

# Lipid Data Analyzer: unattended identification and quantitation of lipids in LC-MS data

Jürgen Hartler<sup>1,2</sup>, Martin Trötzmüller<sup>3</sup>, Chandramohan Chitraju<sup>2</sup>, Friedrich Spener<sup>2</sup>, Harald C. Köfeler<sup>3</sup> and Gerhard G. Thallinger<sup>1,4,\*</sup>

<sup>1</sup>Institute for Genomics and Bioinformatics, Graz University of Technology, <sup>2</sup>Institute of Molecular Biosciences, University of Graz, <sup>3</sup>Core Facility for Mass Spectrometry, Center for Medical Research, Medical University of Graz and <sup>4</sup>Core Facility Bioinformatics, Austrian Center for Industrial Biotechnology (ACIB GmbH), Graz, Austria

Associate Editor: Martin Bishop

## ABSTRACT

**Motivation:** The accurate measurement of the lipidome permits insights into physiological and pathological processes. Of the present high-throughput technologies, LC-MS especially bears potential of monitoring quantitative changes in hundreds of lipids simultaneously. In order to extract valuable information from huge amount of mass spectrometry data, the aid of automated, reliable, highly sensitive and specific analysis algorithms is indispensable.

**Results:** We present here a novel approach for the quantitation of lipids in LC-MS data. The new algorithm obtains its analytical power by two major innovations: (i) a 3D algorithm that confines the peak borders in m/z and time direction and (ii) the use of the theoretical isotopic distribution of an analyte as selection/exclusion criterion. The algorithm is integrated in the Lipid Data Analyzer (LDA) application which additionally provides standardization, a statistics module for results analysis, a batch mode for unattended analysis of several runs and a 3D viewer for the manual verification. The statistics module offers sample grouping, tests between sample groups and export functionalities, where the results are visualized by heat maps and bar charts. The presented algorithm has been applied to data from a controlled experiment and to biological data, containing analytes distributed over an intensity range of  $10^6$ . Our approach shows improved sensitivity and an extremely high positive predictive value compared with existing methods. Consequently, the novel algorithm, integrated in a user-friendly application, is a valuable improvement in the high-throughput analysis of the lipidome.

**Implementation and availability:** The Java application is freely available for non-commercial users at <http://genome.tugraz.at/lda>. Raw data associated with this manuscript may be downloaded from *ProteomeCommons.org Tranche* using the following hash: ZBh3nS5bXk6l/Vn32tB5Vh0qnMpVIW71HBFFQqM0RmdF4/4HcnH3Wggh9kU2teYVotM1JWwHleMHqSS/bc2yYNFmyUAAAAAACIDQ==

**Contact:** Gerhard.Thallinger@tugraz.at

**Supplementary information:** Supplementary data are available from *Bioinformatics* online.

Received on October 1, 2010; revised on December 11, 2010; accepted on December 13, 2010

