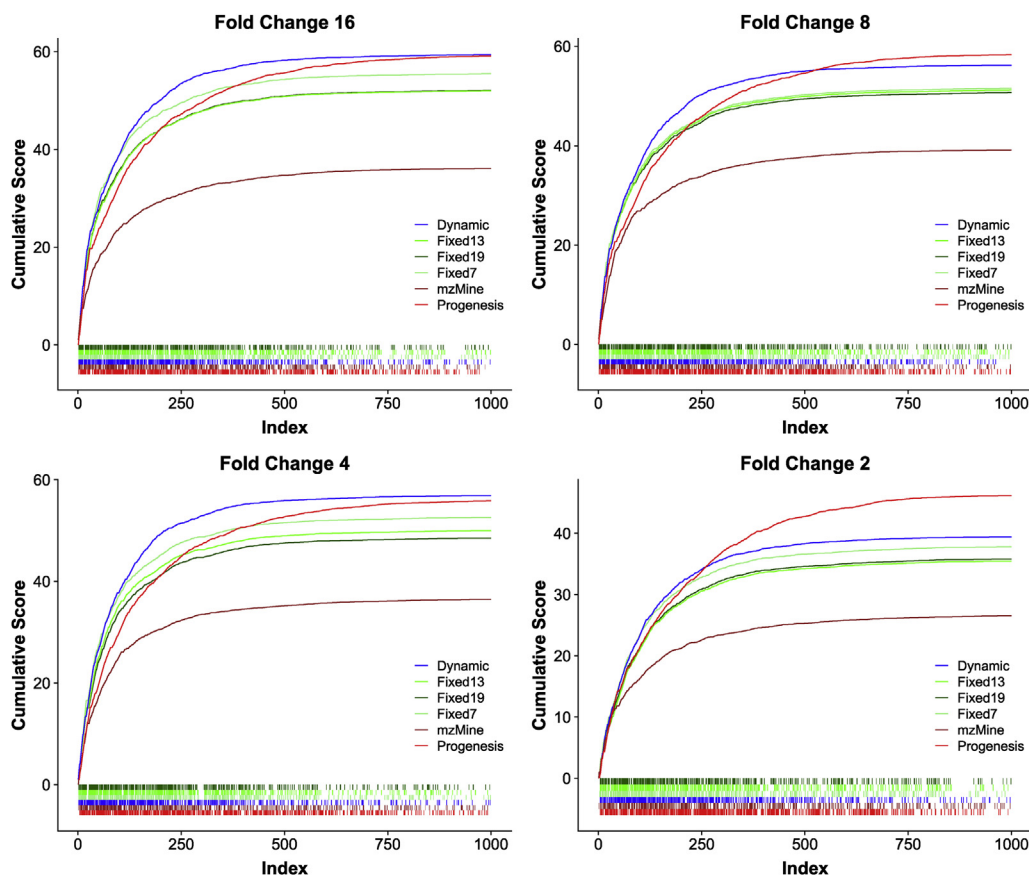
Fig. 6. Plots showing the development of the cumulative quality scores for XCMS with dynamic binning, XCMS with fixed binning (7, 13 and 19 ppm), mzMine and Progenesis allowing to assess the quantification differences of the various LC-MS pre-processing pipelines. The cumulative quality score is based on t-statistics ranked features using samples spiked-in at different levels. The theoretical fold change between spiking levels is indicated in the title of each plot. A binary heat map above the x-axis shows the ranked list of discriminating features from the left (most discriminating) to the right (least discriminating). In this heatmap, matched features related to the IS are shown as colored cells, and the white cells show the features not associated with IS. The features corresponding to the white cells are not contributing to the cumulative quality score increase and will lower the rise of the cumulative score for any subsequent higher rank IS-related features. The dynamic binning method contains more colored cells than white ones (indicating more IS-related features) among the most discriminant features on the left of the heat map.

