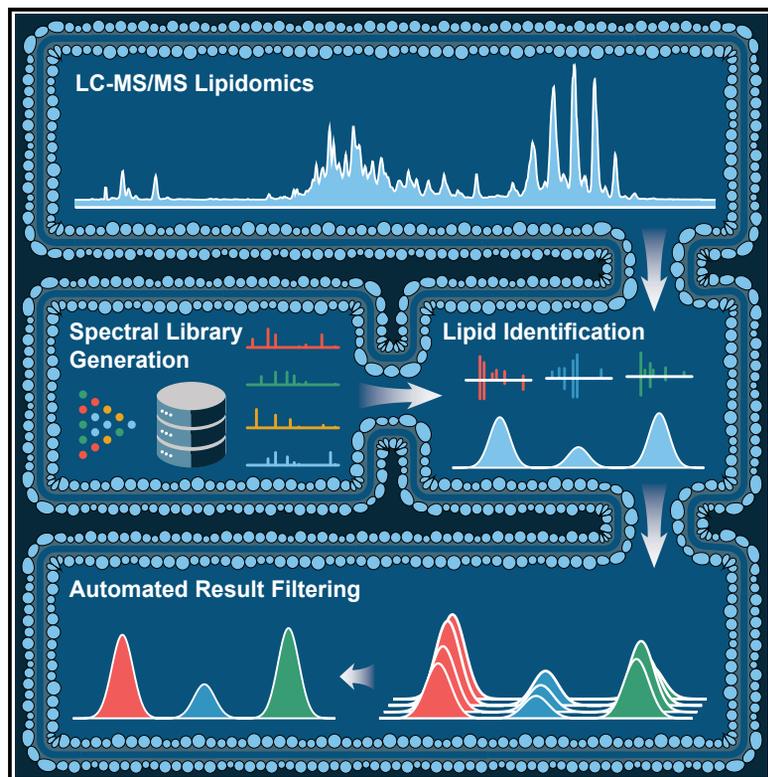


LipiDex: An Integrated Software Package for High-Confidence Lipid Identification

Graphical Abstract



Authors

Paul D. Hutchins, Jason D. Russell,
Joshua J. Coon

Correspondence

jcoon@chem.wisc.edu

In Brief

Hutchins et al. introduce an open-source software suite that automates lipid identification and significantly reduces hands-on processing time for large LC-MS/MS lipidomic datasets. They demonstrate the confident identification of hundreds of diverse lipids from complex lipid extracts using optimized spectral matching techniques and multiple identification filters.

Highlights

- A new open-source software suite for LC-MS/MS lipidomics data analysis
- Flexible *in silico* libraries accurately model fragmentation of diverse lipid types
- Accurate quantitation of co-isolated and co-eluting isobaric lipids
- Streamlined downstream data filtration for high-confidence lipid identifications

LipiDex: An Integrated Software Package for High-Confidence Lipid Identification

Paul D. Hutchins,^{1,3} Jason D. Russell,^{2,3} and Joshua J. Coon^{1,2,3,4,5,*}

¹Department of Chemistry, University of Wisconsin–Madison, Madison, WI 53706, USA

²Morgridge Institute for Research, Madison, WI 53715, USA

³Genome Center of Wisconsin, University of Wisconsin–Madison, Madison, WI 53706, USA

⁴Department of Biomolecular Chemistry, University of Wisconsin–Madison, Madison, WI 53706, USA

⁵Lead Contact

*Correspondence: jcoon@chem.wisc.edu

<https://doi.org/10.1016/j.cels.2018.03.011>

SUMMARY

State-of-the-art proteomics software routinely quantifies thousands of peptides per experiment with minimal need for manual validation or processing of data. For the emerging field of discovery lipidomics via liquid chromatography–tandem mass spectrometry (LC-MS/MS), comparably mature informatics tools do not exist. Here, we introduce LipiDex, a freely available software suite that unifies and automates all stages of lipid identification, reducing hands-on processing time from hours to minutes for even the most expansive datasets. LipiDex utilizes flexible *in silico* fragmentation templates and lipid-optimized MS/MS spectral matching routines to confidently identify and track hundreds of lipid species and unknown compounds from diverse sample matrices. Unique spectral and chromatographic peak purity algorithms accurately quantify co-isolation and co-elution of isobaric lipids, generating identifications that match the structural resolution afforded by the LC-MS/MS experiment. During final data filtering, ionization artifacts are removed to significantly reduce dataset redundancy. LipiDex interfaces with several LC-MS/MS software packages, enabling robust lipid identification to be readily incorporated into pre-existing data workflows.

INTRODUCTION

When coupled with chromatographic separation, the high resolution, mass accuracy, and speed of modern mass spectrometry (MS) systems render them the premier tool for unbiased analysis of diverse biomolecules (Cajka and Fiehn, 2016; Hebert et al., 2014; Patti et al., 2012). Although less mature relative to other -omics technologies (Figure S1), discovery lipidomic analysis by liquid chromatography–tandem MS (LC-MS/MS) is poised to make significant contributions to biomedical research. Already this technology has uncovered novel lipids that improve glucose tolerance in mice (Yore et al., 2014), mapped lipid remodeling during platelet activation (Slatter et al., 2016), and re-

vealed lipid dysregulation antecedent to islet autoimmunity in type 1 diabetes (Oresic et al., 2008). However, on account of lipid structural heterogeneity, co-fragmentation of isobaric lipids, and chromatographic peak (feature) degeneracy, significant data analysis challenges remain (Cajka and Fiehn, 2014; Kuhl et al., 2012; Mahieu and Patti, 2017).

Currently, numerous informatics approaches exist to address one or more of these challenges, including mzMine 2 (Pluskal et al., 2010), LipidBlast (Kind et al., 2013), LipidSearch (Taguchi and Ishikawa, 2010), MS-Dial (Tsugawa et al., 2015), LipidMatch (Koelmel et al., 2017), and Lipid Data Analyzer 2 (Hartler et al., 2017). Accurate lipid identification requires customizable *in silico* spectral libraries, lipid-specific MS/MS matching, accurate assignment of MS/MS identifications to chromatographic peaks, and multi-dimensional downstream data filtration (Tsugawa et al., 2017a). To promote broad usability and acceptance, these elements should be vendor neutral and integrated in graphical user interface (GUI)-based software suites. At present, realizing this ideal workflow for global feature profiling often requires either significant manual validation or integration of multiple software packages. We developed LipiDex to automate and unite these crucial lipid identification steps into a single intuitive software package capable of efficiently processing expansive lipid datasets.

RESULTS

Lipid Identification and Quantitation Workflow

LipiDex comprises a set of lipid spectral library management and data-processing tools optimized for high-resolution MS analysis and linked together in an intuitive GUI (Figure 1A). Developed in Java (JRE v.1.8.0), LipiDex runs efficiently on multiple operating systems, supports vendor-neutral and open-source MS file types, and accepts chromatographic feature tables from several commercial and open-source LC-MS/MS software packages. Here we describe LipiDex's major lipid identification algorithms and demonstrate their performance with multiple datasets. A detailed explanation of the algorithms is provided in STAR Methods, a feature comparison with other software packages is provided in Figure S2, and the software suite and accompanying multimedia tutorials describing the use of LipiDex's major functions are freely available at <http://www.ncqbc.com/resources/software/>.

To achieve confident lipid identification at the sum or molecular composition level (Liebisch et al., 2013), LipiDex integrates

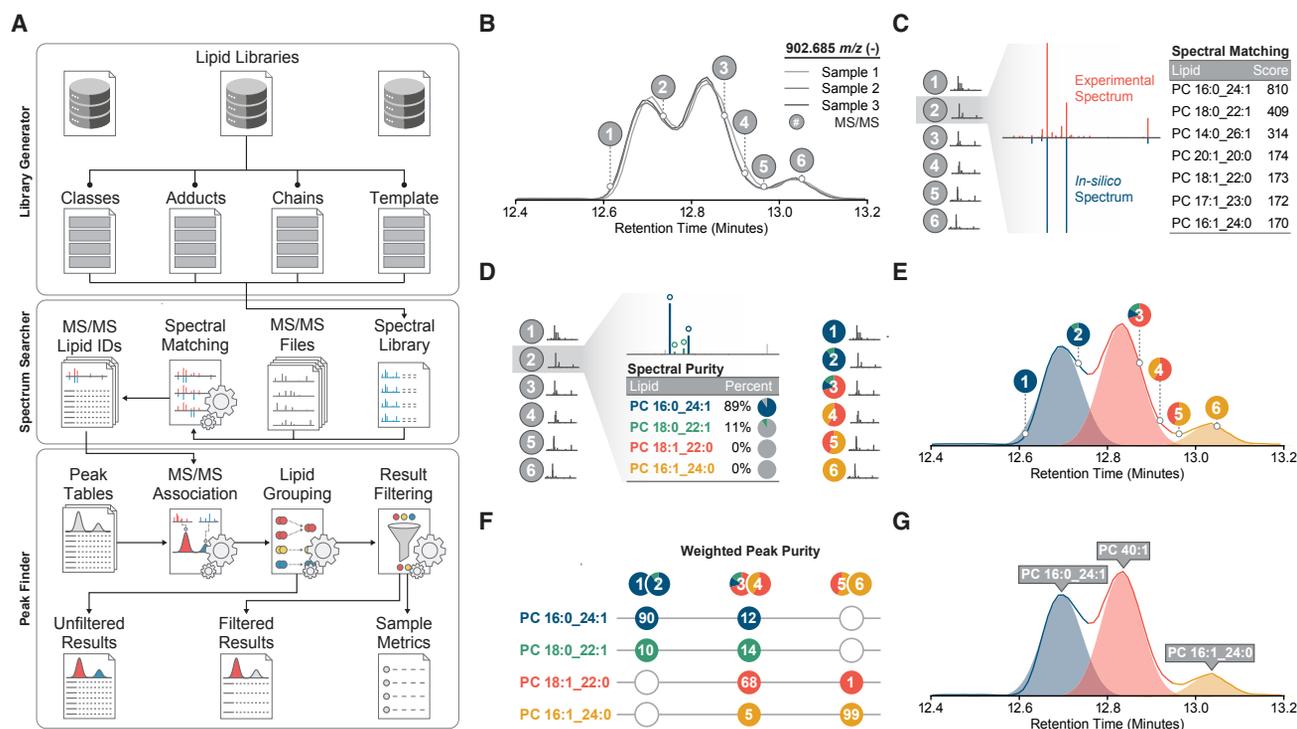


Figure 1. Lipid Identification and Quantitation Workflow

(A) Overview of library construction (Library Generator), lipid identification (Spectrum Searcher), and result filtering (Peak Finder) in LipiDex. (B) Extraction and alignment of lipid chromatographic peaks using external software packages. (C) Identification of MS/MS spectra (gray circles and red spectrum) using *in silico* lipid databases and a modified dot-product scoring algorithm that ranks identifications according to spectral similarity to library entries (blue spectrum). (D) Spectral deconvolution and relative quantitation (colored circles) of co-fragmented isobaric lipid species by sequential matching against and subtracting from the experimental spectra. (E) Accurate association of MS/MS identifications to chromatographic features via Gaussian peak modeling (shaded color). (F) Calculation of total peak purity for each chromatographic peak across all samples using the Gaussian peak profile to modify spectral purity measurements. (G) Final identification of each chromatographic peak using either the sum composition (i.e., PC 40:1) or molecular composition (i.e., PC 16:1_24:0) if the weighted peak purity rises above a user-supplied threshold.

precursor mass-to-charge (m/z), fragmentation spectra, chromatographic retention time, and chromatographic elution profile (Figure 1B). *In silico* lipid spectral libraries for MS/MS identification are first created by Library Generator, which utilizes a set of customizable lipid metadata tables and fragmentation rules to accurately model the m/z and relative intensity of diverse fragment types (Figure S2) (Kind et al., 2014). To facilitate utilization, LipiDex includes extensive pre-built spectral libraries based on numerous authentic reference standards, experimental identifications, and published databases (refer to Tables S1 and S2 for library composition) (Kind et al., 2014; Tsugawa et al., 2017b). Next, Spectrum Searcher generates lipid spectral matches by searching experimental MS/MS spectra against these libraries using a two-step searching algorithm. Experimental MS/MS spectra are first assigned putative identifications via a modified dot-product score (Figure 1C) and are then deconvoluted to quantify lipid co-fragmentation (Figure 1D). Peak Finder links these MS/MS identifications with externally generated feature tables using a Gaussian peak model to accurately assign quantitative abundance values to unique lipid species (Figure 1E). Final result filtering, lipid grouping, and retention time filtering remove degenerate peaks and collapse redundant

identifications (Figure 1F), automating the manual data validation normally required for robust large-scale discovery lipidomics.

