

## Greazy: open-source software for automated phospholipid MS/MS identification

Michael Allen Kochen, Matthew Chase Chambers, Jay D. Holman, Alexey I. Nesvizhskii, Susan T. Weintraub, John T. Belisle, Nurul Islam, and David Lee Tabb

*Anal. Chem.*, Just Accepted Manuscript • DOI: 10.1021/acs.analchem.6b00021 • Publication Date (Web): 17 May 2016

Downloaded from <http://pubs.acs.org> on May 19, 2016

### 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

Analytical Chemistry is published by the American Chemical Society. 1155 Sixteenth Street N.W., Washington, DC 20036

Published by American Chemical Society. Copyright © American Chemical Society. However, no copyright claim is made to original U.S. Government works, or works produced by employees of any Commonwealth realm Crown government in the course of their duties.

# Greazy: open-source software for automated phospholipid MS/MS identification

Michael A. Kochen<sup>1</sup>; Matthew C. Chambers<sup>1</sup>; Jay D. Holman<sup>1</sup>; Alexey I. Nesvizhskii<sup>2</sup>; Susan T. Weintraub<sup>3</sup>; John T. Belisle<sup>4</sup>; Nurul Islam<sup>4</sup>; David L. Tabb<sup>1\*</sup>

<sup>1</sup>Department of Biomedical Informatics, Vanderbilt University, Nashville, TN 37203; <sup>2</sup>Department of Pathology, University of Michigan, Ann Arbor, MI 48109; <sup>3</sup>Department of Biochemistry, UT Health Science Center at San Antonio, San Antonio, TX 78229; <sup>4</sup>Department of Microbiology, Immunology & Pathology, Colorado State University, Fort Collins, CO 80523

\* David L. Tabb, Division of Molecular Biology and Human Genetics, Stellenbosch University Faculty of Medicine and Health Sciences, Cape Town, South Africa 7550. +27 21 938 9403. dtabb@sun.ac.za

## Abstract

Lipid identification from data produced with high-throughput technologies is essential to the elucidation of the roles played by lipids in cellular function and disease. Software tools for identifying lipids from tandem mass (MS/MS) spectra have been developed, but they are often costly or lack the sophistication of their proteomics counterparts. We have developed Greazy, an open source tool for the automated identification of phospholipids from MS/MS spectra, that utilizes methods similar to those developed for proteomics. From user-supplied parameters, Greazy builds a phospholipid search space and associated theoretical MS/MS spectra. Experimental spectra are scored against search space lipids with similar precursor masses using a peak score based on the hypergeometric distribution and an intensity score utilizing the percentage of total ion intensity residing in matching peaks. The LipidLama component filters the results via mixture modelling and density estimation. We assess Greazy's performance against the NIST 2014 metabolomics library, observing high accuracy in a search of multiple lipid classes. We compare Greazy/LipidLama against the commercial lipid identification software LipidSearch and show that the two platforms differ considerably in the sets of identified spectra while showing good agreement on those spectra identified by both. Lastly, we demonstrate the utility of Greazy/LipidLama with different instruments. We searched data from replicates of alveolar type 2 epithelial cells obtained with an Orbitrap and from human serum replicates generated on a Q-TOF. These findings substantiate the application of proteomics derived methods to the identification of lipids. The software is available from the ProteoWizard repository: [<http://tiny.cc/bumbershoot-vc12-bin64>].

## Introduction

The diversity of cellular functions and diseases in which lipids play a role substantiates their importance and the necessity of their study<sup>1-7</sup>. A complete understanding of these roles requires a comprehensive accounting of the lipidome, but this has proven an elusive goal. The study of complex lipid mixtures can be done with the help of rapidly scanning tandem mass spectrometers, producing overwhelming amounts of data that, in turn, require sophisticated software for interpretation. Proteomics has benefitted from well-developed software for years but the development of similar software for lipidomic analysis has been relatively slow.

Current lipidomics tools often cater to specific applications. Some are designed for direct infusion lipidomics where the sample is subjected to electrospray without prior chromatographic separation<sup>8</sup>. AMDMS-SL<sup>9</sup> and LipidView<sup>10</sup> use this technique along with precursor ion and neutral loss scanning to search for ions that are characteristic of a head group or fatty acyl substituent. Tools such as LipidQA<sup>11</sup>, LipidExplorer<sup>12</sup>, and LipidInspector<sup>13</sup> are designed for direct infusion experiments but with data-dependent acquisition scanning. The alternative to direct infusion is liquid chromatography, a method shared with shotgun proteomics. LC/MS can be used to identify a list of possible lipid identifications, often isomers from the same lipid class. Lipid specific software such as Lipid Data Analyzer<sup>14</sup> as well as general metabolomics software like MZmine2<sup>15</sup> are designed for this type of analysis. LC-MS/MS allows for isomeric species to be distinguished through the fragments they produce. The commercial software LipidSearch<sup>16</sup> can analyze such data.

Although lipidomics software has advanced in recent years, several challenges bar the path to routine interpretation of these data sets. In proteomics, peptide sequences are routinely identified from MS/MS spectra through database search algorithms that employ simple rules for predicting peptide fragmentation patterns. Lipids exhibit greater structural diversity than peptides. The LIPID MAPS database categorizes more than 37000 lipids into eight lipid categories that are further divided into dozens of subclasses, each of which ionizes and fragments in a distinct manner<sup>17</sup>. As a result, predicting the appearance of MS/MS spectra for lipids requires greater attention than in proteomics, and the number of fragments seen for a given compound is generally smaller. Publicly available MS/MS spectral libraries for lipids are also far less populated and diverse than those in proteomics, although there has been some recent work in this area with the development of LipidBlast<sup>18</sup>, a collection of predicted MS/MS spectra for many lipid classes. The literature detailing the MS/MS mechanisms for lipids is considerably less fully developed than it is for proteomics. In proteomics, extensive spectral libraries and sequence databases have yielded powerful statistical scoring algorithms and methods for estimating false discovery rates. Leveraging methods developed for proteomics to lipidomics may begin to address the shortage of tools in this space.

We have developed the software “Greazy” for the high-throughput identification of phospholipid MS/MS data. Greazy employs a number of techniques borrowed from proteomics that have the potential to address many of the challenges that arise in lipidomics. Greazy incorporates Proteowizard<sup>19</sup>, enabling it to accept many vendor-specific and open file formats in use today. A set of user defined parameters define a phospholipid search space and a comprehensive literature-derived fragmentation model for each lipid in that space. These predicted spectra are compared to experimental ones through two probability-based scores. Lastly, a mixture modeling approach using density estimation is used to calculate the false discovery rate and to generate a cutoff score. Successful matches are written to the easily reviewed mzTab file format<sup>20</sup>.

## Experimental Section

### Data Sets

**NIST Tandem MS Library.** The 2014 release of the NIST small molecule MS/MS library consists of 193,119 spectra for a number of different compound types including metabolites, drugs, phospholipids, and more. The spectra for these compounds were acquired under a broad range of experimental conditions. Ion trap and collision cell instrument types were used; data were acquired in both positive and negative ion mode. A number of precursor adducts were used, along with a wide range of collision energies, resulting in differing degrees of fragmentation. This data set was split into two files based on the polarity of each entry resulting in a positive ion file containing 161,355 entries and a negative ion file containing 31,764. Next, phospholipids were extracted from these files using a string search. Those entries containing the strings “glycero-3-phos,” “sphingosyl,” or “sphingenyl” were kept and the rest discarded. This reduced the positive ion file to 3,107 entries and the negative ion file to 1,232. Represented in these entries were 141 distinct compounds in the positive ion file and 175 in the negative. The total number of distinct compounds was 214. The resulting files were reformatted as MGF files for use with Greazy.

