Deciphering lipid structures based on platform-independent decision rules

Jürgen Hartler[1–3,12], Alexander Triebl[2,12], Andreas Ziegl[1], Martin Trötzmüller[2,3], Gerald N Rechberger[3,4], Oana A Zeleznik[5,6], Kathrin A Zierler[4], Federico Torta[7], Amaury Cazenave-Gassiot[7], Markus R Wenk[7], Alexander Fauland[8], Craig E Wheelock[8], Aaron M Armando[9], Oswald Quehenberger[9], Qifeng Zhang[10], Michael J O Wakelam[10], Guenter Haemmerle[4], Friedrich Spener[4,11], Harald C Köfeler[2,3] & Gerhard G Thallinger[1,3]

We achieve automated and reliable annotation of lipid species and their molecular structures in high-throughput data from chromatography-coupled tandem mass spectrometry using decision rule sets embedded in Lipid Data Analyzer (LDA; http://genome.tugraz.at/lda2). Using various low- and high-resolution mass spectrometry instruments with several collision energies, we proved the method's platform independence. We propose that the software's reliability, flexibility, and ability to identify novel lipid molecular species may now render current state-of-the-art lipid libraries obsolete.

Lipidomics is a rapidly evolving scientific discipline that provides high-throughput data for elucidating lipid structure, metabolism, and dynamics at cellular and tissue-level scales[1,2]. Liquid chromatography–linked tandem mass spectrometry (LC-MS/MS) enables analyses including simultaneous high-precision quantitative measurements of hundreds to thousands of lipids in complex mixtures[3]. Such profiling can be carried out at six levels of structural information: (i) 'lipid subclass' level, (ii) 'bond type' level, (iii) 'fatty acyl' level, (iv) 'fatty acyl position' level, (v) 'fatty acyl or sphingoid base structure' level, and (vi) 'LIPID MAPS' level—this level comprises full structural elucidation, including double-bond location and geometry[4]. Throughout this paper, the term 'lipid species' refers to lipid subclass (including bond type level), which identifies lipids by the number of carbons and double bonds of constituent fatty acyl and/or alkyl/1-alkenyl chains (e.g., PI 38:4 corresponds to diacylglycerophosphoinositol which contains 38 carbon atoms and four double bonds in the two acyl chains). The term 'lipid molecular species' refers to fatty acyl level (e.g., PI 20:4_18:0 corresponds to diacylglycerophosphoinositol and comprises two fatty acyl chains that contain 18 and 20 carbon atoms, and zero and four double bonds, respectively) and/or fatty acyl position level (e.g., PI 18:0/20:4 corresponds to PI 20:4_18:0, where 18:0 and 20:4 are located at stereochemical numbering (*sn*) positions *sn*-1 and *sn*-2, respectively), in which structural information such as identification of constituent chains and determination of their respective regioselectivities at the glycerol backbone is obtained. In these approaches for lipid profiling, automated lipid annotation currently relies on spectral libraries[5–7]. However, variables such as the type of mass spectrometer, the collision energy applied, the type of adduct ion, and the charge state all cause substantial variation in the MS/MS spectra of lipid molecular species (**Fig. 1**).

Thus, matching of spectral data to experimentally or *in silico* generated spectral libraries is problematic for the following reasons: (i) it is not possible to detect novel lipid molecular species that are absent from spectral libraries (with novel acyl and/or alkyl/1-alkenyl constituents or an unusual *sn* position); (ii) it is challenging to obtain decisive information from low-abundance signals (e.g., fatty acyl and/or alkyl/1-alkenyl chain fragments from phospholipids in positive-ion mode), because the matching algorithms are geared mainly toward intensity signals higher than the base peak; (iii) the *sn* positions of fatty acyl and/or alkyl/1-alkenyl constituents are extremely difficult to determine, because general matching algorithms are not designed to distinguish the intensity relationships of low-abundance fragments that would reveal stereochemistry; (iv) it is not possible to discriminate between isobaric lipid species and between structural isomers of lipid molecular species; (v) users are, to a certain extent, precluded from setting up their own spectral libraries tailored to their platform because of the impracticality of having to generate thousands of *in silico* MS/MS spectra for each adduct of each single lipid subclass.