Generating Confident Lipid Identifications

Although many powerful LC-MS/MS analysis packages identify chromatographic features from user-supplied or public spectral libraries, these packages are not specifically optimized for lipid data analysis. Accordingly, we present several *in silico* and experimental datasets for the optimization and validation of LipiDex's core functions. To generate candidate spectral matches, MS/MS spectra are searched against user-selected spectral libraries leveraging a lipid-optimized dot-product spectral similarity score and the intact precursor m/z as a search constraint. The dot-product score represents the scaled cosine of the angle between the *in silico* and experimental spectral vectors (Stein and Scott, 1994) and contains mass (n) and intensity (m) weighting factors, which can be modulated to tailor the equation to specific applications (Kim et al., 2012). In Figure 2A, we demonstrate the systematic optimization of these terms using a structurally diverse set of 82 lipid reference standard spectra fragmented via higher-energy collisional dissociation (HCD) from the NIST Tandem Mass Spectral Library (National Institute

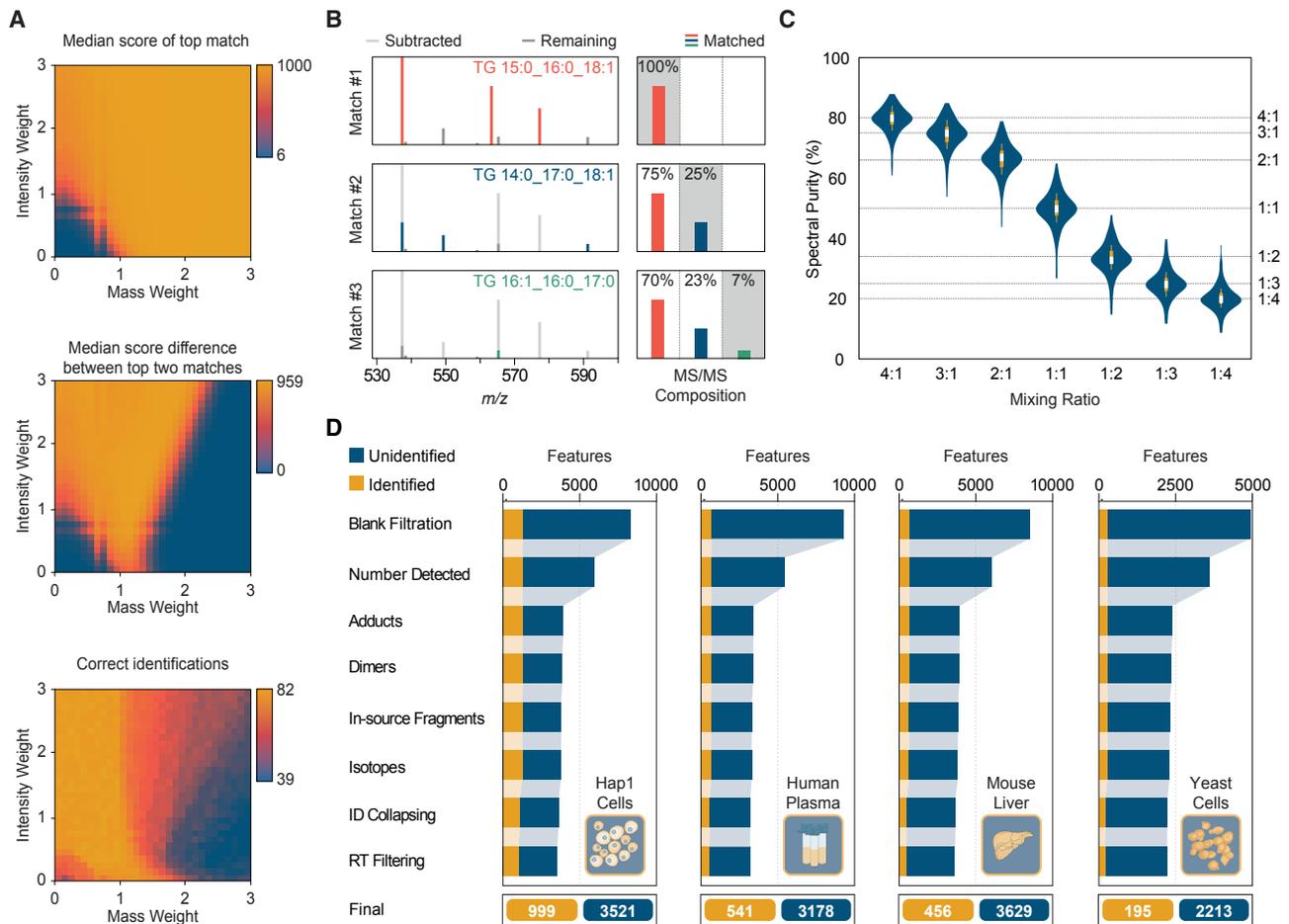


Figure 2. Generating Confident Lipid Identifications

(A) Color-coded values for the median score, median score difference to next match, and correct identifications for the top-scoring spectral matches from 82 lipid reference standard spectra from the NIST Tandem Mass Spectral Library searched using various combinations of intensity and mass weighting factors. (B) Sequential matching (colored lines) and subtraction (light gray lines) from experimental triacylglycerol MS/MS spectrum and representative fragment intensities used to quantify co-fragmentation of isobaric lipids. Note the displayed spectrum is enlarged to highlight the fatty acid-identifying peaks (m/z 530–600). (C) Violin plot showing the measured (blue) and expected (dashed line) spectral purities for 98 isobaric lipid pairs mixed *in silico* at varying ratios. Box plots within each violin demarcate the median (white stripe), the 25th to 75th percentile (interquartile range; box), and 1.5 times the interquartile range (whiskers). (D) Number of identified and unidentified chromatographic feature groups from Hap1 cells, human plasma, mouse liver, and yeast cells analyzed in triplicate after sequential filtering for common LC-MS/MS artifacts and lipid identification filtering.

of Standards and Technology, 2012). The optimal weighting factor set (m : 1.2; n : 0.9) returns a median top match similarity score of 955, a median first and second match score difference of 887, and 82 correct identifications (Figure 2A), and performs similarly well for collision-induced dissociation (CID) spectra (Figure S5) and LC-MS/MS analysis of pure reference standards (Figure S6). Importantly, this large difference between the first and second match scores greatly increases spectral match confidence in the absence of a false discovery rate (FDR) estimate as commonly used in spectral matching for discovery proteomics.

Spectral similarity scoring's accuracy markedly improves upon addition of orthogonal identification filters (Kwiecien et al., 2015), particularly for spectra containing multiple co-fragmented lipid regioisomers (Hartler et al., 2017; Koelme et al., 2017). Spectrum Searcher utilizes a unique spectral deconvolution algorithm to quantify the co-fragmentation of isobaric lipids (Figure S6). In Figure 2B, we demonstrate spectral deconvolution

for a complex experimental triacylglycerol (TG) spectrum where multiple TGs share common chain-identifying fragments. By sequentially subtracting select matched fragments, the deconvolution algorithm avoids assigning the full intensity of fragments shared by multiple isobaric precursors to each species, an inaccuracy inherent to methods that sum the intensity of all matched fragments (Hartler et al., 2017; Koelme et al., 2017). The accuracy of the algorithm is globally demonstrated using a set of pure isobaric phospho-, glycerol-, and sphingolipid spectral pairs extracted from experimental LC-MS/MS data and mixed *in silico* at various ratios to simulate co-fragmentation (Figure 2C). The deconvolution algorithm accurately returns the correct mixing ratio for the vast majority of *in silico* pairs (average spectral purity error of 3.5). Note that the returned values represent a metric of spectral purity and are not used to modify the peak area values for associated features, rendering small inaccuracies in the purity value insignificant.

Removing redundant features that arise from the same analyte (degenerate features) and data artifacts generated from the discovery LC-MS/MS lipidomic workflow is a critical step for robust downstream statistical analysis, which often requires significant manual data validation (Collins et al., 2016; Kuhl et al., 2012; Mahieu and Patti, 2017). To minimize this laborious but essential step, LipiDex automatically detects and removes degenerate identified and unidentified features. In Figure 2D, we present the LC-MS/MS analysis and data filtration of complex lipid extracts of human Hap1 cells, human plasma, mouse liver, and yeast cells. After removal of features present in a blank injection by the peak picking software, each dataset contains thousands of features (8,276, 9,287, 8,494, and 4,921, respectively). LipiDex automatically removes spurious feature detections, adducts, dimers, in-source fragments, and misidentified isotopes, resulting in an average feature reduction of 59%.

After removal of artifacts, LipiDex infers the minimum number of lipid identifications, which accurately explain all identified lipid features and their associated MS/MS identifications (Figure S7) in a process roughly analogous to protein inference (Nesvizhskii and Aebersold, 2005). Although these filters remove the vast majority of incorrect lipid identifications, false positives can remain. Lipid retention time represents a highly specific and orthogonal filter that can greatly improve identification confidence (Aichele et al., 2015; Hartler et al., 2017; Tsugawa et al., 2015, 2017b). Many approaches predict lipid retention time by correlating chemical identifiers to experimentally measured retention times, which inherently requires a static chromatographic setup. LipiDex circumvents this problem by implementing an optional outlier detection algorithm to remove lipid identifications falling outside of the retention time cluster elution pattern of lipid reversed-phase analysis (Figure S8). While these filters do not exhaustively remove degenerate features or false identifications in a sample, the tremendous reduction in candidate features significantly expedites data validation and makes LipiDex a flexible tool for discovery lipidomic analysis.

DISCUSSION

Highly scalable, automated, and robust lipid identification is readily achievable with LipiDex. LipiDex has already been used in multiple lipidomics studies to analyze a wide array of sample types (La-pointe et al., 2017; Reidenbach et al., 2018; Rhoads et al., 2018; Stefely et al., 2016; Veling et al., 2017; Velsko et al., 2017). To extend this identification capability to new biological systems of interest and novel fragmentation techniques, LipiDex's extensive curated lipid databases are easily expanded using its intuitive fragmentation templates. Beyond library generation, the detailed fragmentation metadata provided for each predicted MS/MS fragment empowers accurate spectral matching and deconvolution. Unlike many general small-molecule data analysis packages, LipiDex dynamically modifies the structural resolution encoded within each lipid identification to match the experimental MS/MS data. Moreover, these and other unidentified features are extensively filtered to remove degenerate peaks, redundant lipid identifications, and spurious lipid features. In the absence of a method to calculate an FDR, LipiDex applies multi-dimensional filtering to increase lipid identification confidence. Recently published methods for small-molecule FDR estimation via target-decoy

methods raise the tantalizing possibility of statistically rigorous lipid identification (Palmer et al., 2016; Scheubert et al., 2017). We note that LipiDex's detailed fragmentation rule sets may prove valuable for automatically generating realistic and expandable decoy databases, a vital step for the rapid evolution of discovery lipidomic analysis into an essential systems biology tool.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- METHOD DETAILS
 - Supported Lipidomic Workflows
 - Spectral Similarity Matching
 - Spectral Deconvolution
 - Feature-MS/MS Association
 - Weighted Peak Purity
 - Spurious Peak Detection Filtering
 - Degenerate Peak Filtering
 - Lipid Grouping
 - Retention Time Filtering
 - Hap1 Collection
 - Yeast Collection
 - Hap1 Cell, Yeast, and Mouse Liver Lipid Extraction
 - Human Plasma Lipid Extraction
 - LC-MS/MS Acquisition
- QUANTIFICATIONS AND STATISTICAL ANALYSIS
 - LC-MS/MS Data Analysis
 - NIST MS/MS Dot Product Optimization Dataset
 - Lipid Data Analyzer 2 Comparison Dataset
 - *In-Silico* Co-Fragmentation Dataset
- DATA AND SOFTWARE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information includes eight figures and four tables and can be found with this article online at <https://doi.org/10.1016/j.cels.2018.03.011>.

ACKNOWLEDGMENTS

We gratefully acknowledge support from the National Institutes of Health Grant P41 GM108538 (awarded to J.J.C), the Department of Energy DE-FC02-07ER64494, and the Morgridge Institute for Research Metabolism Theme.

