

Article

## LipidHunter identifies phospholipids by high-throughput processing of LC-MS and shotgun lipidomics datasets

Zhixu Ni, Georgia Angelidou, Mike Lange, Ralf Hoffmann, and Maria Fedorova

*Anal. Chem.*, **Just Accepted Manuscript** • DOI: 10.1021/acs.analchem.7b01126 • Publication Date (Web): 28 Jul 2017

Downloaded from <http://pubs.acs.org> on July 29, 2017

### Just Accepted

"Just Accepted" manuscripts have been peer-reviewed and accepted for publication. They are posted online prior to technical editing, formatting for publication and author proofing. The American Chemical Society provides "Just Accepted" as a free service to the research community to expedite the dissemination of scientific material as soon as possible after acceptance. "Just Accepted" manuscripts appear in full in PDF format accompanied by an HTML abstract. "Just Accepted" manuscripts have been fully peer reviewed, but should not be considered the official version of record. They are accessible to all readers and citable by the Digital Object Identifier (DOI®). "Just Accepted" is an optional service offered to authors. Therefore, the "Just Accepted" Web site may not include all articles that will be published in the journal. After a manuscript is technically edited and formatted, it will be removed from the "Just Accepted" Web site and published as an ASAP article. Note that technical editing may introduce minor changes to the manuscript text and/or graphics which could affect content, and all legal disclaimers and ethical guidelines that apply to the journal pertain. ACS cannot be held responsible for errors or consequences arising from the use of information contained in these "Just Accepted" manuscripts.



ACS Publications

# LIPIDHUNTER IDENTIFIES PHOSPHOLIPIDS BY HIGH-THROUGHPUT PROCESSING OF LC-MS AND SHOTGUN LIPIDOMICS DATASETS.

Zhixu Ni,<sup>†,‡</sup> Georgia Angelidou,<sup>†,‡</sup> Mike Lange,<sup>†,‡</sup> Ralf Hoffmann,<sup>†,‡</sup> Maria Fedorova\*,<sup>†,‡</sup>

<sup>†</sup>Institute of Bioanalytical Chemistry, Faculty of Chemistry and Mineralogy and <sup>‡</sup>Center for Biotechnology and Bio-medicine, Universität Leipzig, Deutscher Platz 5, 04103 Leipzig, Germany.

**ABSTRACT:** Lipids are dynamic constituents of biological systems, rapidly responding to any changes in physiological conditions. Thus, there is a large interest in lipid-derived markers for diagnostic and prognostic applications, especially in translational and systems medicine research. As lipid identification remains a bottleneck of modern untargeted lipidomics, we developed LipidHunter, a new open source software for the high-throughput identification of phospholipids in data acquired by LC-MS and shotgun experiments. LipidHunter resembles a workflow of manual spectra annotation. Lipid identification is based on MS/MS data analysis in accordance with defined fragmentation rules for each phospholipid (PL) class. The software tool matches product and neutral loss signals obtained by collision-induced dissociation to a user-defined white list of fatty acid residues and PL class-specific fragments. The identified signals are tested against elemental composition and bulk identification provided via LIPID MAPS search. Furthermore, LipidHunter provides information rich tabular and graphical reports allowing to trace back key identification steps and perform data quality control. Thereby, 202 discrete lipid species were identified in lipid extracts from rat primary cardiomyocytes treated with a peroxynitrite donor. Their relative quantification allowed the monitoring of dynamic reconfiguration of the cellular lipidome in response to mild nitroxidative stress. LipidHunter is available free for download at <https://bitbucket.org/SysMedOs/lipidhunter>.

Changes in lipid profiles are an important part of cellular response to various pathological conditions. In the past decades, lipid species were typically identified by untargeted profiling using liquid chromatography coupled online to mass spectrometry (LC-MS). However, the untargeted analysis of biological samples are challenged by extremely complex sample matrices, large variation of lipid concentrations, and sophisticated data analysis workflows required to obtain reliable identification and precise quantification results. Extraction from biological samples, chromatographic separation, and analysis by mass spectrometry have been significantly improved<sup>1,2</sup>, whereas bioinformatics tools for lipid identification remain the bottleneck of high-throughput lipidomics studies in translational and systems medicine research<sup>3,4</sup>.

The high structural diversity of known and acknowledged existence of yet unidentified lipids significantly complicates identification in high-throughput studies. Extensive research into the characterisation of main lipid classes (glycerophospholipids, PLs) in MS (preferential ionisation mode and adducts) and MS/MS (fragmentation mechanisms

and relative intensities of characteristic fragment ions) provide certain rules<sup>5-7</sup> that allow lipid identification using computational algorithms to accelerate.

Mainstream software tools identifying PL rely either on spectral matching using reference libraries or on user definable fragmentation patterns. Spectral matching is probably the most popular approach in MS-based lipidomics using structures (LIPID MAPS, HMDB)<sup>8,9</sup> or tandem mass spectra (METLIN)<sup>10</sup> libraries relying on experimental data, or combining them with *in silico* predicted entries. Current resources differ in the number of lipid entries available. LIPID MAPS, the most popular lipidomics database, contained 40,673 distinct lipids including 9,620 PLs based on the experimentally reported and computationally generated species (January 2017). LipidBlast contains 120,000 distinct compounds (including over 33,000 of PLs, usually 5,476 species per class derived by permutation of 74 discrete fatty acid residues) associated with more than 200,000 *in silico* generated tandem mass spectra<sup>11</sup>. Predicted structures allow the identification of previously undetected species, but require

strict control of the results. Furthermore, spectral matching depends greatly on the instrument type and acquisition<sup>11,12</sup>.

Tools relying on the user defined fragmentation patterns such as LipidXplorer<sup>13</sup>, can use highly customised scripts to define fragment ions from a limited range of precursors. This strategy can be adapted to define fragment ions for a wide range of lipid classes. However, the current version of LipidXplorer was designed for shotgun lipidomics and requires additional tunes to be fully functional for LC-MS.

Here, we report a new open source software package, LipidHunter, to assist with the high-throughput analysis of PLs from LC-MS and shotgun DDA experiments resembling procedures of manual spectra annotation. Lipid coverage depends mainly on a user defined white list of fatty acyl and alkyl residues, which can be easily extended or reduced based on the current needs. Furthermore, flexible configuration files allow LipidHunter to be adopted to different instruments or lipid ion types. Importantly, LipidHunter output can be easily evaluated to assess the general quality of data and identification results.

## Experimental Section

**Chemicals** - Phospholipid internal standards were obtained from Avanti Polar Lipids (AL, USA): 1,2-diheptadecanoyl-*sn*-glycero-3-phosphate [PA(17:0/17:0)], 1,2-diheptadecanoyl-*sn*-glycero-3-phosphocholine [PC(17:0/17:0)], 1,2-diheptadecanoyl-*sn*-glycero-3-phosphoethanolamine [PE(17:0/17:0)], 1,2-diheptadecanoyl-*sn*-glycero-3-phospho-(1'-*rac*-glycerol) [PG(17:0/17:0)], and 1,2-diheptadecanoyl-*sn*-glycero-3-phospho-L-serine [PS(17:0/17:0)]. Acetonitrile, methanol, isopropanol and formic acid were purchased from Biosolve (Valkenswaard, Netherlands). Dulbecco's Modified Eagle Medium/Ham's F-12 (DMEM/F12), phosphate buffered saline (PBS), foetal bovine serum (FBS), penicillin-streptomycin, L-glutamine, non-essential amino acids, sodium pyruvate, and gelatine were obtained from Life Technologies GmbH (Darmstadt, Germany). Horse serum, trypsin-EDTA solution, butylated hydroxytoluene (BHT), and Human Serum from human male AB plasma (USA origin, sterile-filtered) were purchased from Sigma-Aldrich GmbH (Taufkirchen, Germany). 3-Morpholininosydnonimine (SIN-1) was purchased from Enzo Life Sciences GmbH (Lörrach, Germany).