Here we describe a universal and flexible solution to the above limitations by introducing 'decision rule sets' for lipid subclass–adduct combinations (referred to as 'subclass/adduct') and an algorithm to apply these rules for the identification of lipid

[1]Institute of Computational Biotechnology, Graz University of Technology, Graz, Austria. [2]Center for Medical Research, Medical University of Graz, Graz, Austria. [3]Omics Center Graz, BioTechMed-Graz, Graz, Austria. [4]Department of Molecular Biosciences, University of Graz, Graz, Austria. [5]Channing Division of Network Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, USA. [6]Department of Medicine, Harvard Medical School, Boston, Massachusetts, USA. [7]Singapore Lipidomics Incubator, National University of Singapore, Singapore. [8]Department of Medical Biochemistry and Biophysics, Karolinska Institutet, Stockholm, Sweden. [9]School of Medicine, University of California San Diego, La Jolla, California, USA. [10]The Babraham Institute, Babraham Research Campus, Cambridge, UK. [11]Department of Molecular Biology and Biochemistry, Medical University of Graz, Graz, Austria. [12]These authors contributed equally to this work. Correspondence should be addressed to G.G.T. (gerhard.thallinger@tugraz.at) or H.C.K. (harald.koefeler@medunigraz.at).

RECEIVED 20 MAY; ACCEPTED 19 SEPTEMBER; PUBLISHED ONLINE 23 OCTOBER 2017; DOI:10.1038/NMETH.4470

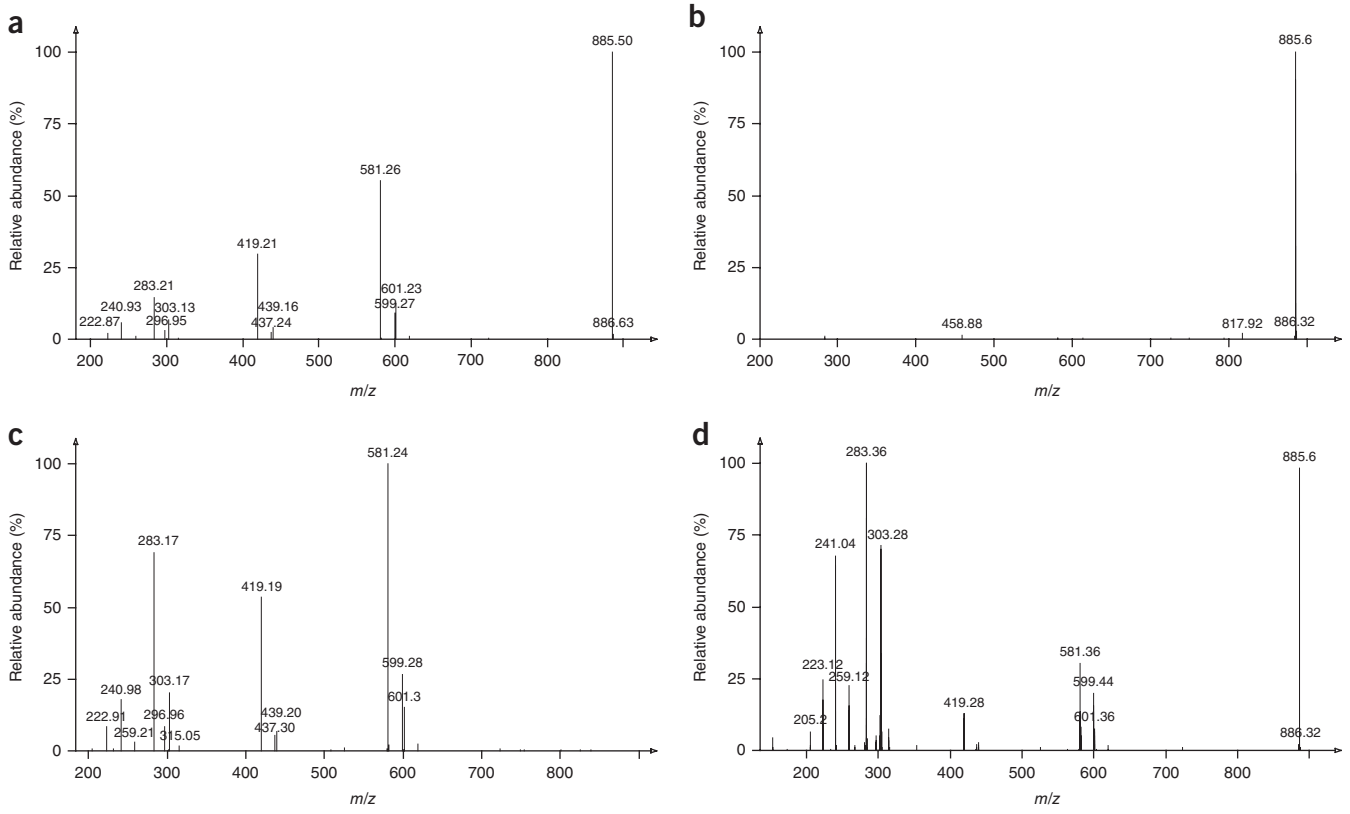

Figure 1 | Tandem mass spectra of lipid molecular species depend on platform and collision energy. Spectra of deprotonated PI 18:0/20:4 from two platforms and two collision energy settings are shown. (**a**) Orbitrap Velos Pro, CID mode, 30%, precursor *m/z* 885.545, damping gas He. (**b**) 4000 QTRAP, CID mode, 30 eV, precursor *m/z* 885.93, collision gas N_2. (**c**) Orbitrap Velos Pro, CID mode, 60%, precursor *m/z* 885.549, damping gas He. (**d**) 4000 QTRAP, CID mode, 60 eV, precursor *m/z* 885.85, collision gas N_2.