AUTHOR CONTRIBUTIONS

P.D.H. developed the software package, curated spectral libraries, performed mass spectrometry analysis, and analyzed lipidomics data. P.D.H. and J.D.R. designed the study and interpreted the results. J.J.C. directed and supervised all aspects of the study. All authors contributed to writing and critical review of the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: November 9, 2017

Revised: January 30, 2018

Accepted: March 14, 2018

Published: April 25, 2018

REFERENCES

- Aicheler, F., Li, J., Hoene, M., Lehmann, R., Xu, G., and Kohlbacher, O. (2015). Retention time prediction improves identification in nontargeted lipidomics approaches. *Anal. Chem.* **87**, 7698–7704.
- Cajka, T., and Fiehn, O. (2014). Comprehensive analysis of lipids in biological systems by liquid chromatography-mass spectrometry. *Trends Analyt. Chem.* **61**, 192–206.
- Cajka, T., and Fiehn, O. (2016). Toward merging untargeted and targeted methods in mass spectrometry-based metabolomics and lipidomics. *Anal. Chem.* **88**, 524–545.
- Collins, J.R., Edwards, B.R., Fredricks, H.F., and Van Mooy, B.A.S. (2016). LOBSTAHS: an adduct-based lipidomics strategy for discovery and identification of oxidative stress biomarkers. *Anal. Chem.* **88**, 7154–7162.
- Hartler, J., Triebel, A., Ziegler, A., Trötz Müller, M., Rechberger, G.N., Zeleznik, O.A., Zierler, K.A., Torta, F., Cazenave-Gassiot, A., Wenk, M.R., et al. (2017). Deciphering lipid structures based on platform-independent decision rules. *Nat. Methods* **14**, 1171–1174.
- Hebert, A.S., Richards, A.L., Bailey, D.J., Ulbrich, A., Coughlin, E.E., Westphall, M.S., and Coon, J.J. (2014). The one hour yeast proteome. *Mol. Cell. Proteomics* **13**, 339–347.
- Kim, S., Koo, I., Wei, X., and Zhang, X. (2012). A method of finding optimal weight factors for compound identification in gas chromatography-mass spectrometry. *Bioinformatics* **28**, 1158–1163.
- Kind, T., Liu, K.-H., Lee, D.Y., DeFelice, B., Meissen, J.K., and Fiehn, O. (2013). LipidBlast in silico tandem mass spectrometry database for lipid identification. *Nat. Methods* **10**, 755–758.
- Kind, T., Okazaki, Y., Saito, K., and Fiehn, O. (2014). LipidBlast templates as flexible tools for creating new in-silico tandem mass spectral libraries. *Anal. Chem.* **86**, 11024–11027.
- Koelmel, J.P., Kroeger, N.M., Ulmer, C.Z., Bowden, J.A., Patterson, R.E., Cochran, J.A., Beecher, C.W.W., Garrett, T.J., and Yost, R.A. (2017). LipidMatch: an automated workflow for rule-based lipid identification using untargeted high-resolution tandem mass spectrometry data. *BMC Bioinformatics* **18**, 331.
- Kuhl, C., Tautenhahn, R., Böttcher, C., Larson, T.R., and Neumann, S. (2012). CAMERA: an integrated strategy for compound spectra extraction and annotation of liquid chromatography/mass spectrometry data sets. *Anal. Chem.* **84**, 283–289.
- Kwiecien, N.W., Bailey, D.J., Rush, M.J., Cole, J.S., Ulbrich, A., Hebert, A.S., Westphall, M.S., and Coon, J.J. (2015). High-resolution filtering for improved small molecule identification via GC/MS. *Anal. Chem.* **87**, 8328–8335.
- Lapointe, C.P., Stefely, J.A., Jochem, A., Hutchins, P.D., Wilson, G.M., Kwiecien, N.W., Coon, J.J., Wickens, M., and Pagliarini, D.J. (2017). Multi-omics reveal specific targets of the RNA-binding protein Puf3p and its orchestration of mitochondrial biogenesis. *Cell Syst.* **6**, 125–135.e6.
- Leys, C., Ley, C., Klein, O., Bernard, P., and Licata, L. (2013). Detecting outliers: do not use standard deviation around the mean, use absolute deviation around the median. *J. Exp. Soc. Psychol.* **49**, 764–766.
- Liebisch, G., Vizcaino, J.A., Köfeler, H., Trötz Müller, M., Griffiths, W.J., Schmitz, G., Spener, F., and Wakelam, M.J. (2013). Shorthand notation for lipid structures derived from mass spectrometry. *J. Lipid Res.* **54**, 1523–1530.
- Mahieu, N.G., and Patti, G.J. (2017). Systems-level annotation of a metabolomics data set reduces 25 000 features to fewer than 1000 unique metabolites. *Anal. Chem.* **89**, 10397–10406.
- Mahieu, N.G., Spalding, J.L., and Patti, G.J. (2015). Warpgroup: increased precision of metabolomic data processing by consensus integration bound analysis. *Bioinformatics* **32**, 268–275.
- National Institute of Standards and Technology. (2012). NIST Tandem Mass Spectral Library (NIST).
- Nesvizhskii, A.I., and Aebersold, R. (2005). Interpretation of shotgun proteomic data: the protein inference problem. *Mol. Cell. Proteomics* **4**, 1419–1440.
- Oresic, M., Simell, S., Sysi-Aho, M., Nääntö-Salonen, K., Seppänen-Laakso, T., Parikka, V., Katajamaa, M., Hekkala, A., Mattila, I., Keskinen, P., et al. (2008). Dysregulation of lipid and amino acid metabolism precedes islet autoimmunity in children who later progress to type 1 diabetes. *J. Exp. Med.* **205**, 2975–2984.
- Palmer, A., Phapale, P., Chernyavsky, I., Lavigne, R., Fay, D., Tarasov, A., Kovalev, V., Fuchser, J., Nikolenko, S., Pineau, C., et al. (2016). FDR-controlled metabolite annotation for high-resolution imaging mass spectrometry. *Nat. Methods* **14**, 57–60.
- Patti, G.J., Yanes, O., and Siuzdak, G. (2012). Innovation: metabolomics: the apogee of the omics trilogy. *Nat. Rev. Mol. Cell Biol.* **13**, 263–269.
- Pluskal, T., Castillo, S., Villar-Briones, A., and Oresic, M. (2010). MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinformatics* **11**, 395.
- Reidenbach, A.G., Kemmerer, Z.A., Aydin, D., Jochem, A., Mcdevitt, M.T., Hutchins, P.D., Wilkerson, E.M., Stark, J.L., Stefely, J.A., Johnson, I.E., et al. (2018). Conserved lipid and small-molecule modulation of COQ8 reveals regulation of the ancient kinase-like UbiB family. *Cell Chem. Biol.* **25**, 154–165.e11.
- Rhoads, T.W., Burhans, M.S., Chen, V.B., Hutchins, P.D., Rush, M.J.P., Clark, J.P., Stark, J.L., McIlwain, S.J., Eghbalnia, H.R., Pavelec, D.M., et al. (2018). Caloric restriction engages hepatic RNA processing mechanisms in rhesus monkeys resource caloric restriction engages hepatic RNA processing mechanisms in rhesus monkeys. *Cell Metab.* **27**, 1–12.
- Scheubert, K., Hufsky, F., Petras, D., Wang, M., Nothias, L.F., Dührkop, K., Bandeira, N., Dorrestein, P.C., and Böcker, S. (2017). Significance estimation for large scale metabolomics annotations by spectral matching. *Nat. Commun.* **8**, 1494.
- Slatter, D.A., Aldrovandi, M., O'Connor, A., Allen, S.M., Brasher, C.J., Murphy, R.C., Mecklemann, S., Ravi, S., Darley-Usmar, V., and O'Donnell, V.B. (2016). Mapping the human platelet lipidome reveals cytosolic phospholipase A2 as a regulator of mitochondrial bioenergetics during activation. *Cell Metab.* **23**, 930–944.
- Stefely, J.A., Licitra, F., Laredj, L., Reidenbach, A.G., Kemmerer, Z.A., Grangeray, A., Jaeg-Ehret, T., Minogue, C.E., Ulbrich, A., Hutchins, P.D., et al. (2016). Cerebellar ataxia and coenzyme Q deficiency through loss of unorthodox kinase activity. *Mol. Cell* **63**, 608–620.
- Stein, S.E., and Scott, D.R. (1994). Optimization and testing of mass-spectral library search algorithms for compound identification. *J. Am. Soc. Mass Spectrom.* **5**, 859–866.
- Taguchi, R., and Ishikawa, M. (2010). Precise and global identification of phospholipid molecular species by an Orbitrap mass spectrometer and automated search engine Lipid Search. *J. Chromatogr. A* **1217**, 4229–4239.
- Tsugawa, H., Cajka, T., Kind, T., Ma, Y., Higgins, B., Ikeda, K., Kanazawa, M., VanderGheynst, J., Fiehn, O., and Arita, M. (2015). MS-DIAL: data-independent MS/MS deconvolution for comprehensive metabolome analysis. *Nat. Methods* **12**, 523–526.
- Tsugawa, H., Ikeda, K., and Arita, M. (2017a). The importance of bioinformatics for connecting data-driven lipidomics and biological insights. *Biochim. Biophys. Acta* **1862**, 762–765.
- Tsugawa, H., Ikeda, K., Tanaka, W., Senoo, Y., Arita, M., and Arita, M. (2017b). Comprehensive identification of sphingolipid species by in silico retention time and tandem mass spectral library. *J. Cheminform.* **9**, 19.
- Veling, M.T., Reidenbach, A.G., Freiburger, E.C., Kwiecien, N.W., Hutchins, P.D., Drahnak, M.J., Jochem, A., Ulbrich, A., Rush, M.J.P., Russell, J.D., et al. (2017). Multi-omic mitoprotease profiling defines a role for Oct1p in coenzyme Q production. *Mol. Cell* **68**, 970–977.e11.
- Velsko, I.M., Overmyer, K.A., Speller, C., Klaus, L., Collins, M.J., Loe, L., Frantz, L.A.F., Sankaranarayanan, K., Lewis, C.M., Martinez, J.B.R., et al. (2017). The dental calculus metabolome in modern and historic samples. *Metabolomics* **13**, 134.
- Yang, X., Neta, P., and Stein, S.E. (2014). Quality control for building libraries from electrospray ionization tandem mass spectra. *Anal. Chem.* **86**, 6393–6400.
- Yore, M.M., Syed, I., Moraes-Vieira, P.M., Zhang, T., Herman, M.A., Homan, E.A., Patel, R.T., Lee, J., Chen, S., Peroni, O.D., et al. (2014). Discovery of a class of endogenous mammalian lipids with anti-diabetic and anti-inflammatory effects. *Cell* **159**, 318–332.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological Samples		
Pooled human plasma, EDTA coagulated	Innovative Research	IPLA-N
Human HAP1 parental control cell line	Horizon	C631
<i>S. cerevisiae</i> WT background strain BY4742 (MAT α his3 1 leu2 0 lys2 0 ura3 0)	GE Dharmacon	Cat#YSC1049
Mouse liver homogenate	Stefely et al., 2016	https://doi.org/10.1016/j.molcel.2016.06.030
Deposited Data		
Thermo Fisher Scientific *.raw Files	Chorus Project	https://chorusproject.org/pages/dashboard.html#/projects/all/1409/experiments
Spectral mixing dataset	This paper	https://github.com/coongroup/LipiDex
Dot-product optimization dataset	This Paper	https://github.com/coongroup/LipiDex
LDA 2 Comparison Dataset	Hartler et al., 2017	http://genome.tugraz.at/lda2/data/ControlExperiment1/Exp1_QExactive.zip
Software and Algorithms		
LipiDex website	This paper	https://github.com/coongroup/LipiDex
LipiDex	This paper	https://github.com/coongroup/LipiDex
Proteowizard msconvert	ProteoWizard Software Foundation	Version 3.0
Compound Discoverer	Thermo Fisher	Version 2.0
MZMine 2	Pluskal et al., 2010	Version 2.28
Java SE 8	Oracle Corporation	SE 8
Other		
Q Exactive HF	Thermo Scientific	IQLAAEGAAPFAL GMBFZ
HESI II heated ESI source	Thermo Scientific	IQLAAEGABBFACNMAGY
Vanquish Binary Pump	Thermo Scientific	VH-P10-A
Vanquish Split Sampler	Thermo Scientific	VF-A10-A
Vanquish Column Compartment	Thermo Scientific	VH-C10-A
ACQUITY UPLC CSH C18 Column	Waters Corporation	186005297

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Joshua J. Coon (jcoon@chem.wisc.edu).

