A MODULAR OPEN SOURCE WORKFLOW FOR LIPIDOMICS USING LIQUID CHROMATOGRAPHY HIGH RESOLUTION TANDEM MASS SPECTROMETRY (UHPLC-HRMS/MS)

By

JEREMY PAUL KOELMEL

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UFNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2017

To my family, mentors, and you, the reader

ACKNOWLEDGMENTS

I first and foremost would like to thank the team of exceptionally talented undergraduates, Berkley Olsen, Jason Cochran, Jordan Zeldin, Nicholas Azcarate, and Nicholas Kroeger, whose dedication and enthusiasm for research is contagious. I specifically would like to thank Nicholas Kroeger, who recoded most of my LipidMatch script, added essential features, and coded the iterative-exclusion script. Here, I would also like to thank Timothy Garrett. The synchronicity of both of us coming up with the idea of iterative-exclusion separately within a few weeks, and his excitement and interest to implement the strategy, catalyzed this work. I also express my appreciation for Jason Cochran, who wrote the code for LipidMatch Quant and code for conversion of MS-DIAL libraries into LipidMatch format. I owe much of my success during the course of my PhD to Jason Cochran's and Nicholas Kroeger's countless hours of work and patience as I continued to add new features and changed old ones. I also appreciate the work of Dr. Yang Li, who combined the entire lipidomics workflow that I developed with my undergraduates (from file conversion to lipid annotation), into a single windows application.

Science is built on the shoulders of giants, and I appreciate the wider lipidomics community for their support. I am grateful to the development team of MS-DIAL, specifically Hiroshi Tsugawa and the LipidBlast development team, for providing open source libraries implemented in LipidMatch. I am grateful to Tamil Selvan, who provided data files for development of oxidized lipid libraries. I would like to acknowledge Christopher Beecher for his help with writing R code to run non-negative matrix factorization (NMF) and interpreting NMF analysis and Lauren McIntyre for help designing the univariate statistical analyses employed in my workflow.

I am inspired by Dr. Richard Yost's mentoring philosophy. I flourished in the freedom he granted me to follow my passion, as well as the resources and superb scientists he connected me with to make my passion a reality. Specifically, Dr. John Bowden was instrumental in my success at University of Florida. Most of my ideas for my PhD were synthesized in late night conversations with John. In addition, John trained me with hands-on experiences in all of my laboratory work. He also worked side-by-side with me during sample preparation, and we processed hundreds of samples together, for which I am very grateful for his time and support. Yost group alumni Dr. Candice Ulmer and Dr. Rainey Patterson, also played an important role in training me to use the Q-Extractive orbitrap used in this dissertation and designed the chromatography method used throughout. All four publications which are the backbone of this dissertation were executed in collaboration with Candice Ulmer, and her work ethic and efficiency are contagious. Other collaborators who deserve specific thanks are Dr. Christina Jones, Emily Gill, and Dr. Tracey Schock, as well as all the Yost group members, who would spend over 3 hours at a time giving in-depth feedback on my oral presentations, posters, and manuscript figures.

This research was done in collaboration between Core 1 and Core 3 of the Southeast Center for Metabolomics (SECIM) < http://secim.ufl.edu/> (NIH Grant #U24 DK097209). I would like to thank Kaitlin Brennan from the SECIM administration core for implementing the website which hosts all the lipidomics software I developed during my PhD <http://secim.ufl.edu/secim-tools/>.

In addition to those who directly supported me in my research, I would also like to thank the scientists and teachers who encouraged me to take the path of analytical

chemistry, specifically Dr. Dulasiri Amarasiriwardena and Dr. M. N. V. Prasad. Their selfless devotion to their students is admirable.

Had it not been for my family and friends, I would be a tense mess by the end of this PhD. My mother, father, sister, partner Harmony Miller, and friends made sure that during my PhD I lived a life of not just working hard, but playing hard. I also express gratitude to my spiritual mentors, who support me in knowing myself and purifying my heart and mind, which allowed me to have even more clarity, focus, and creativity to bring to my research. Specific mentors are Dr. David Wolfe and Marie Glasheen of Satvatove Institute, Mickey Singer of the Temple of the Universe, and S. N. Goenka, teacher of vipassana meditation, who all selflessly serve for the liberation of people from misery to happiness. Finally, I thank those who hosted spiritual programs, helped me relax and connect to myself and others, including Amber and Eliot Larkin of Friday Night Meditations, and Matt Chandler of Aurora.