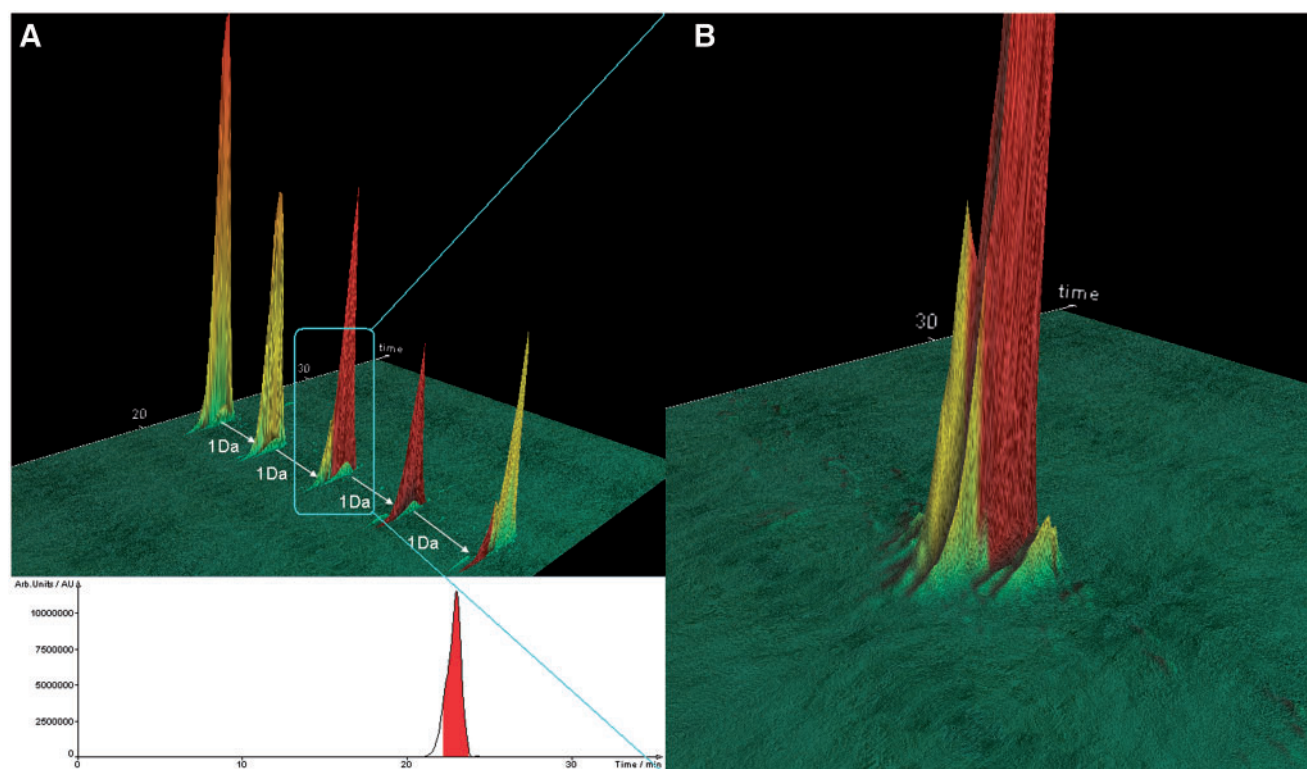
## 1 INTRODUCTION

Advances in lipidomics technologies utilizing mass spectrometry have led to a rapid increase in the number, size and rate at which datasets are generated. LC-MS especially bears the potential for monitoring quantitative changes in hundreds of lipids simultaneously. In order to deal with the acquired data automated and reliable software tools are required.

Currently available lipidomics MS-quantitation tools can be classified in two groups, each extracting 2D subsets of MS data: one group (Ejsing *et al.*, 2006; Haimi *et al.*, 2006, 2009; Leavell and Leary, 2006; Song *et al.*, 2007) requires m/z profiles or m/z spectra; and the other group (Katajamaa *et al.*, 2006; Pluskal *et al.*, 2010) extracts chromatograms for quantitation. A representative of the first group is LIMSAs which can analyse m/z spectra only (Haimi *et al.*, 2006, 2009). It utilizes an additional program (SECD) to extract the m/z spectra. The user has to define manually a trapezoid in the m/z-time map of the LC-MS data containing the lipid series of interest (a collection of lipids having the same amount of C atoms). This extracted m/z spectra serve as input for the LIMSAs algorithm. In contrast, mzMine2 (Katajamaa *et al.*, 2006; Pluskal *et al.*, 2010), a program from the second group, provides several algorithms to automatically extract chromatograms for processing. Both programs select the peaks for an analyte based on the exact m/z value. However, an LC-MS peak is a 3D object, whereupon each of its data points is characterized by retention time, m/z ratio and intensity. Consequently, all of the available tools lose information from a whole dimension (in the first group by the m/z profile extraction and the other group by the chromatogram extraction), and thus decrease the ability to clearly distinguish between closely overlapping peaks (Fig. 1). Furthermore, overlapping peaks occur frequently in lipidomic data due to the particular nature of lipids. The chemical structures of lipids can differ just by one double bond, resulting in a mass difference of 2 Da between the analytes. Thus, the base peak (+0 isotope) of the analyte has almost the same mass as the second isotopic peak (+2 isotope) of the analyte with one additional double bond. Furthermore, elution times are usually very close due to the physicochemical similarity of the analytes.

In this article, we present a novel algorithm that addresses the difficulties specific to lipidomics MS-data quantitation by two novel approaches: (i) exact peak border confinement and (ii) accounting for the theoretical isotopic distribution for peak selection. The novel border confinement increases the ability to detect overlaps and reduces their effects. The theoretical isotopic distribution is

\*To whom correspondence should be addressed.