**Bone Marrow Stem Cells.** In the Department of Biochemistry at the University of Texas Health Science Center at San Antonio lipids were extracted from cultured bone marrow stem cells collected from wild-type mice. The lipid extracts were subjected to reversed-phase liquid chromatography coupled to a Thermo Fisher Q Exactive tandem mass spectrometer. Analysis was done in both positive and negative ion modes resulting in 13362 and 12438 MS/MS spectra respectively. Lipid identifications were made with LipidSearch software version 4.0.14 (Thermo Scientific) and provided to us for comparison purposes. The raw files were converted to mzML files using Proteowizard’s MSconvertGUI tool for analysis with Greazy and LipidLama.

**Alveolar Type 2 Epithelial Cells.** At the Pacific Northwest National Laboratory lipids were extracted from sorted Type 2 epithelial cells obtained via magnetic bead sorting from lung tissue harvested from post-natal day 7 and post-natal day 28 C57BL/6 mice. Lungs from five post-natal day-7 and five post-natal day-28 mice were each pooled, sorted and partitioned into replicates. The replicates were each subjected to lipid extraction using chloroform/methanol (2:1, v/v), and the extracts were analyzed by ultra-performance reversed-phase liquid chromatography coupled to a Thermo Scientific LTQ Orbitrap Velos mass spectrometer. The extracts were analyzed in both positive and negative ion mode.

**Human Serum.** At the Mycobacteria Research Laboratories within the Department of Microbiology, Immunology and Pathology at Colorado State University - Fort Collins lipids were extracted from platelet poor human plasma (Sigma-Aldrich) derived serum using chloroform/methanol (2:1, v/v). The extract was analyzed by high-performance reversed-phase liquid chromatography coupled to an Agilent Q-TOF mass spectrometer. Ten technical replicates were produced with selection of 4 precursors per cycle, five at low collision energy ( $3 * (m/z)/100 + 2$ ) and five at high collision energy ( $5 * (m/z)/100 + 2$ ). All data was acquired in negative ion mode using Agilent’s MassHunter software. These data sets can be found at the MassIVE repository (ID: MSV000079377).

## 1 2 Software 3

4 Greazy is software for the high-throughput identification of phospholipids from MS/MS spectra. Proteowizard<sup>19</sup>  
5 is integrated into the software, enabling Greazy to handle a number of open and vendor specific file formats.  
6 The experimental and search space parameters are input through a graphical user interface (Figure S1) and are  
7 used to build a list of phospholipid precursor ions that are then fragmented in silico for comparison to the  
8 experimental spectra. The software generates theoretical spectra for lipid classes from the glycerophospholipid  
9 and sphingolipid categories, each of which has its own structure and requires a distinct model when building  
10 the search space. Cardiolipins (CL), a member of the glycerophospholipids, require additional attention and are  
11 considered a separate category here. Details of the search space can be found in Supporting Text S1.  
12

13 **Experimental Parameters and Precursors:** Following the construction of the lipid search space, predicted  
14 precursor ions are generated for each included lipid based on the detection polarity (positive or negative) and  
15 the adducts expected to be present. For positive ion mode, available adducts include sodium, potassium,  
16 lithium, and ammonium; chloride, acetate, and formate are included for negative ion mode. Sodium, potassium,  
17 and lithium adducted cardiolipin precursors are also predicted in negative mode. Protonated and deprotonated  
18 precursors are assumed to be present in every positive and negative ion mode search, respectively. Lastly, the  
19 mass accuracy tolerances for the precursor and fragment ions must be supplied in either ppm or Daltons.  
20

21 **Fragmentation Patterns.** Fragmentation patterns for all possible precursors were gleaned from the literature,  
22 or extrapolated based on it<sup>24-35</sup>, and take into account the scanned mass range. To maximize the advantage of  
23 using probabilistic scoring algorithms, the predicted MS/MS spectra for each class of lipids include all possible  
24 fragments found in the literature, differing from systems that model lipids based only on a small set of  
25 characteristic ions.  
26

27 **Glycerophospholipids:** In positive ion mode the glycerophospholipid precursors, and subsequent fragmentation  
28 patterns, are divided into two groups: those with alkali metal adducts,  $[M + Alk]^+$  where Alk = Na, K, or Li (Table  
29 S1: diacyl, Table S2: alkyl and alkenyl)<sup>24-28</sup>, and those with nonmetal adducts including  $[M + H]^+$  and  $[M + NH_4]^+$   
30 (Table S3)<sup>24-27</sup>. Possible precursors in negative ion mode include  $[M - H]^-$ ,  $[M + Cl]^-$ ,  $[M + CHOO]^-$ , and  $[M +$   
31  $CH_3COO]^-$  (Tables S4-S5)<sup>24-27, 29-30</sup>. Negatively charged precursors for PCs are generated only when adducts are  
32 present. These include  $[M + adduct]^-$  and  $[M - CH_3]^-$ .  
33

34 The fragmentation models within the glycerophospholipid classes do not distinguish between positional  
35 isomers. Fragment intensity ratios (carboxylate and loss of carboxylic acid ions) can sometimes be used to  
36 determine fatty acyl positions<sup>9</sup>. These intensities, however, can be affected by a number of factors including  
37 the lipid class, fragment type, presence of adducts, ion mode, instrument type, collisional energy, chain length,  
38 and saturation levels<sup>25</sup>. Greazy is meant to be a tool that can be used under a broad range of experimental  
39 conditions. As such the software does not predict relative ion intensities but instead relies on a probabilistic  
40 intensity score (see below) to retain ion intensity information. To determine positional isomers Greazy outputs  
41 a list of fragments and their intensities for each identification that can be examined to determine the fatty acyl  
42 positions given the experimenters particular setup.  
43

44 **Phosphosphingolipids:** Fragmentation patterns for the phosphosphingolipids in positive ion mode are likewise  
45 divided into metal (Table S6)<sup>25-27, 32-33</sup> and nonmetal (Table S7)<sup>25-26, 31-33</sup> precursor adducts and broken down  
46 further based on the head groups. The same precursors as above are found here. The negative ion mode  
47 fragmentation models for these lipids are sparse, with fragments pertaining only to the head group (Table  
48 S8)<sup>26-27, 33</sup>. Like PCs, negatively charged precursors for sphingomyelins are generated only in adducted forms,  
49 specifically  $[M + adduct]^-$  and  $[M - CH_3]^-$ .  
50

51 **Cardiolipins:** Cardiolipin precursors generated in positive mode include  $[M + H]^+$ ,  $[M + Alk]^+$ ,  $[M - H + 2Alk]^+$ , and  
52  $[M - 2H + 3Alk]^+$  (Table S8)<sup>34</sup>. Negative mode precursors include  $[M - H]^-$ ,  $[M - 2H + Alk]^{1-}$  and  $[M - 2H]^{2-}$  (Tables  
53

1  
2 S10-S12)<sup>27, 35-37</sup>.

