Published in final edited form as:

Rapid Commun Mass Spectrom. 2014 April 30; 28(8): 981–985. doi:10.1002/rcm.6865.

LipidMiner: a software for automated identification and quantification of lipids from multiple liquid chromatography-mass spectrometry data files

Da Meng¹, Qibin Zhang^{2,*}, Xiaoli Gao^{2,#}, Si Wu³, and Guang Lin¹

¹Computational Science and Mathematics Division, Pacific Northwest National Laboratory, Richland, WA 99352, USA

²Biological Sciences Division, Pacific Northwest National Laboratory, Richland, WA 99352, USA

³Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, WA 99352, USA

Dear Editor

Comprehensive understanding of the roles and functions of lipids in cellular physiology and pathology requires unambiguous identification and accurate quantification of individual lipid molecular species. However, the enormous structural diversity of lipids (> 37,000 lipid molecules cataloged in the LIPID MAPS database at www.lipidmaps.org) presents significant challenge in high throughput analysis of lipidomics data.^[1, 2] Traditionally, triple quadrupole or quadrupole time of flight (QTOF) mass spectrometers are utilized for lipids analysis using either precursor ion scanning (PIS) or neutral loss scanning (NLS) methods, which exploit lipid class-characteristic ions or neutral loss fragments generated in collisional induced dissociation of gas phase lipid molecular ions. These methods are readily carried out in direct infusion mode (also known as shotgun lipidomics^[3]) without on-line chromatographic separation of lipids, and software tools such as LIMSA,^[4] Lipid Profiler,^[5] AMDMS-SL^[6] and MS-LAMP^[7] were written to specifically analyze these types of data. Recently, data-dependent acquisition of full scan MS/MS spectra from all detectable precursor ions has gained increasing use in lipid profiling studies. This method is mainly implemented on ion trap, high resolution QTOF or hybrid Orbitrap instruments. The spectra obtained can be considered as the emulation of simultaneous acquisition of an unlimited number of product and neutral loss scans in a single analysis. Accordingly, tools for analyzing this type of data were also being developed, such as LipidQA,^[8] LipidInspector,^[9] and LipidXplorer,^[10] mainly for direct infusion based shotgun lipidomics workflow (Common features of previous lipidomics software tools were summarized in Reference 10). More recently, an *in-silico* tandem mass spectral database, LipidBlast was developed to match with the experimentally acquired MS/MS spectrum,^[11] which provided a new avenue for lipids identification. We and others have found that coupling liquid chromatographic (LC) separation with data-dependent MS/MS adds additional confidence in

*Correspondence to: Q. Zhang, P.O. Box 999, MS K8-98, Richland, WA 99352, USA., qibin.zhang@pnnl.gov.

#Current address: Department of Biochemistry, University of Texas Health Science Center at San Antonio, San Antonio, TX, 78229

lipid isobar/isomer identification, particularly for quantifying lipid molecular species from complex mixtures.^[12, 13] This is due to the fact that even when signals of different species overlap in the mass dimension, they usually can be separated in the retention time domain to a degree that allows reliable assignment and quantification. Although lipID^[14] was developed to handle LC-MS based lipidomics data analysis, it has limited capability in processing LC-MS/MS data. To more effectively analyze this type of multidimensional data (intensity, m/z from MS, m/z from MS/MS and chromatographic retention time) in LC-MS/MS based lipidomics, we developed LipidMiner to facilitate automated identification of lipids from LC-MS data files acquired with data-dependent full-scan MS/MS, and quantification of identified lipid molecular species from multiple data files that are generated either from technical replicates of the same sample or from biological replicates of the same type of biospecimen. The application of LipidMiner to identify lipids from complex samples has been demonstrated in a previous report,^[12] here, we report on the development and functions of this software.