species and lipid molecular species. This method enables lipid annotation in high-throughput data derived from chromatography-coupled tandem mass spectrometry. The tool, which we call Lipid Data Analyzer (LDA), adapts not only to specific parameters of the various MS platforms, but also to changes in collision energies and to different adduct ions. Consequently, lipid annotation is based on well-defined fragments (fragment rules) and their intensity relationships (intensity rules), and this allows for routine profiling of known lipid targets and for detection of novel lipids (see Online Methods). As such, the software flexibly accommodates differences in fragmentation behavior. Importantly, the decision rule sets allow identification of fatty acyl and/or alkyl/1-alkenyl constituents and determination of their respective *sn* positions at the glycerol backbone (in the case of coeluting regioisomers, the position assignment is based on the more abundant regioisomer), even with low-abundance lipid molecular species, as well as the definition of fragments from isobaric or isomeric lipid subclasses for their differentiation.

The basis for the fragment rules is derived from available information about lipid fragmentation[8]. To gather further evidence supporting the reliability of the fragments and to establish the intensity rules, we conducted three control experiments conducted with lipid standards with known constituent fatty acyl and/or alkyl/1-alkenyl chains, including their respective *sn* positions, and one experiment on the lipidome of murine liver samples. In total, we performed more than 600 LC-MS/MS runs on eight different MS/MS platforms (AB Sciex, Agilent Technologies, Thermo Scientific, and Waters; **Supplementary Table 1** and **Supplementary Note 1**); these runs are summarized as follows.

In control experiment 1, which included 78 nonisobaric/nonisomeric standard lipids from 14 lipid subclasses (**Supplementary Table 2**), we generated decision rule sets for each lipid subclass and adduct and successfully validated the algorithm in MS/MS spectra. These rules also allowed the assignment of *sn* positions for lipid molecular species. In control experiment 2, with eight isomeric lipid molecular species (**Supplementary Table 3**), we verified the algorithm's ability to discriminate between isomeric species from different lipid subclasses/adducts in MS and MS/MS spectra (**Supplementary Table 4**). In control experiment 3, with 16 structural isomers of lipid molecular species originating from different subclasses mixed at various concentrations (**Supplementary Table 5**), we demonstrated that the algorithm appropriately assigned the respective structural isomers (**Supplementary Tables 6** and **7**). In the experiment with the murine liver lipidome, we confirmed that LDA can reliably identify lipid molecular species in complex biological samples. The algorithm identified low-abundance species (**Supplementary Fig. 1**), isobaric species, and structural isomers contained in these complex samples (see http://www.ebi.ac.uk/metabolights/MTBLS396). This approach also allowed the identification of 109 novel lipid molecular species and six novel regioisomeric species (**Supplementary Table 8** and **Supplementary Fig. 2**); we consider a lipid molecular species 'novel' if it is not present in LIPID MAPS Structure Database[9], ChEBI[10], CyberLipid (http://www.cyberlipid.org),

Table 1 | Sensitivity and positive predictive value (PPV) of LDA and LipidBlast (LB)

	Total lipid species identified: 1,077			Total lipid molecular species identified: 3,567		
	LDA	LB 450	LB 10	LDA	LB 450	LB 10
Sensitivity (%)[a]	97	36	85	80	15	57
PPV (%)[b]	97	91	70	92	91	58

[a]Sensitivity, percent of total species identified by the software. [b]PPV, percent of correct identifications. Table indicates sensitivity and PPV of LDA and LipidBlast with matching factors 450 (stringent) and 10 (relaxed) in positive-ion mode based on data acquired on Orbitrap Velos Pro in CID mode. The lipidome of murine liver samples was determined five times. Total lipid (molecular) species identified represent the sum of all species manually identified in the five MS runs.

HMDB[11], or YMDB[12]. Details about lipid molecular species identified on the various platforms, including a cross-platform comparison, are given in **Supplementary Tables 9** and **10**.