Apart from finding accurately features with different levels, it is also worth evaluating the precision of the quantification of different methods. Fig. 7 shows the scatter plot of $\log_2$ fold change (x-axis) versus $\log_{10}$ average abundance of a matched features (y-axis). The y-axis range of XCMS and mzMine are similar (5-11), while Progenesis is different (3-9), indicating differences in the quantification metric. To assess if pipeline preserves the quantitative order of peaks, we constructed EICs (Fig. A.9) of three ISs using the raw data. We labeled these three compounds as Compound A (17:0-17:1-17:0 D5 TG), Compound B (18:1-d7 MG), and Compound C (20:0-20:1-20:0 D5 TG), where compound A has the highest area abundance, compound B has the middle area abundance and compound C the lowest. XCMS and mzMine calculated abundances as the area under the EIC curve, while Progenesis used the "raw abundance," which is the sum of all ion intensities in the rectangle defined by the retention time and *m/z* of all isotopes of a compound (Table A.5). The rectangle coordinates were identified in an aggregated ion intensity map combined from all chromatograms after retention time alignment and the area is used in all individual chromatograms to sum up the corresponding intensities. The summed intensities in individual chromatogram are used as quantitative values for the same compound and used for downstream statistical analysis. Compounds A, B, and C shared the same decreasing trend of abundance in XCMS and mzMine (i.e., $26.206 \cdot 10^7$, $3.214 \cdot 10^7$ and $1.980 \cdot 10^7$ for compounds A, B and C

respectively in XCMS). At the same time, we observed no decreasing trend from compound A, B and C in Progenesis (i.e., $12.425 \cdot 10^7$, $0.089 \cdot 10^7$ and $23.802 \cdot 10^7$ for compound A, B and C, respectively). There are many possible reasons for this difference in abundance. One reason may be the different algorithms used to calculate abundance; XCMS and mzMine use the *CAMERA* algorithm for isotope filtration, which usually uses the highest isotope peaks to calculate abundance. In contrast, Progenesis uses the average of detected isotope clusters to calculate abundance. The peak areas of compounds A and B should differ by around one order of magnitude, which is accurately captured by the dynamic peak picking of XCMS and mzMine, while this difference is almost two orders of magnitude in data pre-processed by Progenesis. Another explanation could be that the aggregate LC-MS map is constructed from aligned chromatograms of all fold changes. In this setup, the highest spiking level determines the area rectangle where the ion intensities for all identified isotopes are added together and this rectangle is larger than that should be optimal for lower spiking levels features. Therefore, the rectangle may include ions that do not belong to the spiking features, which may change the order of intensity.

The x-axis of Fig. 7 reflects the fold change of detected features, and each IS-related feature is plotted four times according to FC2, FC4, FC8, and FC16 (labeled in green, red, yellow, and blue, respectively). In the plots, the x-axis of the green, red, yellow, and

blue dots are located around the theoretical $\log_2$ fold change of 1, 2, 3, and 4, respectively. The number of ISs detected by XCMS, mzMine, and Progenesis is shown in the Venn diagram. Progenesis uniquely detected the most ISs (24 ISs), followed by XCMS (5 ISs), and mzMine (1 IS). These numbers are relatively low compared with the number of detected features common to all pipelines (412 ISs). The features from XCMS, mzMine, and Progenesis that are not related to ISs were labeled as plasma-related features (grey dots) and are distributed around $\log_2$ fold change of 0 in the x-axis. The fluctuation of plasma-related features in Progenesis was smaller compared with XCMS and mzMine. This is most probably because the aggregate aligned map was used for feature detection, which detects compounds relatively well when variability between samples is low but may have difficulty detecting compounds when variability between samples is higher. The better performance of the dynamic binning peak picking is shown in Fig. 6; dynamic binning had higher cumulative scores than Progenesis did for the most significantly different features (i.e., features in the left of the plot). The Progenesis peak picking approach uses the aggregate map, which can detect less abundant peaks because the average signal across multiple LC-MS maps is used for peak picking, which helps to identify a larger number of less differential peaks. This is indicated by the higher cumulative scores of Progenesis compared with those of dynamic binning XCMS, taking into account all features for the score calculation.