LipidMiner consists of a graphical user interface (GUI) written in Python and core functions written in C#. A screenshot of the user interface is shown in Figure 1. LipidMiner is composed of three functional modules: detection and quantification of lipid features from each raw data file, as well as assignment of lipid class to detected lipid features (Ion Detection); chromatographic alignment of detected lipid features across multiple data files (Feature Alignment); and identification of lipid features through accurate mass matching with a comprehensive lipid library (Library Match). As outlined in Figure 2, each LC-MS/MS data file containing both MS and MS/MS scans are processed in the Ion Detection module, starting with directly reading all MS/MS scans of each raw file, followed by generating a list for all fragment ions detected in each MS/MS scan above predefined intensity and signal/noise thresholds. The software then calculates the mass difference between the precursor ion and each of its fragment ions to generate a list of neutral loss masses. Based on a user-defined list of signature product ions (for emulated PIS) or neutral loss masses (for emulated NLS) for a list of lipid classes of interest and the mass measurement accuracy of the mass spectrometer used, the Ion Detection module will annotate the precursor ions with their potential lipid classes if matches can be found within the user-defined list. Ion envelopes detected in each MS scan are then deisotoped and only the monoisotopic ion of each isotopic envelope is listed for further quantification. If a precursor ion selected for fragmentation is not the monoisotopic ion as determined from the preceding MS survey scan, then precursor m/z will be replaced with the correct monoisotopic m/z . Typically in data-dependent MS/MS acquisition with dynamic exclusion, MS/MS is not acquired when the precursor ion is at its highest intensity (at the deconvoluted peak apex). For each of the monoisotopic ions detected in the MS scans, an extracted ion chromatogram is constructed in three-dimensional space to find the chromatographic peak by a Gaussian Kernel smoother and Gaussian model fitting, so peak apex and peak boundaries where the signal height reaches a threshold of 2 times the background level or until 10 consequent scans have been identified as noise spikes can be identified. Compared to other similarly sized filters, e.g. mean filter, a Gaussian Kernel smoother can better preserve edges when applied to “bell-shaped” peaks.^[15] Peak area integration is carried out through summing all the ion intensities in the region specified by Gaussian peak edges. In

instances where the same precursor ion triggered multiple MS/MS scanning events within the same chromatographic boundary, these precursor ions and their MS/MS scans will be merged into one feature to eliminate the number of identifications and simplify the peak alignment in the following data processing module. The final output from the Ion Detection module are three comma-delimited text files, with the first file only containing class-annotated lipid features and their abundances, chromatographic peak boundaries, and retention times at peak apexes, and the second file in the same format as the first, but also containing all precursor ions that are being fragmented in MS/MS scan events regardless of their annotation status. The third file contains all MS features (m/z , retention time and abundance) identified in the raw file.

Following lipid feature detection in each individual raw data file, identical annotated features across multiple raw files are chromatographically aligned for their subsequent quantitative comparison. Chromatographic alignment is a two-step process: 1) identification of the alignment relationship using training/reference features from different raw files, and 2) application of this relationship to the target features to generate a quantification table. To obtain a reliable alignment relationship, the annotated and abundant lipid features, which represent a more confident data subset, are used as the training/reference features while the target feature sets can be either detected lipid features or all MS features. A pairwise alignment is used in LipidMiner to align similar features together using a user-defined retention time and mass tolerance, which is different from the LCMSWARP algorithm used in our previous work.^[12] When performing pairwise alignment, the data file with the largest number of features is used as a reference set. In data-dependent MS/MS acquisition with dynamic exclusion, the low abundance ions may not be fragmented at each LC-MS run even if the same sample is being analyzed multiple times, which would result in missing values in quantification if only annotated lipid features were used in the chromatographic alignment. One key feature in our alignment algorithm is that abundance of the quantified ion will be filled in even if this ion was not selected for MS/MS fragmentation within the average chromatographic peak boundaries that this lipid feature appears across all raw files. The fill-in is accomplished by specifically looking for the ion within m/z and retention time tolerance in the LC-MS features table generated in the Ion Detection module. This step greatly alleviates the need for missing value-imputation commonly required for omics data. The final output from the Alignment module is a tab-delimited text file in cross-tab format containing lipid features annotated with average m/z and retention time values, and with abundances (integrated peak area) across all data files.

The last module is for assignment of lipid molecular species information to annotated lipid features by matching to a lipid library. Owing to its comprehensiveness with over 37,000 lipid molecular species and its standardized nomenclature, the downloadable comma-separated values (csv) version (<http://www.lipidmaps.org>) of LIPID MAPs structure database is used as the lipid library. The matching process starts with reducing each monoisotopic lipid ion to its charge neutral form by considering ions detected in the positive electrospray ionization mode as the following potential adducts: $[M + H]^+$, $[M + NH_4]^+$, $[M + Li]^+$, $[M + Na]^+$, $[M + K]^+$; and those detected in the negative mode as $[M - H]^-$ or $[M + CH_3COO]^-$. Then the calculated neutral mass is matched with the accurate mass of each

lipid species in the library according to lipid class (based on the lipid class annotations obtained from the Ion Detection module) within a user-defined mass tolerance. For those classes of lipids that do not have well-defined signature ions representing head groups or structure specific neutral loss fragments, an option is provided for searching the entire list of compounds in the library, albeit with associated risk of increasing false positive identifications. The final output from LipidMiner is cross-tab Excel file containing identified lipid molecular species and their abundances across multiple samples, which can be imported into various software for down-stream statistical analysis. The utilization of LIPID MAPs structure database as the reference library and the molecular lipid identifiers therein also take advantage of the resources and tools associated with LIPID MAPs (<http://www.lipidmaps.org>), such as interpretation of the biochemical pathways of identified lipid species.

