LipidomeDB data calculation environment: Online processing of direct-infusion mass spectral data for lipid profiles.

Zhenguo Zhou, Shantan Reddy Marepally, Daya Sagar Nune, Prashanth Pallakollu, Gail Ragan, Mary R. Roth, Liangjiang Wang, Gerald H. Lushington, Mahesh Visvanathan, and Ruth Welti

LipidomeDB Data Calculation Environment: Online processing of direct-infusion mass spectral data for lipid profiles

Zhenguo Zhou[1,†], Shantan Reddy Marepally[1,†], Daya Sagar Nune[1], Prashanth Pallakollu[2], Gail Ragan[2], Mary R. Roth[2], Liangjiang Wang[3,‡] , Gerald H. Lushington[1], Mahesh Visvanathan[1,*], and Ruth Welti[2,*]

[1]Bioinformatics Core Facility, University of Kansas, Lawrence, KS 66047
[2]Kansas Lipidomics Research Center, Division of Biology, Kansas State University, Manhattan, KS 66506
[3]Division of Biology, Kansas State University, Manhattan, KS 66506

Running title: LipidomeDB Data Calculation Environment

----

[*]To whom correspondence should be addressed at Ruth Welti, Kansas Lipidomics Research Center, Division of Biology, Ackert Hall. Kansas State University, Manhattan, KS 66506-4901 Phone: 785-532-6241, Fax: 785-532-6653, Email: welti@ksu.edu or Mahesh Visvanathan K-INBRE Bioinformatics Core Facility, 2121 Simons Drive, University of Kansas, Lawrence, KS 66047, Phone: 785-864-3337, Email: mvisvanathan@ku.edu

[†]These authors contributed equally.
[‡]Present address: Department of Genetics and Biochemistry, Clemson University, Clemson, SC 29634

Abbreviations: Data Calculation Environment, DCE; mass/charge, m/z; mass spectrometry, MS; triple quadrupole, QqQ.

Abstract

LipidomeDB Data Calculation Environment (DCE) is a web application to quantify complex lipids by processing data acquired after direct infusion of a lipid-containing biological extract, to which a cocktail of internal standards has been added, into an electrospray source of a triple quadrupole mass spectrometer. LipidomeDB DCE is located on the public internet at http://lipidome.bcf.ku.edu:9000/Lipidomics. LipidomeDB DCE supports targeted analyses; analyte information can be entered or pre-formulated lists of typical plant or animal polar lipid analytes can be selected. LipidomeDB DCE performs isotopic deconvolution and quantification in comparison to internal standard spectral peaks. Multiple precursor or neutral loss spectra from up to 35 samples may be processed simultaneously with data input as Excel files and output as tables viewable on the web and exportable in Excel. The pre-formulated compound lists and web access, used with direct-infusion mass spectrometry, provide a simple approach to lipidomic analysis, particularly for new users.

Lipidomics is a developing strategy for functional genomics. Characterizing large numbers of mutants or the results of experimental manipulations applied to large numbers of samples is becoming increasingly feasible. Triple quadrupole (QqQ) mass spectrometers are robust and widely available instruments, and their ability to isolate a signal stemming from the subset of a sample's molecules that produce a specific pair of mass/charge (*m/z*) values provides detection specificity for lipidomic analysis. Direct-infusion QqQ mass spectrometry (MS), using precursor and neutral loss scans, as originally proposed by Brügger et al. [1], offers many advantages as a lipidomics platform. Direct infusion, in comparison to introducing samples through a liquid chromatography interface, is more easily implemented and allows for the rapid analysis of many samples. Another, less recognized advantage of direct-infusion is that the continuous influx of sample allows the user to adjust scanning time to improve signal detection, without concern for elution time. However, a potential bottleneck in the implementation of any lipidomic strategy is the ability to process the data into quantitative lipid profiles. In particular, new users need a simple approach, and many users need to be able to process large datasets.

A number of tools for analysis of mass spectral data from lipids have been developed. Tools designed for direct-infusion QqQ MS analysis include the algorithm developed by Kurvinen et al. [2], LIMSA [3,4], the method developed by Ejsing et al. [5], LipidQA [6] and AMDMS-SL [7]. Each tool is capable of lipid quantification from direct infusion QqQ mass spectral data, and each has special features. Some have the ability to combine data from multiple spectra for identification and analysis of lipids at the true molecular species level [5-7]. Some provide for visualization and statistical analysis of the output data [2-5].