To further evaluate the joint technical variance of LC-MS measurements and data pre-processing of the three methods, the distributions of the coefficient of variations (CV) were calculated based on the intensities of plasma-related features. The full range of detected CV is shown as a violin plot in Fig. A.10. The density plot of CV between 0 and 1.5 (i.e., 150%) is shown in Fig. 8. These figures show that feature quantification was most accurate with Progenesis indicated by lower CVs, indicating lower variability of features in replicates of five individual concentration levels. In contrast, CVs were more widely distributed in mzMine and XCMS. Progenesis had better performance because it used the aligned aggregate map for peak picking and feature detection, which can detect features more accurately with low variance, as discussed previously. Plasma-related features have the same level in each sample, which does not reflect the large biological variability of compounds in clinical samples.

5. Conclusion

Setting accurate PDMT is crucial for accurate peak detection. To this end, we suggested and implemented a dynamic method for a more accurately set the PDMT. Namely, the PDMT is proportional to $\left(\frac{m}{z}\right)^2$ for FTICR, to $\left(\frac{m}{z}\right)^{1.5}$ for Orbitrap, and to m/z for Q-TOF, and is a constant value for Quadrupole. This method improved XCMS performance by reducing the number of missing peaks (Fig. A.7) and by improving the accuracy of quantification (Figs. 3 and 4). As a result, dynamic binning had a higher quality score (Fig. 5, maximum p-value: $9.8 \cdot 10^{-3}$) and cumulative score (Fig. 6) than the fixed binning method. This indicates that the ability of XCMS to find meaningful compounds (i.e., lipid biomarkers) is improved with dynamic binning.

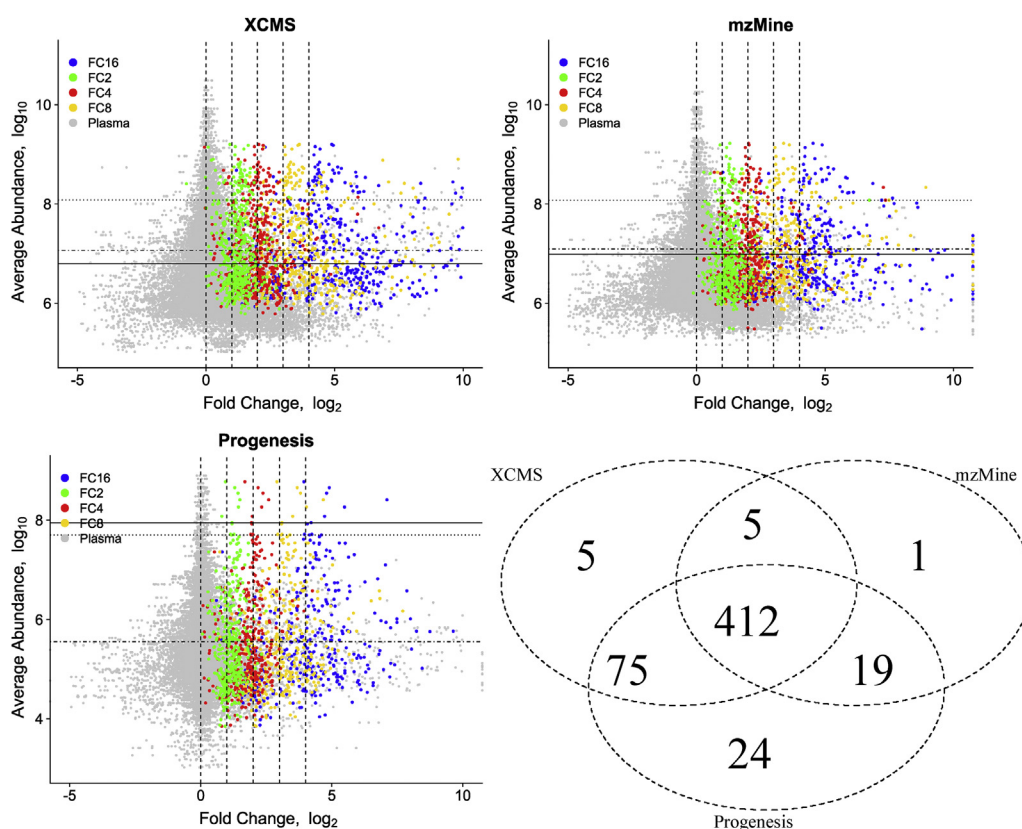Progenesis achieved similarly high cumulative quality scores in