The comprehensiveness of LipidMiner in identification of lipids has been demonstrated previously in profiling of lipids extracted from complex samples, in which 370 and 444 lipid species were identified from a model skin tissue and rat plasma, respectively.^[12] As a demonstration of its quantification capability, multiple LC-MS/MS data files from different concentrations of a mixture of lipid standards were processed. Lipid standards were purchased from Avanti Polar Lipids (Alabaster, AL), and equimolar standard mixtures were prepared and analyzed as previously described.^[12] Briefly, the standard mixture was injected onto an in-house packed capillary column (150 μm i.d. $\times$ 20 cm, with particles of Waters HSS T3, 1.8 μm), and separation was carried out on a Waters NanoAcquity LC system. The column was maintained at 40°C in a column oven, and gradient elution was performed over 90 min with the mobile phases of $\text{CH}_3\text{CN}/\text{H}_2\text{O}$ (40/60) and $\text{CH}_3\text{CN}/\text{IPA}$ (10/90), both containing 10 mM NH_4OAc . The LC system was interfaced to a LTQ-Orbitrap mass spectrometer (Thermo Scientific, San Jose, CA) with electrospray ionization. Data-dependent MS/MS scan events were performed with one high resolution MS survey scan followed by 5 MS/MS scans using higher energy collisional dissociation (HCD). Data-dependent dynamic exclusion was enabled to prevent the same ion from being selected for repeated fragmentation within 60s of its first selection. The precursor ion selection window was set at 3 m/z units. When the signature headgroup ions and neutral loss fragments representing the lipid standards were used in the Ion Detection module, all lipid standards were clearly identified in the final output file after automated peak detection and matching with the LIPID MAPs library. The peak area obtained from automated integration in LipidMiner is correlated very well with the manual integration value obtained from extracted ion chromatogram using XCalibur software (Figure 3). In addition, the alignment of features across multiple data files filled in the missing abundance values for some of the less abundant lipids, as no MS/MS spectra were collected for some precursor ions in those datasets, probably due to the low intensity of these precursor ions when the concentration of standards was very low.

In conclusion, we have developed a tool for automated, high-throughput analysis of multiple LC-MS/MS data files, which greatly simplifies LC-MS based lipidomics analysis. In addition, the workflow implemented in LipidMiner is not limited to identification and quantification of lipids. If a suitable metabolite library is implemented in the library

matching module, LipidMiner could be reconfigured as a tool for general metabolomics data analysis. It is of note that LipidMiner currently is limited to singly charged ions, although it is adequate for the purpose of lipidomics, since lipids are rarely multiply charged, even for the polyphosphoinositides.^[16, 17] Currently, LipidMiner only works on Windows platform and processes LC-MS/MS file formats generated from mass spectrometers from Thermo, i.e. the .RAW data format. In the future, we are planning to accommodate file formats generated by mass spectrometers from other predominant instrument vendors to make this tool more universal.

LipidMiner can be freely downloaded from <http://sourceforge.net/projects/lipidminer/>.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors thank Dr. Thomas Metz on comments of this manuscript. This work was supported by the Office of Science, U.S. Department of Energy, under the Low Dose Radiation Research Program, and by the National Institutes of Health (R21 GM104678). Work by G.L. and D.M. was supported by the U.S. Department of Energy (DOE) Office of Science's Advanced Scientific Computing Research Applied Mathematics program. Portions of this work were performed at the Environmental Molecular Sciences Laboratory, a national scientific user facility sponsored by the Department of Energy's Office of Biological and Environmental Research and located at Pacific Northwest National Laboratory (PNNL) in Richland, Washington. PNNL is a multi-program national laboratory operated by Battelle for the DOE under Contract DE-AC05-76RLO 1830.