Here we describe a new tool for mass spectral data analysis for lipidomics. The strength of LipidomeDB Data Calculation Environment (DCE) is its ability to rapidly and simultaneously analyze multiple spectra from up to 35 samples input in one step. It differs from currently available tools in being on the web. It provides pre-formulated lists of target lipid analytes, including lists for plant lipids, and can accept user-defined target analyte lists.

Materials and Methods

Programming: LipidomeDB DCE has a web interface and a database backend (called LipidomeDB) and scripts to read and perform calculations on input data provided in Excel format. LipidomeDB is a MySQL database that stores the defined lipid species and formulas

and the users' experimental data. Much of the LipidomeDB DCE is coded in JSP. The user interface is implemented primarily in JSP with Javascript used to perform client-side computations, such as those related to isotopic abundances, and with Java used for server-side data analysis, such as searching the database for target lipid species matching uploaded data with the target species lists, and isotopic deconvolution.

Input spectral data: The input data are spectral peak lists acquired by precursor or neutral loss scanning, typically using a triple quadrupole mass spectrometer in the multiple channel analyzer (MCA) mode (a mode that averages or sums the signal from multiple scans to produce a spectrum). Data processing, i.e., baseline subtraction, smoothing, and centroiding (or integration), are performed with the mass spectrometer's acquisition software. Spectra, each as a list of *m/z* vs. signal, are exported from the acquisition program to Excel files for upload to LipidomeDB DCE.   AB Sciex programmers have written a script, called "MultiplePeriodProcessing", which exports data in the proper format; however, the format may be replicated manually or through other scripts. The format of the Excel spreadsheets for upload is specified in the documentation accompanying LipidomeDB DCE and example files are provided.

Algorithms and functions: Algorithms and functions are described briefly below. Full details are provided in the documentation (tutorial) for LipidomeDB DCE accessed from its home page: http://lipidome.bcf.ku.edu:9000/Lipidomics

Target lipid information: The mass of target lipids and abundances of their isotopic variants are calculated from the chemical formula by adding the masses of the component atoms. The adduct formed by the target lipids is specified by the user. Possible adduct choices, in the negative mode, are $[M - H]^-$, $[M - CH_3]^-$, $[M + Cl]^-$, and $[M + C_2H_3O_2]^-$ (i.e., $[M + acetate]^-$), and in the positive mode, $[M + H]^+$, $[M + NH_4]^+$, $[M + Li]^+$, $[M + Na]^+$, and $[M + K]^+$.

Lipid search: The peaks corresponding to specified adduct ions of the internal standards and target lipids are located in the input data by *m/z* and the value of each corresponding signal is collected. Any mass that is within the specified mass tolerance window (specified compound mass ± the mass tolerance) is considered a candidate mass for the specified target lipid or internal standard. The search algorithm provides three options, in case multiple masses are found within the mass tolerance window. The three options are: "Nearest Mass", which selects

4

the *m/z* closest to the lipid ion's *m/z*, "Highest Signal", which selects the *m/z* having the highest corresponding signal value, and "Sum of Signals", which sums signals within the specified mass tolerance target window.

Isotopic deconvolution (correction) of signal and lipid analyte quantification: Quantification is essentially the method described by Brügger et al. [1] and used in the LIMSA analytical tool [3,4]. Isotopic deconvolution to remove signals due to "A+1" and "A+2" isotopic variants (in LipidomeDB DCE, called a_1 and a_2), in order to detect the "A" peaks of other compounds at the same nominal mass, is performed in a manner similar to the "subtraction" method in the LIMSA analytical tool [3].

After the corrected signals are established by deconvolution, the amount of each target lipid is calculated. The amount is calculated from the corrected signals for the target lipids, the signals for 1-3 internal standards, and the corresponding internal standard amounts specified by the user. A linear calibration curve (signal vs. *m/z*) fits the signals of the internal standards [3] and is used to correct for mass-dependent variation in instrument response. If any internal standard is not detected, an error message appears on an output page for the individual spectrum (i.e., the spectrum for one target list and one sample); if this error is present, the target analytes are not quantified in the sample with the missing standards. If the internal standards are detected, data calculations are shown on the output page for the individual spectrum.

