LipidLynxX: lipid annotations converter for large scale lipidomics and epilipidomics datasets

Zhixu Ni [1,2] and Maria Fedorova[1,2]*

[1]Institute of Bioanalytical Chemistry, Faculty of Chemistry and Mineralogy, University of Leipzig, Germany; [2]Center for Biotechnology and Biomedicine, University of Leipzig.

Corresponding author:

* Dr. Maria Fedorova, Institut für Bioanalytische Chemie, Faculty of Chemistry and Mineralogy, Biotechnologisch-Biomedizinisches Zentrum, Leipzig University, Deutscher Platz 5, 04103 Leipzig, Germany. E-mail: maria.fedorova@bbz.uni-leipzig.de.

Abstract

Summary

Modern high throughput lipidomics provides large-scale datasets reporting hundreds of lipid molecular species. However, cross-laboratory comparison and systems biology integration of published datasets remain challenging due to the high diversity of used lipid annotation systems. To support lipidomics data integration and usability of publicly available lipid identification datasets, we developed LipidLynxX, a software to convert diverse lipid annotations to unified identifiers and cross-ID matching.

Availability and implementation

LipidLynxX is available for download at https://github.com/SysMedOs/LipidLynxX.

Introduction

Lipids are now recognized as important biomolecules with a variety of biological functions far beyond simple energy storage units (Wenk, 2005; Shevchenko and Simons, 2010). Our current understanding of lipid metabolism and signalling is, at least in part, determined by the significant developments in mass spectrometry (MS) based lipidomics capable to provide inventory of natural lipidomes for a variety of tissues, cells and conditions at qualitative and even quantitative levels (Rustam and Reid, 2018). Availability of large-scale datasets reporting hundreds of lipid species allows to perform systems biology integration of

lipidomics data including network reconstructions, pathway mapping and enrichment (Subramaniam *et al.*, 2011). However, despite several suggestions for unified shorthand notations of lipid identifiers in lipidomics studies (Liebisch *et al.*, 2013; Pauling *et al.*, 2017), reported lipid annotations remain very diverse and usually reflect the output style of the software tools used for lipid identification and/or authors personal preferences. Furthermore, in contrast to genomics, transcriptomics, and proteomics, lipid annotations obtained by MS analysis can reflect different levels of identification confidence (e.g. PC(38:4) vs PC(18:0_20:4) vs PC(18:0/20:4)). The diversity of lipid identifiers in publicly available datasets makes it difficult to compare, compile and integrate lipidomics data derived from different sources. Here we introduce LipidLynxX, a converter of lipid annotations capable to crossmatch and provide unified IDs for lipid species reported using different shorthand notation systems.

Materials and Methods

The LipidLynxX is developed using Python 3.7. The software provides both API access for advanced users (RESTful) and simple web-based GUI for clients without programming skills. A LipidLynxX data exchange file structure is developed using JSON file format together with corresponding schema for the strict validation of each data field. The current LipidLynxX version can be used to set up a local web service. LipidLynxX is compatible with Linux, macOS and Windows platforms and able to provide both API access and user-friendly GUI. LipidLynxX is freely available for academic use under the GNU General Public License version 2 (https://www.gnu.org/licenses/old-licenses/gpl-2.0.en.html) from the following GitHub repository: https://github.com/SysMedOs/LipidLynxX. A pre-compiled Windows installation file is provided for Windows platform users together with corresponding source code for all supported platforms.

Results

LipidLynxX includes two functional modules – Converter and Equalizer for main mammalian lipid classes. LipidLynxX supports translation of annotations for both unmodified and modified lipids (e.g. oxidized lipids) from lipidomics and epilipidomics datasets (Figure 1). LipidLynxX is an open-source software equipped with an easy to use graphical user interface allowing to use the tool as a desktop converter, via API access for

professional users, and can be easily extended to website-based service. Applied license allows to integrate the tool into other applications or use it for lipid libraries generation.