**Fig. 1.** LDA 3D-view on MS data. (A) Shows the 3D view at the top of the picture and the 2D chromatogram view at the bottom. In the 3D view, the isotopic peaks (+0, +1 and +2) of TG56:6 and TG56:7, plus the +0 peak of TG56:5 are depicted, whereupon the peaks stained in red correspond to the quantified peaks of TG56:6. The 2D view shows the extracted chromatogram at the m/z value of TG56:6. In the chromatogram the overlapping +2 isotopic peak of TG56:7 is not distinguishable from the main peak of TG56:6; resulting in one peak. (B) Shows the zoomed 3D view of the overlap of the +2 peak of TG56:7 with the +0 peak of TG56:6. In 3D the two peaks are clearly discernable. The red stained part of the overlapping peaks indicates the part that has been used for the quantitation of TG56:6.

used as an exclusion and selection criterion and as such increases specificity. These two principles allow accurate and highly specific peak identification, which is required for the unattended high-throughput analysis of lipids. To assess the performance of the new algorithm, we (i) conducted a controlled experiment to test the accuracy of the quantitation and (ii) applied the algorithm to biological data to test its sensitivity and positive predictive value.

## 2 SYSTEM AND METHODS

The Lipid Data Analyzer (LDA) software is a stand-alone, platform-independent Java application, using Java3D (<https://java3d.dev.java.net>) for visualization.

### 2.1 Theoretical isotopic distribution

The theoretical isotopic distribution is calculated for each mass shift with the aid of probability theory, so that the probability for an isotopic mass peak is the sum of the mutual combinations of atomic isotopes, which result in the +x mass shift:

$$p(+x) = \sum p(\text{mutualCombination}) \quad (1)$$

For the theoretical isotopic distribution, the reference values are then determined based on the measured base peak according to the formula (Supplementary Section 1):

$$\text{Area}(+x) = \frac{p(+x)}{p(+0)} \text{Area}(+0) \quad (2)$$

### 2.2 Controlled experiment

The controlled experiment consisted of five mixtures (prepared as triplicates) of the analytes TG54:0, TG54:1, TG54:2 and TG54:3, and TG48:0 as reference value, which was present at the same concentration (5  $\mu\text{M}$ ). In the first mix, the analytes were present at the same concentration (2  $\mu\text{M}$ ) to test the ionization efficiencies of the analytes, while in mix 2–5 the analytes were mixed at different concentrations to test the quantitation at different intensities. The relative concentrations of the analytes were: TG54:3 : TG54:2 = 2 : 1; TG54:2 : TG54:1 = 5 : 1; TG54:1 : TG54:0 = 10 : 1. Mixes 2–5 correspond to a dilution series with the following ratios: mix2 : mix3 = 2 : 1; mix2 : mix4 = 5 : 1; mix2 : mix5 = 10 : 1.

Data were acquired by MS (more details about sample preparation and MS can be found in Supplementary Section 2) and analysed with LDA, LIMSA and mzMine2. The standard parameters of LIMSA and mzMine2 were inappropriate for high-resolution MS data (too few analytes were identified), and thus the parameters were adapted accordingly (Supplementary Section 3). The analytes showed ammonium and ammonium-acetonitrile adducts, and the sum of both adducts was taken as the quantitative measure. The analysis with LDA was performed in an automated way relying only on the m/z value. In contrast, LIMSA required manual extraction of the profiles for the automated quantitation, and mzMine2 information on retention time. Following that, the ratios of the dilution series were calculated and the average deviation from the expected ratio was taken as quality parameter:

$$\text{AverageDeviation} = \sum_{i=1}^n \frac{\text{ratio}(i)}{n \cdot \text{idealRatio}(i)} \quad (3)$$