TABLE OF CONTENTS

page

LIST OF TABLES

LIST OF FIGURES

LIST OF OBJECTS

LIST OF ABBREVIATIONS

AGC	Automated gain control
AIF	All ion fragmentation
ANOVA	Analysis of variance
APCI	Atmospheric pressure chemical ionization
APPI	Atmospheric pressure photoionization
BEH	Ethylene bridged hybrid
BFF	Blank feature filtering
CcO	Cytochrome c oxidase
CE	Cholesterol ester
Cer	Ceramide
CID	Collision-induced dissociation
CLA	Conjugated linoleic acid
Csv	Comma separated values
D	Heat treated samples (with denator device)
D1hr	Heat treated samples (with denator device), and samples sat on ice for one hour prior to extraction
Da	Dalton
DC	Direct current
DDA	Data-dependent analysis
ddMS2-topN	Data-dependent top N
DG	Diacylglyceride
DIA	Data-independent analysis
EDTA	Ethylenediaminetetraacetic acid
EIC	Extracted ion chromatogram

ELIFE	European Lipidomics Initiative: shaping the life sciences
ESI	Electrospray ionization
ether-LPC	Plasmenyl and/or plasmanyl lysophosphatidylcholine
ether-LPE	Plasmenyl and/or plasmanyl lysophosphatidylethanolamine
ether-PC	Plasmenyl and/or plasmanyl phosphatidylcholine
ether-PE	Plasmenyl and/or plasmanyl Phosphatidylethanolamine
ether-TG	Plasmenyl and/or plasmanyl triglyceride
FA	fatty acid
FDR	False discovery rate
FT-ICR	Fourier-transform ion cyclotron resonance
FWHM	Full width at half maximum
GalCer	Galactosylceramide
GC	Gas chromatography
Glog	Generalized log transformation
GPL	Glycerophospholipid
HCD	Higher energy collision-induced dissociation
HESI	Heated electrospray ionization
HF	Ultra-high field orbitrap
HILIC	Hydrophilic interaction chromatography
HMDB	Human Metabolome Database
HPLC	High-performance liquid chromatography
HRMS	High-resolution mass spectrometry
IACUC	The Institutional Animal Care and Use Committee
IE	Iterative exclusion
IM	Ion mobility

InChI	International Chemical Identifier
IS	Internal standard
KEGG	Kyoto Encyclopedia of Genes and Genomes
LC	Liquid chromatography
LDA	Lipid Data Analyzer
LIPID MAPS	Lipid Metabolites and Pathways Strategy
LMQ	LipidMatch Quant
LPC	Lysophosphatidylcholine
LPE	Lysophosphatidylethanolamine
LPL	Lysoglycerophospholipid
LPS	Lysophosphatidylserine
LTQ	Linear ion trap
m/z	Mass-to-charge ratio
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
N	Flash frozen samples (not heat treated)
N1hr	Flash frozen samples (not heat treated), and samples sat on ice for one hour prior to extraction
NCE	Normalized collision energy
NIST	National Institute of Standards and Technology
NL	Neutral loss
OxLPC	Oxidized lysophosphatidylcholines
OxPC	Oxidized phosphatidylcholine
OxPE	Oxidized phosphatidylethanolamine
OxTG	Oxidized triglyceride

OzID	Ozone-induced dissociation
PA	phosphatidic acid
PC	Phosphatidylcholine
PC1	Principle component 1
PC2	Principle component 2
PCA	Principle components analysis
PE	Phosphatidylethanolamine
PG	Phosphatidylglycerol
PI	Phosphatidylinositol
PLA	Phospholipase A
PLC	Phospholipase C
PLD	Phospholipase D
PMe	Phosphatidylmethanol
ppm	Parts per million
PS	Phosphatidylserine
qTOF	Quadrupole time-of-flight
RC	Red Cross
RF	Radio frequency
rpm	Revolutions per minute
RSD	Residual standard deviation (also termed coefficient of variation (CV))
SECIM	Southeast Center for Integrated Metabolomics
SLIM	Structures for lossless ion manipulation
SM	Sphingomyelin
SRM	Standard reference material

t_0	Time of collection
TG	Triglyceride
UHPLC	Ultra-high-performance liquid chromatography
UV	Ultraviolet
v:v	Volume to volume
v:v:v	Volume to volume to volume