METHOD DETAILS

Supported Lipidomic Workflows

LipiDex bases lipid identifications on accurate precursor m/z and matching of high or low resolution MS/MS spectra against theoretical in-silico libraries. These identifications are then extensively filtered to ensure datasets with minimal artefactual and redundant entries. Although the software was initially developed and benchmarked using RP-LC-MS/MS data collected on the Q-Exactive platform, it supports numerous common lipidomic workflows from diverse LC-MS/MS setups including acetate and formate mobile phase systems, reversed phase and HILIC separations, HCD and CID fragmentation spectra, and fast polarity switching and separate polarity data acquisition. Additionally, the flexible spectral library generation tools allow the software package to be readily expanded to cover novel lipid classes and fragmentation templates. Currently, LipiDex does not support data collected using data-independent acquisition (DIA), all ion fragmentation (AIF), or infusion-based methods.

Spectral Similarity Matching

Candidate lipid identifications are assigned to experimental MS/MS spectra using a modified dot-product spectral similarity score. First, user-selected *in-silico* lipid spectra are extracted from generated .msp files and binned by precursor *m/z* for efficient retrieval prior to spectral matching. Experimental MS/MS spectra in the Mascot Generic Format (MGF) or mzXML format are then extracted and the fragment intensities normalized to 1000. All library spectra whose intact precursor *m/z* falls within a user-defined tolerance of the experimental MS/MS precursor *m/z* are then compared to the experimental MS/MS using the dot-product score. The following dot-product equation is used to measure spectral similarity:

$$\text{Dot Product} = 1000 * \frac{(\sum A_{\text{exp}} * A_{\text{lib}})^2}{\sum A_{\text{exp}}^2 * \sum A_{\text{lib}}^2}$$

$$A = (\text{Relative Intensity})^m * (m/z)^n$$

The optimized intensity ($m = 1.2$) and mass ($n = 0.9$) weighting factors give greater emphasis to higher *m/z* and intensity fragments which often identify individual fatty acid moieties of the lipid. To reduce the deleterious effect of co-fragmented lipid species on spectral scoring, the intensities of peaks present in the experimental spectrum but not in the library entry are reduced by 50% (Tsugawa et al., 2015). For the reverse dot-product score, only peaks present in both experimental and library spectra (within the user-defined *m/z* tolerance) are considered for spectral similarity scoring. *Peak Finder* utilizes both the standard and reverse dot-product scores to filter out potential false identifications whose spectral similarity scores fall below a user-supplied threshold.

Spectral Deconvolution

Accurate quantification of co-fragmentation of isobaric lipids and total spectral purity follows the logic flow detailed in Figure S2. For a given MS/MS spectrum, all spectral matches are ranked according to their dot-product score and added to the scan queue (SQ). Next, fragmentation template entry for each candidate lipid species is queried to find the fatty acid-identifying fragment type which is assigned the highest relative intensity. The experimental spectrum is then searched for fragments which correspond to this rule type. If all of the predicted highest intensity chain-identifying fragments are not found, the candidate identification is discarded. If all predicted highest intensity chain-identifying fragments are found, the relative intensity of matched fragments is corrected using any matching *m/z* values found in the correction list (CL). The corrected *m/z* values are then assigned the median intensity of the matched fatty acid-identifying fragments and added to the CL. Depending on the number of fatty acids for the identified lipid, either the median or maximum matched intensity is added to the intensity list (IL). This intensity value is used to quantify the relative abundance of the lipid in the MS/MS spectrum. This process is repeated until no candidate spectra remain. All intensities found in the IL are then scaled to the sum of all matched species' intensities.

Feature-MS/MS Association

LipiDex accepts chromatographic feature tables generated by Compound Discoverer (Thermo Fisher Scientific) or MZmine 2 (v. 2.30) (Pluskal et al., 2010). To associate lipid spectral matches with their parent features, each feature is modeled using a Gaussian function generated from the feature metadata in the imported table. The following function is used for Gaussian modeling:

$$f(x) = ae^{-\frac{(x-b)^2}{2 * \left(\frac{FWHM}{2.3548}\right)^2}}$$

The parameter *a* is the apex height of the feature, *b* is the apex retention time, and *FWHM* is the full peak width at half of the maximum intensity. MS/MS identifications whose dot-product score is greater than a user-supplied threshold are then linked to chromatographic features if they stem from the same sample and if they share the same charge, precursor *m/z* (within user-supplied tolerance), and retention time. As MS/MS are not necessarily acquired at the feature apex, spectra are assigned to features if the retention time falls within a user-specified factor of the feature's peak width (FWHM). To accurately associate spectra for partially co-eluting isobaric peaks, each spectrum is assigned to the peak which has the highest Gaussian-modeled intensity at that retention time.

Weighted Peak Purity

To quantify co-elution of isobaric features, LipiDex combines the Gaussian feature profile and the deconvoluted spectral purity values to calculate a weighted purity value for the feature across all samples (feature group). All identified lipid spectral components generated from spectral deconvolution and assigned to a specific feature group are weighted according to their proximity to the peak apex using the Gaussian peak model. These weighted values are then used to calculate a weighted average for each component assigned to the feature group.

Spurious Peak Detection Filtering

When analyzing complex LC-MS data, feature detection software routinely identifies chromatographic peaks which cannot be reliably quantified across many samples. These features are often of low abundance, suffer from poor chromatographic resolution, or are artifacts of the specific peak detection routine (Mahieu et al., 2015). To remove these features, LipiDex discards all feature groups which were not detected in at least a user-supplied number of LC-MS experiments from the dataset.

Degenerate Peak Filtering

Following assignment of MS/MS identifications to features, co-eluting feature groups are filtered to remove redundant adduct, multimers, misidentified isotopes, and in-source fragments. Feature groups are considered to co-elute if the apex retention time difference is less than half the average peak width (FWHM). In this way, small differences in apex retention time, which can arise from polarity-switching LC-MS/MS experiments, do not erroneously disqualify potential degenerate feature groups. Co-eluting peaks are considered adducts or dimers of the same parent compound if they satisfy the following equations within a user-defined m/z tolerance:

$$\text{Adduct: } m_1 - a_x = m_2 - a_y$$

$$\text{Dimmer: } (m_1 - a_x) * 2 = m_2 - a_x$$

Here, m refers to the precursor m/z for feature groups 1 or 2 and a denotes the m/z of adducts x or y . All potential adduct combinations for the correct polarity are taken from a user-supplied list in *Peak Finder*. If two features are determined to stem from the same parent compound, the feature with the largest area is retained and the other discarded. If only one of the features is identified as a lipid from its MS/MS spectra, that feature is retained and the feature without an MS/MS spectral identification is discarded. A feature group is considered an in-source fragment if its precursor m/z is found in the assigned MS/MS identification *in-silico* spectrum for any co-eluting feature group. Occasionally, LC-MS feature extraction tools will misidentify the M+1 isotope as the M0 isotope. LipiDex removes these errors by searching for co-eluting peaks whose m/z differs by the mass of a neutron within a user-defined mass tolerance.

Lipid Grouping

Lipid grouping refers to the collapsing of redundant lipid identifications and modulation of the structural resolution embedded in the lipid identification (sum vs. molecular composition) based on the assigned MS/MS identifications. In this way, the minimum set of lipid identifications which supports the observed features and MS/MS are reported. Common lipid grouping scenarios are outlined in Figure S2. After lipid grouping, feature groups are identified with the sum composition (i.e. PC 34:1) when the MS/MS spectra support multiple isobaric identifications (Figure S2C) or contain significant co-fragmentation of isobaric lipids (Figure S2E). Lipids identifications are given the molecular composition (i.e. PC 16:0_18:1) if, after combining redundant adducts, the MS/MS spectra contain fatty-acid identifying fragments (Figures S2A and S2D) with minimal co-fragmentation.

Retention Time Filtering

Lipid retention times (RT) for a specific lipid class roughly approximate a normal distribution in reversed-phase chromatographic separation. In LipiDex, lipid identifications are optionally filtered via retention time using an LC system-agnostic outlier detection. First, apex retention times for all identified feature groups are extracted and grouped together by lipid class. If there are at least four identifications for the specific lipid class, the median absolute RT deviation from the median (MAD) is calculated. MAD represents a more robust method for outlier detection than the standard deviation alone (Leys et al., 2013). Any identification which falls outside of a user-defined multiple of this value (3.5 by default) is considered to be a spurious identification and removed. In cases where there are very few identified lipids from a given class or non-endogenous lipid internal standards are used, true identifications may fall outside the class retention time window. In these or other instances of erroneous outlier detection, the user can either toggle the retention time filtering off or find the removed lipid in the Unfiltered Peak List CSV file.

Hap1 Collection

HAP1 cells were maintained in high glucose (25 mM) Iscove's Modified Dulbecco's Medium supplemented with 10% (v/v) fetal bovine serum and penicillin-streptomycin (IMDM, FBS, PS) in a humidified incubator (37°C, 5% CO₂). Cells were subcultured by trypsinization (0.05% Trypsin-EDTA). Two million HAP1 cells were plated into a 10 cm dish with IMDM, FBS, PS (37°C, 10 mL) and incubated (37°C, 5% CO₂) for 46 h (until the cultures were ~50–60% confluent). The media was gently aspirated and replaced with glucose-free DMEM, FBS, PS with galactose (10 mM) and incubated (37°C, 5% CO₂, 24 h). Immediately prior to cell harvest, the media was aspirated and the cells were gently washed with PBS (~2 mL, ~20°C). The PBS wash was aspirated, PBS (700 μL, 4°C) was added, and the cells were harvested by scraping and transferring the resuspended cells into a 1.5 mL tube on ice. The cells were isolated by centrifugation (5000 g, 1 min, 4°C). The supernatant was aspirated, and the cell pellets were snap frozen in N₂(l) and stored at –80°C.

Yeast Collection

An individual colony of yeast was used to inoculate a starter culture (3 mL pABA-, G, D) and incubated (30°C, 230 rpm, 10–15 h). Media (100 mL media at ambient temperature in a sterile 250 mL Erlenmeyer flask) was inoculated with 2.5×10^6 yeast cells and incubated (30°C, 230 rpm). Samples were harvested 25 h after inoculation. 1×10^8 yeast cells were harvested by centrifugation (3,000 g, 3 min, 4°C), the supernatant was removed, and the cell pellet was flash frozen in liquid nitrogen. Pellets were stored at –80°C prior to extraction.

Hap1 Cell, Yeast, and Mouse Liver Lipid Extraction

Yeast cell pellets, mouse liver homogenates (100 µg), and HAP1 cell pellets were thawed on ice and each sample was mixed with glass beads (0.5 mm diameter, 100 µL). $\text{CHCl}_3/\text{MeOH}$ (1:1, v/v, 4°C) (900 µL) was added and vortexed (2 × 30 s). HCl (1 M, 200 µL, 4°C) was added and vortexed (2 × 30 s). The samples were centrifuged (5,000 g, 2 min, 4°C) to complete phase separation. 500 µL of the organic phase was transferred to a clean tube and dried under Ar(g). The remaining organic residue was reconstituted in ACN/IPA/H₂O (65:30:5, v/v/v) (100 µL) for LC-MS analysis.

Human Plasma Lipid Extraction

Human plasma (Innovative Research, (20 µL) was thawed on ice and vortexed. MeOH (4°C) (225 µL) was added and vortexed (30 s). Methyl tert-butyl ether (750 µL, 4°C) was added and vortexed (30 s). The samples were mixed using an orbital shaker (6 mins) to extract lipids. H₂O (225 µL, 4°C) was added and vortexed (30 s) to induce phase separation. Samples were centrifuged (14,000 g, 2 min, 4°C) to complete phase separation. 200 µL of the organic phase was transferred to a clean vial and dried in a vacuum centrifuge (45 min). The organic residue was reconstituted in MeOH/Toluene (9:1, v/v) (100 µL) for LC-MS/MS analysis.

LC-MS/MS Acquisition

Sample analysis was performed on an Acquity CSH C18 column held at 50°C (2.1 × 100 mm × 1.7 µm particle size; Waters Corporation) using a Vanquish Binary Pump (400 µL/min; Thermo Scientific). Mobile phase A consisted of 10 mM ammonium acetate in ACN/H₂O (70:30, v/v) containing 250 µL/L acetic acid. Mobile phase B consisted of 10 mM ammonium acetate in IPA/ACN (90:10, v/v) with the same additives. Initially, mobile phase B was held at 2% for 2 min and then increased to 30% over 3 min. Mobile phase B was then further increased to 85% over 14 min and then raised to 99% over 1 min and held for 7 min. The column was then re-equilibrated for 5 min before the next injection. Ten microliters of lipid extract were injected by a Vanquish autosampler (Thermo Scientific). The LC system was coupled to a Q Exactive HF mass spectrometer (Thermo Scientific) by a HESI II heated ESI source (Thermo Scientific). The MS was operated in positive and negative mode during sequential injections collecting both MS¹ and MS² spectra. For detailed MS acquisition parameters, reference [Table S3](#).