We used data from control experiment 1 and the murine liver lipidome experiment (acquired on Orbitrap Velos Pro in CID mode and on 4000 QTRAP with collision-energy settings of +50% and −50%, and +45 electron volts (eV) and −45 eV, respectively) to verify our approach and to benchmark the LDA algorithm against the state-of-the-art *in silico* library LipidBlast[7] (see Online Methods and **Supplementary Note 2**). LDA identified considerably more lipid (molecular) species with higher confidence in all but one case, in which LDA either identified considerably more species with lower confidence (LB 450) or identified fewer species with a much higher confidence (LB 10) (**Table 1** and **Supplementary Tables 11–13**). Data at lipid-species level revealed that 'stringent' LipidBlast conditions identified only one-third of the lipid species identified by LDA (n = 1,041 lipid species; 97% of 1,077 manually identified species). When we used 'relaxed' LipidBlast settings, the number of correctly identified lipid species increased at the cost of drastically reduced positive predictive values. More dramatic were the findings at the level of lipid molecular species, as LDA identified 2,862 (80% of 3,567) lipid molecular species (see **Table 1**), which underlined its power to discriminate lipid structural details. In addition to its broader scope to quantitatively analyze lipid molecular species[13] even at low abundance (**Supplementary Fig. 1**), a further advantage of LDA is its ability to detect unanticipated fatty acyl and/or alkyl/1-alkenyl combinations (**Supplementary Table 8** and **Supplementary Fig. 2**). Moreover, LDA unambiguously assigned the *sn* positions for almost all standards (positive-ion mode, 104/110; negative-ion mode, 105/105); whereas LipidBlast under 'relaxed' settings consistently reported an erroneous positional isomer in addition to the correct species (http://www.ebi.ac.uk/metabolights/MTBLS397). In the case of coeluting regioisomeric lipid molecular species, the assignment reports the more abundant regioisomer (**Supplementary Figs. 3** and **4**); chromatographic approaches exist to solve this issue[14].

Sophisticated software programs for lipid identification in direct infusion (shotgun) MS have been developed[15–18]; however, unlike LDA, they do not support chromatography-linked approaches, which are now frequently used in lipidomics[19]. The LDA approach correctly identifies isobaric and isomeric lipids, and structural isomers, and this identification enables their use as diagnostic markers in routine analyses and as key indicators of healthy versus aberrant metabolism. Owing to the high sensitivity attainable with LDA, information derived from low-abundance fragments (e.g., in positive-ion mode) is now made accessible and can be converted into lipid structures. Moreover, the software reports structural annotations based solely on spectral evidence (**Supplementary Figs. 5** and **6**) and avoids misleading structural overdetermination[20].

LDA offers platform independence and flexibility by circumventing the need for experimental and *in silico* spectral libraries. Indeed, users can easily adapt existing decision rule sets or generate new ones (even for other metabolite classes), as LDA features a graphical user interface for such rule definition that provides direct visual feedback on acquired spectra (see Online Methods and **Supplementary Fig. 7**). Further decision rule set development should be based on measured standards and subsequent validation in pertinent biological settings; but rule set development can also be performed on biological data directly, if the lipid subclasses/adducts are sufficiently separated by chromatography.

LDA currently offers reliable decision rule sets for various adducts of 14 major lipid subclasses acquired using platforms from multiple major instrument vendors. LDA annotations reflect the level of structural details inherent to the analyzed spectra while avoiding reporting of unsubstantiated structural details. The simplicity of defining and handling decision rule sets allows for easy application of LDA by bioinformaticians and mass spectrometrists.

Note: Requests for materials should be directed to G.G.T. (gerhard.thallinger@tugraz.at) or H.C.K. (harald.koefeler@medunigraz.at).

METHODS

Methods, including statements of data availability and any associated accession codes and references, are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

Support by the Austrian Science Fund (FWF Project Grant P26148 to G.G.T.) and the Austrian Ministry for Science, Research and Economy (HSRSM Grant Omics Center Graz, BioTechMed-Graz to G.G.T.) is gratefully acknowledged. M.J.O.W. and Q.Z. were funded by the BBSRC (UK; Grant BBS/E/B/000C0415). We thank R. Salek for his extensive help in MetaboLights upload. Furthermore, we thank AB Sciex, Agilent Technologies, Bruker Daltonics, and Thermo Fisher Scientific for providing permission to distribute the WiffReader SDK, the MassHunter DAC, the CompassXtract, and the MSFileReader libraries in the software.