3  
4 **Preprocessing and Scoring Algorithms.** Preprocessing of an MS/MS spectrum begins with removal of peaks  
5 within the predefined tolerance of the precursor m/z. The total ion current is then normalized and the peak  
6 intensities are accordingly adjusted to sum to 1000. When the experimental and search space precursor m/z  
7 values are within the user provided tolerance, the associated MS/MS spectrum is scored against the  
8 corresponding theoretical tandem spectrum using the scoring algorithm below. The number of matching peaks  
9 and the proportion of total ion intensity found in those peaks are considered separately. These scores are then  
10 combined to determine a final score. The scoring algorithms of Greazy are designed to find the single best  
11 match for each MS/MS spectrum. When isomeric lipid species are present the highest scoring isomer is  
12 returned.  
13

14 *Peak Score:* The peak score uses the hypergeometric distribution (HGD) to calculate the probability of randomly  
15 matching a given number of peaks. The HGD is used when one is sampling without replacement from a discrete  
16 population for which members are one of two types. Consider, for example, a jar containing N marbles, M of  
17 which are green with the remainder white. We would like to know the probability of finding x green marbles  
18 when selecting K marbles from the jar without replacement. This can be calculated using the HGD which takes  
19 the form  
20

$$P(x|N, M, K) = \frac{\binom{M}{x} \binom{N-M}{K-x}}{\binom{N}{K}}$$

21 In order to use the HGD for peak scoring, the continuous m/z range is divided into a discrete number of bins  
22 based on the scan range of the experiment and the provided mass tolerance. We can then use the HGD to  
23 compute the probability of finding i matching peaks given N bins, K spectrum peaks, and M theoretical peaks.  
24 Given x matching peaks between a predicted and experimental spectrum we calculate the probability of at  
25 least x+1 matches occurring by chance by summing from x+1 to M and obtain the peak score  
26

$$S_1 = \sum_{i=x+1}^M P(i|N, M, K).$$

27 *Intensity Score:* Consider a spectrum with k matching peaks that contain a proportion of the total ion intensity  
28 for that spectrum. The intensity score is the probability of finding more ion intensity than the matched peaks  
29 when randomly choosing k peaks from the spectrum. Let i be the summed intensity of the matched peaks and  $i_j$   
30 be the summed intensity of the jth combination of k peaks. The intensity score is defined as the fraction of  
31 combinations that hold more ion intensity than the matched peaks. This is given by the equation  
32

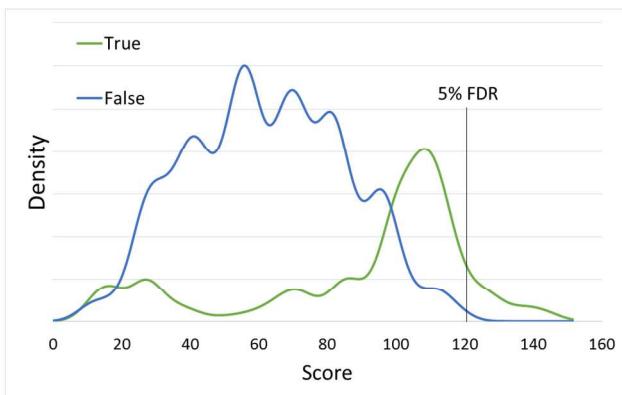
$$S_2 = \frac{\sum_{j=1}^n p_j}{\binom{n}{k}}, \quad p_j = \begin{cases} 1 & \text{if } i_j > i \\ 0 & \text{if } i_j \leq i \end{cases}$$

33 *Combined Score:* Fisher's method is used to combine the peak and intensity scores into a final score<sup>38</sup> resulting  
34 in a  $\chi^2$  statistic that is given by  
35

$$\chi^2 = -2 \sum_{i=1}^2 \ln(S_i).$$

36 **Error Estimation.** Greazy produces a list of lipid identifications by first matching the precursor m/z and then  
37 producing a score as detailed above. Low scoring identifications may match a limited number of predicted  
38 fragments and are likely to be false discoveries. LipidLama was developed to eliminate such low quality  
39 matches. LipidLama, a software component within Greazy, employs a mixture modeling approach to the  
40 problem of false discoveries (Figure 1). It is assumed that the distribution of scores produced by Greazy is a  
41 combination of distributions for true and false matches. LipidLama uses the kernel density estimation method  
42 to fit two such distributions to the data. In this method the distribution of true matches is estimated using the  
43

1  
2 distribution of high scores and a known distribution that is assumed to be representative of the false matches.  
3 LipidLama uses the distribution of second highest scores as the model for the distribution of false matches<sup>39</sup>.  
4  
5



19  
20 Figure 1. The kernel density estimation algorithm  
21 separates the score distribution from Greazy into  
22 components for the True (green) and False (blue)  
23 positives. A cutoff score for a desired FDR is then  
24 calculated.

25 The algorithm used by LipidLama largely follows Choi et al.,<sup>40</sup> with a few exceptions. Bandwidth selection, for  
26 example, is calculated using the method of Sheather and Jones<sup>41</sup>. Also, the proportion of the high score  
27 distribution that is accounted for by the false match distribution must be estimated empirically. Let  $i$  represent  
28 evenly spaced points along the Greazy score axis,  $f_1$  and  $f_2$  represent the estimated distributions for the highest  
29 and second highest scores respectively, and  $w_i$  be the difference between these two distributions at the point  $i$ .  
30 Then the proportion  $p$  of the full distribution that is accounted for by the false hits is estimated to be  
31

$$\min_p \sum_{\substack{i \\ f_{2,i} > f_{1,i}}} w_i (pf_{2,i} - f_{1,i})^2.$$

32 Once the distributions are constructed, the tail areas of the true and combined distributions are used to  
33 determine a cutoff score for any desired confidence level. Figure 1 shows the cutoff at an estimated 5% FDR.  
34

35 The mixture modeling approach assumes that the score distribution is a composite of two distributions: one  
36 true and one false. In practice there could be many distinct true and false distributions that represent different  
37 lipid classes, or adduct additions. In LipidLama this is accounted for by applying the algorithm to each of these  
38 classes separately and computing a distinct cutoff for each one. If, for example, PCs are included in a multiple  
39 class search, the top two PC scores for each spectrum will be recorded and used by LipidLama to calculate a  
40 cutoff for that class. Lipids with multiple charge states are similarly dealt with. In the case of  
41 glycerophospholipids, isomers of the correctly identified lipid will often have similar spectra but lower scores  
42 reported by Greazy. This is the scenario for which LipidLama was designed. For cardiolipins there are multiple  
43 ways in which the same set of fatty acyl chains can be configured and for sphingolipids there is only one chain,  
44 which often leads to a single possible identification made by Greazy. The mixture modeling algorithm is not  
45 suitable for either of these situations, thus LipidLama uses a percentage cutoff, specified by the user, to filter  
46 the results for these two classes.  
47  
48

49 **Output.** Greazy produces an intermediary text format that is read by LipidLama. LipidLama presents the final  
50 identification results in the form of a summary mzTab file<sup>20</sup>. The mzTab file consists of a list of identified  
51 precursor ions. For each precursor in this list the scan number, retention time, and score for each spectrum  
52 whose best match was to that precursor are presented. The list is in descending order based on the score of  
53

the best matching spectrum for each lipid. The final set of identified precursors are those that were a best match for at least one spectrum in the data set. Additional information for each identified precursor includes the chemical formula, experimental and expected precursor m/z values, and the type of precursor that was analyzed.