QUANTIFICATIONS AND STATISTICAL ANALYSIS

LC-MS/MS Data Analysis

The resulting LC-MS data were processed using Compound Discoverer 2.0 (Thermo Scientific). For a detailed description of all data processing parameters used, reference [Table S4](#). Briefly, all MS¹ scans were extracted (*Select Spectra* node) and aligned (*Align Retention Times* node) across all LC-MS/MS experiments. Chromatographic features were subsequently detected from the aligned scans (*Detect Unknown Compounds* node) and grouped according to common *m/z* and elution profile (*Group Unknown Compounds* node). Each LC-MS/MS experiment was then re-searched to minimize feature groups which contained missing values (*Fill Gaps* node). Finally, feature groups which were present in the blank injection were removed (*Mark Background Compounds* node). To generate an unaligned dataset, MS¹ scans were extracted (*Select Spectra* node) and chromatographic features were subsequently detected from the aligned scans (*Detect Unknown Compounds* node) using the same parameters as the aligned dataset. MS/MS spectra were converted to the MGF format using Proteowizard 3.0 (Proteowizard Software Foundation v3.0) and searched against the LipiDex_HCD_Acetate, LipiDex_HCD_Hydroxy, and LipiDex_HCD_ULCFA libraries using *Spectrum Searcher*. In *Peak Finder*, lipid spectral matches were assigned to chromatographic features via precursor *m/z* and retention time. Lipid identifications were further filtered via MS/MS spectral purity, retention time, and chromatographic profile to remove artifacts and redundant identifications.

NIST MS/MS Dot Product Optimization Dataset

MS/MS lipid reference standard spectra which were collected under similar MS/MS conditions as those used in the LC-MS/MS dataset were extracted from the NIST Tandem Mass Spectral Library. For the HCD dataset, 82 phospho-, glycerol-, and sphingo-lipid MS/MS spectra which were fragmented using higher-energy collisional dissociation (HCD) and a collision energy between 40–50 eV were extracted. For the CID dataset, 89 phospho-, glycerol-, and sphingo-lipid MS/MS spectra which were fragmented using ion trap collision induced dissociation were extracted. In *Spectrum Searcher*, the spectra were analyzed using the dot-product equation described above with varying mass and intensity weights (0.1 to 3.0) and the search parameters listed in [Table S3](#). Lipid identifications were deemed correct if they identified the correct fatty acid moieties or, if in the case of indistinguishable spectra (i.e. PC [M+H]⁺ spectra), the correct sum identification.

Lipid Data Analyzer 2 Comparison Dataset

To effectively compare the MS/MS identification performance of LipiDex against the recently published Lipid Data Analyzer 2 (Hartler et al., 2017), a dataset comprising 7 LC-MS/MS analyses of a lipid reference standard mix was downloaded from the LDA 2 software homepage (http://genome.tugraz.at/lda2/data/ControlExperiment1/Exp1_QExactive.zip) and analyzed using both LDA 2 and LipiDex. Lipids were identified in LDA 2 using precursor mass matching parameters of 2 matching isotopes, the Q Exactive positive and negative mass lists from the Biological Experiment dataset (http://genome.tugraz.at/lda2/data/Biological/Bio_QExactive.zip), and the OrbiTrap_exactive_neg and OrbiTrap_exactive_MS/MS rule sets. In LipiDex, MS/MS files were searched against the Lipidex_Formic spectral library using a 0.02 Th MS1 search tolerance, 0.02 MS2 Th search tolerance, 1 max search results returned, and a 61.0 Th MS2 low mass cutoff. To directly compare MS/MS identification performance, the spectral identifications from both programs were matched to the LDA 2-generated feature list and compared at the sum composition and molecular composition level for all lipid class excluding the deuterium-labelled TG standards which are not present included in both programs' libraries For reference standard feature identifications, identifications between LipiDex and LDA 2 were matched if both programs returned the same molecular composition identification. For identification of all lipid features, identifications were matched if both programs returned identical sum lipid identifications. MS/MS signal-to-noise was calculated for each spectrum using the ratio of the most intense spectral peak to the median spectral peak intensity (Yang et al., 2014).

In-Silico Co-Fragmentation Dataset

Lipid co-fragmentation was simulated by mixing experimental isobaric lipid spectra *in-silico* across varying ratios and analyzing the resulting spectra using the spectral deconvolution algorithm. MS/MS identifications were generated from LC-MS/MS data as above and only high-scoring and high-purity MS/MS (dot-product>890, purity>98%, maximum intensity>1.0E⁵) were retained for *in-silico* mixing. Isobaric spectra were then added together and their intensities modulated to mix spectra at various ratios (1:5, 1:4, 1:3, 1:2, 1:1, 2:1, 3:1, 4:1, and 5:1). The mixed spectra were then re-searched using *Spectrum Searcher* to generate spectral purity values.

DATA AND SOFTWARE AVAILABILITY

LipiDex source code, compiled program, manuscript datasets, and test dataset are available via GitHub: <http://www.ncqbc.com/resources/software/>

Proteowizard for MS/MS file conversion is available here: <http://proteowizard.sourceforge.net/downloads.shtml>

Compound Discoverer 2.0 for feature finding is available here: <https://www.thermofisher.com/order/catalog/product/OPTON-30783>

MZmine 2 for feature finding is available via GitHub: <http://mzmine.github.io/>

Raw MS data files for multi-tissue dataset available via Chorus (ID: 1409): <https://chorusproject.org>

Cell Systems, Volume 6

Supplemental Information

**LipiDex: An Integrated Software Package
for High-Confidence Lipid Identification**

Paul D. Hutchins, Jason D. Russell, and Joshua J. Coon

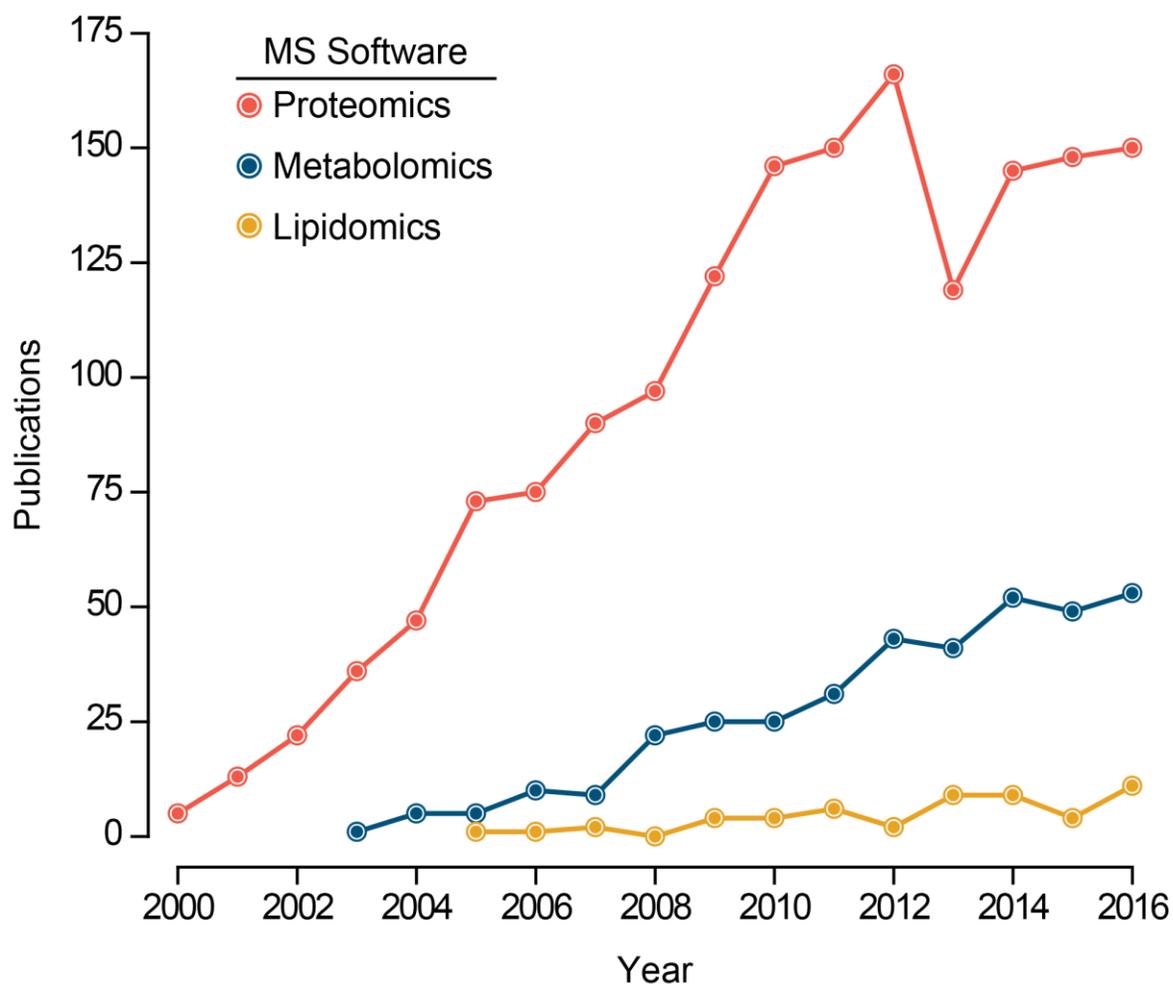


Figure S1. Omics Software Publications, Related to STAR Methods

Number of omics mass spectrometry software manuscripts published between 2000 and 2016 for proteomics (red), metabolomics (blue), and lipidomics (yellow). Note: here we simply aim to show the development of mass spectrometry software and the recent rise but relative scarcity of lipidomic software packages. A rigorous categorization of omic mass spectrometry software is infeasible. As a proxy, we counted publications which contained the Web of Science topics “mass spectrometry”, “software”, and the specific omic category. For reference, 28 of the 52 lipidomics software publications shown here present novel lipidomic software or informatics concepts.

Lipid Software	Peak Detection	Unknown Tracking	Adduct Filtering	Lipid Identification		
				Precursor <i>m/z</i>	MS/MS Matching	Variable Structural Resolution
LipidEx		█	█	█	█	█
Greazy/LipidLama <i>Kochen et al., 2016</i>			█	█	█	
Lipid Data Analyzer 2 <i>Hartler et al., 2017</i>	█		█	█	█	█
LipidFinder <i>O'Connor et al., 2017</i>			█	█		
LipidFrag <i>Witting et al., 2017</i>				█	█	
LipidMatch <i>Koelmel et al., 2017</i>				█	█	█
LIQUID <i>Kyle et al., 2017</i>	█			█	█	
LOBSTAHS <i>Collins et al., 2017</i>			█	█		
MS-DIAL <i>Tsugawa et al., 2015</i>	█	█	█	█	█	
MZmine 2 <i>Pluskal et al., 2010</i>	█	█	█	█		

Figure S2. Major Data Processing Features of Open Source Lipidomics Packages, Related to STAR Methods

Major data processing features (green bars) of common LC-MS software packages including peak detection, result filtering, and lipid identification. Note that although care was taken to carefully classify each package, this table is not meant to be an exhaustive list of either all lipidomics software or all possible data processing routines.