Results and Discussion

The primary function of LipidomeDB DCE is to search for *m/z* and signal for entered target lipid species in integrated (centroided) mass spectra uploaded by the user, to isotopically deconvolute the signals, and to quantify the signals in comparison to internal standard signals and amounts, in order to calculate the amount of each target lipid species in the samples. LipidomeDB DCE has three phases, shown in Figure 1, data collection, data analysis, and result output and collection.

Getting started. To use LipidomeDB DCE, a login ID can be acquired by following instructions on the home page. A "tutorial" (i.e., documentation), accessible from the home page includes complete instructions for the use of LipidomeDB DCE. The instructions include brief directions for performing the lipidomic experiments, including suggestions for sample preparation, internal

standards and their addition, acquisition of the mass spectral data, and formatting of the mass spectral data for input into LipidomeDB DCE. The intent of this part of the tutorial is to simplify and describe the process of performing lipidomic analyses. Also available, on the home page and by following links in the tutorial, are example input data files, acquired on mouse intestinal lipid extracts, and the output file that should be obtained when the data are processed according to the instructions, allowing users to test the interface function and verify correct usage. Step-by-step directions for using LipidomeDB DCE (as shown in Table 1) with screenshots are provided in the tutorial. Also in the tutorial are sections discussing normalization to tissue metrics and practical considerations about the limitations of the quantitative capabilities of direct-infusion electrospray ionization QqQ MS-based analysis.

Input data. The input to LipidomeDB consists of lists of target lipids, i.e. lipid species being analyzed, lists of internal standards utilized, internal standard amounts, and the mass and signal lists from precursor or neutral loss scan(s) (Figure 1). Either pre-formulated lipid species target lists, user-specified lipid target lists, or a combination of the two types of lists can be used. If the user composes his own sheets, they can be saved for his later use. The pre-formulated target lists for "plants" and "animals" are based on lipid species detected by the Kansas Lipidomics Research Center's analytical laboratory by electrospray ionization QqQ MS in a number of higher plant and mammalian species. Figure 2 is a screenshot of the interface for indicating the target lipids and the corresponding internal standards. Similar pages are used for each group of lipids to be analyzed (i.e. for each precursor or neutral loss scan). The pre-formulated lists provide a suitable and quick starting point for many users. Amounts of internal standards are entered, and mass spectral data are uploaded in Excel files. The interface for uploading the mass spectral data is shown in Figure 3. LipidomeDB DCE searches the *m/z* of mass spectral data for the *m/z* values of lipids and internal standards in the target lists.

Data calculation and output. Data calculation is described briefly in the Materials and Methods section, and in more detail in the tutorial. The output of LipidomeDB DCE can be viewed on (1) a sheet showing calculations for one target list for one sample, (2) as a summary showing calculated data for one target compound set from all samples, or (3) as a summary of data for all targeted compound sets from all samples (Figure 4). The sheet showing calculations for one target list for one sample (i.e., data from one spectrum) allows the user to confirm detection of the internal standards, to view the corrections for isotopic variants (isotopic deconvolution

6

results), and to view the results of the quantification. The summary sheets provide the data in several convenient forms. All sheets can be viewed on the web or exported in Excel format.

Summary. LipidomeDB DCE is a well-documented, web application for analysis of direct infusion electrospray ionization QqQ mass spectral data, with capability to analyze multiple scans from up to thirty five samples simultaneously. The pre-formulated lists of plant and animal polar lipids provide a quick start for lipidomic data analysis that should provide for ease in performing lipidomics analyses. In the long term, the web-accessible lipidomics data calculation system and database can be developed for storage and retrieval of lipidomics data in the context of experimental metadata.

Acknowledgments

References

1. Brügger B, Erben G, Sandhoff R, Wieland FT, Lehmann WD (1997) Quantitative analysis of biological membrane lipids at the low picomole level by nano-electrospray ionization tandem mass spectrometry. Proc Natl Acad Sci USA 94 : 2339-2344.

2. Kurvinen J-P, Aaltonen J, Kuksis A, Kallio H (2002) Software algorithm for automatic interpretation of mass spectra of glycerolipids. Rapid Commun Mass Spectrom 16 : 1812-1820.

3. Haimi P, Uphoff A, Hermansson M, Somerharju P (2006) Software tools for analysis of mass spectrometric lipidome data. Anal Chem 78 : 8324-8331.

4. Haimi P, Chaithanya K, Kainu V, Hermansson M, Somerharju P (2009) Instrument-independent software tools for the analysis of MS-MS and LC-MS lipidomics data. In Armstrong D (ed) Lipidomics, Methods in Molecular Biology, Vol 580, Humana Press, Totowa, New Jersey, pp 285–294.