Greazy also produces a series of spectral visualization files for further analysis (Figure 2). Because Greazy's scoring algorithms do not consider relative ion intensities, determining the correct positional isomer for the fatty acyl chains of a glycerophospholipid requires manual inspection. For each spectrum with at least one lipid match, the spectrum is reproduced along with the theoretical spectrum of the top scoring lipid. Peaks that match between the two spectra are colored green. The image also incorporates the fragment descriptions for the theoretical spectrum, along with their m/z values and the intensity of any fragments matched to the experimental spectrum. The precursor description and mass are also given along with the final score. These files are labeled by their spectrum number and are written as SVG embedded html files.

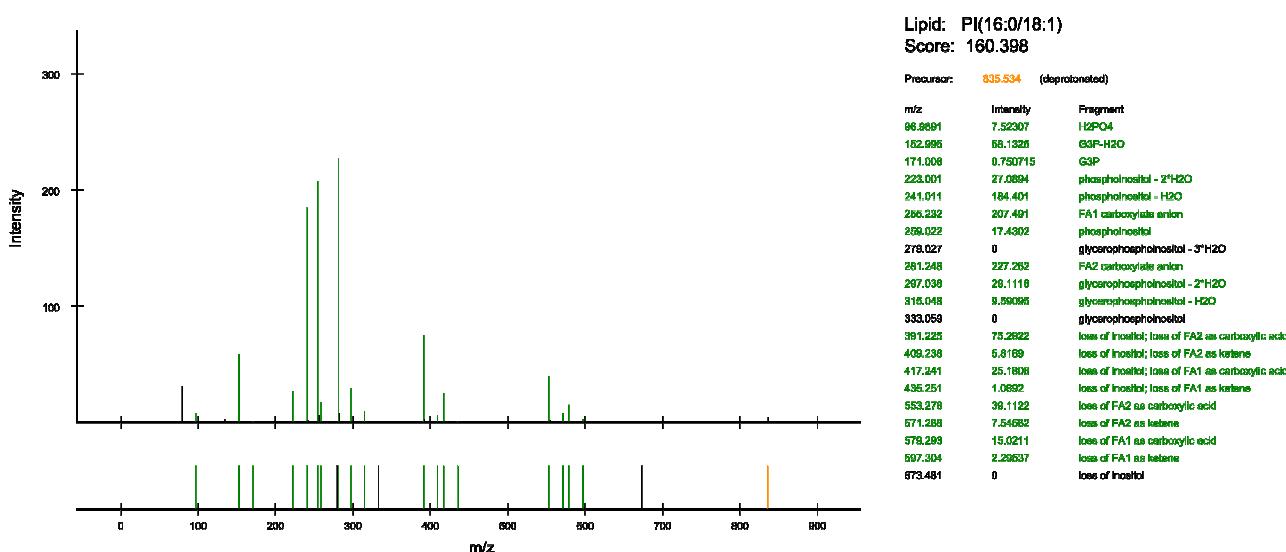


Figure 2. For each spectrum with a match, the spectrum is reproduced (top left) along with the theoretical spectrum of the lipid (bottom left). For each of the peaks in the theoretical lipid spectrum a description of the associated fragment is provided as is the intensity of any of those peaks that match (green). The name of the lipid, the score, and the precursor m/z and description are also given.

## Results and Discussion

Greazy is designed to make high-confidence identifications of phospholipids from high throughput MS/MS data sets. We used the NIST 2014 small molecule MS/MS library as a set of reference spectra with which to validate the fragmentation models and scoring algorithms within Greazy. We then compared the identifications made by Greazy to those of LipidSearch version 4.0.14, a leading commercial lipid identification software platform, in an experimental data set derived from a bone marrow cell line. Lastly, we gauged the software's performance on different platforms with the analysis of data from biological replicates of human alveolar type 2 epithelial cells obtained with an Orbitrap and from technical replicates of human plasma generated on a Q-TOF.

## Characterization of the Search Space

The definition of the search space is the greatest difference between proteomic database search and a lipid search with Greazy. In Greazy the search space is defined by the parameters set by the user and reflects the experimental setup and the lipid composition expected for a given sample. The size of the search space should also be controlled, as it grows exponentially for certain parameters.

Fatty acyl configurations impact the search space size most significantly. Supporting Figure S3a illustrates the effect of increasing the number of possible chain lengths for one, two, and four fatty acyl chains while holding all other parameters constant. The search space for lipids with one fatty acyl chain grows linearly, but the search space for lipids with two chains grows with the square of possible chain lengths. The search space for cardiolipins with their four fatty acyl chains grows at the rate of  $\frac{1}{2}(N^4 + N^2)$  where N is the number of possible fatty acyl chain lengths. This reflects the elimination of identical species that arise due to the symmetry of this class of lipid. The number of double bonds in the fatty acyl chains has a similar impact (Figure S3b). Typical configurations without CL generate search spaces of between 10,000 and 100,000 distinct lipids.

## Validation against the NIST 2014 small molecule MS/MS Library

We sought to validate the fragmentation models and scoring algorithms of Greazy by testing the software against the NIST 2014 small molecule MS/MS library. We filtered the library for spectra that correspond to phospholipids and split the data into two sets based on the polarity of the ions (see Methods). The chosen search space covered the majority of phospholipids known to be in the NIST data set. Searches were performed in both positive and negative ion mode. Glycerophospholipids in the negative mode search included glycerophosphoethanolamines (PE), glycerophosphoglycerols (PG), glycerophosphoserines (PS), glycerophosphates (PA), glycerophosphoinositols (PI), and phosphoinositides (PIP) with all fatty acyl chains ranging from 4 to 24 carbons and 0 to 6 double bonds per chain, allowing lyso-glycerophospholipids as well. Also included were CLs with fatty acyl chains of 16-20 carbons and 0-4 double bonds. In total, the negative mode search included 315,673 distinct lipids. The majority of the glycerophosphocholines (PC) and sphingomyelins (SM) in the NIST database were present as protonated precursors. As such, the search for these lipids was done in positive ion mode. The positive mode search space included the PCs with the same fatty acyl configurations as the glycerophospholipids in negative mode as well as SMs with an 18 carbon sphingosine base, 12-24 fatty acyl carbons and 0-1 double bonds, for a total of 15,032 distinct lipids. The mass accuracy tolerances for these searches were 0.5 Daltons for both precursor and fragment ions because the data included spectra from ion trap, QTOF, and FTMS instruments. Only protonated and deprotonated precursors were included in the search.

*Overall Identification Results.* Of the 214 distinct lipids with MS/MS spectra in at least one of the positive or negative ion data sets, 170 were included within the search space for Greazy. While considering only the best scoring spectrum for each lipid on the list, Greazy correctly identified 157 of these lipids (92%). Across lipid

classes Greazy correctly identified at least 87.5% of the detectable precursors for each class in the search space (Figure 3).