References

1. Fahy E, Subramaniam S, Brown HA, Glass CK, Merrill AH Jr, Murphy RC, Raetz CR, Russell DW, Seyama Y, Shaw W, Shimizu T, Spener F, van Meer G, VanNieuwenhze MS, White SH, Witztum JL, Dennis EA. *J Lipid Res.* 2005; 46:839. [PubMed: 15722563]
2. Fahy E, Subramaniam S, Murphy RC, Nishijima M, Raetz CR, Shimizu T, Spener F, van Meer G, Wakelam MJ, Dennis EA. *J Lipid Res.* 2009; 50(Suppl):S9. [PubMed: 19098281]
3. Han X, Gross RW. *Mass Spectrom Rev.* 2005; 24:367. [PubMed: 15389848]
4. Haimi P, Uphoff A, Hermansson M, Somerharju P. *Anal Chem.* 2006; 78:8324. [PubMed: 17165823]
5. Ejsing CS, Duchoslav E, Sampaio J, Simons K, Bonner R, Thiele C, Ekroos K, Shevchenko A. *Anal Chem.* 2006; 78:6202. [PubMed: 16944903]
6. Yang K, Cheng H, Gross RW, Han X. *Anal Chem.* 2009; 81:4356. [PubMed: 19408941]
7. Sabareesh V, Singh G. *J Mass Spectrom.* 2013; 48:465. [PubMed: 23584940]
8. Song H, Hsu FF, Ladenson J, Turk J. *J Am Soc Mass Spectrom.* 2007; 18:1848. [PubMed: 17720531]
9. Schwudke D, Oegema J, Burton L, Entchev E, Hannich JT, Ejsing CS, Kurzchalia T, Shevchenko A. *Anal Chem.* 2006; 78:585. [PubMed: 16408944]
10. Herzog R, Schwudke D, Schuhmann K, Sampaio JL, Bornstein SR, Schroeder M, Shevchenko A. *Genome Biol.* 2011; 12:R8. [PubMed: 21247462]
11. Kind T, Liu KH, Lee do Y, Defelice B, Meissen JK, Fiehn O. *Nat Methods.* 2013; 10:755. [PubMed: 23817071]
12. Gao X, Zhang Q, Meng D, Isaac G, Zhao R, Fillmore TL, Chu RK, Zhou J, Tang K, Hu Z, Moore RJ, Smith RD, Katze MG, Metz TO. *Anal Bioanal Chem.* 2012; 402:2923. [PubMed: 22354571]
13. Nie H, Liu R, Yang Y, Bai Y, Guan Y, Qian D, Wang T, Liu H. *J Lipid Res.* 2010; 51:2833. [PubMed: 20526000]

14. Hubner G, Crone C, Lindner B. J Mass Spectrom. 2009; 44:1676. [PubMed: 19816875]
15. Aurich, V.; Weule, J. Mustererkennung; DAGM-Symposium; 1995; Springer-Verlag; 1995. p. 538
16. Milne SB, Ivanova PT, DeCamp D, Hsueh RC, Brown HA. J Lipid Res. 2005; 46:1796. [PubMed: 15897608]
17. Wenk MR, Lucast L, Di Paolo G, Romanelli AJ, Suchy SF, Nussbaum RL, Cline GW, Shulman GI, McMurray W, De Camilli P. Nat Biotechnol. 2003; 21:813. [PubMed: 12808461]

Figure 1.

Representative screen shots from LipidMiner illustrating the main menu and operations within each module. For each module (Ion Detection, Feature Alignment and Library Match), users can select input raw dataset files, list of signature product ions and neutral losses specific to each lipid class (shown as Query File), and module specific parameters.

Figure 2.

A schematic of the workflow used in LipidMiner for identification and quantification of lipid molecular species from multiple LC-MS/MS raw data files. Multiple RAW datasets can be selected for sequential processing in the Ion Detection module. Once Ion Detection is completed for all raw datasets, the lipid features can be aligned simultaneously in the Feature Alignment step, followed by Library Match. Details of Feature Identification can be found in supplementary material.

Figure 3.

Calibration curves and correlation plot of lipid standard PC(14:0/0:0). Data files were acquired with data dependent LC-MS/MS from lipid mixture containing this standard at 4 different concentrations of 0.01, 0.05, 0.1 and 0.5 nM. Data were processed by A) automated compound identification and peak area integration using LipidMiner through emulated PIS of m/z 184.07 and matching with LIPID MAPs compound library; B) manual peak area integration using XCalibur through extracted ion chromatogram of m/z 486.31 ($[M + H]^+$) with a mass range tolerance window of 0.1 Da; and C) correlation of peak areas obtained from LipidMiner and XCalibur. ▲, ■, and ◆ in A) and B) represent three replicated runs of the same sample.