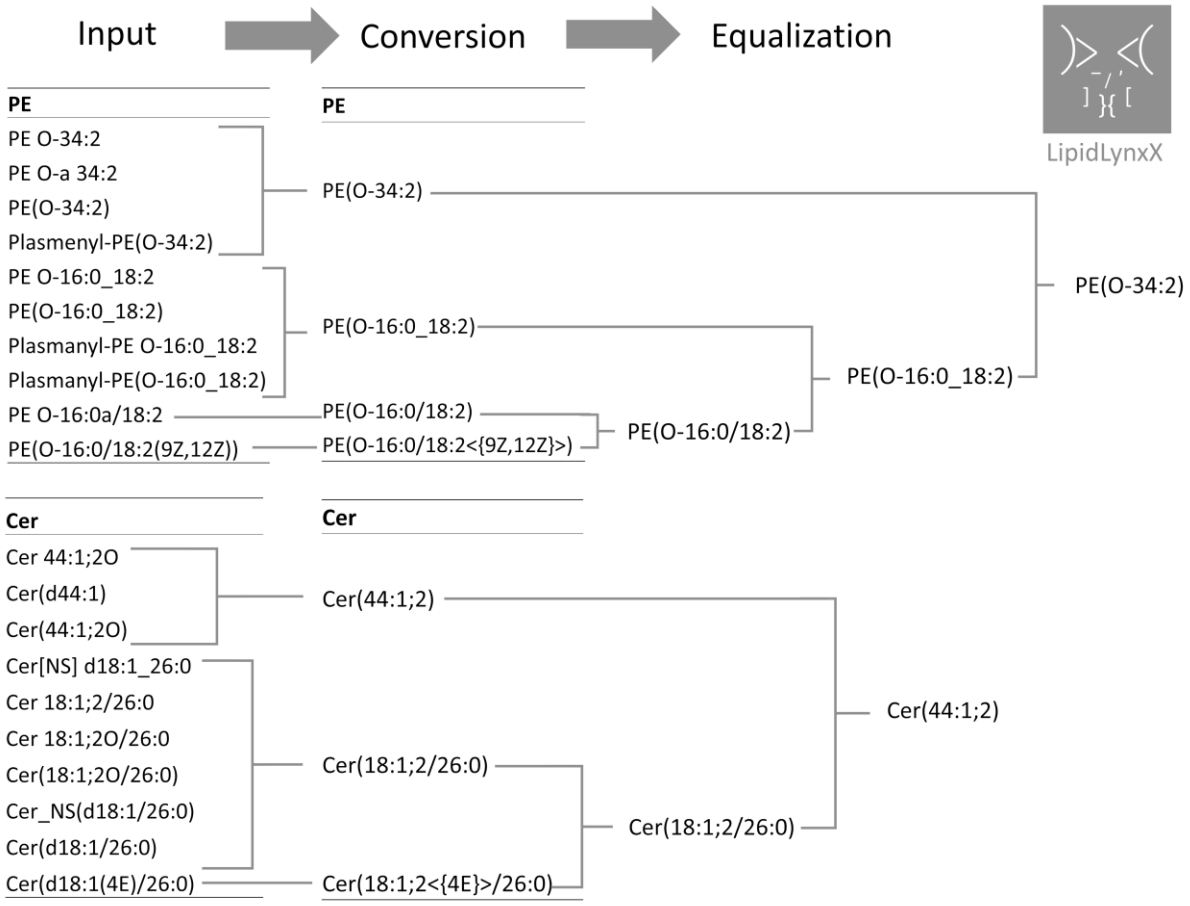

Figure 1. Schematic representation of LipidLynxX functionalities on the example of conversion and cross-level matching of diverse lipid annotations for ether-phosphatidylethanolamine PE(O-34:2) and ceramide Cer(44:1;2) lipids into unified ID.

1. **Conversion of lipid annotations.** LipidLynxX Converter tool converts different lipid annotations to a unified identifier based on the community accepted shorthand notation system introduced by Liebisch et al (Liebisch *et al.*, 2013). LipidLynxX identifiers utilize hierarchical brackets system (Figure S1) to support computational processing of large lipidomics datasets. Currently LipidLynxX Converter can translate lipid abbreviations from 13 most popular lipid identification software (e.g. ALEX, LDA2, LipidSearch, LipidHunter, Lipostar, LPPtigger, MS_DIAL etc (Pauling *et al.*, 2017; Hartler *et al.*, 2017; Ni, Angelidou, Lange, *et al.*, 2017; Goracci *et al.*, 2017; Ni, Angelidou, Hoffmann, *et al.*, 2017; Tsugawa *et al.*, 2020)) and lipid databases (LIPID MAPS, HMDB (Fahy *et al.*, 2009; Wishart *et al.*, 2017)) via extendable predefined regular expressions provided as configuration files in JSON

format. LipidLynxX framework is designed to support further extensions and customize input/output identifies using provided configuration file templates.

LipidLynxX support the conversion of 25 lipid classes from five major lipid categories defined by LIPID MAPS including fatty acids (FA01-FA04), glycerophospholipids (GP01-GP12), glycerolipids (GL01-GL03), sphingolipids (SP01-SP03, SP05, SP06), sterols and sterol esters (ST01). It also supports annotation of isotopically labelled lipids both at the acyl chains, glycerol and head group moieties. LipidLynxX ID at the full structural annotation level (e.g. PE(16:0/20:4 <{5Z,9E,12E,15E}>)) is defined via strictly controlled JSON format and thus can be used by other tools for lipid structure visualization, elemental composition calculation, and libraries generation. Each LipidLynxX identifier can be exported using JSON file format.

2. **Cross-level matching.** Based on (LC)MS method used for the identification, lipids can be reported at different levels of structural annotation – e.g. PE(38:5) vs PE(18:1_20:4) vs PE(18:1<{9}>/20:4<{5,9,12,15}>). Comparison and integration of datasets reporting lipids at different structural levels is still possible but would require matching of all reported lipid IDs to the least specific annotation level reported (e.g. PE(38:5)). LipidLynxX Equalizer tool allows to bring lipid identifiers to the same level of annotation and perform cross-level matching between different datasets.