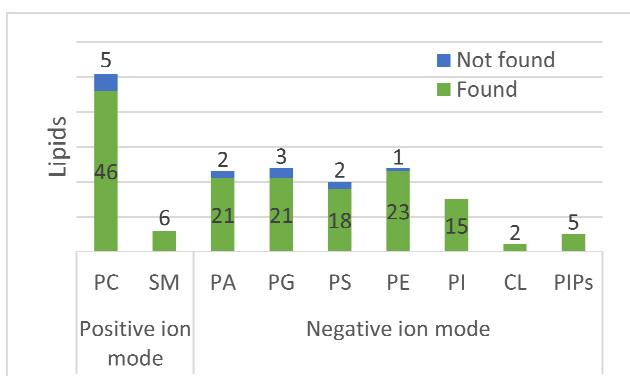


Figure 3. Greazy identified 157 of the 170 lipids in the filtered NIST small molecule MS/MS library.

**Combined and Peak Only Scoring Accuracy.** The final score employed by Greazy to match predicted spectra to experimental ones combines a peak matching score based on the hypergeometric distribution (HGD) with an optional intensity score that takes into account the percentage of the total ion current (TIC) accounted for in those peaks that match. In practice, the HGD produces lower p-values than the TIC and comprises the majority of the final score. The NIST database searches were repeated using only HGD. Figure 4 displays the number of correct and incorrect lipid identifications at different score thresholds. Using only the HGD score increased the number of false compounds identified (by 21.5% and 31.9% in positive and negative ion mode, respectively) while reducing the number of true compounds identified in negative ion mode (by 4.8%). Including the TIC score increased accuracy most notably in negative ion mode.

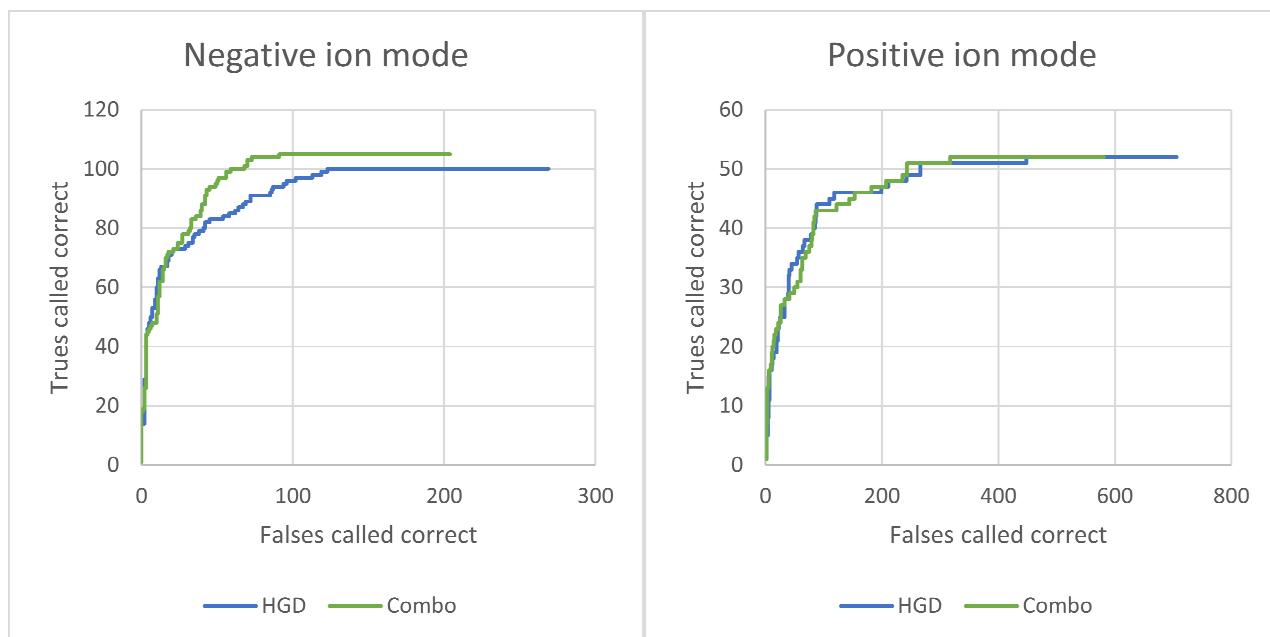


Figure 4. Correct vs incorrect matches for Greazy search and scoring of the NIST library using both the combined and HGD-only scoring algorithms.

*Spectrum Identification Accuracy.* Greazy's final results, used in the above analysis, reflect the best evidence, or highest scoring spectrum, for a given lipid. In this test, however, we evaluate how well Greazy's combined score finds the correct match for each spectrum. Figure 5 shows the rank of the correct matches for identifiable spectra (those with identities in the current search space), where rank 1 hits reflect the spectra that produce the highest score when matched to the correct lipids. The correct matches for 86% of spectra in negative mode and 59% in positive mode were ranked as the top match. In positive mode, a substantial number of spectra produced tie scores for multiple lipids; if the correct identity was tied with other lipids for the top score, it was treated as first rank. Spectra that contain a limited number of fragment ions may match many isomeric species equally well. If, for example, the correct match for a spectrum is PC(16:0/16:0) but only a head group fragment was matched, then a series of lipids (PC(14:0/18:0), PC(15:0/17:0), PC(16:0/16:0), etc.) will have the same score and thus the same rank in this test. Filtering at 5% greatly reduces the number of the identified spectra while increasing the percentage of those with a rank of 1 to 100% for positive mode and 90% for negative mode (Figure S4).

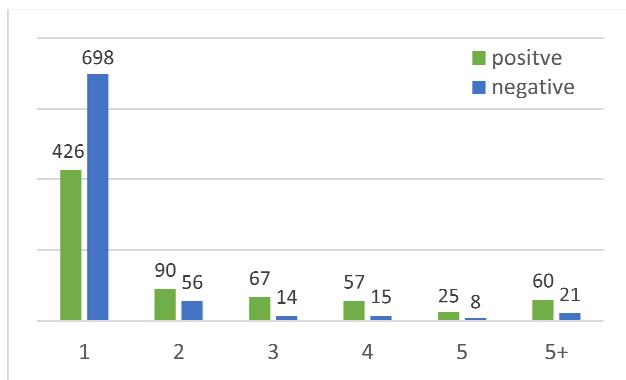


Figure 5. Correct match ranks for spectra in the NIST library whose identity falls within the search space.

*Greazy at Various Energy Levels.* Many of the phospholipids within the NIST metabolomics library were measured at various collision energies in a Thermo Finnigan LTQ Orbitrap Elite by ion trap collision-induced dissociation (CID) and higher-energy collisional dissociation (HCD), allowing us to test Greazy under different energy levels. This analysis can be found in Supporting Text S2.

## Comparison of Greazy/LipidLama to LipidSearch

While the above tests were intended to evaluate the ability of Greazy to match the correct lipid composition to spectra, tools are also needed to determine which spectra have been successfully identified. The LipidLama tool has the same relationship to Greazy that PeptideProphet<sup>21</sup> does to SEQUEST<sup>22</sup> and other database search tools developed for peptide MS/MS data. Because the generation of decoy lipids is infeasible we used the second best spectra scores as a substitute for modeling the false hit distribution. As such, LipidLama can be quite conservative in assessing lipid-spectrum matches (LSMs). LSMs are evaluated separately by lipid class since scores for each may have different distributions.