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

A MODULAR OPEN SOURCE WORKFLOW FOR LIPIDOMICS USING LIQUID CHROMATOGRAPHY HIGH RESOLUTION TANDEM MASS SPECTROMETRY (UHPLC-HRMS/MS)

By

Jeremy Paul Koelmel

December 2017

Chair: Richard A. Yost
Major: Chemistry

Lipidomics, the comprehensive measurement of lipids within a biological system or substrate, is an emerging field with significant potential for improving clinical diagnosis and our understanding of health and disease. The clinical utility of lipidomics is derived from the diverse biological roles of lipids. This diversity in lipid function is enabled by the diversity of lipid structures and the wide dynamic range of lipid concentrations, thus making the execution of lipidomic experiments analytically challenging. Advances in ultra-high performance liquid chromatography and high resolution tandem mass spectrometry (UHPLC-HRMS/MS) enable the acquisition of data covering a large portion of the lipidome in a single experiment.

Data processing tools for mining this information have only recently emerged, and strategies are still needed to more accurately and comprehensively annotate and quantify lipid ions from UHPLC-HRMS/MS data. Therefore, we developed a suite of open source lipidomics software tools and novel strategies which cover the majority of the lipidomics workflow, including: increasing lipid fragmentation coverage, improving peak picking, removing background ions, combining redundant ion information,

annotating lipid species, and quantifying lipid species using class specific lipid standards.

For improved lipid coverage, we introduce LipidMatch, a lipid identification software leveraging the largest open source *in-silico* fragmentation library. LipidMatch covers over 250,000 lipid species, spanning over 56 lipid types. To further increase lipid coverage, we employ an R script tool (termed iterative exclusion) for generating exclusion lists from ions selected for fragmentation in previous injections, resulting in 69 % more lipid identifications in positive mode analysis of human plasma. For improved data-quality, we developed a tool based on an algorithm for removing peaks present in extraction blanks, and a unique peak picking strategy using MZmine which reduced the residual standard deviation between samples. For automated relative quantification of the lipidome, we developed another tool LipidMatch Quant, which matches standards to analytes based on lipid class, adduct, and retention time; hence accounting for both lipid structural and regional chromatographic effects on ionization efficiency.

Through increased lipid coverage and more accurate measurements, these tools can increase the probability of lipid biomarker discovery and understanding disease etiology. Parameters for each tool allow the user to optimize and tailor tools depending on the vendor, instrument, and experimental design employed. Since each of these is built as a separate modular tool, users can integrate them into other lipidomics software packages to design a workflow based on their expertise, application, finances, and instrumentation. Finally, we have integrated the tools into a single software application, LipidMatch Flow, which has been designed to cover all steps of the data-processing workflow up to lipid annotation for more robust and rapid analysis.

CHAPTER 1
INTRODUCTION

Lipidomics, an Overview

A comprehensive study of biomolecules and biopolymers, such as metabolites and proteins, can be useful in answering key biological questions, including questions regarding the mechanism of a disease. While genomics and transcriptomics generally describe the cellular machinery in terms of “what could happen”, metabolites, which are the small molecules generated through metabolism, are the most direct measure of “what is happening” as they relate directly to the phenotype.[1] Metabolomic studies are based on the concept that global analyses of metabolites represent the physiological state of an organism at a given point in time and that the metabolite fluxes are sensitive to cellular change.[2] Metabolites that are up or down-regulated can be used directly as biomarkers for disease or exposure; additionally, they can be used to understand the influence of disease on biological pathways.[3] One key subset of metabolites are lipids, which are either hydrophobic or amphiphilic molecules. Lipidomics is a sub-division of metabolomics. Lipids are ubiquitous and play essential and diverse biological roles in nearly all organisms, which makes them likely candidates for biomarkers and understanding disease etiology.

The roles of lipids in biological systems include membrane structure and function[4], modulators of immune system function[5], energy storage[6], signaling[7–9], as targets of oxidation[10], alveoli functioning[11], and water retention in the skin[12] and eyes[13]. Therefore, lipids are likely biomarkers in omics analyses due to their integration in multiple biological pathways and ubiquity across virtually all organisms and biological substrates. For example, possible lipid biomarkers have been discovered

in cardiovascular disease[14], chronic kidney disease[15], eating patterns, specific cancers[16], and as general markers for oxidative damage[10, 17].