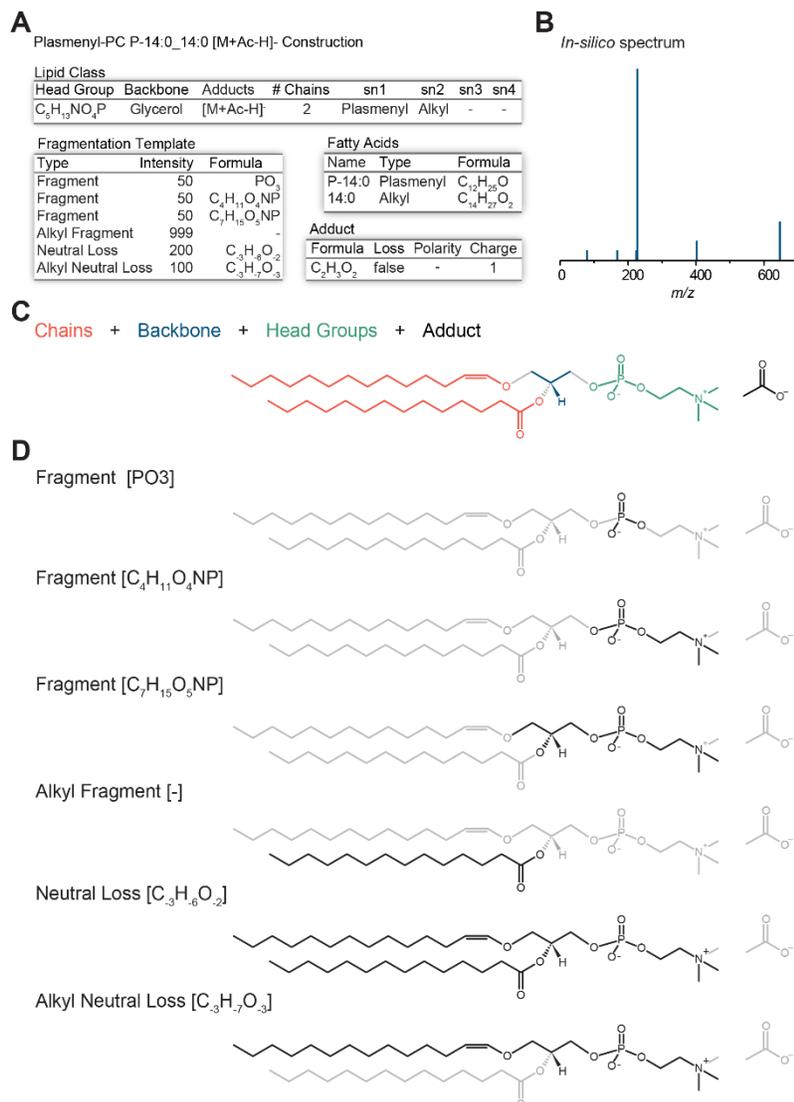


Figure S3. Lipid Fragmentation Modelling using Fragmentation Templates, Related to STAR Methods

(A) Summary lipid class metadata and fragmentation template tables used for *in-silico* generation of theoretical of Plasmenyl-PC P-14:0_14:0 [M+Ac-H]⁻ fragments.

(B) *In-silico* generated spectrum for Plasmenyl-PC P-14:0_14:0 [M+Ac-H]⁻.

(C) Lipid structural component classification used for lipid metadata tables and fragmentation rule.

(D) Structural representation of Plasmenyl-PC P-14:0_14:0 [M+Ac-H]⁻ fragmentation rules. Note these fragment structures are used for illustration of fragment types and are not meant to reflect gas-phase structures.

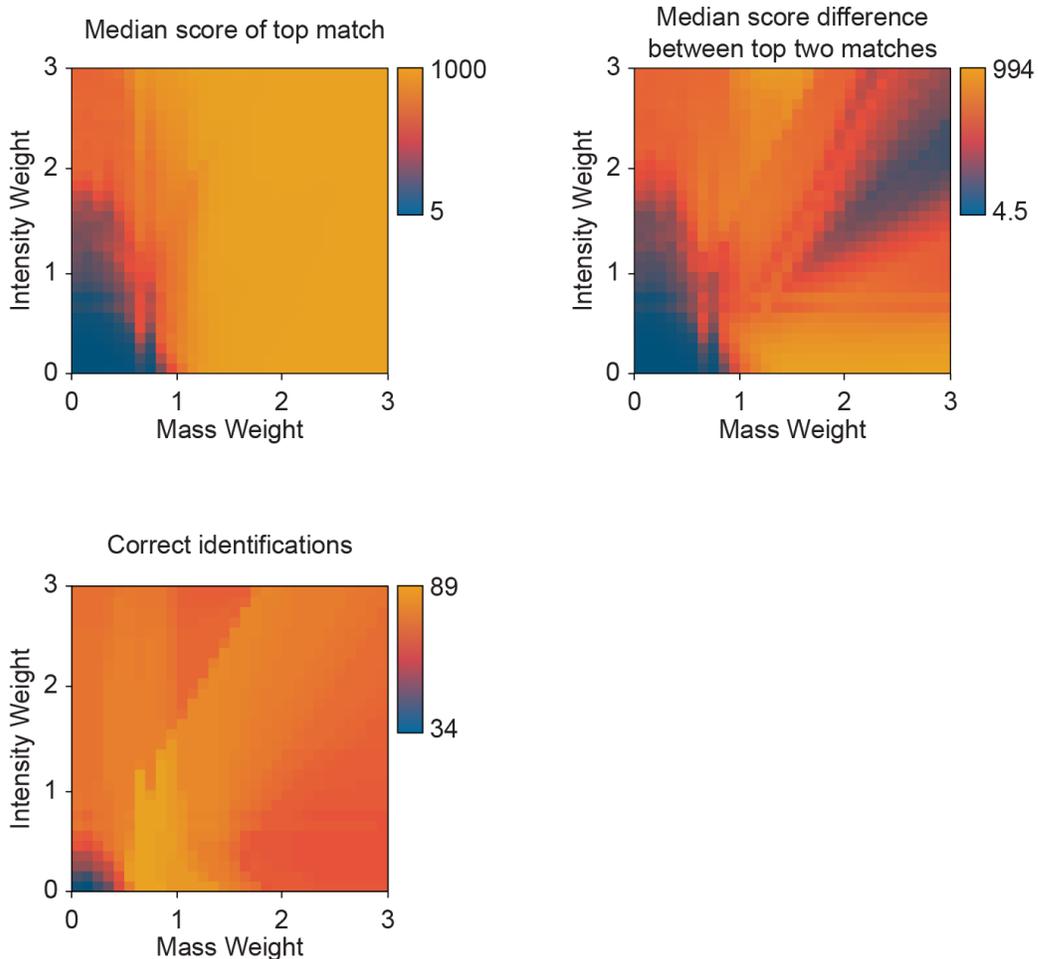


Figure S4. Dot-Product Weighting Factor Optimization for ion trap CID, Related to Figure 1

Color coded values for the median score, median score difference to next match, and correct identifications for the top-scoring spectral matches from 89 lipid reference standard spectra from the NIST Tandem Mass Spectral Library collected using resonant excitation ion trap CID and searched using various combinations of intensity and mass weighting factors.

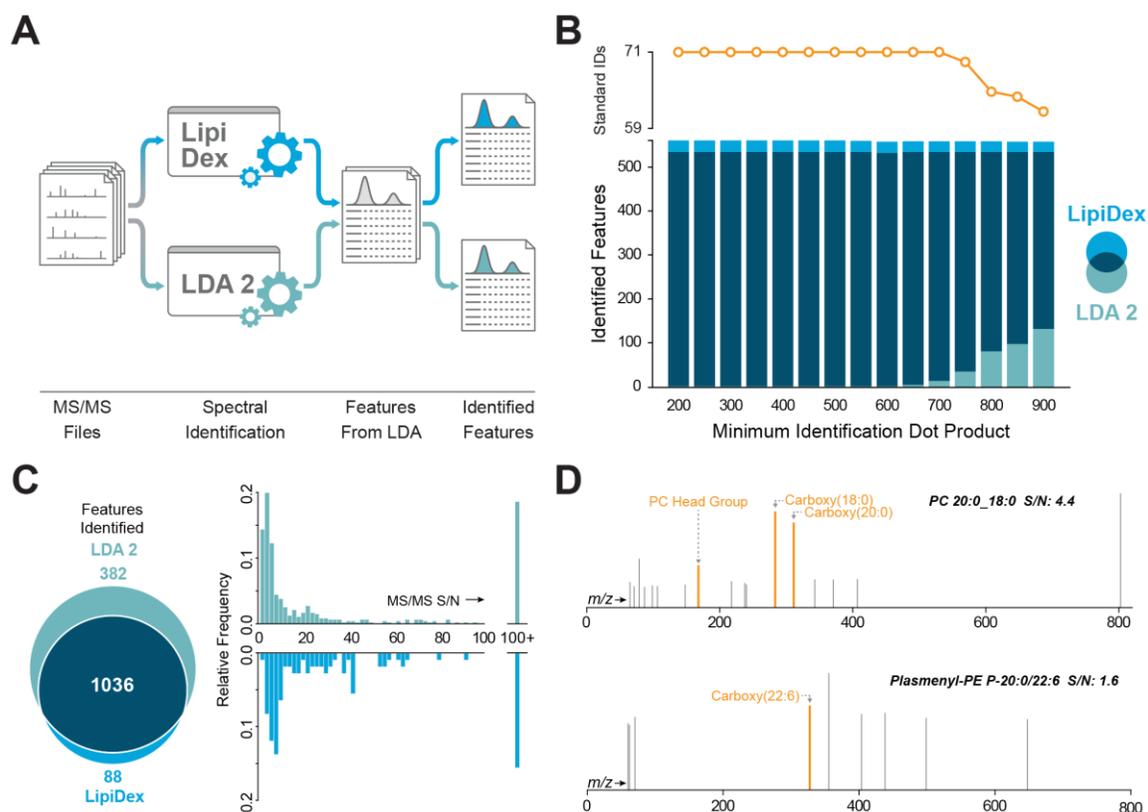


Figure S5. Identification Performance Comparison to Lipid Data Analyzer 2, Related to STAR Methods

(A) Overview of identification comparison workflow between LipiDex and Lipid Data Analyzer 2 (LDA 2).

(B) Reference standard feature sum identification overlap (dark blue) between LipiDex (light blue) and LDA 2 (aqua blue) and number of unique reference standards identified at the molecular composition level by LipiDex (orange dots) out of 71 possible across varying minimum dot product thresholds. Note that MS/MS identification performance does not deteriorate until a very conservative dot-product threshold (>750) is used.

(C) Feature sum identification overlap (dark blue) between LipiDex (light blue) and LDA 2 (aqua blue) for all detected features and the MS/MS signal-to-noise (S/N) ratio for the unique features identified by each package. Note the rule-based matching technique of LDA 2 identifies lipids from MS/MS spectra with lower S/N ratios than LipiDex due to noise peaks artificially reducing the dot-product score.

(D) Example spectra, returned identification, and S/N ratio for two unique features identified by LDA 2 which demonstrate appropriate lipid identification (upper panel) and spurious lipid detection (lower panel) at low S/N levels. Although rule-based matching often returns more lipid identifications from noisy spectra than the more conservative dot-product score, manual verification is needed to ensure these identifications are warranted