**Cell culture** - Primary rat cardiomyocytes (Innoprot, Elealde Derio, Spain) were cultured in gelatine-coated 6-well-plates (CELLSTAR®, Greiner Bio-One GmbH, Frickhausen, Germany) in DMEM/F12 medium supplemented with 20% FBS, 5% horse serum, 2 mmol/L l-glutamine, 3 mmol/L sodium pyruvate, 0.1 mmol/L non-essential amino acids, 100 U/mL penicillin and 100 µg/mL streptomycin at 37°C (humidified atmosphere of 5% CO<sub>2</sub> and 95% air). When cells reached 80% confluence, medium was replaced by serum-free medium (DMEM/F12 supplemented with 100 U/mL penicillin and 100 µg/mL streptomycin), and cells were treated with 10 µmol/L SIN-1. After 15 min, 30 min, 70 min, or 16 h, plates were placed on ice, and cells were washed with cold PBS containing BHT (0.01%, w/v) and scraped into 1 mL of cold methanol containing acetic acid (3%, v/v) and BHT (0.01% w/v)<sup>14</sup>. Samples were spiked with internal standards [PC(17:0/17:0), PE(17:0/17:0), PS(17:0/17:0), PG(17:0/17:0),

PA(17:0/17:0); 100 ng per sample), dried, and resuspended in 50 µL of mixture of water, isopropanol, acetonitrile and methanol (5:3:1:1, by volume).

**Lipid extraction from human serum** - Lipids from 5 µL of human male AB serum were extracted using MTBE protocol<sup>15</sup>. Dried extracts were resuspended in 50 µL of mixture of water, isopropanol, acetonitrile and methanol (5:3:1:1, by volume).

**RPLC-MS** - Acquity UPLC M-class (Waters GmbH, Eschborn, Germany) was coupled online to a Synapt G2-Si mass spectrometer equipped with an ESI source (Waters GmbH, Eschborn, Germany) operating in negative ion mode. Eluent A was a mixture of water and acetonitrile (90:10, v/v) containing formic acid (0.1% v/v), and eluent B was a mixture of isopropanol, acetonitrile, and methanol (60:20:20, v/v/v) containing formic acid (0.1% v/v). Lipids (1 µL in 50 % B; each sample in triplicate) were loaded onto a C18-column Acquity UPLC® CSH™ C18, (internal diameter 1.0 mm, length 100 mm, particle diameter 1.7 µm) and eluted with linear gradients from 50 to 90% eluent B (30 min) and to 99% B (1 min) and held for 10 min. Column temperature was set to 50°C and the flow rate to 60 µL/min.

Sampling cone voltage was set to 40 V, source offset to 60 V, source temperature to 120°C, cone gas flow to 30 L/h, desolvation gas flow to 650 L/h, desolvation temperature to 250°C, nebuliser gas pressure of 6 bar and an ion spray voltage of -2.0 kV. Data were acquired in negative ion data-dependent (DDA) and independent (MS<sup>E</sup>) resolution modes. In the DDA mode, precursor ion survey scans (scan time 0.5 sec) were acquired for *m/z* 200 to 1200. Tandem mass spectra (ramp collision energy: LM CE start/end 10-40 and HM CE start/end 20-60) were recorded (scan time 0.25 sec) for the 12 most abundant ions in each survey scan using a dynamic exclusion for 30 sec. For MS<sup>E</sup> low (3 V) and high collision energy scans (ramp from 20 to 50 V, scan time 0.6 sec) were acquired for *m/z* 50 to 1200. The signal of Leu-enkephalin (554.26151) was acquired as lock mass.

**Software assisted data analysis** - DDA datasets were converted into mzML format using the MSConvert module from ProteoWizard project (version 3.0.9134 64bit)<sup>16</sup> and processed by an in-house developed software LipidHunter to assist high-throughput PL identification. LipidHunter was developed using Python programming language (version 2.7.3) with pymzML package<sup>17</sup> to process mzML files and other packages, such as Numpy, Pandas, Scipy, and Matplotlib for data processing and visualisation. LipidHunter provides a cross platform Graphic User Interface (GUI) that can be executed on Windows, Linux and macOS platforms, the project repository is hosted on Bitbucket (<https://bitbucket.org/SysMedOs/lipidhunter>).

Bulk PL were identified using LipidMAPS online tools ([http://www.lipidmaps.org/tools/ms/glycerophospholipids\\_batch\\_bulk.html](http://www.lipidmaps.org/tools/ms/glycerophospholipids_batch_bulk.html))<sup>8</sup>.

**Lipid abbreviations** - Lipids were abbreviated based on the nomenclature proposed by LIPID MAPS consortium<sup>18</sup>. Thus PC(16:0\_18:2) indicates a discrete FA composition without providing positional information.

**Relative label free quantification** - Datasets acquired in MS<sup>E</sup> mode were imported into Progenesis QI (version 2.1.0,

Nonlinear, Newcastle, UK). PLs identified by LipidHunter from DDA data were matched to the corresponding features ( $m/z$  values and retention time) in aligned MS<sup>E</sup> datasets. Spiked internal standards were used for normalization.

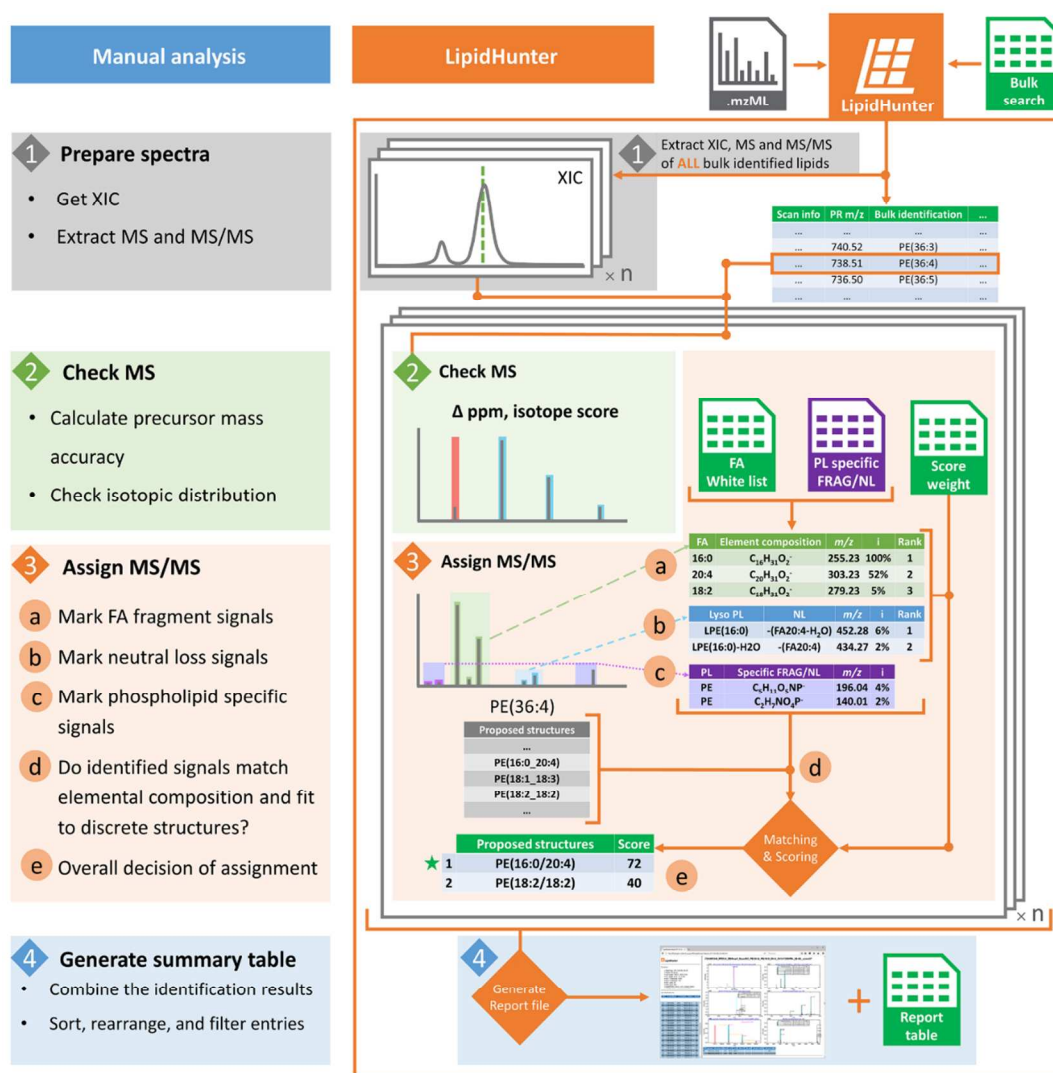
PLs detected with significantly different intensities (ANOVA  $p$  value  $\leq 0.05$ ) were exported and further analysed by EZinfo (version 1.0, MKS Instruments, Crewe, UK) and Genesis (version 1.7.7)<sup>19</sup>.