5. Ejsing CS, Duchoslav E, Sampaio J, Simons K, Bonner R, Thiele C, Ekroos K, Shevchenko A (2006) Automated identification and quantification of glycerolipid molecular species by multiple precursor scanning. Anal Chem 76 : 6202-6214.

6. Song H, Hsu F-F, Ladenson J, Turk J (2007) Algorithm for processing raw mass spectrometric data to identify and quantitate complex lipid molecular species in mixtures by data-dependent scanning and fragment ion database searching. J Am Soc Mass Spectrom 18 :1848-1858.

7. Yang K, Cheng H, Gross RW, Han X (2009) Automated lipid identification and quantification by multi-dimensional mass spectrometry-based shotgun lipidomics. Anal Chem 81 : 4356-4368.

Figure Legends

Figure 1. Scheme indicating the workflow of the system and relationships of the three phases of LipidomeDB DCE: data collection, data analysis, and result output and collection.

Figure 2. Screenshot of target compound input sheet. The screenshot shows the pre-formulated list for "Animal LPC".

Figure 3. Screenshot of spectral files upload sheet.

Figure 4. Screenshot of summary. A summary of data for all targeted compound sets from all samples can be exported as an Excel file.

Table 1: Categories of step-by-step directions in the LipidomeDB DCE tutorial

Step 1. Assemble the data to be entered
Step 2. User log-in
Step 3. Enter data about the samples
Step 4. Enter the first target compound set
      a. Enter a target compound set from an existing (pre-formulated) list
      b. Enter your own target compound set
Step 5. Enter any additional target compound sets
Step 6. Provide data about internal standard amounts
Step 7. Upload the input data generated from the experiments in Excel format
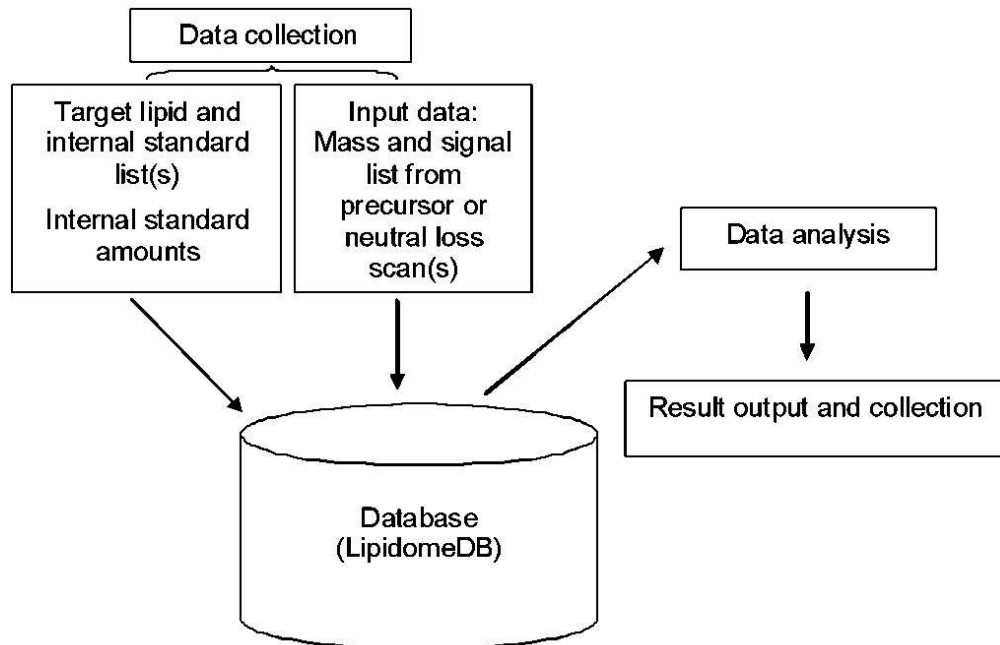Step 8. Collect the results

Figure 1. Scheme indicating the workflow of the system and relationships of the three phases of LipidomeDB DCE: data collection, data analysis, and result output and collection.
86x56mm (300 x 300 DPI)

Animal LPC

Internal Standards

Mass of Detected Ion	Chemical Formula of M	Compound Name	Adduct	a_1	a_2
454.28	C21H44O7PN	LPC(13:0)	[M + H] ⌄	0.242(	0.022
538.37	C27H56O7PN	LPC(19:0)	[M + H] ⌄	0.309:	0.036
			⌄		
Fragment Formula	C5H15O4PN ⌄	phosphocholine (184+) ⌄			