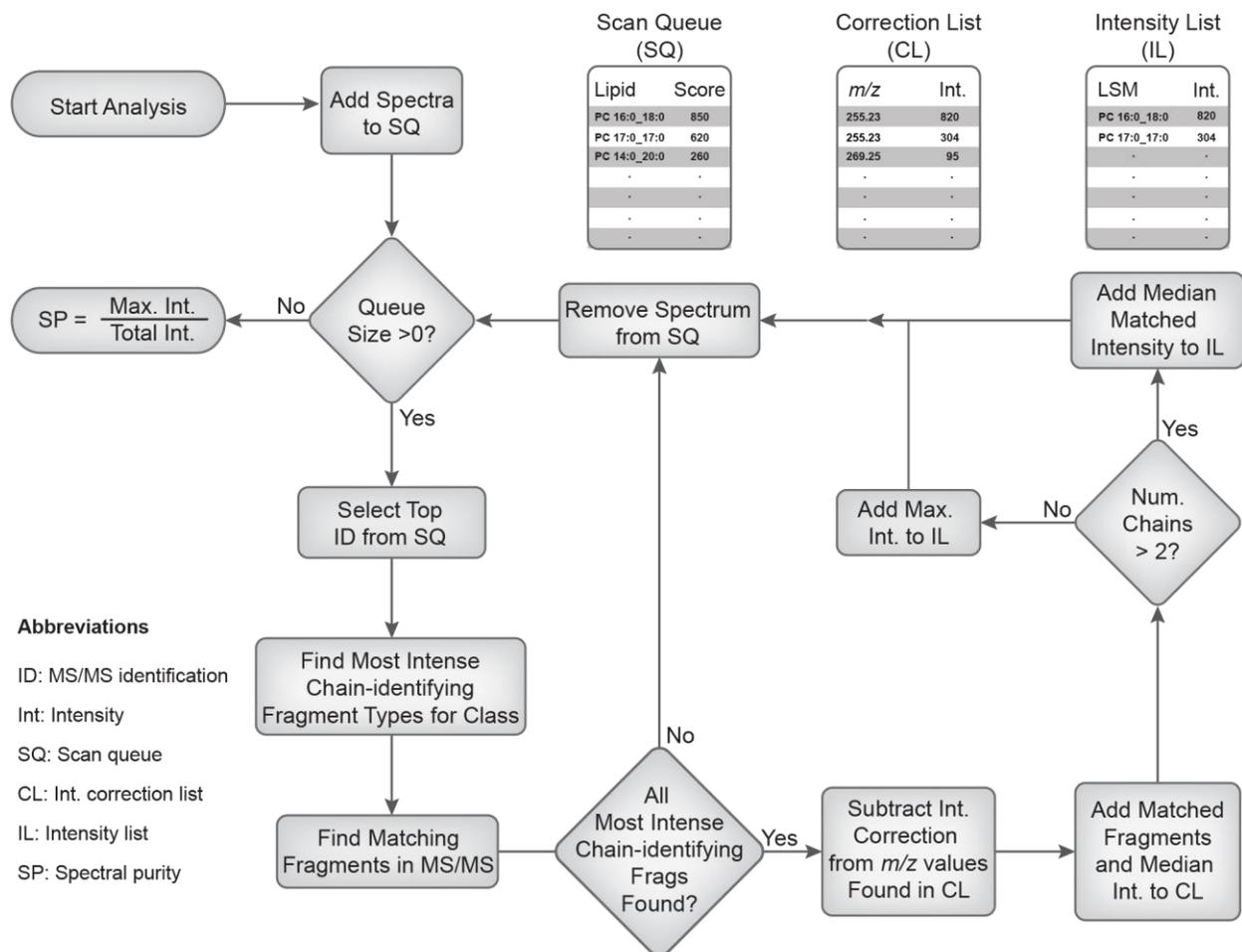


Figure S6. Spectral Deconvolution Logic Flow, Related to Figure 1

The Scan Queue (SQ) is first populated with the spectral matching results and is rank-ordered by the dot-product score. MS/MS peaks corresponding to the most intense fatty acid-identifying fragment type are found and subtracted from the spectrum using the correction factor found in the Correction List (CL). Depending on the number of fatty acid moieties, a specific matched fragment (either the maximum or median intensity chain identifying fragment depending on the lipid) is added to the Intensity List (IL) and used to calculate the relative abundance of the species in the spectrum. This process is repeated until all candidate spectral matches are processed and a spectral purity value is calculated.

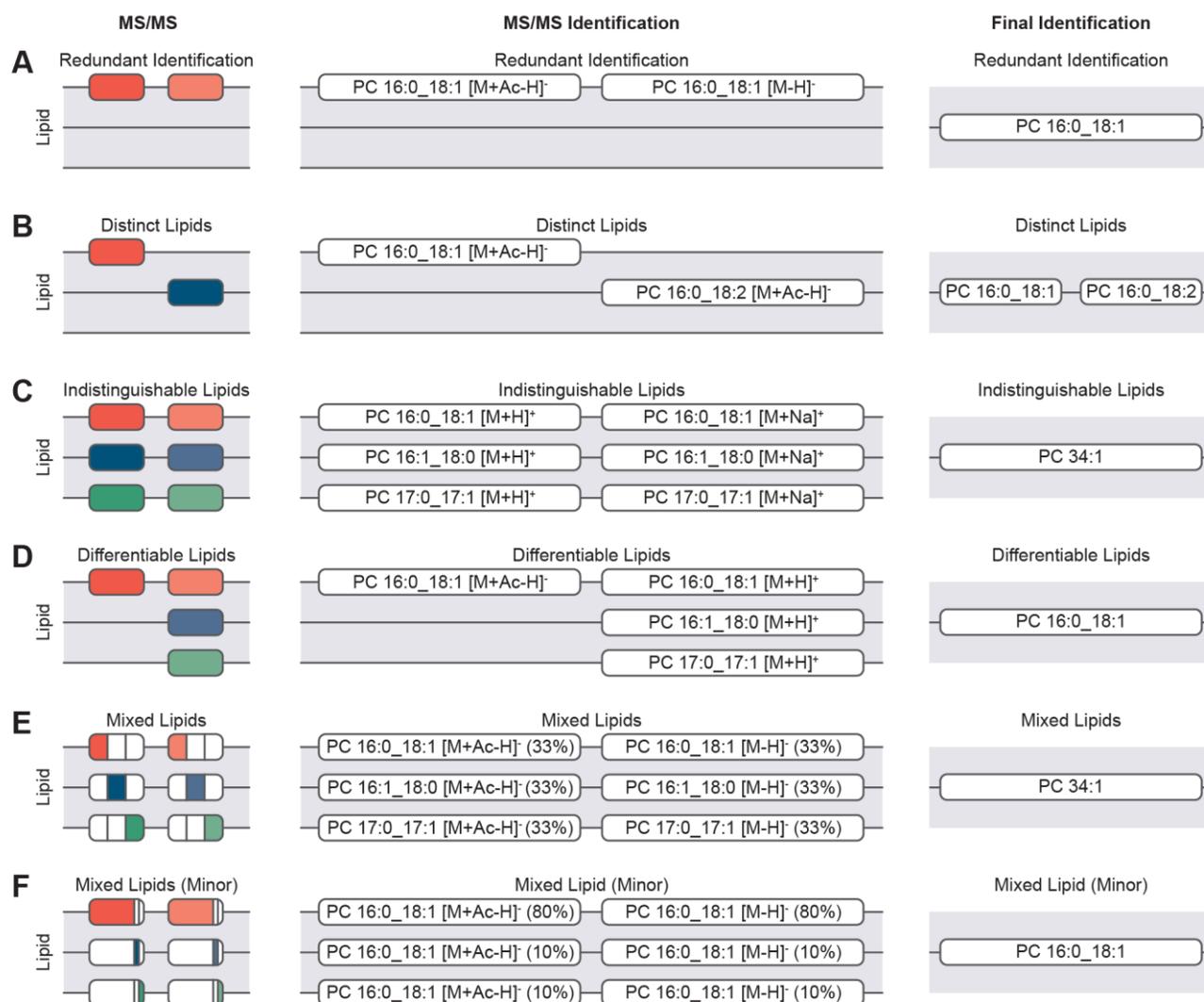


Figure S7. Common Lipid Grouping Scenarios and Final Identifications, Related to Figure 2

- (A) Redundant adduct (lighter shade) identification annotated with molecular lipid composition.
- (B) Distinct lipid identifications annotated separately with molecular lipid composition.
- (C) Indistinguishable spectra which contain no fatty acid-specific fragments identified with sum molecular composition.
- (D) Lipids differentiated with MS/MS which contains fatty-acid specific fragment annotated with molecular lipid composition.
- (E) Mixed lipid spectra which contain major co-fragmentation of isobaric species annotated with sum composition.
- (F) Mixed lipid spectra which contain minor co-fragmentation of isobaric species annotated with molecular composition.

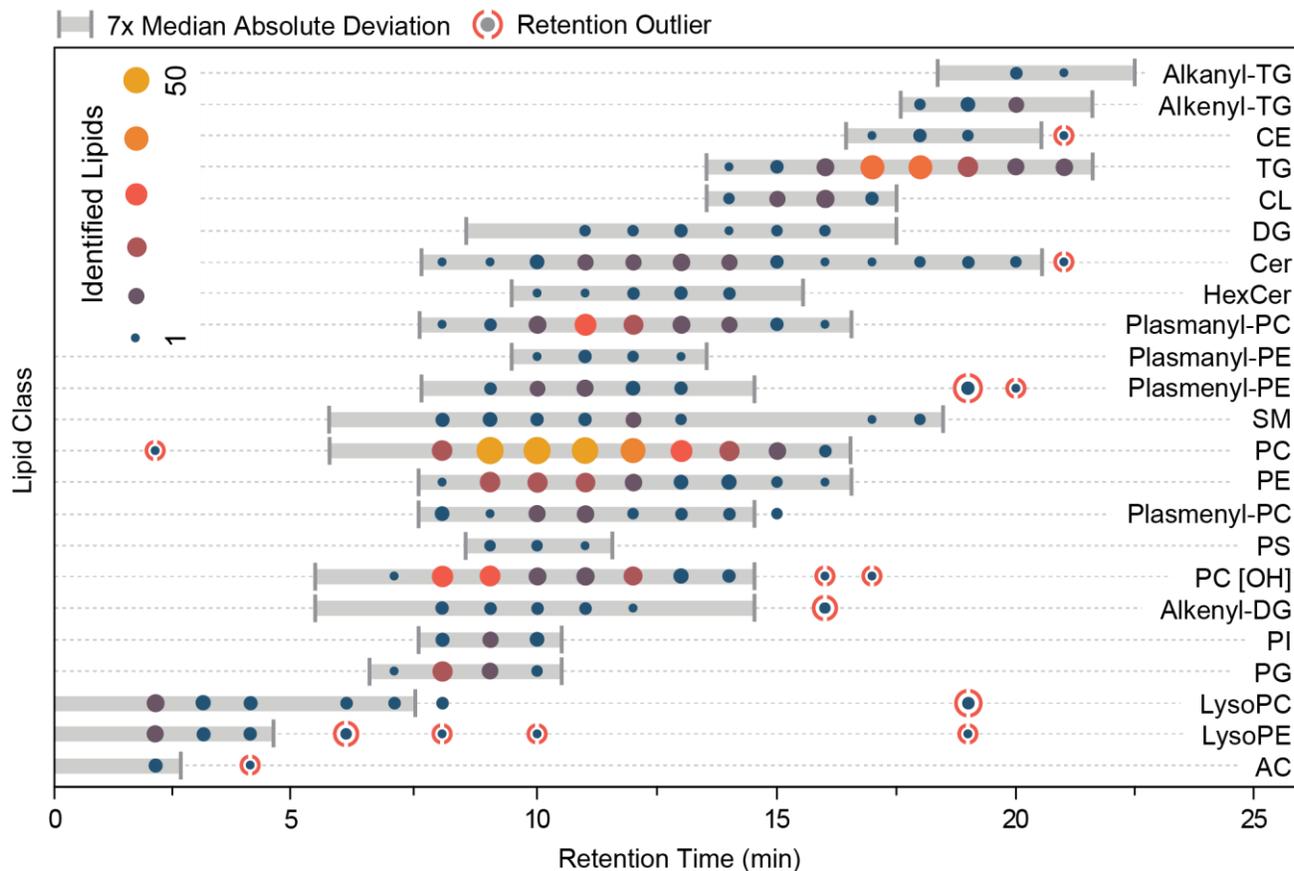


Figure S8. Lipid Retention Time Outlier Filtering, Related to Figure 2

Lipid class retention time distributions are shown for the human Hap1 cell dataset. Each cluster corresponds to the binned retention time of all identified lipids from unique lipid class and bubble size and color denote the number of identified lipids per retention time bin. Grey bars represent the default allowed retention time window corresponding to 7x the median absolute retention time deviation for each lipid class.