3. **Conversion and matching for modified lipids IDs.** Similar to other levels of biological organization (genome, transcriptome and proteome), the lipidome is also subjected to different enzymatic and non-enzymatic modifications. Modifications of lipids including oxidation, nitration, sulfation and halogenation, constitute a new level of complexity of lipidomes (epilipidome) necessary for the regulation of complex biological functions (Ni *et al.*, 2019). Increasing number of datasets reporting modified (mostly oxidized) lipids are published nowadays (Kagan *et al.*, 2017; Gil de la Fuente *et al.*, 2018; Godzien *et al.*, 2019; Striesow *et al.*, 2020). However, unified identifiers for modified lipids were not available till now. LipidLynxX introduce a new identifiers system for modified lipids to support datasets compatibility in epilipidomics studies. New system is based on ordered controlled vocabularies and support five levels of structural annotations, including modification mass shift (1), elemental composition (2), type (3), position (4), and stereochemistry (5) for bulk (B), discrete (D), and *sn*-specific (S) levels reflecting a variety of methodological approaches used to identify epilipids (Figure S2). All currently supported modifications were converted into a list of controlled vocabularies and organized in strict order (Figure S3). Commonly used abbreviation and alias for oxidized fatty acids (e.g. HETE) can also be converted.

Conclusion

LipidLynxX is an open-source easy to use tool for conversion and cross-level matching of lipid identifies to support comparison and integration of large-scale lipidomics datasets.

Acknowledgments

Financial support from the German Federal Ministry of Education and Research (BMBF) within the framework of the e:Med research and funding concept for SysMedOS project (to MF) are gratefully acknowledged. We thank Prof. Ralf Hoffmann (Institute of Bioanalytical Chemistry, University of Leipzig) for providing access to his laboratory.

Conflict of interests

The authors declare no conflict of interests.

References

Fahy,E. *et al.* (2009) Update of the LIPID MAPS comprehensive classification system for lipids. *J. Lipid Res.*, **50**, S9–S14.

Gil de la Fuente,A. *et al.* (2018) Characterization and annotation of oxidized glycerophosphocholines for non-targeted metabolomics with LC-QTOF-MS data. *Anal. Chim. Acta*, **1037**, 358–368.

Godzien,J. *et al.* (2019) Oxidized glycerophosphatidylcholines in diabetes through non-targeted metabolomics: Their annotation and biological meaning. *J. Chromatogr. B*, **1120**, 62–70.

Goracci,L. *et al.* (2017) Lipostar, a Comprehensive Platform-Neutral Cheminformatics Tool for Lipidomics. *Anal. Chem.*, **89**, 6257–6264.

Hartler,J. *et al.* (2017) Deciphering lipid structures based on platform-independent decision rules. *Nat. Methods*, **14**, 1171–1174.

Kagan,V.E. *et al.* (2017) Oxidized arachidonic and adrenic PEs navigate cells to ferroptosis. *Nat. Chem. Biol.*, **13**, 81–90.

Liebisch,G. *et al.* (2013) Shorthand notation for lipid structures derived from mass spectrometry. *J. Lipid Res.*, **54**, 1523–30.

Ni,Z. *et al.* (2019) Computational solutions in redox lipidomics – Current strategies and future perspectives. *Free Radic. Biol. Med.*, **144**, 110–123.

Ni,Z., Angelidou,G., Lange,M., *et al.* (2017) LipidHunter Identifies Phospholipids by High-

Throughput Processing of LC-MS and Shotgun Lipidomics Datasets. *Anal. Chem.*, **89**, 8800–8807.

Ni,Z., Angelidou,G., Hoffmann,R., *et al.* (2017) LPPtiger software for lipidome-specific prediction and identification of oxidized phospholipids from LC-MS datasets. *Sci. Rep.*, **7**.

Pauling,J.K. *et al.* (2017) Proposal for a common nomenclature for fragment ions in mass spectra of lipids. *PLoS One*, **12**, e0188394.

Rustam,Y.H. and Reid,G.E. (2018) Analytical Challenges and Recent Advances in Mass Spectrometry Based Lipidomics. *Anal. Chem.*, **90**, 374–397.

Shevchenko,A. and Simons,K. (2010) Lipidomics: Coming to grips with lipid diversity. *Nat. Rev. Mol. Cell Biol.*, **11**, 593–598.

Striesow,J. *et al.* (2020) Oxidative modification of skin lipids by cold atmospheric plasma (CAP): A standardizable approach using RP-LC/MS2 and DI-ESI/MS2. *Chem. Phys. Lipids*, **226**, 104786.

Subramaniam,S. *et al.* (2011) Bioinformatics and systems biology of the lipidome. *Chem. Rev.*, **111**, 6452–6490.

Tsugawa,H. *et al.* (2020) MS-DIAL 4: accelerating lipidomics using an MS/MS, CCS, and retention time atlas. *bioRxiv*, 2020.02.11.944900.

Wenk,M.R. (2005) The emerging field of lipidomics. *Nat. Rev. Drug Discov.*, **4**, 594–610.

Wishart,D.S. *et al.* (2017) HMDB 4.0: the human metabolome database for 2018. *Nucleic Acids Res.*, **46**, D608–D617.