Target Compounds

Lipid No.	Mass of Detected Ion	Chemical Formula of M	Compound Name	Adduct	a_1	a_2
0	494.31	C24H48O7PN	LPC(16:1)	[M + H] ⌄	0.275	0.028
1	496.33	C24H50O7PN	LPC(16:0)	[M + H] ⌄	0.275(	0.028
2	518.31	C26H48O7PN	LPC(18:3)	[M + H] ⌄	0.297	0.033
3	520.33	C26H50O7PN	LPC(18:2)	[M + H] ⌄	0.297(	0.033
4	522.34	C26H52O7PN	LPC(18:1)	[M + H] ⌄	0.297:	0.033
5	524.36	C26H54O7PN	LPC(18:0)	[M + H] ⌄	0.298(	0.033
6	542.31	C28H48O7PN	LPC(20:5)	[M + H] ⌄	0.319	0.039:
7	544.33	C28H50O7PN	LPC(20:4)	[M + H] ⌄	0.319(	0.039:

Figure 2. Screenshot of target compound input sheet. The screenshot shows the pre-formulated list for "Animal LPC".
86x82mm (300 x 300 DPI)

Spectral Files Upload

Upload Target Compound Set: **Animal LPC**   C:\Documents and Setti [Browse...]

Upload Target Compound Set: **Animal PC-ePC-SM**   C:\Documents and Setti [Browse...]

Upload Target Compound Set: **Animal LPE**   C:\Documents and Setti [Browse...]

Upload Target Compound Set: **Animal PE-ePE-PECer** C:\Documents and Setti [Browse...]

Upload Target Compound Set: **Animal PA**   C:\Documents and Setti [Browse...]

Upload Target Compound Set: **Animal PG**   C:\Documents and Setti [Browse...]

Upload Target Compound Set: **Animal PI**   C:\Documents and Setti [Browse...]

Upload Target Compound Set: **Animal PS**   C:\Documents and Setti [Browse...]

[Continue] [Clear All]

Figure 3. Screenshot of spectral files upload sheet.
86x55mm (300 x 300 DPI)