Fig. 7. Ratio distributions between two spiking levels for each LC-MS pre-processing method. The figure shows the scatter plot of fold change (x-axis, in $\log_2$) versus average abundance in the whole data set (y-axis, in $\log_{10}$). The y-axis range is similar in XCMS and mzMine (5-11 orders of magnitude), but different in Progenesis (3-9 orders of magnitude). In the x-axis, each IS is plotted four times with FC2, FC4, FC8, and FC16 labeled in green, red, yellow, and blue, respectively. The green, red, yellow, and blue dots are located around $\log_2$ fold change of 1, 2, 3, and 4, respectively, on the x-axis. The number of ISs detected by XCMS, mzMine, and Progenesis is shown in the Venn diagram (bottom right plot). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)
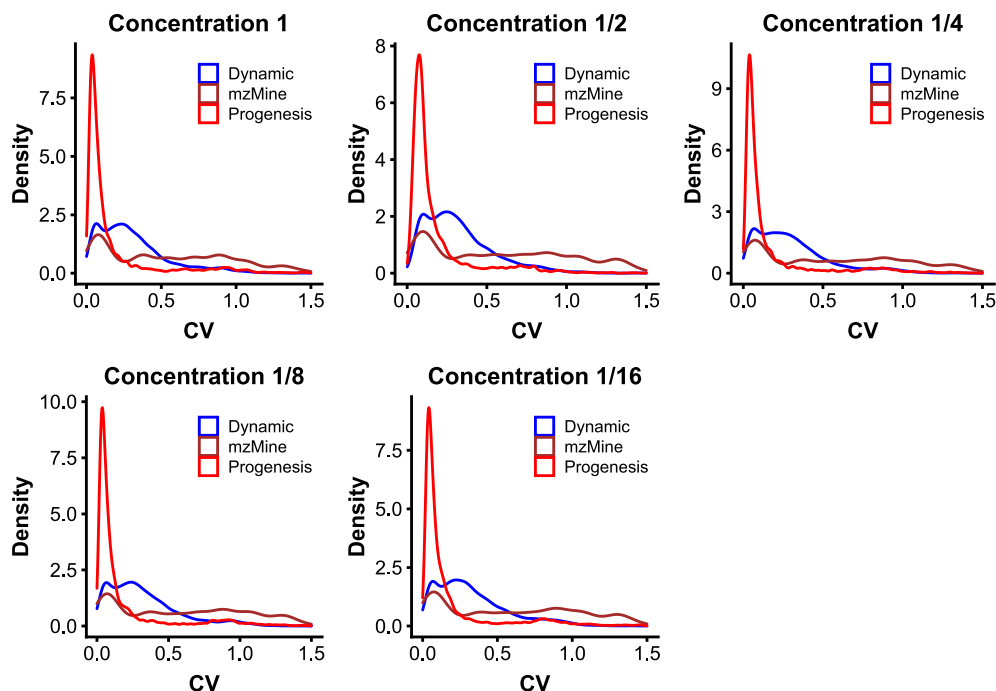
Fig. 8. Density plot of coefficient of variation (CV) for 3 different LC-MS pre-processing methods using plasma-related features. The plot shows the density of CV between 0 and 1.5 (150%). Progenesis shows the most precise feature quantification indicated by lower CVs in replicates of five individual concentration levels, while mzMine and XCMS have a wider CV distribution.