**LipidBlast PL identification**<sup>11</sup> – Three replicates of SIN-1\_70min datasets were converted to .mgf using ProteoWizard MSconvert and searched against LipidBlast using negative mode spectra library and NIST MS PepSearch tool ([http://chemdata.nist.gov/dokuwiki/doku.php?id=peptide:ms\\_pepsearch](http://chemdata.nist.gov/dokuwiki/doku.php?id=peptide:ms_pepsearch)). Identified entries were manually controlled using NIST MS search and MassLynx for the identification accuracy, XIC, and the isotope distribution pattern of the precursor  $m/z$ .

## Results and Discussion

**LipidHunter** – LipidHunter was designed as a tool for high-throughput lipid identification resembling manual spectra annotation processes and providing full control and transparency of each identification and verification step (Figure 1). This tool processes mzML files obtained from LC-MS/MS and shotgun analysis in DDA mode.

LipidHunter workflow is subdivided into five main steps. The first two steps (i.e., Extract MS level and Extract Scan info), extract MS and MS/MS information from mzML files. MS information is used for bulk lipid identification in LIPID MAPS database (step III, Bulk Search). Bulk matches are linked to the corresponding MS and MS/MS scans (step IV, Link lipids to scans), for assigning discrete lipids based on the product ions (step V, Hunt for ID). Results are exported as.xlsx tables and HTML files containing six-panel graphical images allowing the evaluation and verification of each step.



**Figure 1.** Schematic representation of the LipidHunter algorithm comparing manual and software assisted lipid identification. Workflow illustrates the main steps of conventional data analysis including extraction of corresponding MS, MS/MS, and XIC (1), evaluation of

precursor mass accuracy, and isotope distribution of MS scan (2), assignment of MS/MS using lipid specific fragment signals (3), and the generation of a report (4).

The built-in cross-platform GUI (available for Windows, Linux and macOS), guides users through each step and allows all parameters to be changed directly without specific knowledge of command line operations (Figure S-1). Details on the workflow design and implemented algorithms are provided below:

**< I > Extract MS Level:** LipidHunter extracts MS survey scan information ( $m/z$  values and intensities) from a single or multiple .mzML files. In case of multiple datasets, extracted information is merged to a single file and duplicated  $m/z$  values are removed. Retention time window,  $m/z$  range and intensity threshold can be applied to reduce processing time. A combined peak list will be used for bulk lipid identification in the LIPID MAPS database in step III.

**< II > Extract Scan info:** All precursors selected for fragmentation are linked to the corresponding DDA rank, MS/MS scan number and scan time.

**< III > Bulk Search:** The combined peak list generated in step I is submitted to LIPID MAPS online tool for bulk identification of lipids ([http://www.lipidmaps.org/tools/ms/glycerophospholipids\\_batch\\_bulk.html](http://www.lipidmaps.org/tools/ms/glycerophospholipids_batch_bulk.html)). The results should be saved as .xlsx files. In principle, any other lipid database can be used as long as the resulting .xlsx files correspond to the template.

**< IV > Link lipids to Scans:** Precursors identified by bulk search are assigned to the corresponding MS/MS scans and associated information from step II, and mzML file. All precursors unmatched to the bulk search results are discarded. Furthermore, an optional fast head group specific filter for PC, PE, and PS species can be applied<sup>6,7,20</sup>.

**< V > Hunt for ID:** The core function of LipidHunter uses the linked bulk identifications and scan information (step IV) to assign the MS/MS using an automatic eight steps workflow applied to each bulk identified feature:

(i) Using the information from step IV, peak lists (fragment ions  $m/z$  values with corresponding intensities) are extracted for all MS/MS scans matching precursors identified by bulk search.

(ii) Each MS/MS scan is linked to the corresponding MS scan of the same DDA duty cycle. Precursor  $m/z$  value and elemental composition obtained by bulk identification are used to calculate isotopic distribution score and mass accuracy.

(iii) A precursor-specific elution profile (XIC) is generated from mzML file using pymzML module<sup>17</sup>.

(iv) Three-stage verification of precursor isotopic distribution for each extracted MS scan considering elemental composition obtained by bulk identification is performed for each identification. First, the monoisotopic signal of a precursor ion is treated as a pseudo [M+1] ion, and the  $m/z$  of the corresponding pseudo [M+0] signal is calculated. If intensities of all signals within a user defined mass accuracy range for this signal are lower than the intensity of precursor ion, the precursor signal is assigned as monoisotopic and considered for consecutive identification steps.

Second, theoretical isotope distribution of the proposed elemental composition is calculated for [M+0] to [M+2] by

means of binomial and McLaurin expansion as described before<sup>12,21</sup>. For instance, for [PE(36:4)-H] with an elemental composition of C<sub>41</sub>H<sub>73</sub>NO<sub>8</sub>P is as follows (phosphorus as monoisotopic element is not considered in the equation).

$$\begin{aligned} & ({}^{12}\text{C} + {}^{13}\text{C})^{41} ({}^1\text{H} + {}^2\text{H})^{73} ({}^{14}\text{N} + {}^{15}\text{N}) ({}^{16}\text{O} \\ & \quad + {}^{17}\text{O} + {}^{18}\text{O})^8 \\ & = ({}^{12}\text{C}_{41} {}^1\text{H}_{73} {}^{14}\text{N}_{16} {}^{16}\text{O}_8) \left( 1 \right. \\ & \quad + \frac{{}^{13}\text{C}}{{}^{12}\text{C}} \Big)^{41} \left( 1 + \frac{{}^2\text{H}}{{}^1\text{H}} \right)^{73} \left( 1 \right. \\ & \quad + \frac{{}^{15}\text{N}}{{}^{14}\text{N}} \Big) \left( 1 + \frac{{}^{17}\text{O}}{{}^{16}\text{O}} + \frac{{}^{18}\text{O}}{{}^{16}\text{O}} \right)^8 \end{aligned}$$

However, for most PL species, <sup>13</sup>C contributes dominantly to the isotope pattern of [M+0] to [M+2] signals. Thus, a “<sup>13</sup>C only” mode was also implemented with the binomial expansion of <sup>13</sup>C to accelerate the processing speed. In “<sup>13</sup>C only” mode, the equation for PE(36:4), for example, is simplified as follows:

$$({}^{12}\text{C} + {}^{13}\text{C})^{41} = {}^{12}\text{C}_{41} \left( 1 + \frac{{}^{13}\text{C}}{{}^{12}\text{C}} \right)^{41}$$

The “<sup>13</sup>C only” mode reduced the time to calculate the isotopic distribution by 2/3 and result in an overall runtime reduction of 30% in comparison to the full elements mode. The preferred mode can be selected via GUI.

Finally, an isotope score is generated by calculating the similarity between experimental and theoretical isotope distributions using an adapted similarity equation<sup>12</sup>.  $I_M$  and  $I_{M+i}$  refer to the signal intensities of the [M+0] and [M+i] peaks, respectively.