AUTHOR CONTRIBUTIONS

J.H., A.T., M.T., F.S., G.H., H.C.K., and G.G.T. designed the study. J.H., A.T., M.T., G.N.R., and H.C.K. designed the experiments. K.A.Z. and G.H. provided the biological samples. A.T., M.T., G.N.R., F.T., A.C.-G., M.R.W., A.F., C.E.W., A.M.A., O.Q., Q.Z., and M.J.O.W. designed and performed the mass spectrometric experiments. J.H. and A.Z. implemented the algorithm and the software. J.H., A.T., O.A.Z., and H.C.K. developed the decision rule sets. J.H. and A.T. benchmarked the algorithm in comparison to LipidBlast and prepared the spectral evidence for the novel species. J.H. and H.C.K. prepared and uploaded the data in MetaboLights. J.H., A.T., F.S., and G.G.T. wrote the manuscript in cooperation with all contributing authors.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Wenk, M.R. *Cell* **143**, 888–895 (2010).
2. Quehenberger, O. & Dennis, E.A. *N. Engl. J. Med.* **365**, 1812–1823 (2011).
3. Dove, A. *Science* **347**, 788–790 (2015).
4. Liebisch, G. *et al. J. Lipid Res.* **54**, 1523–1530 (2013).
5. Song, H., Hsu, F.F., Ladenson, J. & Turk, J. *J. Am. Soc. Mass Spectrom.* **18**, 1848–1858 (2007).
6. Taguchi, R. & Ishikawa, M. *J. Chromatogr. A* **1217**, 4229–4239 (2010).
7. Kind, T. *et al. Nat. Methods* **10**, 755–758 (2013).
8. Hsu, F.F. & Turk, J. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* **877**, 2673–2695 (2009).
9. Sud, M. *et al. Nucleic Acids Res.* **35**, D527–D532 (2007).
10. Degtyarenko, K. *et al. Nucleic Acids Res.* **36**, D344–D350 (2008).
11. Wishart, D.S. *et al. Nucleic Acids Res.* **37**, D603–D610 (2009).
12. Jewison, T. *et al. Nucleic Acids Res.* **40**, D815–D820 (2012).
13. Hartler, J. *et al. Bioinformatics* **27**, 572–577 (2011).
14. Holčapek, M., Jandera, P., Zderadička, P. & Hrubá, L. *J. Chromatogr. A* **1010**, 195–215 (2003).
15. Han, X., Yang, K. & Gross, R.W. *Mass Spectrom. Rev.* **31**, 134–178 (2012).
16. Herzog, R. *et al. PLoS One* **7**, e29851 (2012).
17. Yang, K., Cheng, H., Gross, R.W. & Han, X. *Anal. Chem.* **81**, 4356–4368 (2009).
18. Husen, P. *et al. PLoS One* **8**, e79736 (2013).
19. Cajka, T. & Fiehn, O. *Trends Analyt. Chem.* **61**, 192–206 (2014).
20. Liebisch, G., Ejsing, C.S. & Ekroos, K. *Clin. Chem.* **61**, 1542–1544 (2015).

ONLINE METHODS

Decision rule sets. MS/MS spectra of lipids vary greatly depending on the type of mass spectrometer used, the collision energy applied, the adduct ions, and the charge state. Taking these factors into account, we developed flexible decision rule sets that enable automated annotation of lipids in the generally accepted format[4] at multiple levels of structural detail (**Supplementary Figs. 5** and **6**).

A decision rule set for a lipid subclass/adduct consists of the section [GENERAL]—which pertains to general lipid subclass information—and of the three sections [HEAD], [CHAINS], and [POSITION]—which correspond to information concerning the structural details. The latter three sections do not apply to subclasses/adducts that lack a head group or chain fragment. In these three sections, fragment rules and intensity rules reflect the pattern of MS/MS spectra, as we show for deprotonated diacylglycerophosphoinositol (PI) in **Supplementary Figure 8**. This figure demonstrates fragment rules ('!FRAGMENTS') that include an arbitrary name, a chemical formula (for *m/z* value calculation), the charge state, the MS^n level where the fragment might be observed ('2' corresponds to MS/MS), and whether the presence of a fragment is required for positive identification at a certain structural level. Moreover, the parameter 'formula' allows for the placeholder '$PRECURSOR' (corresponding to the mass of the precursor) to define neutral losses. '$CHAIN' designates any possible fatty acyl chain, and '$ALKYLCHAIN'/'$ALKENYLCHAIN' any alkylated/1-alkenylated forms, respectively. Previously defined fragments can be reused (e.g., the section '[CHAINS]'). The parameter 'mandatory' is set to 'true' for characteristic fragments, such as the neutral loss of the phosphoethanolamine head fragment (neutral loss of 141 Da) in spectra of protonated diacylglycerolphosphoethanolamines (PE). The parameter 'mandatory' is set to 'false' for fragments observed infrequently. Even though the presence of such fragments is not essential to any annotated structural level, usage of infrequent fragments in intensity rules considerably improves the reliability of annotations. A third option for this parameter is 'other', which designates fragments originating from isobaric or isomeric lipid species not belonging to the lipid subclass of the rule set. This parameter setting is used to discard false-positive identifications.

Intensity rules (!INTENSITIES') consist of 'equation' parameters representing allowed intensity relationships of fragments and the parameter 'mandatory'. The parameter 'equation' utilizes any previously defined fragments, including a placeholder called '$BASEPEAK', to define a minimum intensity for fragments. Furthermore, an optional number in square brackets defines the *sn* position of the fragment. For the parameter 'mandatory', only the values true and false are allowed. The effect of this parameter depends on the section in which it is used, as will be discussed in the following paragraphs.