	A	B	C	D	E	F	G	H	I
1									
2									
3	Mass	Compound Forr	Compound Na	std1	mouse 1	mouse 2	mouse 3	mouse 4	mouse 5
4	494.3	C24H48O7PN	LPC(16:1)	0.0000	0.0030	0.0021	0.0006	0.0056	0.0018
5	496.3	C24H50O7PN	LPC(16:0)	0.0003	0.3276	0.2801	0.0798	0.4288	0.2339
6	518.3	C26H48O7PN	LPC(18:3)	0.0000	0.0010	0.0004	0.0002	0.0015	0.0008
7	520.3	C26H50O7PN	LPC(18:2)	0.0001	0.0548	0.0406	0.0174	0.0550	0.0294
8	522.3	C26H52O7PN	LPC(18:1)	0.0000	0.0372	0.0310	0.0100	0.0448	0.0290
9	524.4	C26H54O7PN	LPC(18:0)	0.0010	0.1933	0.1831	0.0608	0.2633	0.2265
10	542.3	C28H48O7PN	LPC(20:5)	0.0001	0.0026	0.0011	0.0008	0.0024	0.0010
11	544.3	C28H50O7PN	LPC(20:4)	0.0000	0.0099	0.0051	0.0027	0.0088	0.0055
12	546.3	C28H52O7PN	LPC(20:3)	0.0000	0.0021	0.0011	0.0003	0.0022	0.0017
13	548.4	C28H54O7PN	LPC(20:2)	0.0000	0.0011	0.0010	0.0002	0.0009	0.0010
14	550.4	C28H56O7PN	LPC(20:1)	0.0000	0.0028	0.0019	0.0006	0.0038	0.0024
15	552.4	C28H58O7PN	LPC(20:0)	0.0001	0.0009	0.0017	0.0003	0.0029	0.0019
16	568.3	C30H50O7PN	LPC(22:6)	0.0000	0.0068	0.0037	0.0012	0.0055	0.0043
17	570.3	C30H52O7PN	LPC(22:5)	0.0000	0.0002	0.0003	0.0000	0.0004	0.0000
18									
19	Mass	Compound Forr	Compound Na	std1	mouse 1	mouse 2	mouse 3	mouse 4	mouse 5
20	676.5	C36H70O8PN	PC(28:1)	0.0000	0.0010	0.0013	0.0005	0.0012	0.0005
21	701.5	C39H77N2O6P	SM(16:1)	0.0000	0.0045	0.0031	0.0040	0.0036	0.0039
22	703.6	C39H79N2O6P	SM(16:0)	0.0000	0.2742	0.2091	0.2805	0.2352	0.2313
23	704.5	C38H74O8PN	PC(30:1)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
24	705.6	C39H81N2O6P	DSM(16:0)	0.0000	0.0253	0.0186	0.0287	0.0198	0.0183
25	706.5	C38H76O8PN	PC(30:0)	0.0000	0.0198	0.0155	0.0195	0.0164	0.0201
26	714.5	C40H76O7PN	ePC(32:3)	0.0000	0.0003	0.0005	0.0006	0.0007	0.0005
27	716.6	C40H78O7PN	ePC(32:2)	0.0000	0.0014	0.0017	0.0014	0.0025	0.0010
28	718.6	C40H80O7PN	ePC(32:1)	0.0000	0.0072	0.0080	0.0095	0.0082	0.0078
29	720.6	C40H82O7PN	ePC(32:0)	0.0000	0.0167	0.0179	0.0242	0.0215	0.0243
30	729.6	C41H81N2O6P	SM(18:1)	0.0000	0.0033	0.0029	0.0029	0.0032	0.0039
31	730.5	C40H76O8PN	PC(32:2)	0.0000	0.0072	0.0073	0.0099	0.0077	0.0087
32	731.6	C41H83N2O6P	SM(18:0)	0.0000	0.0476	0.0323	0.0490	0.0336	0.0442
33	732.5	C40H78O8PN	PC(32:1)	0.0000	0.0408	0.0426	0.0579	0.0513	0.0558
34	733.6	C41H85N2O6P	DSM(18:0)	0.0000	0.0024	0.0000	0.0000	0.0000	0.0001
35	734.6	C40H80O8PN	PC(32:0)	0.0003	0.1719	0.1610	0.2222	0.1711	0.1973
36	740.6	C42H78O7PN	ePC(34:4)	0.0000	0.0007	0.0009	0.0008	0.0012	0.0004
37	742.6	C42H80O7PN	ePC(34:3)	0.0000	0.0035	0.0040	0.0039	0.0036	0.0046
38	744.6	C42H82O7PN	ePC(34:2)	0.0000	0.0261	0.0258	0.0276	0.0282	0.0251
39	746.6	C42H84O7PN	ePC(34:1)	0.0000	0.0239	0.0273	0.0307	0.0288	0.0313
40	748.6	C42H86O7PN	ePC(34:0)	0.0000	0.0109	0.0123	0.0115	0.0110	0.0149
41	754.5	C42H76O8PN	PC(34:4)	0.0000	0.0041	0.0038	0.0043	0.0045	0.0034
42	756.5	C42H78O8PN	PC(34:3)	0.0001	0.0585	0.0600	0.0693	0.0644	0.0592
43	758.6	C42H80O8PN	PC(34:2)	0.0002	2.4140	1.9600	2.3389	1.9092	1.4283
44	760.6	C42H82O8PN	PC(34:1)	0.0000	0.3872	0.3884	0.4623	0.4166	0.4690
45	762.6	C42H84O8PN	PC(34:0)	0.0000	0.0450	0.0355	0.0495	0.0381	0.0544
46	766.6	C44H80O7PN	ePC(36:5)	0.0000	0.0125	0.0103	0.0122	0.0111	0.0136
47	768.6	C44H82O7PN	ePC(36:4)	0.0000	0.0234	0.0214	0.0237	0.0225	0.0206
48	770.6	C44H84O7PN	ePC(36:3)	0.0000	0.0122	0.0129	0.0161	0.0134	0.0135
49	772.6	C44H86O7PN	ePC(36:2)	0.0000	0.0481	0.0643	0.0647	0.0532	0.0495
50	774.6	C44H88O7PN	ePC(36:1)	0.0000	0.0192	0.0220	0.0224	0.0225	0.0267

new sheet

Figure 4. Screenshot of summary. A summary of data for all targeted compound sets from all samples can be exported as an Excel file.
86x114mm (300 x 300 DPI)