Lipidex Library

Abbreviation	Lipid Class	Adducts
AC	Acylcarnitines	[M+H] ⁺
Alkanyl-TG	Alkanyl Triacylglycerol	[M+NH ₄] ⁺
Alkanyl-DG	Alkanyl Diacylglycerol	[M+H] ⁺
Alkenyl-TG	Alkenyl Triacylglycerol	[M+NH ₄] ⁺
Alkenyl-DG	Alkenyl Diacylglycerol	[M+H] ⁺
CE	Cholesteryl ester	[M+NH ₄] ⁺
Cer [AP]	CeramideAP	[M-H] ⁻ ; [M+Ac-H] ⁻
Cer [AS]	CeramideAS	[M+Ac-H] ⁻ ; [M-H] ⁻
Cer [BS]	CeramideBS	[M-H] ⁻ ; [M+Ac-H] ⁻
Cer [NP]	CeramideNP	[M-H] ⁻ ; [M+Ac-H] ⁻
Cer [NS]	CeramideNS	[M+H] ⁺ ; [M+Ac-H] ⁻ ; [M-H] ⁻ ; [M+H-H ₂ O] ⁺
CerP	Ceramide-1-Phosphate	[M+H] ⁺ ; [M-H] ⁻
CL	Cardiolipin	[M-H] ⁻ ; [M] ⁺ ; [M-2H] ²⁻
DG	Diacylglycerol	[M+NH ₄] ⁺
HexCer [AP]	Hexosyl CeramideAP	[M+Ac-H] ⁻ ; [M+FA-H] ⁻
HexCer [NS]	Hexosyl CeramideNS	[M+H] ⁺ ; [M-H] ⁻ ; [M+Ac-H] ⁻
LysoPC	Lysophosphatidylcholine	[M+H] ⁺ ; [M+Ac-H] ⁻
LysoPE	Lysophosphatidylethanolamine	[M+H] ⁺ ; [M-H] ⁻
LysoPG	Lysophosphatidylglycerol	[M-H] ⁻
LysoPI	Lysophosphatidylinositol	[M-H] ⁻
LysoPS	Lysophosphatidyl serine	[M-H] ⁻
LysoSM	Lysosphingomyelin	[M+H] ⁺ ; [M+Ac-H] ⁻
Methyl-PA	Methylphosphatidic Acid	[M-H] ⁻
MG	Monoacylglycerol	[M+NH ₄] ⁺
PA	Phosphatidic acid	[M+NH ₄] ⁺ ; [M-H] ⁻
PC	Phosphatidylcholine	[M+H] ⁺ ; [M+Ac-H] ⁻
PE	Phosphatidylethanolamine	[M+H] ⁺ ; [M-H] ⁻
PE-NMe	Monomethyl Phosphatidylethanolamine	[M+H] ⁺ ; [M-H] ⁻
PE-NMe2	Dimethyl Phosphatidylethanolamine	[M+H] ⁺ ; [M-H] ⁻
PG	Phosphatidylglycerol	[M+NH ₄] ⁺ ; [M-H] ⁻
PI	Phosphatidylinositol	[M-H] ⁻ ; [M+NH ₄] ⁺ ; [M+H] ⁺
Plasmanyl-PC	Plasmanyl Phosphatidylcholine	[M+H] ⁺ ; [M+Ac-H] ⁻
Plasmanyl-PE	Plasmanyl Phosphatidylethanolamine	[M+H] ⁺ ; [M-H] ⁻
Plasmenyl-PC	Plasmenylphosphatidylcholine	[M+H] ⁺ ; [M+Ac-H] ⁻ ; [M+FA-H] ⁻
Plasmenyl-PE	Plasmenylphosphatidylethanolamine	[M+H] ⁺ ; [M-H] ⁻
PS	Phosphatidylserine	[M+H] ⁺ ; [M-H] ⁻
S1P	Sphingosine-1-Phosphate	[M+H] ⁺ ; [M-H] ⁻
SHexCer	Sulfatides	[M+H] ⁺ ; [M-H] ⁻
SM	Sphingomyelin	[M+H] ⁺ ; [M+Ac-H] ⁻ ; [M+FA-H] ⁻
SP	Sphingosine	[M+H] ⁺ ; [M+H-H ₂ O] ⁺
TG	Triacylglycerol	[M+NH ₄] ⁺ ; [M+Na] ⁺

Table S1. Lipidex Lipid Library Composition, Related to Figure 1

Lipidex includes a pre-built library which were generated internally. The Lipidex library was generated exclusively from HCD spectra collected on the Q-Exactive (Thermo Fisher) platform.

LipidBlast Library

Abbreviation	Lipid Class	Adducts
AC	Acylcarnitines	[M+H] ⁺ ; [M] ⁺
BMP	Bis-monoacylglycero-phosphate	[M+NH ₄] ⁺
Cer [ADS]	CeramideADS	[M-H] ⁻ ; [M+Ac-H] ⁻ ; [M+FA-H] ⁻
Cer [AP]	CeramideAP	[M-H] ⁻ ; [M+Ac-H] ⁻ ; [M+FA-H] ⁻
Cer [AS]	CeramideAS	[M-H] ⁻ ; [M+Ac-H] ⁻ ; [M+FA-H] ⁻
Cer [BDS]	CeramideBDS	[M-H] ⁻ ; [M+Ac-H] ⁻ ; [M+FA-H] ⁻
Cer [BS]	CeramideBS	[M-H] ⁻ ; [M+Ac-H] ⁻ ; [M+FA-H] ⁻
Cer [NDS]	CeramideNDS	[M-H] ⁻ ; [M+Ac-H] ⁻ ; [M+FA-H] ⁻
Cer [NP]	CeramideNP	[M-H] ⁻ ; [M+Ac-H] ⁻ ; [M+FA-H] ⁻
Cer [NS]	CeramideNS	[M-H] ⁻ ; [M+Ac-H] ⁻ ; [M+FA-H] ⁻
DG	Diacylglycerol	[M+NH ₄] ⁺
GlcCer [ADS]	HexosylCeramideADS	[M-H] ⁻ ; [M+Ac-H] ⁻ ; [M+FA-H] ⁻
GlcCer [AP]	HexosylCeramideAP	[M-H] ⁻ ; [M+Ac-H] ⁻ ; [M+FA-H] ⁻
GlcCer [AS]	HexosylCeramideAS	[M-H] ⁻ ; [M+Ac-H] ⁻ ; [M+FA-H] ⁻
GlcCer [BDS]	HexosylCeramideBDS	[M-H] ⁻ ; [M+Ac-H] ⁻ ; [M+FA-H] ⁻
GlcCer [BS]	HexosylCeramideBS	[M-H] ⁻ ; [M+Ac-H] ⁻ ; [M+FA-H] ⁻
GlcCer [NDS]	HexosylCeramideNDS	[M-H] ⁻ ; [M+Ac-H] ⁻ ; [M+FA-H] ⁻
GlcCer [NP]	HexosylCeramideNP	[M-H] ⁻ ; [M+Ac-H] ⁻ ; [M+FA-H] ⁻
GlcCer [NS]	HexosylCeramideNS	[M-H] ⁻ ; [M+Ac-H] ⁻ ; [M+FA-H] ⁻
LysoPA	Lyso phosphatidic acid	[M+NH ₄] ⁺ ; [M-H] ⁻
LysoPS	Lysophosphatidyl serine	[M-H] ⁻
LysoPC	Lysophosphatidylcholine	[M+H] ⁺ ; [M+Ac-H] ⁻ ; [M+FA-H] ⁻
LysoPE	Lysophosphatidylethanolamine	[M+H] ⁺ ; [M-H] ⁻
LysoPG	Lysophosphatidylglycerol	[M-H] ⁻
LysoPI	Lysophosphatidylinositol	[M-H] ⁻
LysoSM	Lysosphingomyelin	[M+H] ⁺ ; [M+Ac-H] ⁻ ; [M+FA-H] ⁻
MG	Monoacylglycerol	[M+NH ₄] ⁺
PA	Phosphatidic acid	[M+NH ₄] ⁺ ; [M-H] ⁻
PC	Phosphatidylcholine	[M+H] ⁺ ; [M+Ac-H] ⁻ ; [M+FA-H] ⁻
PE	Phosphatidylethanolamine	[M+H] ⁺ ; [M-H] ⁻
PG	Phosphatidylglycerol	[M+NH ₄] ⁺ ; [M-H] ⁻
PI	Phosphatidylinositol	[M-H] ⁻ ; [M+NH ₄] ⁺
PS	Phosphatidylserine	[M+H] ⁺ ; [M-H] ⁻
Plasmenyl-PC	Plasmenylphosphatidylcholine	[M+H] ⁺ ; [M+Ac-H] ⁻ ; [M+FA-H] ⁻
Plasmenyl-PE	Plasmenylphosphatidylethanolamine	[M+H] ⁺ ; [M-H] ⁻
SM	Sphingomyelin	[M+H] ⁺ ; [M+Ac-H] ⁻ ; [M+FA-H] ⁻
TG	Triacylglycerol	[M+NH ₄] ⁺ ; [M+Na] ⁺

Table S2. LipidBlast Lipid Library Composition, Related to Figure 1

Lipidex includes an imported library which was generated from the MS-Dial LipidBlast fork (Tsugawa et al., 2015). The LipidBlast library was generated from diverse MS/MS setups and is most performant when used with MS/MS spectra fragmented using collision-induced dissociation (CID).

Mouse liver, human plasma, and yeast cell**Hap1 Cells**

Positive Polarity		Positive Polarity	
Parameter	Value	Parameter	Value
Polarity	positive	Polarity	positive
Full MS Resolution	60000	Full MS Resolution	60000
Full MS AGC Target	3.00E+06	Full MS AGC Target	3.00E+06
Full MS Max. IT (ms)	100	Full MS Max. IT (ms)	100
Full MS Scan Range	<i>m/z</i> 400 - 1200	Full MS Scan Range	<i>m/z</i> 240 - 1200
MS/MS Resolution	15,000	MS/MS Resolution	15,000
MS/MS AGC Target	1.00E+06	MS/MS AGC Target	5.00E+05
MS/MS Max. IT (ms)	160	MS/MS Max. IT (ms)	80
MS/MS Loop Count	5	MS/MS Loop Count	10
MS/MS Isolation Window	1.4	MS/MS Isolation Window	1.4
MS/MS Stepped NCE	20, 25	MS/MS Stepped NCE	20, 25
Dynamic Exclusion (s)	7.5	Dynamic Exclusion (s)	7.5

Negative Polarity		Negative Polarity	
Parameter	Value	Parameter	Value
Polarity	negative	Polarity	negative
Full MS Resolution	60000	Full MS Resolution	60000
Full MS AGC Target	3.00E+06	Full MS AGC Target	3.00E+06
Full MS Max. IT (ms)	100	Full MS Max. IT (ms)	100
Full MS Scan Range	<i>m/z</i> 240 - 1600	Full MS Scan Range	<i>m/z</i> 240 - 1600
MS/MS Resolution	15,000	MS/MS Resolution	15,000
MS/MS AGC Target	1.00E+06	MS/MS AGC Target	5.00E+05
MS/MS Max. IT (ms)	160	MS/MS Max. IT (ms)	80
MS/MS Loop Count	5	MS/MS Loop Count	10
MS/MS Isolation Window	1.4	MS/MS Isolation Window	1.4
MS/MS Stepped NCE	20, 25	MS/MS Stepped NCE	20, 30

Table S3. LC-MS/MS Analysis Parameters, Related to Figure 2

Detailed data collection and parameters for the experimental dataset presented in the manuscript. Note that, although optimized, the parameters represent a suitable starting point for lipid analysis for the QE HF and should be systematically optimized for users' specific LC-MS/MS setup.

Select Spectra		Detect Unknown Compounds	
Parameter	Value	Parameter	Value
Precursor Selection	Use MS(n - 1) Precursor	Mass Tolerance (ppm)	10 ppm
Lower RT Limit	0.4	Intensity Tolerance (%)	100
Upper RT Limit	0	S/N Threshold	3
First Scan	0	Min. Peak Intensity	100000
Last Scan	0	Ions	[M+H] ⁺ 1; [M-H] ⁻ 1
Ignore Specified Scans		Min. Element Counts	C H
Lowest Charge State	0	Max. Element Counts	C100 H190 N10 Na2 O15 P2 S2
High Sharge State	0		
Min. Precursor Mass	100 Da	Group Unknown Compounds	
Max. Precursot Mass	5000 Da	Parameter	Value
Total Intensity Threshold	0	Mass Tolerance	10 ppm
Minimum Peak Count	1	RT Tolerance [min]	0.2
Mass Analyzer	Any	Rule #1	Unspecified
MS Order	Any	Rule #2	Unspecified
Activation Type	Any	Preferred MS Order	MS1
Min. Collision Energy	0	Preferred Ion	[M+H] ⁺ 1
Max. Collision Energy	1000		
Scan Type	Any	Fill Gaps	
Polarity Mode	(Not specified)	Parameter	Value
S/N Threshold (FT-only)	1.5	Mass Tolerance	10 ppm
Unrecognized Charge Replacement	1	RT Tolerance [min]	0.2
Unrecognized Mass Analyzer Replacement	FTMS	S/N Threshold	1.5
Unrecognized MS Order Replacements	MS1		
Unrecognized Activation Type Replacements	HCD	Mark Background Compounds	
Unrecognized Polarity Replacements	+	Parameter	Value
Unrecognized MS Resolution Replacements	60000	Max. Sample/Blank	3
Unrecognized MSn Resolutino Replacements	15000	Max. Blank/Sample	0
		Hide Background	True
Spectrum Searcher		Peak Finder	
Parameter	Value	Parameter	Value
Library	Lipidex_HCD_Acetate Lipidex_HCD_Hydroxy Lipidex_HCD_ULCFA	File Type	Compound Discoverer
MS1 Search Tolerance	0.01 Th	Min. Lipid Spectral Purity	75%
MS2 Search Tolerance	0.01 Th	Min. MS2 Search Dot Prod.	500
		Min MS2 Search Rev. Dot Product	700
		FWHM Window Multiplier	2

Table S4. Data Analysis Parameters, Related to Figure 2

Detailed processing parameters for the experimental dataset presented in the manuscript.