The algorithm processes the sections in the aforementioned order. A decision rule set is applied on a consolidated spectrum—i.e., a spectrum consisting of the sum of denoised spectra (**Supplementary Note 3**) within a detected MS^1 peak. Quantification of MS^1 peaks and removal of isotopic peaks is performed as described by Hartler *et al.*[13]. Moreover, for lipid species verified by reverse-phase LC-MS/MS, LDA offers a nonlinear fitting approach to predict retention times of lipid species determined by MS^1 only. This feature allows for automatic removal of peaks that have implausible retention times (**Supplementary Note 4**).

Starting with the [HEAD] section, the algorithm calculates the *m/z* values of the fragments and interrogates the consolidated spectrum for their presence (**Supplementary Fig. 9a**). When mandatory fragments cannot be detected in the spectrum, the algorithm discards the associated MS^1 peak. Otherwise, fragment intensities are checked for compliance with the intensity rules (**Supplementary Fig. 9b**). Again, if a mandatory intensity rule is not fulfilled in the [HEAD] section, the MS^1 identification will be discarded. [HEAD] section rules are the primary check for verification of a lipid subclass/adduct. Note that in cases where subclasses/adducts lack head-group-specific fragments (e.g., ammoniated triacylglycerols), false-positive MS^1 identification will be discarded by the spectrum coverage. The spectrum coverage is controlled by an adjustable threshold for the percentage of annotated fragment intensities.

For the [CHAINS] section, the algorithm computes all possible chain combinations pertaining to the total number of carbon atoms and double bonds of the particular lipid species; e.g., PI 18:1_20:3 is appropriate for PI 38:4. The same procedure as for head group is applied for each potential chain (**Supplementary Fig. 10** shows an example for PI 38:4 containing a 20:4 residue). The algorithm will typically only report chain combinations if all chains in the combination comply with the decision rules. However, there are subclasses/adducts where acyl- or alkyl/1-alkenyl chains at certain positions show low-abundance fragments only. An example is deprotonated 1-(1*Z*-alkenyl),2-acylglycerophosphoethanolamine, where the deprotonated alkenyl chain is of extremely low intensity (on account of resonance stabilization of the carboxylate anion). For such cases, a parameter is available in section [GENERAL] to allow acceptance of a certain combination with only one verified chain. If low/high abundance chains comply with the rules, the algorithm will advance to the [POSITION] rules.

[POSITION] rules consist of intensity comparisons of previously defined fragments (**Supplementary Fig. 11a**). If mandatory intensity rules are defined, all of them must be fulfilled for *sn*-position assignment; whereas for optional rules ('mandatory' false), a majority already suffices for an assignment (**Supplementary Fig. 11b**). This approach is preferable, because in some cases the most reliable position information is derived from low-abundance, rare fragments. If these fragments are present, they are decisive by a mandatory intensity rule; otherwise, position assignment is based on less reliable, optional intensity rules.

The decision rule sets and the algorithm for their interpretation allow for utmost flexibility, such as inclusion of isotopically labeled standards (used in TG rule development, **Supplementary Table 2**) and even the detection of coeluting lipid molecular species, which are encountered frequently (**Supplementary Fig. 12**). Although the rules use a syntax easily comprehensible for mass spectrometrists, we recognized the need for adapting and extending the existing rules provided thus far and for generating rules for further lipid subclasses/adducts. Consequently, we implemented a graphical user interface for rule definition which provides direct visual feedback on acquired spectra (**Supplementary Fig. 7**).

Experiments carried out for rule development and verification. The experimental execution is described in detailed protocols presented in **Supplementary Note 1** and **Supplementary Table 1**. Data obtained are discussed in the main text and shown in **Supplementary Tables 2–10** and **Supplementary Figure 13**. No

statistical testing was applied. Further information is provided in the **Life Sciences Reporting Summary** published alongside this paper. Detailed data and further details on results are available online (see "Data availability statement").

Application 1. Collision energies in mass spectrometry are considered optimal for a subclass/adduct when both the head group and chain fragments are equally well represented. Since these energies vary depending on the subclass/adduct, as a tradeoff we selected energies which delivered the best overall result with the platform we used for the given lipidome in control experiment 1.

Application 2. The basic fragment rules are based on published results[8,21]. The rules were adapted and extended by visual inspection of spectra from control experiment 1 and biological data. We determined detectable fragments, identified mandatory fragments, derived intensity rules, and extracted decisive differences for many isobaric/isomeric subclasses/adducts. Further, we determined intensity relationships characteristic for *sn*-position assignment. Finally, we found novel fragment ion relationships, such as the relative intensity of the sodiated form of a carboxylated chain fragment that allowed for differentiation between 1,2- and 1,3-diacylglycerols at optimal collision energies, and we demonstrated the software's capability to distinguish regioisomers under certain chromatographic conditions (**Supplementary Fig. 4** and **Supplementary Table 14**).