## 2.3 Biological experiment

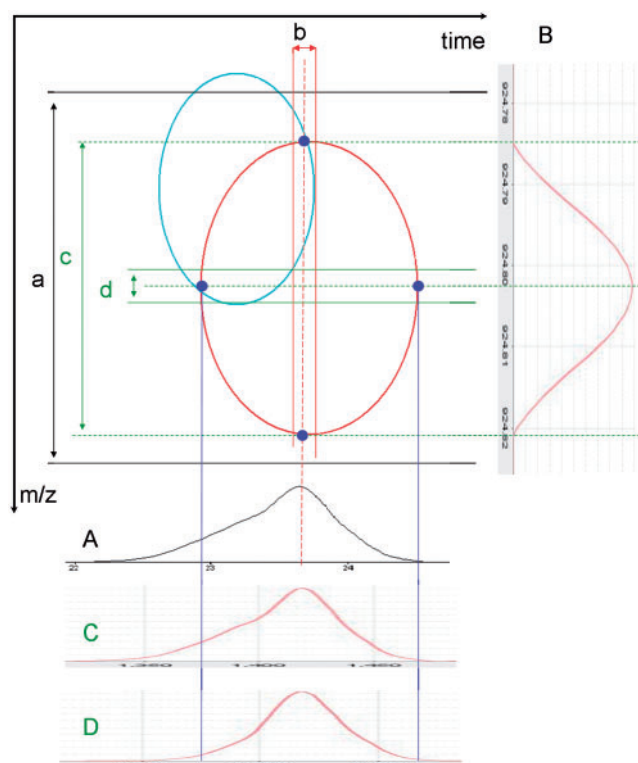
For the biological data, three groups of C57BL male mice ( $n = 3$  each) aged 12 weeks were maintained on a regular light-dark cycle (14 h light, 10 h dark) and fed on either standard laboratory chow diet (4.5% w/w crude fat) or high fat diet (34% w/w crude fat). For fasted samples, mice fed on chow diet were fasted over night for 16 h. The aim was to characterize the TG species of lipid droplets (LD) isolated from primary hepatocytes of mice subjected to the different feeding regimes. Primary hepatocytes were isolated by collagenase digestion according to Riccalton-Banks *et al.* (2003) with some modifications (Supplementary Section 4). Primary hepatocytes were lysed by nitrogen cavitation and LDs were isolated by ultra centrifugation. Lipids were extracted according to Folch *et al.* (1957). The sample was split into two parts, since LDs contain mainly TG (Blouin *et al.*, 2010) and the chromatographic separation and peak shape would be very poor for TGs without dilution, whereas other lipids would be hardly detectable in a diluted sample. One part of the extract was diluted 1:66 and served for the analysis of TGs, and the other fraction remained undiluted for the analysis of phospholipids and sphingomyelin. The extracts were dissolved in  $\text{CHCl}_3/\text{MeOH}$  (1:1, v/v) and acquired by LC-MS as described in Supplementary Section 2. First, the peaks were manually verified (visualization with the implemented 3D viewer). A peak was considered valid if it occurs at the expected retention time (each additional double bond in a molecular species lowers the retention time by about 0.5–1.5 min). Subsequently, TG data were automatically analysed by LDA and mzMine2 and the results were compared. While LDA performs the automated quantitation based on the  $m/z$  value of the analyte only, mzMine2 requires additionally information about the retention time (cf. controlled experiment). The data from undiluted samples were analysed with LDA only to show the applicability to molecule classes other than TG.

## 3 ALGORITHM IMPLEMENTATION

### 3.1 3D algorithm

The novel algorithm for accurate peak confinement works on the basis of several chromatogram and profile extractions, in which the chromatogram/profile extraction is based on an adapted version (Hartler *et al.*, 2007) of the ASAPRatio algorithm (Li *et al.*, 2003) with a changed peak border detection method. This new method detects peak borders based on abrupt changes in the gradient of a peak curve (see Section 5 and SFig. 1 in the Supplementary Material). The chromatogram and profiles are used to determine four peak border points. An ellipse through these border points is fitted, forming the border for the peak (Supplementary Section 6). Raw intensities inside the ellipse are used to calculate the area of the peak. This approach minimizes the effect of potential overlaps and does not contain smoothing artefacts, since it works on raw intensities.

Figure 1 shows two overlapping peaks which cannot be separated by the conventional 2D methods. In Figure 2, these two peaks are depicted from a top-view on the  $m/z$  map and the novel 3D algorithm is illustrated schematically. First, the algorithm extracts a chromatogram with a broad  $m/z$  range and identifies the time point at the peak summit. Second, at the time point of the summit, an  $m/z$  profile with a narrow time range is extracted. The algorithm determines the peak summit again and detects the two border points of the profile peak, which already form points for the ellipse fit in  $m/z$  direction. Then, two chromatograms are extracted; the first one uses the two profile border points as  $m/z$  range for the extraction; the second one uses a very narrow range around the  $m/z$  value of the profile peak summit. Now, in both chromatograms peak border



**Fig. 2.** Illustration of the novel algorithm from a top-view on the  $m/z$  map. The bigger ellipse represents the peak to be quantified; the smaller one is the overlapping one. First, a standard chromatogram 'A' with a broad  $m/z$  range 'a' is extracted (no overlap distinguishable). Second, at the time point with the highest intensity an  $m/z$ -profile 'B' with a narrow range 'b' is extracted. Third, the  $m/z$  borders and the  $m/z$  value with the highest intensity are used to extract chromatogram 'C' with range 'c' (uses borders found in 'B') and 'D' with narrow range 'd'. 'C' and 'D' are used to identify borders in the time range. Now four border points of the peak are determined (dots on the borders of the bigger ellipse) and an ellipse is fitted through them. The ellipse is used as the peak border. Just intensities inside the ellipse contribute to the total peak intensity.

points are detected, which can be used for the ellipse fit in the time direction.