In order to serve numerous biological roles, lipid structures are highly diverse [18]. Recently, biomedical-related international initiatives have begun to establish databases to organize and document the diversity of lipids, including Japan's LipidBank founded in 1989 [19], Lipid Metabolites and Pathways Strategy (LIPID MAPS) in the USA founded in 2003 [20, 21], and the European Lipidomics Initiative: shaping the life sciences (ELIFE) founded in 2004 [22]. One of the major lipid databases, LipidMaps, contains over 45,000 lipids (both computer simulated and experimentally confirmed) [23, 24]. However, these databases only begin to scratch the surface of the lipidome diversity. The lipidome – the entire collection of individual lipid species in cells, tissues, or biofluids – has been estimated to be composed of 1,000 to more than 180,000 molecular lipid species [25, 26], but many of these species are likely very low in abundance or have not been observed [27]. Furthermore, these calculations and experimental measurements do not take into account all possible double bond positions, backbone substitutions, and stereochemistry [25]. As a result, when all subtle structural differences are accounted for, several million potential lipids are possible. It is therefore not surprising that a significant portion of the human genome is devoted to synthesize, metabolize and regulate this lipid diversity [28].

The diversity of common fatty acids (FAs) and their potential arrangement on the backbone gives rise to an enormous array of lipid structures (see Figure 1-1 for phospholipids). For example, triglycerides (TGs) contain 3 fatty acyl constituents attached to a glycerol backbone and with 38 experimentally verified fatty acyl

constituents, 9,880 TG structures can be proposed without accounting for positional isomers or differences in double bond location. This calculation is a combination (order does not matter, because *sn* position is unknown) with repetition, shown in Equation 1-1, where n is the number of possible fatty acids and r is the number fatty acyl chains on a specific lipid (3 in the case of triglycerides).

$$Lipid\ Species = \frac{(n + r - 1)!}{r!\,(n - 1)!} \quad (1\text{-}1)$$

Each lipid is characterized by the specific backbone and corresponding fatty acyl constituents, for example PC(16:0/18:1(9Z)), can be read as a phosphatidylcholine (PC) head group (PO_4-$CH_2CH_2N(CH_3)_3$), and two fatty acyl constituents attached, one with 16 carbons and no double bonds (16:0) in the *sn*1 position and the other with 18 carbons and one *cis*-double bond on the 9th carbon from the backbone (18:1(9Z)) in the *sn*2 position. A generalized structure for phospholipids is shown in Figure 1-1.

The structural diversity of lipids poses a challenging analytical problem for lipidomics analyses. For example, both identification and quantification in metabolomics is routinely done with labeled external or internal standards for each metabolite of interest. However, in lipidomics, internal standards have not yet been synthesized to cover even a small portion of the lipidome and purchasing the diverse species that are available can be prohibitively expensive. Another challenge is the separation of numerous lipids with only subtle differences in chemical and physical properties. The separation and unique characterization of these molecules is also important for identification and quantification. Therefore, lipid diversity and lack of standards poses a unique challenge to comprehensively and accurately measuring lipids.

Lipidomics Workflow

Each step in the lipidomics workflow influences the coverage and accuracy of lipid measurements. The lipidomics workflow consists of four major steps, which are sample preparation, data acquisition, data processing, and statistics/interpretation (Figure 1-2).

Sample Preparation

Sample preparation consists of the following stages: sample collection, storage, homogenization, and extraction. During these steps, a major factor influencing the accuracy of measurements is sample degradation due to physical factors such as oxidation during exposure to light and air [29] and biological factors such as enzymatic biotransformation [30]. Therefore, enzyme inhibitors,[30] oxidation inhibitors, [31] heat treatment,[32] sample preparation and storage under cryogenic temperatures, and/or other techniques for stabilizing the lipidome are necessary; however, these techniques are not often used.