We defined more than 1,000 decision rule sets for lipid subclasses/adducts for various MS platforms and experimental conditions. These decision rule sets cover the major lipid subclasses and mass spectrometers commonly used today and will serve as an entry point for investigators unfamiliar with lipid data analysis. The direct visual feedback particularly provides an easy introduction to fragmentation patterns of lipids. Importantly, decision rule sets developed are provided along with software for the algorithm, which can be downloaded from http://genome.tugraz.at/lda2. In addition, raw data, results of detailed analysis, comments about information content that can be derived from the various adduct ions, and suggestions about optimal collision energies for subclasses/adducts are available.

Application 3. In general, isobars or isomers from different lipid subclasses/adducts are only slightly chromatographically separated, and such separation cannot be judged from MS^1 spectra. Thus, we expanded our algorithm to separate MS^1 peaks consisting of pairs of isobaric or isomeric subclasses/adducts. The algorithm extracts ion chromatograms from the absolute intensities of distinct fragments belonging solely to one or the other species, and it computes the retention time (RT) maximum of either lipid species. A weighted mean (based on abundances of the fragments) is used to estimate the RT of such maxima. If there is at least one MS/MS scan between the maxima, or if the maxima are in the vicinity of two different adjacent MS/MS spectra (in the range of 20% of the distance between the spectra), the mean of the two RTs is defined as the position of the split of the MS^1 peak. If the RT cannot be determined using absolute intensities, the same procedure is applied to relative intensities. If this also fails, the MS^1 peak intensity is distributed according to the intensities of the distinct fragments. However, for isobars/isomers of different subclasses/adducts (e.g., protonated PC and PE), this approach is highly inaccurate and should be avoided, because intensities of the fragments typically do not reflect the MS^1 intensities. To verify this, we pooled lipid standards of PC, PE, LPC, and LPE subclasses, where isomeric species can be observed as protonated and sodiated adduct ions, and we deliberately modified chromatography parameters to generate overlapping MS^1 peaks (**Supplementary Fig. 14**). This experiment revealed that a successful peak split primarily depends on the availability of MS/MS scans. Whereas successful splits were frequent for PC/PE species and for platforms with high MS/MS scan rates, less well-separated LPC/LPE species were often left unsplit (**Supplementary Table 4**). Nonetheless, the presence of either isomer was detected in almost all cases (97%).

Application 4. LDA detects and assigns structural isomers of the same lipid subclass/adduct from a shared MS^1 peak, whereupon the abundance of the MS^1 peak is split according to intensities detected in the MS/MS spectra (**Supplementary Fig. 12**). To this end we determined detection rate, accuracy, and variability of results obtained from experiments with structural isomers mixed in concentration ratios up to 1:20. As expected, results varied depending on MS ion mode and ionization mode (**Supplementary Tables 6** and **7**; **Supplementary Fig. 13**). Whereas negative-ion mode generally produced results with low coefficients of variation, positive-ion mode led to results with high coefficients of variation due to low-abundance chain fragments. In fact, at higher concentration ratios, chain fragments of low-abundance species were not detectable at all. An interesting showcase for a potential pitfall on account of a low MS/MS sampling rate was found for PE acquired on QTOF in positive-ion mode. Whereas chromatographically separated isomeric PE 36:4 species produced excellent lipid molecular species ratios, mixed PE 36:2 species produced much higher abundances for PE 18:0/18:2 in comparison to PE 18:1/18:1. The reason is that, in this particular case, the QTOF instrument reported MS/MS spectra only at the end of the MS^1 peak; therefore, the slightly earlier eluting PE 18:1/18:1 yielded lower 18:1 chain intensities. Interestingly, chain fragments of some species reflect the true ratios quite well (e.g., PC 34:0 in negative-ion mode), while others usually underestimate the true ratio (e.g., PE 36:4). Generally, to derive absolute intensities of pairs of structural isomers from the same MS^1 peak, calibration curves are strongly recommended.

Application 5. In a benchmark test of LDA versus LipidBlast[7], we used data from both the first control experiment and the murine liver lipidome experiment; both sets of data were acquired on Orbitrap Velos Pro in CID +50% and −50% and on 4000 QTRAP +45 eV and −45 eV, respectively. For LipidBlast evaluation, we used the recommended MSPepSearchGUI (http://peptide.nist.gov/software/ms_pep_search_gui/MSPepSearch.html). The same *m/z* tolerances were applied in both LDA and LipidBlast. The specificity and sensitivity of LipidBlast depend on a so-called matching factor[22], a value ranging from 0–999. Using the default setting of 450 for the matching factor, many lipid standards in control experiment 1 were not detected. Consequently, the matching factor was lowered to 10, in which case LipidBlast detected almost all of the lipid standards in negative-ion mode. Further reduction did not improve the sensitivity of LipidBlast. In positive-ion mode, irrespective of the matching factor setting, LipidBlast was not able

doi:10.1038/nmeth.4470

to identify as many lipid molecular species as was LDA. Details about the LipidBlast parameters are given in **Supplementary Note 2**. In this benchmark, we used only lipid subclasses/adducts that both LDA and LipidBlast are able to detect. Correct assignment of lipid species and lipid molecular species identified in liver lipidomes was verified by manual inspection of the spectra and by determining whether the observed retention time of a lipid species was within the expected range[23].