In general, the border points of the chromatogram with the smaller  $m/z$  range are used for the ellipse. However, the chromatogram with the narrow  $m/z$  range could possibly cut just a small part of the overlapping peak, resulting in an undetectable change of the gradient in the curve, whereas the extracted broader chromatogram detects the change. Hence, the broader chromatogram returns a much smaller time range than the narrower one. In this case, the border points of the broad chromatogram are taken for the ellipse fit. The ellipse is then fitted with the two border points in  $m/z$  direction and the two border points in time direction (blue points in Fig. 2), and only intensities within this ellipse contribute to the final peak area.

### 3.2 Theoretical isotopic distribution

Due to the nature of lipids, there is in general more than one candidate peak available for a distinct  $m/z$  value. In order to select the correct peak, LDA takes the theoretical isotopic distribution into account, which is used as a selection/exclusion criterion. First,



the algorithm checks whether the peak belongs to another isotopic distribution (exclusion). Second, it checks whether the isotopic peaks related to the analyte match (in a certain range) its theoretically calculated values (selection). If both criteria are fulfilled the peak is accepted. In contrast with the other existing algorithms, the LDA has no deisotoping step; every isotopic peak is quantified separately and the sum is taken as the quantitative measure.

### 3.3 Robust standardization

In addition to the novel algorithm the software features a statistics section with standardization. Three different standardization methods are provided: (i) correction based on a single standard; (ii) a median method; (iii) standardization on an internally developed methodology. The normalization on a single standard is trivial: each standard forms a reference value. The median method uses the median of the standards in a sample as a reference. The third method has been developed to provide a very robust method of standardization because suppression effects are quite common in MS. First, statistics from each standard are calculated (median and coefficient of variation). Second, the ‘most robust’ standard is selected by a two-step procedure. In the first step, standards are selected only if they are found at least 90% of the most often found standard. The best standard is then the one with the highest intensity which is within 5% of the standard with the best coefficient of variation. If there is none within the 5%, the one with the best coefficient of variation is taken. Third, applicable standards are selected for each experiment by a generous outlier removal. A standard is an outlier only if it is less than half of the lower quartile value or if it is higher than the double of the upper quartile value. Fourth, a ‘global reference area’ is defined, which is the ‘most robust’ standard area of an experiment whose value is closest to the mean value of the ‘most robust’ standard. Fifth, for each experiment a reference ratio between each applicable standard for the experiment and the experiment chosen in the previous step is calculated.

$$r_1 = \frac{s_{1\text{exp}_n}}{s_{1\text{exp}_{\text{ref}}}}, r_2 = \frac{s_{2\text{exp}_n}}{s_{2\text{exp}_{\text{ref}}}}, \dots, \quad (4)$$

The reference for the experiment is then the ‘global reference’ times the mean of the single standard ratios (a comparison of the standard median method and the new method can be found in Supplementary Section 7):

$$\text{refArea}_{\text{exp}_n} = \text{globalRefArea} * \text{mean}(r_1, r_2, r_3, \dots) \quad (5)$$