Greazy/LipidLama and LipidSearch (Thermo Scientific)<sup>16</sup> are both intended to discover the best match of a lipid to explain each MS/MS spectrum. We compared the algorithms using LC-MS/MS data from samples of mouse bone marrow stem cells acquired on a Q Exactive mass spectrometer (see Methods). Two raw files, one for positive mode and one for negative, were analyzed by both tools. The Greazy search space was chosen to capture the majority of the lipids identified by LipidSearch. The search space for the positive and negative runs

were identical. Glycerophospholipids in the search space included the PCs, PEs, PIs, PGs, PSs, and PAs with fatty acyl chains of 12-24 carbons and 0-6 double bonds per chain. Lyso-glycerophospholipids and lipids with an ether-linked hydrocarbon chain at carbon-1 of glycerol were also included. SMs in the search space could incorporate a sphingosine or sphinganine base containing 16-24 carbons and a fatty acyl chain of length 14-28 with 0-4 double bonds. CLs in the search space had fatty acyl chains with 16-20 carbons and 0-4 double bonds. NH<sub>4</sub> and CH<sub>3</sub>COO adducts were allowed in positive and negative mode, respectively. The mass accuracy tolerances for these searches were set to 20 ppm for both precursor and fragment ions. LipidLama was configured for an estimated 10% FDR, while the LipidSearch software was set to an m-score threshold of 5.0 (the m-score measures the fit between experimental and expected product ion spectra and is based on the number of matching fragments).

*Comparison at the spectrum level.* LipidSearch reported LSMs within Greazy's search space for 388 negative ion spectra and 267 positive. In negative ion mode an unfiltered Greazy search returned 2352 identified spectra, agreeing in 297 of the 388 LSMs (76.5%) identified by LipidSearch. In positive ion mode Greazy returned 1569 spectral identifications, agreeing in 132 of the 267 LipidSearch identifications (49.4%). Filtering Greazy's search results with LipidLama returned 586 and 398 LSMs in negative and positive ion mode, respectively. Figure 6a illustrates the overlap between identifications in Greazy/LipidLama and LipidSearch. For those spectra with an identification reported by both platforms, 91.3% of LSMs agreed in negative ion mode and 76.5% in positive. When Greazy's results were filtered to produce the same overall number of identifications as LipidSearch, a similarly small overlap in identified spectra was obtained (Supporting Figure S8). Supporting Figure S9a and S9b separate the spectra by lipid class. In negative mode the median score for those spectra identified only by Greazy was 112.58 while the median for those agreed upon by both was 114.69. In positive mode these medians were 60.52 and 60.68 respectively. There does not appear to be a substantial difference between the two groups of spectra that would explain the small overlap in reported identifications. Algorithmic differences, however, may explain this divergence. For a given lipid, LipidSearch appears to be reporting a subset of spectra for which that lipid was a best match. Of the 413 negative ion and 283 positive ion LSMs reported only by Greazy/LipidLama 278 (67.3%) and 183 (64.7%) matched to lipids that both systems had matched to other spectra. LipidSearch can report identifications based only on an accurate mass and retention time. Greazy, however, requires the presence of MS/MS spectral data for identification; this difference would result in many LipidSearch-only identifications.

*Comparison of lipid identifications.* LipidSearch identified 386 distinct lipids while Greazy/LipidLama yielded 286. They agreed on 187 of these lipids. Figure 6b gives the class breakdown of these results. Agreement between the two platforms is much better for the glycerophospholipids than for the CLs or SMs, whether one is making spectral or lipid identifications. There were several CLs identified by Greazy with the same complement of fatty acyl substituents but in different positions than those identified by LipidSearch. Greazy's fragmentation model for CL is position-specific with regard to the fatty acyl chains while those for glycerophospholipids are not. The added complexity and rigid matching criteria for Greazy may hinder agreement between the two platforms. The SM fragmentation model predicts far fewer fragment ions in these MS/MS spectra, decreasing the discrimination possible for this class. It should also be noted that LipidLama employs an alternative approach for CL and SM; the comparison of the distribution of top match scores and second-best match scores is not viable. Instead, LipidLama was configured to accept the top 10 percent of LSMs by score for these two classes.

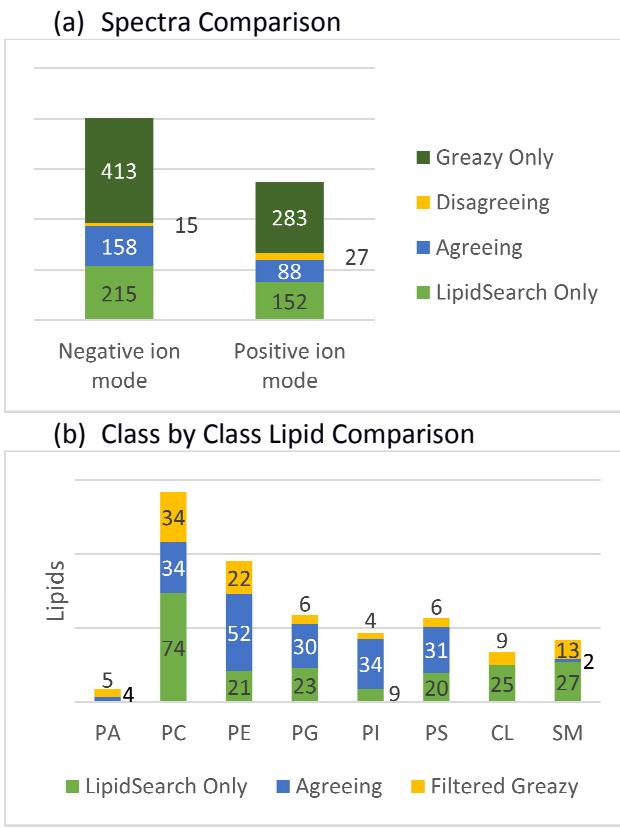


Figure 6. (a) Overlap of spectral identifications made by LipidSearch and Greazy. Greazy's results were filtered by LipidLama at a 10% FDR. (b) Overlap of class specific lipid identifications made by LipidSearch and Greazy.

## Analysis of Biological and Technical Replicates with Greazy and LipidLama

We sought to demonstrate Greazy's utility by analyzing experimental data from different platforms. Data sets included technical replicates from human plasma obtained with a Q-TOF instrument and biological replicates of alveolar type 2 epithelial cells obtained from an orbitrap. Details of the epithelial analysis can be found in Supporting Text S3.

*Analysis of human plasma lipids.* The lipid search space for the human plasma replicates, five each for high and low collision energy, included the glycerophospholipids PC, PE, PI, PG, PS, and PA. Greazy contains fragmentation models for PC in negative ion mode only when acetate adducts are in the search space. The lyso- counterparts to these lipids were included as were those with ether and vinylether linkages. Fatty acyl chain lengths ranged from 12 to 24 carbons (even numbered only), and the number of double bonds allowed was 0-1 at sn-1 and 0-6 at sn-2. A 20 ppm mass accuracy was used for both precursor and fragment ions. LipidLama was applied at a 10% estimated FDR.