all four fold changes as the dynamic binning XCMS method because aligned aggregate LC-MS was used for feature detection. In addition to the cumulative quality score, we evaluated the quantification stability of XCMS, mzMine, and Progenesis using the $\log_2$ fold change distribution (Fig. 7) and CV distribution (Fig. 8) based on the intensities of detected features. Progenesis had the lowest CVs for plasma-related features, indicating lower variability of features in replicates of five individual concentration levels, followed by XCMS and mzMine. The lower variability of Progenesis (Fig. A.10) may be due to the unique quantification algorithm. However, this also altered the order of quantitative values of ISs (Table A.5 and Fig. A.9), which compounds have higher variability between samples. The use of aligned aggregate LC-MS maps for feature detection areas may help to quantify peaks with low variability but may also lead to artifacts if LC-MS maps are misaligned and/or if features have larger variability. Dynamic binning had a similar peak picking performance in XCMS as Progenesis, without using aligned aggregate maps and without the aforementioned artifacts. Therefore, we recommend the dynamic binning approach for quantitative profiling in lipidomics and metabolomics. In this research, we only generated a spike-in dataset using Orbitrap LC-MS/MS and demonstrated the improvement of quantification of XCMS with dynamic binning. According to our dynamic binning theory, the peak width in m/z is linearly related to m/z for Q-TOF data. Therefore, setting PDMT for EIC construction in ppm in the original XCMS should provide accurate quantification for Q-TOF data. In this work, we did not generate spiked-in dataset with FTICR and quadrupole mass analyzer but have provided the theory of considering peak width change and setting EIC construction tolerance (PDMT) properly for accurate quantification.

CRediT authorship contribution statement

Xiaodong Feng: Software, Formal analysis, (XCMS dynamic binning and statistical postprocessing) was performed, Writing and revision was made, Visualization of figure was made. **Wenxuan Zhang:** Investigation, (human serum collection, lipid dataset sample preparation and LC-MS/MS analysis) was made, Writing and revision was made. **Folkert Kuipers:** Writing and revision was made, Supervision of XF was made. **Ido Kema:** Writing and revision was made, Supervision of XF was made. **Andrei Barcaru:** Conceptualization, Methodology, (concept of dynamic binning) of the study was developed, Writing and revision was made, Supervision of XF was made. **Péter Horvatovich:** Conceptualization, Methodology, (concept of dynamic binning) of the study was developed, Writing and revision was made, Supervision of XF was made, Visualization of figure was made.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendices. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.aca.2021.338674.

References

[1] C.A. Smith, E.J. Want, G. O'Maille, R. Abagyan, G. Siuzdak, XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification, Anal. Chem. 78 (2006) 779–787, https://doi.org/10.1021/ac051437y.

[2] N.G. Mahieu, J.L. Genenbacher, G.J. Patti, A roadmap for the XCMS family of software solutions in metabolomics, Curr. Opin. Chem. Biol. 30 (2016) 87–93, https://doi.org/10.1016/j.cbpa.2015.11.009.

[3] T. Cajka, O. Fiehn, LC–MS-Based lipidomics and automated identification of lipids using the LipidBlast in-silico MS/MS library, in: S.K. Bhattacharya (Ed.), Lipidomics, Springer New York, New York, NY, 2017, pp. 149–170, https://doi.org/10.1007/978-1-4939-6996-8_14.

[4] T. Hu, J.-L Zhang, Mass-spectrometry-based lipidomics, J. Separ. Sci. 41 (2018) 351–372, https://doi.org/10.1002/jssc.201700709.

[5] P.S. Niemelä, S. Castillo, M. Sysi-Aho, M. Orešič, Bioinformatics and computational methods for lipidomics, J. Chromatogr., B 877 (2009) 2855–2862, https://doi.org/10.1016/j.jchromb.2009.01.025.