Sample extraction can also influence the physical and enzymatic transformations of lipids. For example, extractions which are highly acidic or alkaline can induce the hydrolysis of lipids,[33] and primary alcohols such as methanol and ethanol can form methylated and ethylated glycerophospholipids in the presence of phospholipase D [34]. Additionally, because lipids can range from acidic to neutral, zwitterionic to uncharged, and polar to relatively non-polar, the extraction conditions and solvents may preferentially extract certain lipid classes or species over others.[33, 35–37]

The Bligh-Dyer [38] and Folch extraction procedures [39] are among the most common extraction procedures used in lipidomics, and while multiple comparisons of lipid extraction methods have been made, the benefits of each extraction procedure is

often sample dependent.[33, 35–37] Both Bligh-Dyer and Folch extractions capture a majority of the lipidome, with Folch extractions having higher extraction efficiencies, but using more solvent and being slightly more time-intensive. For Bligh-Dyer and Folch extractions, highly acidic or highly polar lipids such as phosphatidic acid (PA) and fatty acids, respectively, are often not extracted in high enough concentrations for detection. For PA and other lipids, which are not easily driven into the organic phase, salts such as NaCl can be used to drive the lipids into the organic layer.[40] However, the use of these salts can reduce ion signals through competitive adduct formation. For fatty acids either derivatization and analysis by GC-MS or analysis of polar fractions in LC-MS are often used.[41] Therefore, the Bligh–Dyer extraction protocol (1:2, v:v, chloroform:methanol) and Folch extraction protocol (8:4:3, v:v, chloroform:methanol:water) were used in this work.

Data Acquisition: Reverse-Phase Liquid Chromatography

The second step of the lipidomics workflow is data acquisition (Figure 1-2). As with extraction, the choice of column, elution gradient, mobile phase solvents, and mobile phase additives can drastically alter the coverage of the lipidome and the amount of information obtained for each lipid species. For example, reverse phase chromatography is the most commonly employed technique for lipidomics. Reverse phase chromatography's popularity in lipidomics stems from its ability to retain analytes according to non-polar interactions with the stationary phase, enabling the separation of analytes by polarity. Therefore, lipid species can be separated by lipid class, fatty acyl constituents, and even regioisomers.[42] On the contrary, more polar lipids such as fatty acids are not retained as well in reverse phase chromatography, and hence methods such as normal phase chromatography [43] or hydrophilic interaction chromatography

(HILIC) [44] are often used. While reverse phase and normal phase chromatography can separate out lipids with different fatty acyl constituents, techniques such as HILIC separate lipids out by lipid class. Therefore, quantification by lipid class is simplified in HILIC, at the expense of identification and quantification of lipids at the molecular level. When molecular species information is desired, HILIC can be placed in tandem with ion mobility (IM), which separates ions by their collision cross section, a property that is characteristic of ions size, shape, and charge for given experimental conditions.[45] In this work reverse phase chromatography using a C18 column was employed, because it is able to separate most lipid species (as compared to other chromatographic techniques) under short chromatographic runs.

Another important factor for determining the lipid coverage in a lipidomics workflow is the choice of additives and mobile phase conditions, which in turn affect which lipids will ionize in electrospray ionization (ESI). For example, most polar lipids readily form positively charged protonated ions, and negatively charged deprotonated ions, depending on the ESI polarity. Neutral lipids often do not form protonated ions at high efficiencies, and will best be observed with the addition of cations (lithium) or as ammoniated ions with the addition of ammonium salts to the mobile phase. In addition, lipids such as phosphatidylcholine and many sphingolipids do not form deprotonated ions easily, but form acetate, formate, and chlorinated adducts readily.[46] In this work, ammonium formate was added as a buffer to control pH and to form ammonium or formate adducts for species which are not readily protonated or deprotonated.

Data Acquisition: Volatizing and Ionizing Lipids with Electrospray

One of the most important aspects of mass spectrometry is the ionization source, which determines which ions will and will not be observed. The parameters and choice

of the ionization source impact linear range of detection, sensitivity, and other important factors. Ionization sources which can ionize and volatilize neutral relatively high molecular mass compounds are necessary in LC-MS analysis of lipids. In the 1970s and 1980s UV-detection with reverse phase chromatography was used for lipid analysis, because an ionization source conducive to lipid analysis lipid analysis was not widely available [42]. It was not until the late 1980s, with the advent of electrospray ionization (ESI) and atmospheric pressure chemical ionization (APCI), that lipids could be analyzed by mass spectrometry. Mass spectrometry has become the predominant technique for lipidomics because of its ability to detect low abundant species (high sensitivity), to quantify lipids (proportional signal, with minimal variance across classes compared to UV), and to assign specific structures to co-eluting isobaric species (high specificity due to using exact mass and collision-induced fragmentation) [47].