## 4 EXPERIMENTAL VALIDATION

In order to test the accuracy of the algorithm, we conducted a controlled experiment (see Section 2) and compared the results of LDA, mzMine2 (Katajamaa *et al.*, 2006; Pluskal *et al.*, 2010) and LIMS (Haimi *et al.*, 2006, 2009). First, samples were prepared containing all analytes at the same concentration to test the ionization efficiency. LDA and mzMine2 showed nearly the same intensities for TG54:0-2, but TG54:3 and the standard TG48:0 showed smaller intensities than expected (Supplementary Table TS1). TG54:3 and TG48:0 eluted exactly at the same retention times (25.9–26.1min) in all of the samples (Supplementary Table TS1), and suppressed each other. Consequently, we did not use TG48:0 as a standard, since it would have skewed the results. Second, we compared analytes over a dilution series (mix 2–5), from which we calculated a ratio relative to the areas found in mix 3 (Table 1 and for more details see Supplementary Table TS1). In this experiment, the suppression effects on TG54:3 could just be noticed at mix2, because here the intensity of TG54:3 was highest. In 8 of 12 possible ratios, the LDA was closer to the ideal ratio than mzMine2, and in 10 out of 12 ratios than LIMS. The average deviation from the ideal ratio was 9.7% for LDA, 10.5% for mzMine2 and 36.3% for LIMS. The LDA performed just slightly better (0.8%) than mzMine2 in this experiment, since the peaks in the controlled experiment were quite well separated, and no overlap occurred. However, the intention of this experiment was just to show the accuracy of the results. Nevertheless, the areas of the LDA could be taken for further analysis without manual review, whereas for mzMine2 manual selection of the correct hit from a list of ambiguous peaks was necessary in 25 of the 150 possible cases.

The performance of the algorithms on biological data was assessed without LIMS, because of the unsatisfactory results of the controlled experiment. The comparison with mzMine2 was conducted solely based on TG data, since it allowed for a more objective manual identification of a correct peak due to the regular distribution of TG peaks. For this data, LDA showed higher sensitivity (86.1% → 93.3%) and a significantly improved positive predictive value (89.4% → 99.1%) compared with mzMine2 (Table 2 and for more details see Supplementary Table TS2), which is extremely important for automated high-throughput analysis. Regarding sensitivity, LDA correctly identified 678 (93.3%) of the 727 possible peaks, whereas mzMine2 identified 626 peaks (85.4%). However, 168 of the mzMine2 identifications returned more than one candidate peak (ambiguous identifications), and for LDA only 21. Nineteen of those 21 peaks were located in the peak-tail belonging to the analyte with <5% intensity of the main

**Table 1.** Comparison of quantitation accuracy of the controlled experiment

Lipid	mix2 : mix3				mix4 : mix3				mix5 : mix3			
	Ideal	LDA	mzMine2	LIMS	Ideal	LDA	mzMine2	LIMS	Ideal	LDA	mzMine2	LIMS
TG54:0	2	2.3	2.5	0.9	0.4	0.42	0.43	0.57	0.2	0.20	0.20	0.24
TG54:1	2	2.3	2.4	3.8	0.4	0.38	0.38	0.31	0.2	0.15	0.15	0.16
TG54:2	2	2.3	2.2	0.9	0.4	0.40	0.40	0.35	0.2	0.17	0.17	0.17
TG54:3	2	1.8	1.8	1.0	0.4	0.43	0.43	0.38	0.2	0.19	0.19	0.12

The values are calculated relative to mix3. ‘Ideal’ corresponds to the expected ratio.

**Table 2.** Comparison of the identification performance of the LDA and mzMine2

	LDA	mzMine2
Peaks unambiguously identified (A)	657	458
Peaks ambiguously identified (B)	21	168
Wrong peak identified (C)	4	71
Additional wrong peak (no peak there) (D)	2	3
Peaks not identified (E)	45	30
Peaks present in sample (F=A+B+C+E)	727	727
Total peaks identified (G=A+B+C+D)	684	700
Sensitivity ((A+B)/F)	93.3%	86.1%
Sensitivity unambiguous (A/F)	90.4%	63.0%
Positive predictive value ((A+B)/G)	99.1%	89.4%

The sensitivity and the positive predictive value are the final quality parameters.

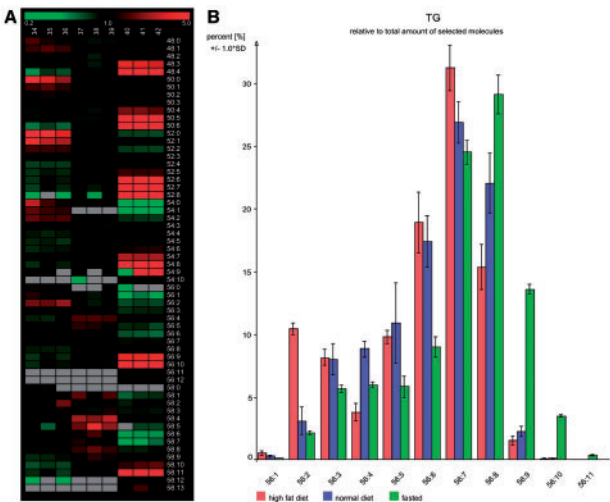
peak, or a peak appearing with two summits was considered as two possible identifications. The remaining two peaks would not be viable for analysis and a manual decision would have been necessary. Regarding positive predictive value, 678 (99.1%) of the 684 reported LDA identifications were correct, while 626 (89.4%) of the 700 mzMine2 hits were true positives, with the same amount of ambiguous identifications as in the sensitivity evaluation. This high positive predictive value demonstrates the extremely high reliability of the algorithm over a large intensity range, since the intensity ratio between highest and lowest valid identifications is almost 10<sup>6</sup>.