### Isotope Score

$$\begin{aligned} & = 100 \\ & \times \left( 1 - \sum |r_{obs,i} - r_{lib,i}| \right) \text{ with } r_i \\ & = \frac{I_{M+i}}{I_M}, 1 \leq i \leq 2 \end{aligned}$$

Only precursors with the isotope scores above the user-defined threshold (e.g. > 90) will be transferred to the next step. The possible overlap of isotope distributions of PLs with FAs differing by one double bond [e.g. PC(34:1) and PC(34:2)], was addressed by an additional deconvolution algorithm adapted from Meija et al<sup>22</sup>.

(v) Discrete lipid species are identified by comparing each signal of a MS/MS peak list with a predefined list of possible FA and PL head group specific fragment/neutral loss signals (Tables S-1 and S-2)<sup>6,7,11</sup>. The 10 most intense matched signals of deprotonated FA or FA neutral loss ions are ranked

according to their intensities (RankIndex parameter 1 to 10). Rank factor  $R_{frag}$  is calculated for each matched signal.

LipidHunter allows to calculate  $R_{frag}$  in a “relative intensity mode”, where  $R_{frag}$  directly corresponds to the relative intensity of the signal, and a “signal rank mode” using following equation:

$$R_{frag} = \frac{(10 - (RankIndex - 1))}{10} \times 100\%$$

The “Signal rank mode” is more suitable to identify as many lipids as possible in complex samples with partially co-eluting (and thus co-fragmenting) structural isomers. The “relative intensity mode” will provide better identification for the most intense isomer.

Signals specific to other PL classes identified in MS/MS will be marked and included in the output table.

(vi) For each type of fragment ion, a weight factor ( $W_{frag}$ ) reflecting its significance for overall structural assignment is provided (Table S-3).  $W_{frag}$  for each product ion type was assigned based on the literature review<sup>11,12</sup> and manual data interpretation considering the type of the mass spectrometer used for data acquisition. The configuration file defining  $W_{frag}$  can be easily customised for other instruments, ionisation modes and lipid ion adducts.

(vii) Independent from the previous step, LipidHunter use the whitelist of possible FA residues (Table S-1) to propose the combination, which will fit a bulk identification. For instance, PE (36:4) can be assigned to PE(16:0\_20:4), PE(16:1\_20:3), PE(18:1\_18:3) or PE(18:2\_18:2).

(viii) If step vi results are matched to the structures proposed in step vii, LipidHunter score is calculated considering  $W_{frag}$  and  $R_{frag}$  of each fragment ion:

$$LipidHunter\ Score = \sum W_{frag} \times R_{frag}$$

LipidHunter provides each dataset an .xlsx table, which summarizes lipid identities (bulk identification, proposed discrete structure, elemental composition, theoretical and observed  $m/z$  values, mass accuracy, and retention time), identification metrics (LipidHunter and isotope scores, relative intensities of matched fragments, PL specific and unspecific signals), and data specific details (DDA rank, scan number). Furthermore, a six-panel image for each identified lipid is generated. Graphical representation of identified species allows fast data quality control and includes precursor specific XICs (Figure 2A), corresponding MS scan (Figure 2B), zoomed MS scan to illustrate the precursor isotope pattern (Figure 2C), MS/MS scan used for identification (Figure 2D), and zoomed region of fatty acid fragments (Figure 2E) and neutral loss signals (Figure 2F). All images are integrated and indexed in an informative HTML report file (Figure S-2) for manual reviewing. Availability of a graphical data representation and organized report files allows the fast evaluation of identification results. The report HTML file with its built-in identification table and a log file of the corresponding LipidHunter parameters, allow a simple solution for data tracking and storage.

A detailed step-by-step LipidHunter manual is provided as Supplementary Information (Supplementary File 1) and available at <https://bitbucket.org/SysMedOs/lipidhunter/>.

#### Identification of PLs in SIN-1 treated cardiomyocytes -

LipidHunter was evaluated by analysing PL from rat primary cardiomyocytes treated with peroxynitrite donor SIN-1 for different periods of time (control, 15, 30, 70 min, and 16 h). Low levels of oxidative stress accompany numerous human disorders and were correlated with oxidative modifications of cellular biomolecules<sup>23-25</sup>. An *in vitro* model of mild nitroxidative stress was recently developed by our group to characterise changes in oxidized lipids (nitrated fatty acids, low molecular weight carbonyls) and proteins (protein carbonylation and modifications by lipid peroxidation products)<sup>26,27</sup>. Here, the same cellular model was used to identify the diversity and detect possible reconfiguration of PL lipidome upon mild nitroxidative stress. Lipid extracts were analysed by RPC-ESI-MS/MS on a Q-TOF instrument in negative ion mode using DDA and DIA methods. The top 12 DDA and MS<sup>E</sup> methods were compared in terms of MS and MS/MS data quality for identification and quantification purposes (Figure S-3). DDA and MS<sup>E</sup> data were selected for identification and quantification, respectively.

Fifteen DDA files (five experimental time points in three biological replicates) were processed by LipidHunter. One example of the summary tables is provided in Supplementary Information (Table S-4).

LipidHunter reliably identified 202 discrete lipid species from six PL classes (Table 1 and Table S-5), which were mostly (84%) identified for all five time points. Most identified lipids were PC (91), followed by PE (54), PS (25), PG (14), PI (12), and PA (6), generally reflecting their abundances in biological systems. Interestingly, relatively high numbers of species with ether linked fatty acid residues were identified among PC and PE lipids, i.e., 11 alkyl and 15 alkenyl (plasmalogens) PC and 2 alkyl and 13 alkenyl PE were reliably identified. The high numbers and contents of ether lipids are characteristic for heart tissue<sup>28</sup>.

LipidHunter nicely discriminated isomeric lipid species. Thus, 91 discrete PC species were represented by 51 unique precursor  $m/z$  including 40 isomers. For instance, 822.5233 with bulk identification PC(36:6) was assigned by LipidHunter to three discrete PC lipids eluting at different retention times based on three tandem mass spectra (Figure S-4), i.e., PC(18:3\_18:3), PC(16:1\_20:5), and PC(14:0\_22:6) eluted at 21.5, 22.3 and 22.6 min, respectively.

Furthermore, LipidHunter provided good specificity for isobaric PL species from different lipid classes (Figure S-5). It was possible to distinguish PC, PE, and PS lipids isobaric above the mass accuracy limit of 75 ppm. Such specificity of LipidHunter, although not required for high resolution and mass accuracy mass analysers, can be very useful for lipidomics datasets acquired on ion trap and quadrupole MS instruments.

To validate LipidHunter performance towards different PL classes and precursor ion abundances, identification results for different PLs within three orders of magnitude of precursor intensities were demonstrated and compared by manual spectra annotation (Supplementary File 2). For each PL class,

low, medium, and high abundant precursors were selected, manually identified, and compared with LipidHunter results. No bias towards PL class or precursor abundance was shown (Supplementary File 2).

To perform cross-validation of LipidHunter results with well-known lipid identification software LipidBlast<sup>11</sup>, LC-MS/MS data from three biological replicates of CM treated with SIN-1 for 70min (SIN\_70min) were used. LipidHunter identified 197 PLs (Table S-4). 89 PLs were identified by LipidBlast of which 79 were in common with the LipidHunter results (Table S-6). LipidBlast utilize experimental or predicted spectra libraries for PL identification. Thus, main differences observed for the identification of PC lipids are due to the absence of PC plasmalogens spectra library for the negative mode. Furthermore, spectra library might be instrument dependent and PL fragmentation patterns on Q-TOF instrument used in this study may differ from the one used to generate the LipidBlast spectra library. The bottom-up identification strategy used by LipidHunter is not limited by existing libraries, thus more PLs can be identified in contrast to the library based methods. Moreover, identification outputs generated by NIST MS PepSearch do not consider the isotope distribution and don't provide precursor associated extracted ion chromatograms, and manual revision is often necessary to ensure correct identification. Thus, out of 133 PL originally identified by LipidBlast only 89 passed manual inspection.