Supporting Figure S13 reports the distribution of lipid identifications found in 1, 2, 3, 4, or 5 replicates for the high and low collision energies. A total of 95 lipids were identified in the high energy data sets with 17 (18%) found in all five replicates. Of the 106 lipids identified from the low energy data, 21 (20%) were found in all five replicates. In both sets more than half of the lipids identified were found in only one replicate; this fraction

would be expected to increase as the number of replicates grows. A full breakdown of the identified lipids are in Supporting Tables S14 and S15. The lipid classes most prominently represented were the PCs, PEs, and PIs. All lipids that were found in at least four replicates, in both the high and low energy groups, came from these classes. PGs, PSs, and PAs were also present but with less overlap between the replicates. Lysophospholipids and lipids with ether or vinyl-ether linkages were also represented. An analysis of pairwise identification overlap was also made. For each group of five replicates 10 pairwise comparisons were possible. Boxplots of the overlapping identification counts for these comparisons are given in Supporting Figure S14. The means and variances of these two groups were not significantly different as measured by a two-sample t-test (equal variance) and an F-test. Overall, it appears that the low energy data had a slight advantage in the total number of identifications, the number of lipids common to all five replicates, and the pairwise overlap count.

*Comparison of spectra via clustering.* Greazy/LipidLama identified 802 spectra in total from the 10 replicates, representing 139 lipid-adduct species. 412 of these identified spectra originated from low energy replicates with the remaining 390 coming from high energy data. To gauge the relatedness of the identified spectra the spectra from all ten replicates were clustered using the PRIDE clustering algorithm<sup>42</sup> with a clustering FDR of 10%. For those spectra identified by Greazy this produced 119 distinct clusters of spectra. Of the 139 species identified by Greazy/LipidLama, 66 were represented by a single spectrum; the fact that 66 of the identifiable lipid spectra were singletons in clustering may explain why so many of the identifications were not repeated in other replicates. The number of spectra representing the remaining 73 species ranged from 2 to 42. A comparison of the Greazy identifications with the cluster data showed that 15 of the identified species were split into multiple clusters. In general, the clusters corresponding to each of these species consisted of a single dominant cluster along with one or more minor clusters. For example, the species PE(0:0/18:0)-H was represented by 30 spectra in 9 clusters distributed as (20, 3, 1, 1, 1, 1, 1, 1, 1). A total of 160 spectra were identified as one of these 15 multi-cluster species. Counting only those spectra in the top cluster for each of these species gives 124 spectra. If we assume each cluster represents a single lipid-adduct species and that the top cluster for each species is the correct one, then these 124 of 160, and 766 of all 802 spectra (95.5%) are correctly identified. A complete breakdown of the cluster groupings is in Table S16.

## Conclusions

Greazy and LipidLama are intended to remove a key bottleneck to lipidomic experiments by automating the process of identification from MS/MS data. Their approach borrows heavily from comparable algorithms in proteomics. Even when 100,000 distinct lipids are included in the search space, the algorithm can complete its analysis of an LC-MS/MS experiment in much less time than was required to collect the data at the mass spectrometer. Tests in the gold-standard NIST MS/MS library establish that the software is highly effective in identifying lipids. A relatively low intersection of identifications with the commercial LipidSearch software likely reflects differences in data processing. Nevertheless, agreement was high within the set of spectra for which both algorithms claimed an identification. The release of a quality open-source codebase for lipid identification may be a powerful enabler for laboratories seeking to apply lipidomics to biological problems.

## Associated Content

**Supporting Information Available:** GUI layout; Lipid search space description; Tables of fragmentation patterns; Search space characterization figures; Additional figures and text for the NIST and LipidSearch data; Additional figures, tables, and text for the replicate data; Table and text comparing Greazy to LipidBlast. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## References

- (1) Wenk, M. R. *Nat Rev Drug Discov* **2005**, *4*, 594–610.
- (2) van Meer, G.; Voelker, D. R.; Feigenson, G. W. *Nat. Rev. Mol. Cell Biol.* **2008**, *9*, 112–124.
- (3) Di Paolo, G.; De Camilli, P. *Nature* **2006**, *443*, 651–657.
- (4) Sorensen, C. M.; Ding, J.; Zhang, Q.; Alquier, T.; Zhao, R.; Mueller, P. W.; Smith, R. D.; Metz, T. O. *Clin. Biochem.* **2010**, *43*, 948–956.
- (5) Han, X.; Rozen, S.; Boyle, S. H.; Hellegers, C.; Cheng, H.; Burke, J. R.; Welsh-Bohmer, K. A.; Doraiswamy, P. M.; Kaddurah-Daouk, R. *PLoS ONE* **2011**, *6*, e21643.
- (6) Yan, Y.; Kang, B. *J. Mol. Biol. Res.* **2012**, *2*, 1–11.
- (7) Fernandis, A. Z.; Wenk, M. R. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* **2009**, *87*, 2830–2835.
- (8) Han, X.; Gross, R. W. *Mass Spectrom Rev* **2005**, *24*, 367–412.
- (9) Yang, K.; Cheng, H.; Gross, R. W.; Han, X. *Anal. Chem.* **2009**, *81*, 4356–4368.
- (10) Ejsing, C. S.; Duchoslav, E.; Sampaio, J.; Simons, K.; Bonner, R.; Thiele, C.; Ekroos, K.; Shevchenko, A. *Anal. Chem.* **2006**, *78*, 6202–6214.
- (11) Song, H.; Hsu, F.-F.; Ladenson, J.; Turk, J. *J. Am. Soc. Mass Spectrom.* **2007**, *18*, 1848–1858.
- (12) Herzog, R.; Schwudke, D.; Schuhmann, K.; Sampaio, J. L.; Bornstein, S. R.; Schroeder, M.; Shevchenko, A. *Genome Biol.* **2011**, *12*, R8.
- (13) Schwudke, D.; Oegema, J.; Burton, L.; Entchev, E.; Hannich, J. T.; Ejsing, C. S.; Kurzchalia, T.; Shevchenko, A. *Anal. Chem.* **2006**, *78*, 585–595.
- (14) Hartler, J.; Trötzmüller, M.; Chitraju, C.; Spener, F.; Köfeler, H. C.; Thallinger, G. G. *Bioinformatics* **2011**, *27*, 572–577.
- (15) Pluskal, T.; Castillo, S.; Villar-Briones, A.; Orešič, M. *BMC Bioinformatics* **2010**, *11*, 395.
- (16) Taguchi, R.; Nishijima, M.; Shimizu, T. *Meth. Enzymol.* **2007**, *432*, 185–211.
- (17) Schmelzer, K.; Fahy, E.; Subramaniam, S.; Dennis, E. A. *Meth. Enzymol.* **2007**, *432*, 171–183.
- (18) Kind, T.; Liu, K.-H.; Lee, D. Y.; DeFelice, B.; Meissen, J. K.; Fiehn, O. *Nat. Methods* **2013**, *10*, 755–758.