The data from the undiluted extract were used to demonstrate the general applicability of the software to samples containing minor amounts of phospholipids, i.e. phosphatidyl choline (PC), phosphatidyl ethanolamine (PE) and sphingomyelin (SM), in the presence of bulk TG. Even though the average intensity of a PC peak is around 5% and an average SM peak is around 1% of the intensity of an average TG peak (Supplementary Table TS3), the positive predictive values attained were 97.2% for PC, 93.4% for PE, 97.3% for SM and respective sensitivities were 92.5% for PC, 84.0% for PE and 85.7% for SM (Supplementary Table TS4).

5 SOFTWARE DESCRIPTION

The application supports mzXML (Pedrioli *et al.*, 2004) as raw data format. In addition, the Thermo RAW format and the Waters MassLynx raw data format are supported via the freely available software ReAdW and massWolf (Keller *et al.*, 2005), which can be directly integrated in the program. Before quantitation, the mzXML files are translated into an internal indexed chrom file format, which provides rapid access to the data (especially useful for data visualization). The molecules to be identified are defined in a Microsoft Excel file, containing, for each analyte, the name and the molecular formula and the mass. The m/z must be entered manually, since it is possible to define several adducts or modifications for one lipid. The quantitation itself can be performed in batch mode. The calculation of the theoretical isotopic distribution is based on the list of chemical elements and the probabilities of their isotopes provided in an extendable XML file.

The software features a statistics section with normalization, grouping, visualization and export functionalities. The data can be normalized on ‘internal standards’, which are added before the MS data acquisition and should compensate instrument-specific



**Fig. 3.** LDA result visualization. Three groups of C57BL male mice (*n* = 3 each) were analysed in this experiment (fed/high fat diet 34–36; fed/normal diet 37–39; fasted/normal diet 40–42). The values are normalized on the total TG content. Changes of single molecules are detected at glance with the heat map (A). The bar charts (B) allow a more detailed comparison (here TG 56:1–11 is depicted; the values are normalized on the total content of TG56). The heat map cells provide direct links to the 3D-chromatogramm viewer (Fig. 1). In Supplementary Material, a detailed list of the lipid assignments can be found in Table TS2, sheet ‘assignments’.

variations, as well as on ‘external standards’, which are added before sample preparation and should compensate losses throughout sample preparation. The standardization procedures are described at Section 3.3.

The results can be displayed in heat maps and/or bar charts (Fig. 3), in which several display options are available (normalization on standards, base peak, total class content, content of a specific group, etc.), and an overview of the content of the total classes. The results can be exported into Microsoft Excel or tab-delimited format, and the heat maps and bar charts in PNG or SVG format.

In the LDA, we implemented a sophisticated 3D viewer for the manual verification of the results (Fig. 1), because the extracted profiles/chromatograms are the result of data processing procedures and provide a constrained view on the data. The 3D viewer offers adjustable m/z display ranges and resolution levels. Furthermore, hits can be deleted, added or corrected with several available quantitation methods (3D method, standard ASAPRatio method, the method described in Hartler *et al.* (2007), and the method described in Section 5 and SFig. 1 in the Supplementary Data).

6 DISCUSSION

The LDA proved to be a powerful and reliable tool for high-throughput analysis of the lipidome using LC-MS. In a controlled experiment, we showed that the results are more accurate than with other available tools (average deviation from the ideal ratio was 9.7% for LDA, 10.5% for mzMine2 and 36.3% for LIMSA). Moreover on biological data, the novel algorithm proved its real strength: the sensitivity increased from 86.1 to 93.3% and the positive predictive value from 89.4 to 99.1%

compared with mzMine2. The high positive predictive value is particularly important for an automated analysis with minimal human intervention. Moreover, we demonstrated the applicability of LDA to low abundant lipid species in lipid droplets, like phospholipids and sphingomyelin. Although intensities measured were close to noise levels, our algorithm still achieved good to excellent positive predictive values. Although the experimental data in this study was generated on an FT-MS instrument, the application is not limited to high-resolution data. The parameters for the 3D algorithm have already been adapted for QTOF and low-resolution QTRAP.