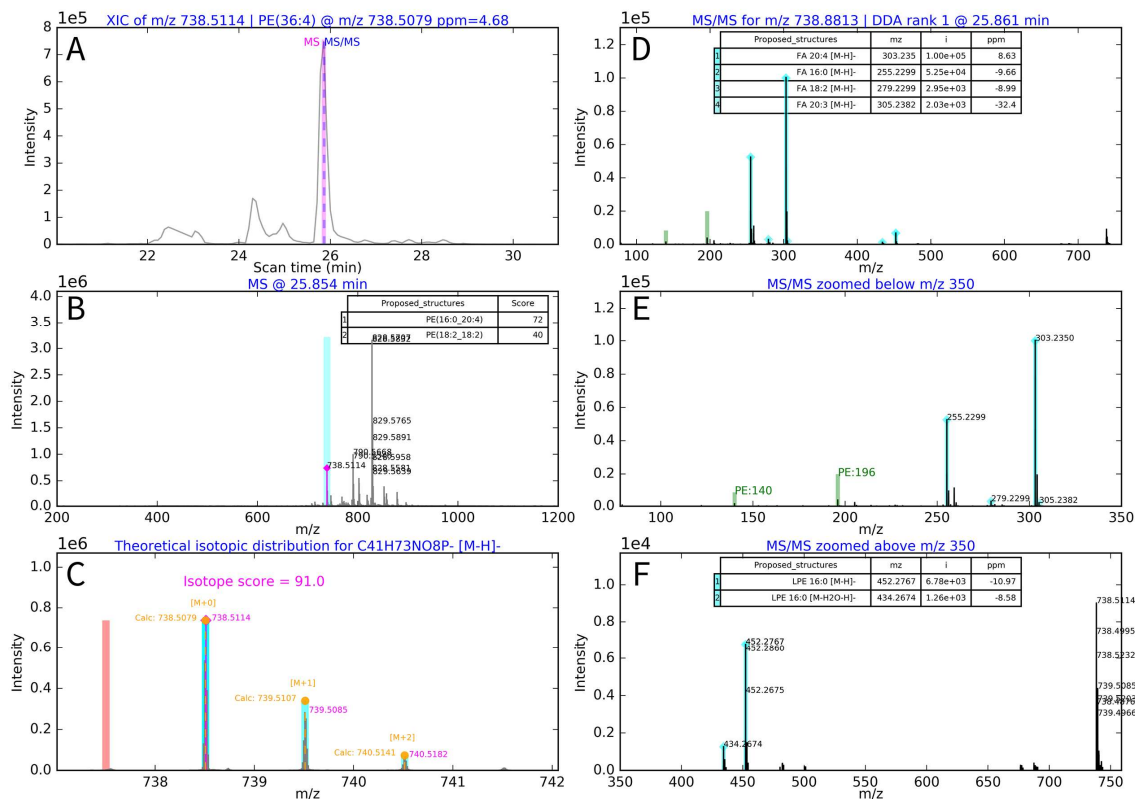
**Relative quantification of PLs in cardiomyocytes** – PLs identified by LipidHunter were relatively quantified by Progenesis QI based on the corresponding MS<sup>E</sup> datasets. Overall, 93 PLs showed significant differential regulation (ANOVA  $p \leq 0.05$ ) upon SIN-1 treatment. Among them were 53 PC, 19 PE, 9 PS, 6 PG, 5 PI, and 1 PA lipid (Table 1 and

Table S-7). Unfortunately, Progenesis QI was not always able to assign discrete structures. For instance, it was not possible to distinguish 15 out of 53 differentially regulated discrete PC species. Other peak alignment and integration tools such as XCMS<sup>29</sup> can be used for relative quantification in cooperation with LipidHunter as well.

PC lipids with long chain highly unsaturated FA were among the most down-regulated PL species (Table S-7 and Figure S-6, S7). Thus, PC(22:6\_22:6), PC(20:4\_20:4), PC(20:4\_22:6), PC(20:4\_22:5), and PC(18:3\_18:3) gradually decreased upon SIN-1 treatment reaching the lowest levels after 16 h. In contrast, two adrenic (docosatetraenoic) acid-containing PCs, i.e., PC(20:3\_22:4) and PC(P-16:0\_22:4) were among the most up-regulated lipids. Interestingly, several PG and PS species were also strongly up-regulated.

**Table 1. Summary of PL species identified by LipidHunter and relatively quantified in lipid extracts from SIN-1 treated cardiomyocytes.**

	PC	PE	PS	PG	PI	PA	Total
Unique precursors $m/z$	51	33	15	11	9	5	124
Discrete lipid species	91	54	25	14	12	6	202
Differentially regulated	53	19	9	6	5	1	93





**Figure 2. Image generated by LipidHunter for precursor at  $m/z$  738.5114 identified as PE(16:0\_20:4) lipid. Panel A:** Extracted ion chromatogram (XIC) for ion observed at  $m/z$  738.5114. Results of bulk identification [PE(36:4)], theoretical (738.5079) and observed (738.5114)  $m/z$  values, mass accuracy error (4.68 ppm) are listed in the header. Positions of MS (magenta line) and MS/MS (blue dashed line) scans used for the identification are indicated in the XIC. **Panel B:** Mass spectrum with the corresponding precursor signal marked (magenta diamond indicating the apex of the monoisotopic signal) within defined mass accuracy range (cyan bar). Scan time (25.854 min) is listed at the header. **Panel C:** Zoomed Mass spectrum illustrating precursor isotopic distribution. Elemental composition used for the calculation is listed in the header. Pseudo [M+0] signal marked with a magenta bar. Theoretically calculated and observed  $m/z$  values for the isotopes signals are shown in orange and magenta text, respectively. **Panel D:** Tandem mass spectrum used for assignment.  $m/z$  value of the precursor listed for the corresponding MS/MS scan (738.8813), DDA rank (1) and MS/MS scan time (25.861 min) are listed in the header. Product ions assigned to specific signals are marked with cyan (FA product and neutral loss ions) and green (PL class specific ions). Top 10 FA specific product ions (in this example, four signals were identified and illustrated) ranked according to their intensities (used for Rank factor  $R_{frag}$  calculations) are listed in the table (proposed structure,  $m/z$ , intensity and mass accuracy error). **Panel E:** MS/MS spectrum zoomed below  $m/z$  350 for better illustration of FA (cyan), PL class specific (green) and unspecific (magenta lines) product ions. **Panel F:** MS/MS spectrum zoomed above  $m/z$  350 for better illustration of FA (cyan), PL class specific (green) and unspecific (magenta lines) neutral loss ions. Top 10 FA specific neutral loss ions ranked according to their intensities (used for Rank factor  $R_{frag}$  calculations) are listed in the table (proposed structure,  $m/z$ , intensity and mass accuracy error).

Principal component analysis and clustering of significantly regulated PL species demonstrated a correlation of experimental groups with the duration of nitroxidative stress (Figures S6 and S7). Thus, control and 16 h groups were clearly separated, whereas short-term treated samples (15, 30, and 70 min) were related to each other. Two PS [PS(16:0\_20:4), PS(18:2\_20:4)], one PA [PA(16:0\_18:1)] and one plasmalogen PC [PC(P-16:0\_22:4)] constituted the cluster with strongly up-regulated species showing a gradual increase in quantities over all time points, where PC(22:6\_22:6) and PS(18:1\_20:3) were the most down-regulated of the lipids included. Relative changes in the cardiomyocyte lipidome upon oxidative stress-related pathology were reported previously<sup>30</sup>. However, the high specificity of LipidHunter allowed the identification of discrete PL species affected by metabolic reconfiguration.