Three major techniques are currently used as ion sources for lipidomics: atmospheric pressure photoionization (APPI), atmospheric pressure chemical ionization (APCI), and electrospray ionization (ESI). ESI is the ionization source employed in this work, and has a higher sensitivity for many lipid compounds in comparison to APCI and APPI when using chemical modifiers such as ammonium formate [48, 49]; however, significantly lower linear dynamic range and correlation coefficients for calibration curves have been found in ESI LC-MS of fatty acid and glycerol species [48].

Chech and Enke [50] provide an in-depth review of the mechanics and consequences of ESI fundamentals. Briefly, a negative or positive high voltage (2-5 kV) is applied to a capillary containing a dilute solution moving at a slow flow rate (0.1-10 µL/min for high performance systems). The charge partitioning produced by this electric-

field generates a Taylor cone where liquid protrudes from the capillary tip. Once the Rayleigh limit is reached where the surface tension of the liquid is overcome by the columbic repulsion of the high density of charges, the droplet disintegrates into smaller droplets, which are directed to the capillary sampling orifice of the mass spectrometer by ion lenses after the capillary and electrical gradients. As the smaller droplets evaporate rapidly at atmospheric pressure within the source, ions are formed.

The exact mechanism of ion formation is unknown, although several theories exist including charge residue [51], ion evaporation [52], or chain ejection [52]. Evidence exists showing that for large molecules, such as proteins, ions are predominantly formed by the charge residue mechanism.[53] In the charge residue mechanism, droplets continue to undergo fission as an effect of the Rayleigh limit being reached, until a single charged molecule or no molecules exist in each droplet. Continued evaporation leads to single ions in the gas phase. For proteins chain ejection is another mechanism shown for extended or unfolded conformations. In this mechanism, the protein migrates towards the droplet surface due to hydrophobic groups on the protein, and a terminus gets expelled into the vapor phase, followed step-wise by the remainder of the protein.[54]

For small molecules, evidence supports that ion formation is predominantly by ion evaporation.[53] Therefore for lipids, which are relatively small molecules, the ionization mechanism is most likely via ion evaporation. In ion evaporation, the field strength at the surface of the droplet becomes strong enough to eject ions after evaporation increases charge density at the surface of the droplet. Lipids have a higher propensity for ejection via ion evaporation than other metabolites, because they are

relatively non-polar, and hence accumulate at the surface of droplets between the air-droplet interface.[55] The ionization efficiency, which is the number of ions produced in proportion to the number of molecules introduced to the ESI source, can drastically differ for different lipid species and the concentration of lipids. One can imagine that at high concentrations certain lipids may out-compete other lipids for migration to the air-droplet interface, depending on the lipid structure. Additionally, lipids readily form aggregates and micelles in solution phase, especially for non-polar lipid species.[56] These aggregates will less readily be ejected from the droplets via ion evaporation, hence decreasing ionization efficiency of the ions involved in aggregate formation.

Aggregate formation can have a significant impact on accurate quantification and sensitivity of various methods (see the section titled: Processing of LC-HRMS/MS data). For example, increasing concentration of lipid samples will increase overall signal intensity, but will lead to aggregate formation dependent on fatty acyl constituents and lipid class.[56] Therefore, the degree of ionization for each lipid species can differ, and despite this variation, internal standards do not exist to correlate ion signal to concentration for most species observed. In addition, the concentration of lipids within ESI droplets depends on the chromatographic region of elution, and hence aggregate formation will depend on the concentrations of coeluting lipids. Therefore ionization efficiency based on aggregate formation is difficult to predict in LC-MS, as aggregate formation may be controlled by coeluting lipids and additional factors other than a specific lipid species structure and concentration.

A heated electrospray ionization (HESI II) probe was used as the ion source in this work. HESI II provides higher auxiliary gas flows and temperatures than traditional

ESI due to the absence of a heat exchanger. The heated capillary aids in solvent declustering and removal of neutrals through a higher energy and more frequent collisions. Hence, less solvent clusters with analytes of interest are observed, which simplifies spectra, and ionization efficiencies may be increased.[55]

Data Acquisition: Orbitrap Mass Spectrometers for Lipidomics

While the lipidomics community is moving towards the use of high resolution mass spectrometers for lipidomics, as employed in this work, traditionally ESI has been used in tandem with triple quadrupole mass spectrometers for untargeted and targeted studies of the lipidome. This technique is still used due to its high sensitivity and relatively high specificity for targeted studies. Scanning modes and considerations for lipidomics employing triple quadrupoles are reviewed in Han et al. [57].