[6] Progenesis_QI_User_Guide_2_3.pdf, (n.d.). http://storage.nonlinear.com/webfiles/progenesis/qi/v2.3/user-guide/Progenesis_QI_User_Guide_2_3.pdf (accessed May 20, 2018).

[7] T. Pluskal, S. Castillo, A. Villar-Briones, M. Orešič, MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data, BMC Bioinf. 11 (2010) 395, https://doi.org/10.1186/1471-2105-11-395.

[8] M. Katajamaa, J. Miettinen, M. Orešič, MZmine: toolbox for processing and visualization of mass spectrometry based molecular profile data, Bioinformatics 22 (2006) 634–636, https://doi.org/10.1093/bioinformatics/btk039.

[9] X. Du, A. Smirnov, T. Pluskal, W. Jia, S. Sumner, Metabolomics data preprocessing using ADAP and MZmine 2, in: S. Li (Ed.), Comput. Methods Data Anal. Metabolomics, Springer US, New York, NY, 2020, pp. 25–48, https://doi.org/10.1007/978-1-0716-0239-3_3.

[10] H. Tsugawa, T. Cajka, T. Kind, Y. Ma, B. Higgins, K. Ikeda, M. Kanazawa, J. VanderGheynst, O. Fiehn, M. Arita, MS-DIAL: data independent MS/MS deconvolution for comprehensive metabolome analysis, Nat. Methods 12 (2015) 523–526, https://doi.org/10.1038/nmeth.3393.

[11] H. Tsugawa, K. Ikeda, M. Takahashi, A. Satoh, Y. Mori, H. Uchino, N. Okahashi, Y. Yamada, I. Tada, P. Bonini, Y. Higashi, Y. Okazaki, Z. Zhou, Z.-J. Zhu, J. Koelmel, T. Cajka, O. Fiehn, K. Saito, M. Arita, M. Arita, A lipidome atlas in MS-DIAL 4, Nat. Biotechnol. 38 (2020) 1159–1163, https://doi.org/10.1038/s41587-020-0531-2.

[12] S. Liggi, C. Hinz, Z. Hall, M.L. Santoru, S. Poddighe, J. Fjeldsted, L. Atzori, J.L. Griffin, KniMet: a pipeline for the processing of chromatography–mass spectrometry metabolomics data, Metabolomics 14 (2018), https://doi.org/10.1007/s11306-018-1349-5.

[13] M. Rurik, O. Alka, F. Aicheler, O. Kohlbacher, Metabolomics data processing using OpenMS, in: S. Li (Ed.), Comput. Methods Data Anal. Metabolomics, Springer US, New York, NY, 2020, pp. 49–60, https://doi.org/10.1007/978-1-0716-0239-3_4.

[14] B. Wen, Z. Mei, C. Zeng, S. Liu, metaX: a flexible and comprehensive software for processing metabolomics data, BMC Bioinf. 18 (2017), https://doi.org/10.1186/s12859-017-1579-y.

[15] J.P. Koelmel, N.M. Kroeger, C.Z. Ulmer, J.A. Bowden, R.E. Patterson, J.A. Cochran, C.W.W. Beecher, T.J. Garrett, R.A. Yost, LipidMatch: an automated workflow for rule-based lipid identification using untargeted high-resolution tandem mass spectrometry data, BMC Bioinf. 18 (2017) 331, https://doi.org/10.1186/s12859-017-1744-3.

[16] J. Chong, J. Xia, MetaboAnalystR: an R package for flexible and reproducible analysis of metabolomics data, Bioinformatics (2018), https://doi.org/10.1093/bioinformatics/bty528.

[17] R. Tautenhahn, G.J. Patti, D. Rinehart, G. Siuzdak, XCMS Online: a web-based platform to process untargeted metabolomic data, Anal. Chem. 84 (2012) 5035–5039, https://doi.org/10.1021/ac300698c.