Lipidomics studies addressing cardiovascular diseases (CVDs) usually focus on the role of lipid classes rather than exact molecular species or report only bulk PL compositions<sup>31,32</sup>. Knowledge of FA carbon chain length and the number of double bounds is important to reveal the underlying mechanisms of metabolic reconfiguration under pathological conditions, and thus can provide more specific biomarkers and therapeutic targets. In high-throughput lipidomics, LipidHunter can significantly facilitate the discovery of molecular species correlated with disease progression.

**LipidHunter application to serum lipidomics** – To demonstrate LipidHunter performance for well-studied lipidomes, lipids extracted from human serum were analysed by DDA LC-MS/MS. LipidHunter was capable to identify 175 discrete PL species including 67 PC, 43 PE, 27 PS, 15 PG, 14 PI, and 9 PA (Table S-8). Comparing results of bulk identification by LipidHunter (134 PLs) with previously published data from NIST plasma lipidome (139 bulk PL identifications)<sup>33</sup>, 75% (101 PLs) were common for both datasets.

**LipidHunter application to shotgun datasets** – Finally LipidHunter was validated for shotgun lipidomics using data provided by Herzog et al<sup>34</sup> for bovine heart lipids. This dataset was acquired on an LTQ-Orbitrap MS in negative ion mode using DDA mode. Using enlarged FA configuration file and precursor list available from the raw data, LipidHunter successfully identified all PL structures provided by LipidXplorer, with 35 additionally identified PLs. In total, 107

PLs from all six classes, i.e., 22 PC, 58 PE, 5 PG, 10 PI, 6 PS, and 6 PA (Table S-9), were identified.

## Conclusion

High-throughput software assisted analysis of a large number of MS datasets, is currently a bottleneck in untargeted lipidomics studies in translational and systems medicine research. The most popular methods rely on the availability of experimental or *in silico* created libraries of tandem mass spectra limiting their flexibility. LipidHunter provides an alternative approach for the automated rapid identification of lipids using typical steps of manual spectra assignment. It makes LipidHunter intuitively clear and easy to use.

LipidHunter relies on a set of configuration files, which can be easily adjusted to an instrument, ionisation, and ion adduct types without any need to reference libraries. Furthermore, configuration files/white lists for FA and PL specific fragment ions can be extended, based on the needs (e.g. can include oxidized, isotopically or otherwise labelled lipids etc.). LipidHunter provides full control over identification results via detailed tables and graphical reports. This allows assessing identification results and monitoring data quality (e.g., position of MS/MS scans relative to precursor XIC might indicate the efficiency of DDA intensity threshold as well as the optimal length of dynamic exclusion).

LipidHunter was validated for the identification of PLs in lipidome of cardiomyocytes obtained from the cell culture model of nitroxidative stress. Using the complex and dynamically changing lipidome, LipidHunter reliably identified 202 discrete lipid species, 93 of which were further shown to be differentially regulated upon stress induction.

The current version of LipidHunter (free available at <https://bitbucket.org/SysMedOs/lipidhunter> under GPLv2 license) was optimised for PL identification. However, applications to lysoPL, sphingolipids, and glycerolipids are under development.

## ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge on the ACS Publications website.

**Table S-1.** Configuration file used to define the white list for fatty acid residues. **Table S-2.** Configuration file used to define PL



class specific ions. **Table S-3.** Configuration file used to define weight factors  $W_{frag}$ . **Table S-4.** LipidHunter output tables for PC, PE, PS, PI, PG, and PA lipids exemplified for 70min\_SIN sample. **Table S-5.** Summary of PLs identified by LipidHunter in lipid extracts from cardiomyocytes treated with SIN-1. **Table S-6.** Comparison between LipidHunter and LipidBlast results for PL identification. **Table S-7.** Summary of relative quantification for discrete PLs identified by LipidHunter in extracts from SIN-1 treated cardiomyocytes using Progenesis Q1. **Table S-8.** Human serum PLs identified by LipidHunter. **Table S-9.** PLs identified by LipidHunter in shotgun DDA datasets. **Supplementary File 1.** LipidHunter User Guide. **Supplementary File 2.** Examples of LipidHunter identification results for lipids from different PL classes with low, medium and high abundances of precursor ions. **Figure S-1.** Screenshot of LipidHunter graphic user interface (GUI). **Figure S-2.** Screenshot of LipidHunter .html report. **Figure S-3.** Comparison of MS and MS/MS information between DDA and MS<sup>E</sup> acquisition modes. **Figure S-4.** LipidHunter specificity towards isomeric lipid species of one PL class. **Figure S-5.** LipidHunter specificity towards isobaric lipid species from different PL classes. **Figure S-6.** Clustering analysis of differentially regulated PLs in SIN-1 treated cardiomyocytes. **Figure S-7.** Principal component analysis of differentially regulated PLs in SIN-1 treated cardiomyocytes.

## AUTHOR INFORMATION

### Corresponding Author

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

### Author Contributions

MF and ZN contributed to the idea conception and design of the LipidHunter algorithms. ZN contributed with source code design and implementation. GA assisted with the computational method design. ML assisted with software validation. The manuscript was written through the contributions of all authors (MF, ZN, RH, GA, ML). All authors have given approval to the final version of the manuscript.

### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENT

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 is gratefully acknowledged. We acknowledge financial support from the Deutsche Forschungsgemeinschaft (DFG; FE-1236/3-1 to M.F., INST 268/289-1 FUGG to R.H.), and the European Regional Development Fund (ERDF, European Union and Free State Saxony; 100146238 and 100121468 to M.F).

## REFERENCES

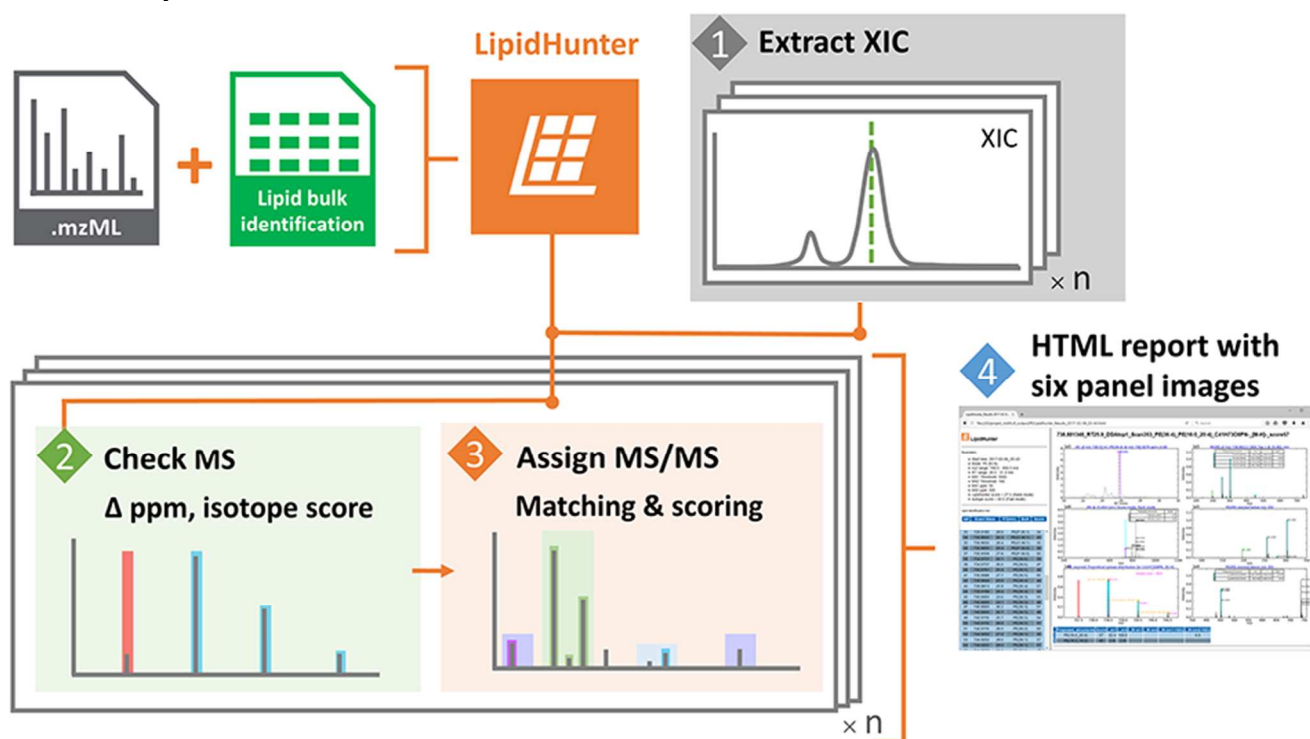
- (1) Cajka, T.; Fiehn, O. *TrAC Trends in Anal. Chem.* **2014**, *61*, 192-206.
- (2) Cajka, T.; Fiehn, O. *Anal. Chem.* **2016**, *88*, 524-545.
- (3) Niemelä, P. S.; Castillo, S.; Sysi-Aho, M.; Orešič, M. *J. Chromatogr. B: Anal. Technol. Biomed. Life Sci.* **2009**, *877*, 2855-2862.