LDA is not limited to the analysis of lipids, but can be extended to other singly charged analytes by providing the respective molecule definition file. Support for multiple charged molecules like peptides would require the extraction of more chromatograms for the theoretical isotopic distribution as selection/exclusion criterion (see Section 3.2). For example, when the algorithm checks whether or not the peak belongs to another isotopic distribution, a chromatogram at  $m/z$  (analyte) – mass (neutron) is extracted. To check a doubly charged molecule a chromatogram at  $m/z$  (analyte) – (mass (neutron))/2 is required and so on. The smoothing of a chromatogram is quite time consuming, thus a pre-screening method that works on the raw data are expected to analyse multiply charged molecules in a rapid manner.

## ACKNOWLEDGEMENTS

The authors thank the mass spectrometry department of Gerald N. Rechberger of the University of Graz for providing the QTOF data, the group of Bernd Helms, Department for Biochemistry and Cell Biology of the University of Utrecht for providing the QTRAP data, Anita Eberl of the Core Facility for Mass Spectrometry, Center for Medical Research, Medical University of Graz for the fruitful discussions and Ravi Tharakan, Bayview Proteomics Center, Johns Hopkins University for critically reading the manuscript.

**Funding:** LipidomicNet, an EU Framework 7 project (grant no. 202272); the Austrian Ministry of Science and Research, GEN-AU project Bioinformatics Integration Network.

**Conflict of Interest:** none declared.

## REFERENCES

- Blouin, C.M. *et al.* (2010) Lipid droplet analysis in caveolin-deficient adipocytes: alterations in surface phospholipid composition and maturation defects. *J. Lipid Res.*, **51**, 945–956.
- Ejlsing, C.S. *et al.* (2006) Automated identification and quantification of glycerophospholipid molecular species by multiple precursor ion scanning. *Anal. Chem.*, **78**, 6202–6214.
- Folch, J. *et al.* (1957) A simple method for the isolation and purification of total lipides from animal tissues. *J. Biol. Chem.*, **226**, 497–509.
- Haimi, P. *et al.* (2006) Software tools for analysis of mass spectrometric lipidome data. *Anal. Chem.*, **78**, 8324–8331.
- Haimi, P. *et al.* (2009) Instrument-independent software tools for the analysis of MS-MS and LC-MS lipidomics data. *Methods Mol. Biol.*, **580**, 285–294.
- Hartler, J. *et al.* (2007) MASPECTRAS: a platform for management and analysis of proteomics LC-MS/MS data. *BMC Bioinformatics.*, **8**, 197.
- Katajamaa, M. *et al.* (2006) MZmine: toolbox for processing and visualization of mass spectrometry based molecular profile data. *Bioinformatics*, **22**, 634–636.
- Keller, A. *et al.* (2005) A uniform proteomics MS/MS analysis platform utilizing open XML file formats. *Mol. Syst. Biol.*, **1**, 2005.
- Leavell, M.D. and Leary, J.A. (2006) Fatty acid analysis tool (FAAT): An FT-ICR MS lipid analysis algorithm. *Anal. Chem.*, **78**, 5497–5503.
- Li, X.J. *et al.* (2003) Automated statistical analysis of protein abundance ratios from data generated by stable-isotope dilution and tandem mass spectrometry. *Anal. Chem.*, **75**, 6648–6657.
- Pedrioli, P.G. *et al.* (2004) A common open representation of mass spectrometry data and its application to proteomics research. *Nat. Biotechnol.*, **22**, 1459–1466.
- Pluskal, T. *et al.* (2010) MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinformatics*, **11**, 395.
- Riccalton-Banks, L. *et al.* (2003) A simple method for the simultaneous isolation of stellate cells and hepatocytes from rat liver tissue. *Mol. Cell Biochem.*, **248**, 97–102.
- Song, H. *et al.* (2007) Algorithm for processing raw mass spectrometric data to identify and quantitate complex lipid molecular species in mixtures by data-dependent scanning and fragment ion database searching. *J. Am. Soc. Mass Spectrom.*, **18**, 1848–1858.