[18] J. Chong, O. Soufan, C. Li, I. Caraus, S. Li, G. Bourque, D.S. Wishart, J. Xia, MetaboAnalyst 4.0: towards more transparent and integrative metabolomics analysis, Nucleic Acids Res. 46 (2018) W486–W494, https://doi.org/10.1093/nar/gky310.

[19] Y. Gloaguen, F. Morton, R. Daly, R. Gurden, S. Rogers, J. Wandy, D. Wilson, M. Barrett, K. Burgess, PiMP my metabolome: an integrated, web-based tool for LC-MS metabolomics data, Bioinformatics 33 (2017) 4007–4009, https://doi.org/10.1093/bioinformatics/btx499.

[20] F. Giacomoni, G. Le Corguillé, M. Monsoor, M. Landi, P. Pericard, M. Pétéra, C. Duperier, M. Tremblay-Franco, J.-F. Martin, D. Jacob, S. Goulitquer, E.A. Thévenot, C. Caron, Workflow4Metabolomics: a collaborative research infrastructure for computational metabolomics, Bioinformatics 31 (2015) 1493–1495, https://doi.org/10.1093/bioinformatics/btu813.

[21] E.M. Forsberg, T. Huan, D. Rinehart, H.P. Benton, B. Warth, B. Hilmers, G. Siuzdak, Data processing, multi-omic pathway mapping, and metabolite activity analysis using XCMS Online, Nat. Protoc. 13 (2018) 633–651, https://doi.org/10.1038/nprot.2017.151.

[22] C. Kuhl, R. Tautenhahn, C. Böttcher, T.R. Larson, S. Neumann, CAMERA: an integrated strategy for compound spectra extraction and annotation of LC/MS data sets, Anal. Chem. 84 (2012) 283–289, https://doi.org/10.1021/ac202450g.

[23] J.B. Coble, C.G. Fraga, Comparative evaluation of preprocessing freeware on chromatography/mass spectrometry data for signature discovery, J. Chromatogr. A 1358 (2014) 155–164, https://doi.org/10.1016/j.chroma.2014.06.100.

[24] R. Tautenhahn, C. Böttcher, S. Neumann, Highly sensitive feature detection for high resolution LC/MS, BMC Bioinf. 9 (2008) 504, https://doi.org/10.1186/1471-2105-9-504.

[25] C.J. Conley, R. Smith, R.J.O. Torgrip, R.M. Taylor, R. Tautenhahn, J.T. Prince, Massifquant: open-source Kalman filter-based XC-MS isotope trace feature detection, Bioinformatics 30 (2014) 2636–2643, https://doi.org/10.1093/bioinformatics/btu359.

[26] A. Rafiei, L. Sleno, Comparison of peak-picking workflows for untargeted liquid chromatography/high-resolution mass spectrometry metabolomics data analysis, Rapid Commun. Mass Spectrom. 29 (2015) 119–127, https://doi.org/10.1002/rcm.7094.

[27] O.D. Myers, S.J. Sumner, S. Li, S. Barnes, X. Du, One step forward for reducing false positive and false negative compound identifications from mass spectrometry metabolomics data: new algorithms for constructing extracted ion chromatograms and detecting chromatographic peaks, Anal. Chem. 89 (2017) 8696–8703, https://doi.org/10.1021/acs.analchem.7b00947.

[28] P. Du, W.A. Kibbe, S.M. Lin, Improved peak detection in mass spectrum by incorporating continuous wavelet transform-based pattern matching, Bioinforma. Oxf. Engl. 22 (2006) 2059–2065, https://doi.org/10.1093/bioinformatics/btl355.

[29] F. Suits, B. Hoekman, T. Rosenling, R. Bischoff, P. Horvatovich, Threshold-avoiding proteomics pipeline, Anal. Chem. 83 (2011) 7786–7794, https://doi.org/10.1021/ac201332j.