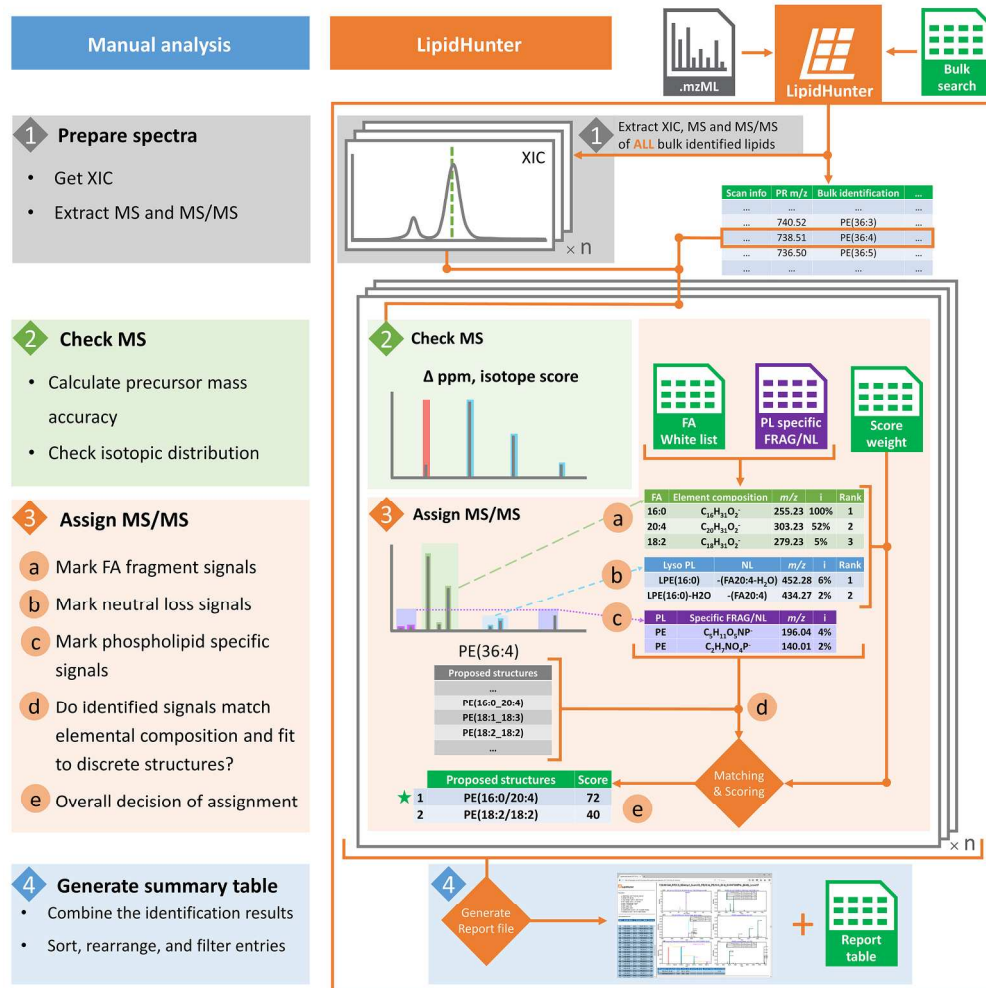
- (4) Hartler, J.; Tharakan, R.; Köfeler, H. C.; Graham, D. R.; Thallinger, G. G. *Briefings Bioinf.* **2013**, *14*, 375-390.
- (5) Pulfer, M.; Murphy, R. C. *Mass Spectrom. Rev.* **2003**, *22*, 332-364.
- (6) Milne, S.; Ivanova, P.; Forrester, J.; Alex Brown, H. *Methods (Amsterdam, Neth.)* **2006**, *39*, 92-103.
- (7) Godzien, J.; Ciborowski, M.; Martínez-Alcázar, M. P.; Samczuk, P.; Kretowski, A.; Barbas, C. *J. Proteome Res.* **2015**, *14*, 3204-3216.
- (8) Sud, M.; Fahy, E.; Cotter, D.; Brown, A.; Dennis, E. A.; Glass, C. K.; Merrill, A. H.; Murphy, R. C.; Raetz, C. R. H.; Russell, D. W.; Subramaniam, S. *Nucleic Acids Res.* **2007**, *35*, D527-D532.
- (9) Wishart, D. S.; Jewison, T.; Guo, A. C.; Wilson, M.; Knox, C.; Liu, Y.; Djoumbou, Y.; Mandal, R.; Aziat, F.; Dong, E.; Bouatra, S.; Sinelnikov, I.; Arndt, D.; Xia, J.; Liu, P.; Yallou, F.; Bjorn Dahl, T.; Perez-Pineiro, R.; Eisner, R.; Allen, F., et al. *Nucleic Acids Res.* **2013**, *41*, D801-D807.
- (10) Smith, C. A.; O'Maille, G.; Want, E. J.; Qin, C.; Trauger, S. A.; Brandon, T. R.; Custodio, D. E.; Abagyan, R.; Siuzdak, G. *Ther. Drug Monit.* **2005**, *27*, 747-751.
- (11) Kind, T.; Liu, K.-H.; Lee, D. Y.; DeFelice, B.; Meissen, J. K.; Fiehn, O. *Nat. Methods* **2013**, *10*, 755-758.
- (12) Tsugawa, H.; Cajka, T.; Kind, T.; Ma, Y.; Higgins, B.; Ikeda, K.; Kanazawa, M.; VanderGheynst, J.; Fiehn, O.; Arita, M. *Nat. Methods* **2015**, *12*, 523-526.
- (13) Herzog, R.; Schwudke, D.; Shevchenko, A. In *Current Protocols in Bioinformatics*; John Wiley & Sons, Inc.: Hoboken, NJ 07030, U.S.A., 2013.
- (14) Gruber, F.; Bicker, W.; Oskolkova, O. V.; Tschachler, E.; Bochkov, V. N. *J. Lipid Res.* **2012**, *53*, 1232-1242.
- (15) Matyash, V.; Liebisch, G.; Kurzchalia, T. V.; Shevchenko, A.; Schwudke, D. *J. Lipid Res.* **2008**, *49*, 1137-1146.
- (16) Kessner, D.; Chambers, M.; Burke, R.; Agus, D.; Mallick, P. *Bioinformatics* **2008**, *24*, 2534-2536.
- (17) Bald, T.; Barth, J.; Niehues, A.; Specht, M.; Hippler, M.; Fufezan, C. *Bioinformatics* **2012**, *28*, 1052-1053.
- (18) Liebisch, G.; Vizcaino, J. A.; Köfeler, H.; Trotzmüller, M.; Griffiths, W. J.; Schmitz, G.; Spener, F.; Wakelam, M. J. *J. Lipid Res.* **2013**, *54*, 1523-1530.
- (19) Sturn, A.; Quackenbush, J.; Trajanoski, Z. *Bioinformatics* **2002**, *18*, 207-208.
- (20) Taguchi, R.; Houjou, T.; Nakanishi, H.; Yamazaki, T.; Ishida, M.; Imagawa, M.; Shimizu, T. *J. Chromatogr. B: Anal. Technol. Biomed. Life Sci.* **2005**, *823*, 26-36.
- (21) Valkenburg, D.; Mertens, I.; Lemièr, F.; Witters, E.; Burzykowski, T. *Mass Spectrom. Rev.* **2012**, *31*, 96-109.
- (22) Meija, J.; Caruso, J. A. *J. Am. Soc. Mass Spectrom.* **2004**, *15*, 654-658.
- (23) Finkel, T.; Holbrook, N. J. *Nature (London, U. K.)* **2000**, *408*, 239-247.
- (24) Madamanchi, N. R.; Vendrov, A.; Runge, M. S. *Arterioscler., Thromb., Vasc. Biol.* **2005**, *25*, 29-38.
- (25) Harrison, D.; Griendling, K. K.; Landmesser, U.; Hornig, B.; Drexler, H. *Am. J. Cardiol.* **2003**, *91*, 7-11.