- 1  
2 (19) Chambers, M. C.; Maclean, B.; Burke, R.; Amodei, D.; Ruderman, D. L.; Neumann, S.; Gatto,  
3 L.; Fischer, B.; Pratt, B.; Egertson, J.; Hoff, K.; Kessner, D.; Tasman, N.; Shulman, N.; Frewen,  
4 B.; Baker, T. A.; Brusniak, M.-Y.; Paulse, C.; Creasy, D.; Flashner, L.; Kani, K.; Moulding, C.;  
5 Seymour, S. L.; Nuwaysir, L. M.; Lefebvre, B.; Kuhlmann, F.; Roark, J.; Rainer, P.; Detlev, S.;  
6 Hemenway, T.; Huhmer, A.; Langridge, J.; Connolly, B.; Chadick, T.; Holly, K.; Eckels, J.;  
7 Deutsch, E. W.; Moritz, R. L.; Katz, J. E.; Agus, D. B.; MacCoss, M.; Tabb, D. L.; Mallick, P.  
8 *Nat. Biotechnol.* **2012**, *30*, 918–920.  
9  
10  
11 (20) Griss, J.; Jones, A. R.; Sachsenberg, T.; Walzer, M.; Gatto, L.; Hartler, J.; Thallinger, G. G.;  
12 Salek, R. M.; Steinbeck, C.; Neuhauser, N.; Cox, J.; Neumann, S.; Fan, J.; Reisinger, F.; Xu, Q.-  
13 W.; Del Toro, N.; Pérez-Riverol, Y.; Ghali, F.; Bandeira, N.; Xenarios, I.; Kohlbacher, O.;  
14 Vizcaíno, J. A.; Hermjakob, H. *Mol. Cell Proteomics* **2014**, *13*, 2765–2775.  
15  
16  
17 (21) Keller, A.; Nesvizhskii, A. I.; Kolker, E.; Aebersold, R. *Anal. Chem.* **2002**, *74*, 5383–5392.  
18  
19 (22) Eng, J. K.; McCormack, A. L.; Yates III, J. R. *J. Am. Soc. Mass Spectrom.* **1994**, *5*, 976–989.  
20  
21  
22 (23) Tabb, D. L.; Vega-Montoto, L.; Rudnick, P. A.; Variyath, A. M.; Ham, A.-J. L.; Bunk, D. M.;  
23 Kilpatrick, L. E.; Billheimer, D. D.; Blackman, R. K.; Cardasis, H. L.; Carr, S. A.; Clauser, K. R.;  
24 Jaffe, J. D.; Kowalski, K. A.; Neubert, T. A.; Regnier, F. E.; Schilling, B.; Tegeler, T. J.; Wang,  
25 M.; Wang, P.; Whiteaker, J. R.; Zimmerman, L. J.; Fisher, S. J.; Gibson, B. W.; Kinsinger, C. R.;  
26 Mesri, M.; Rodriguez, H.; Stein, S. E.; Tempst, P.; Paulovich, A. G.; Liebler, D. C.; Spiegelman,  
27 C. *J. Proteome Res.* **2010**, *9*, 761–776.  
28  
29  
30 (24) Hsu, F.-F.; Turk, J. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* **2009**, *877*, 2673–2695.  
31  
32 (25) Pulfer, M.; Murphy, R. C. *Mass Spectrom Rev* **2003**, *22*, 332–364.  
33  
34 (26) Murphy, R. C.; Fiedler, J.; Hevko, J. *Chem. Rev.* **2001**, *101*, 479–526.  
35  
36 (27) Han, X.; Gross, R. W. *J. Am. Soc. Mass Spectrom.* **1995**, *6*, 1202–1210.  
37  
38  
39 (28) Ho, Y.-P.; Huang, P.-C.; Deng, K.-H. *Rapid Commun. Mass Spectrom.* **2003**, *17*, 114–121.  
40  
41  
42 (29) Hsu, F.-F.; Turk, J. *J. Am. Soc. Mass Spectrom.* **2000**, *11*, 986–999.  
43  
44  
45 (30) Milne, S. B.; Ivanova, P. T.; DeCamp, D.; Hsueh, R. C.; Brown, H. A. *J. Lipid Res.* **2005**, *46*,  
46 1796–1802.  
47  
48 (31) Byrdwell, W. C.; Perry, R. H. *J Chromatogr A* **2006**, *1133*, 149–171.  
49  
50 (32) Hsu, F. F.; Turk, J. *J. Am. Soc. Mass Spectrom.* **2000**, *11*, 437–449.  
51  
52  
53 (33) Hsu, F.-F.; Turk, J.; Zhang, K.; Beverley, S. M. *J. Am. Soc. Mass Spectrom.* **2007**, *18*, 1591–  
54 1604.  
55  
56  
57 (34) Hsu, F.-F.; Turk, J. *J Am Soc Mass Spectrom* **2006**, *17*, 1146–1157.  
58  
59  
60

- 1  
2  
3 (35) Hsu, F.-F.; Turk, J.; Rhoades, E. R.; Russell, D. G.; Shi, Y.; Groisman, E. A. *J. Am. Soc. Mass*  
4 *Spectrom.* **2005**, *16*, 491–504.  
5  
6 (36) Hsu, F.-F.; Turk, J. *J. Am. Soc. Mass Spectrom.* **2006**, *17*, 420–429.  
7  
8 (37) Beckedorf, A. I.; Schäffer, C.; Messner, P.; Peter-Katalinić, J. *J Mass Spectrom* **2002**, *37*, 1086–  
9 1094.  
11  
12 (38) Mosteller, F.; Fisher, R. A. *Am. Stat.* **1948**, *2*, 30–31.  
13  
14 (39) Lei X.; Baozhen S.; Bin M. Determining the False Discovery Rate for Peptide Identification  
15 without a Decoy Database. Presented at the 58st ASMS Conference on Mass Spectrometry  
16 and Allied Topics, Salt Lake City, Utah, May 23–27, 2010; Bioinformatics I - 025.  
17  
18 (40) Choi, H.; Ghosh, D.; Nesvizhskii, A. I. *J. Proteome Res.* **2008**, *7*, 286–292.  
19  
20 (41) SJ Sheather, M. J. *J. R. Stat. Soc. Series B Stat. Methodol.* **1991**, *53*, 683–690.  
21  
22 (42) Griss, J.; Foster, J. M.; Hermjakob, H.; Vizcaíno, J. A. *Nat Meth* **2013**, *10*, 95–96.  
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

## Acknowledgements

This work was supported in part by the NLM Training Grant T15LM007450. The Type 2 epithelial cell data was generated with funding from NIH/NHLBI under the grants HL122642 to Cincinnati Children's Hospital Medical Center and HL122703 to Pacific Northwest National Laboratory. A.I.N. was partially supported by the NIH grants R01GM107148 and U24DK097153. We appreciate access to the NIST 14 MS/MS Library (2014), courtesy of the NIST Mass Spectrometry Data Center. We would also like to thank Tomas O. Metz and Charles Ansong from the Pacific Northwest National Laboratory for their input and contribution of data. Mass spectrometry analyses of the bone marrow stem cells and data processing by LipidSearch were conducted by Dr. Xiaoli Gao in the metabolomics component of the UTHSCSA Mass Spectrometry Laboratory, supported by UTHSCSA and by NIH grant 1S10RR031586-01 (STW). The human serum data was generated with funding the NIH/NIAID grant U01AI115619.

1  
2 For Table of Contents Only  
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