[30] O.D. Myers, S.J. Sumner, S. Li, S. Barnes, X. Du, Detailed investigation and comparison of the XCMS and MZmine 2 chromatogram construction and chromatographic peak detection methods for preprocessing mass spectrometry metabolomics data, Anal. Chem. 89 (2017) 8689–8695, https://doi.org/10.1021/acs.analchem.7b01069.

[31] B. Hoekman, R. Breitling, F. Suits, R. Bischoff, P. Horvatovich, msCompare: a framework for quantitative analysis of label-free LC-MS data for comparative candidate biomarker studies, Mol. Cell. Proteomics 11 (2012), https://doi.org/10.1074/mcp.M111.015974. M111.015974.

[32] Orbitrap LC-MS Comparison Chart - NL, (n.d.). https://www.thermofisher.com/uk/en/home/industrial/mass-spectrometry/liquid-chromatography-mass-spectrometry-lc-ms/lc-ms-systems/orbitrap-lc-ms/orbitrap-lc-ms-comparison-chart.html (accessed September 1, 2019).

[33] E. de Hoffmann, V. Stroobant, Mass Spectrometry: Principles and Applications, in: J. Wiley (Ed.), third ed., 2007. Chichester, West Sussex, England ; Hoboken, NJ.

[34] K.K. Murray, R.K. Boyd, M.N. Eberlin, G.J. Langley, L. Li, Y. Naito, Definitions of terms relating to mass spectrometry (IUPAC Recommendations 2013), Pure Appl. Chem. 85 (2013) 1515–1609, https://doi.org/10.1351/PAC-REC-06-04-06.

[35] A. Makarov, Electrostatic axially harmonic orbital Trapping: A high-performance technique of mass analysis, Anal. Chem. 72 (2000) 1156–1162, https://doi.org/10.1021/ac991131p.

[36] P. Lössl, J. Snijder, A.J.R. Heck, Boundaries of mass resolution in native mass spectrometry, J. Am. Soc. Mass Spectrom. 25 (2014) 906–917, https://doi.org/10.1007/s13361-014-0874-3.

[37] R.H. Perry, R.G. Cooks, R.J. Noll, Orbitrap mass spectrometry: instrumentation, ion motion and applications, Mass Spectrom. Rev. 27 (2008) 661–699, https://doi.org/10.1002/mas.20186.

[38] A.G. Brenton, A.R. Godfrey, Accurate mass measurement: terminology and treatment of data, J. Am. Soc. Mass Spectrom. 21 (2010) 1821–1835, https://doi.org/10.1016/j.jasms.2010.06.006.

[39] E.W. Weisstein, Gaussian Function, (n.d.). http://mathworld.wolfram.com/GaussianFunction.html (accessed November 6, 2019).

[40] A. Gil, W. Zhang, J.C. Wolters, H. Permentier, T. Boer, P. Horvatovich, M.R. Heiner-Fokkema, D.J. Reijngoud, R. Bischoff, One- vs two-phase extraction: re-evaluation of sample preparation procedures for untargeted lipidomics in plasma samples, Anal. Bioanal. Chem. (2018) 1–12, https://doi.org/10.1007/s00216-018-1200-x.

[41] V. Matyash, G. Liebisch, T.V. Kurzchalia, A. Shevchenko, D. Schwudke, Lipid extraction by methyl-*tert*-butyl ether for high-throughput lipidomics, J. Lipid Res. 49 (2008) 1137–1146, https://doi.org/10.1194/jlr.D700041-JLR200.

[42] C.W.N. Damen, G. Isaac, J. Langridge, T. Hankemeier, R.J. Vreeken, Enhanced lipid isomer separation in human plasma using reversed-phase UPLC with ion-mobility/high-resolution MS detection, J. Lipid Res. 55 (2014) 1772–1783, https://doi.org/10.1194/jlr.d047795.

[43] A cross-platform toolkit for mass spectrometry and proteomics | Nature Biotechnology, (n.d.). https://www.nature.com/articles/nbt.2377 (accessed February 4, 2020).