- (26) Milic, I.; Hoffmann, R.; Fedorova, M. *Anal. Chem.* **2012**, *85*, 156-162.
- (27) Griesser, E.; Vemula, V.; Raulien, N.; Wagner, U.; Reeg, S.; Grune, T.; Fedorova, M. *Redox Biol.* **2017**, *11*, 438-455.
- (28) McHowat, J.; Creer, M. H. *Am. J. Physiol.: Heart Circ. Physiol.* **2000**, *278*, H1933-H1940.
- (29) Tautenhahn, R.; Patti, G. J.; Rinehart, D.; Siuzdak, G. *Anal. Chem.* **2012**, *84*, 5035-5039.
- (30) Sousa, B.; Melo, T.; Campos, A.; Moreira, A. S. P.; Maciel, E.; Domingues, P.; Carvalho, R. P.; Rodrigues, T. R.; Girão, H.; Domingues, M. R. M. *J. Cell. Physiol.* **2016**, *231*, 2266-2274.
- (31) Stegemann, C.; Pechlaner, R.; Willeit, P.; Langley, S. R.; Mangino, M.; Mayr, U.; Menni, C.; Moayyeri, A.; Santer, P.; Rungger, G.; Spector, T. D.; Willeit, J.; Kiechl, S.; Mayr, M. *Circulation* **2014**, *129*, 1821-1831.

- (32) Würtz, P.; Havulinna, A. S.; Soininen, P.; Tynkkynen, T.; Prieto-Merino, D.; Tillin, T.; Ghorbani, A.; Artati, A.; Wang, Q.; Tiainen, M.; Kangas, A. J.; Kettunen, J.; Kaikkonen, J.; Mikkilä, V.; Jula, A.; Kähönen, M.; Lehtimäki, T.; Lawlor, D. A.; Gaunt, T. R.; Hughes, A. D., et al. *Circulation* **2015**, *131*, 774-785.
- (33) Quehenberger, O.; Armando, A. M.; Brown, A. H.; Milne, S. B.; Myers, D. S.; Merrill, A. H.; Bandyopadhyay, S.; Jones, K. N.; Kelly, S.; Shaner, R. L.; Sullards, C. M.; Wang, E.; Murphy, R. C.; Barkley, R. M.; Leiker, T. J.; Raetz, C. R. H.; Guan, Z.; Laird, G. M.; Six, D. A.; Russell, D. W., et al. *J. Lipid Res.* **2010**, *51*, 3299-3305.
- (34) Herzog, R.; Schwudke, D.; Schuhmann, K.; Sampaio, J. L.; Bornstein, S. R.; Schroeder, M.; Shevchenko, A. *Genome Biol.* **2011**, *12*, R8.

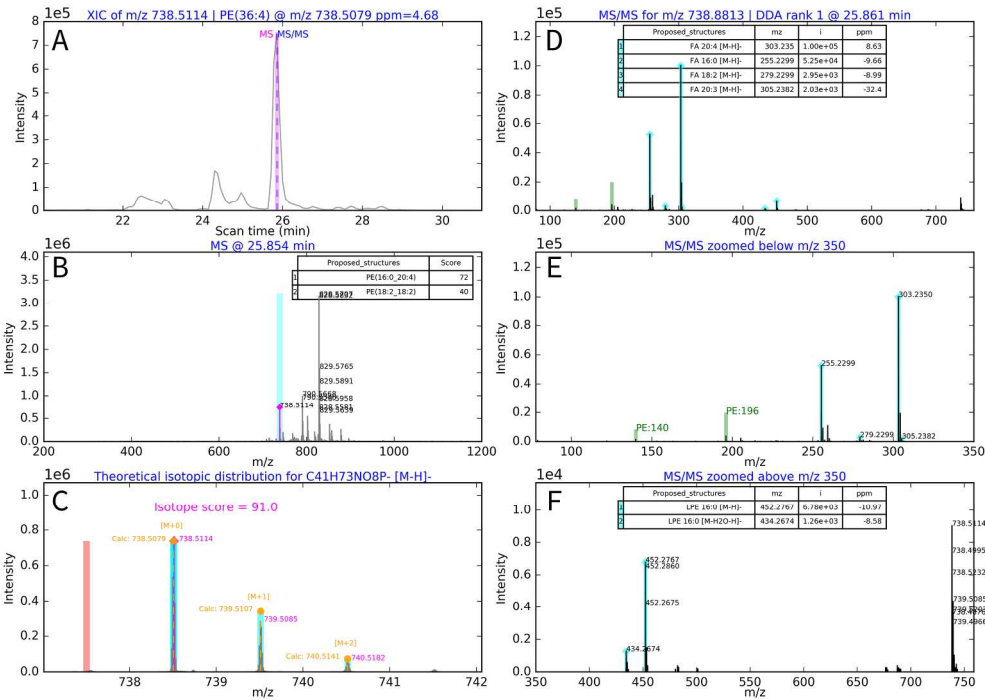
For TOC Only



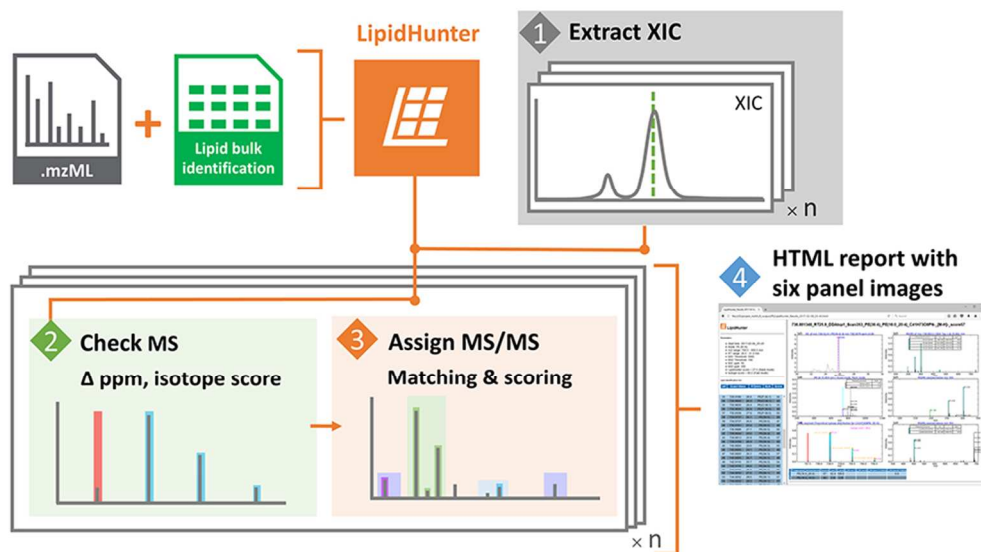


215x217mm (300 x 300 DPI)

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



215x152mm (300 x 300 DPI)



82x46mm (300 x 300 DPI)