Code availability. The algorithm presented is embedded in the Java software package LDA (version 2.5.2) which performs MS^1 peak deconvolution[13], and it supports several operating systems such as Windows, MacOS, Linux, and other Unix-based systems. Calculations were performed on a 64-bit Windows 7 desktop PC equipped with an Intel Core i7-2600 CPU at 3.4 GHz and 16 GB RAM under Windows 7. Decision rule sets were tested on the following MS/MS platforms: 4000 QTRAP and QTRAP 6500 from AB Sciex; G6550A QTOF from Agilent Technologies; Orbitrap Elite, Orbitrap Velos Pro in CID and HCD mode, and Q Exactive from Thermo Fisher Scientific; and SYNAPT G1 HDMS QTOF from Waters. The primary raw data format is mzXML[24]; however, the software allows for direct processing of vendor formats from AB Sciex, Agilent Technologies, Bruker Daltonics, and Thermo Fisher Scientific by an integrated version of msConvert[25], as we obtained permission for redistribution of vendor-provided libraries from respective mass spectrometer manufacturers. For Waters '.raw' directories, installation of Mass++ (http://masspp.jp) is required. LDA and the decision rule sets are freely available from http://genome.tugraz.at/lda2. The source code is released under a GNU GPL v3 license and is available from https://github.com/ThallingerLab/LDA2/releases/tag/2.5.2.

Data availability statement. Data and analysis of results from control experiments 1–3, the murine liver lipidome experiment, LipidBlast benchmarking, HCD characterization, and detection of regioisomers are available from the MetaboLights[26] repository with accession numbers MTBLS394 (control experiment 1), MTBLS391 (control experiment 2), MTBLS398 (control experiment 3), MTBLS396 (murine liver lipidome experiment), MTBLS397 (benchmarking), and MTBLS462 (HCD characterization and regioisomers). Raw data and results are also available from the authors' website (http://genome.tugraz.at/lda2). Detailed data documentation can be found in **Supplementary Note 5**.

21. Murphy, R.C. & Axelsen, P.H. *Mass Spectrom. Rev.* **30**, 579–599 (2011).
22. Stein, S.E. & Scott, D.R. *J. Am. Soc. Mass Spectrom.* **5**, 859–866 (1994).
23. Fauland, A. *et al. J. Lipid Res.* **52**, 2314–2322 (2011).
24. Pedrioli, P.G. *et al. Nat. Biotechnol.* **22**, 1459–1466 (2004).
25. Chambers, M.C. *et al. Nat. Biotechnol.* **30**, 918–920 (2012).
26. Haug, K. *et al. Nucleic Acids Res.* **41**, D781–D786 (2013).

nature research

Corresponding author(s): Gerhard G Thallinger
Harald C Köfeler

☐ Initial submission ☐ Revised version ☒ Final submission

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

Experimental design

1. Sample size

   Describe how sample size was determined. — not applicable

2. Data exclusions

   Describe any data exclusions. — not applicable

3. Replication

   Describe whether the experimental findings were reliably reproduced. — All attempts of replicating the experimental findings were successful.

4. Randomization

   Describe how samples/organisms/participants were allocated into experimental groups. — not applicable

5. Blinding

   Describe whether the investigators were blinded to group allocation during data collection and/or analysis. — not applicable

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
☐	☒ A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ A statement indicating how many times each experiment was replicated
☒	☐ The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)
☒	☐ A description of any assumptions or corrections, such as an adjustment for multiple comparisons
☒	☐ The test results (e.g. *P* values) given as exact values whenever possible and with confidence intervals noted
☐	☒ A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
☐	☒ Clearly defined error bars

See the web collection on statistics for biologists for further resources and guidance.

Nature Methods: doi:10.1038/nmeth.4470

Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study.	Online Methods, 'Code availability and technical details' subsection and Supplement, 'Supplementary Note 2'.

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* guidance for providing algorithms and software for publication provides further information on this topic.

Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.	not applicable

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).	not applicable

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.	not applicable
b. Describe the method of cell line authentication used.	not applicable
c. Report whether the cell lines were tested for mycoplasma contamination.	not applicable
d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.	not applicable

Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.	Supplement, 'Supplementary Note 1', 'Sample preparation' subsection

Policy information about studies involving human research participants

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.	not applicable

Nature Methods: doi:10.1038/nmeth.4470