The complexity of the lipidome and multiple arrangements of the same constituent building blocks to form different lipids leads to a high degree of overlap both in fragmentation and precursor ion mass-to-charge (*m/z*) values for differing lipid species. This difficulty imposed by the overlap of fragment and precursor ions is exacerbated by the low resolution obtained by triple quadrupole instruments. The advent of high-resolution hybrid mass spectrometers (e.g., time of flight and orbitrap mass spectrometers [58, 59]), has allowed for enhanced lipid identification and structural annotation, when compared to unit resolution mass spectrometers because of improved specificity, sensitivity, and reproducibility [60, 61]. High mass accuracy (often sub-ppm) can reduce the list of possible molecular formulas by providing the isotopic structure detail of precursor lipid ions. The addition of resolved isobaric fragment ions reduces false positive and negative molecular identities [62]. The advantages of high-

resolution mass spectrometry for accurately annotating lipids have led to its increased use in lipidomics.

Fourier-transform ion cyclotron resonance (FT-ICR) mass spectrometers provide the highest resolution of mass spectrometers used for lipidomics. Introduced in 1974, FT-ICR mass spectrometers are expensive and have a large footprint, due to the use of a large magnet. For example, systems with 7 Tesla and 12 Tesla magnets cost about $800,000 and $2,000,000 dollars respectively, and weigh on the order of tens of tons.[63] Orbitrap mass spectrometers provide the second highest level of resolution, and under the same scan speeds can provide the same resolutions as 7 Tesla FT-ICR instruments at a lower cost and significantly smaller footprint. Orbitrap mass spectrometers have the unique characteristic that the mass resolution scales inversely to the size of the detector (for reasons described in the section titled: Measuring high-resolution with orbitrap mass spectrometers), which is contrary to time-of-flight instruments where resolution is proportional to ion flight path (often achieved though increasing the length of the flight tube) and FT ICR instruments where resolution is proportional to the strength of the magnet (often increased by increasing the size of the magnet). Therefore, to achieve resolution of 240,000 at a 786 ms transient for *m/z* 400, an orbitrap with inner dimensions of 20 mm is used, as apposed to the multi-ton 12 tesla FT ICR acquiring data at the same transient with magnitude-mode spectra obtaining a resolution of 170,000 [64].

Due to the small size of the orbitrap, the actual resource cost to manufacture orbitrap mass spectrometers is low compared to that of time of flight and FT ICR instruments. As demand increases and technologies improve, it can be predicted that

instruments containing orbitrap detectors will decrease in price at the same time as there is an increase in performance.

Overcoming the difficulty of using ESI with orbitrap mass spectrometers

The components of a Q Exactive used in this work are shown in Figure 1-3, which can be referenced for all lenses, ion transmission elements, and detectors mentioned hereafter. From atmospheric pressure in the ion max housing, a rotary vane pump establishes mbar pressure, followed by two turbo pumps, with a final pressure less than $8x10^{-8}$ mbar in the orbitrap. Low pressures allow for higher ion transmission efficiencies and longer mean free paths. Therefore, ion dynamics measurements, based on fundamentals such as mass-to-charge ratios (*m/z*), are not perturbed by collisions. The bent 90º flatapole decreases the number of neutral molecules entering the instrument further reducing noise and improving sensitivity.

Initially, orbitraps were considered unsuitable for continual sources such as ESI due to the necessity of injecting short (less than one μs) ion packets for high resolution analysis. In 2003, Hardman and Makarov proved that ion traps could be used for the collection and injection of these tight ion packets, developing an interface between orbitrap analyzers and ESI sources.[65] A curved RF only linear trap quadrupole (C-Trap) is applied in the Q Exactive [59], which uses N_2 or other cooling gases to reduce ion kinetic energies and collect ions in a thin thread, confined axially by applying 200 V to both the split lens (lens 7) and C-Trap entrance lens (lens 8). The C-Trap can also be used for automated gain control (AGC), where ions are accumulated in the C-Trap until a certain user defined intensity is obtained (or until a user defined time is reached), after which, ions are ejected into the orbitrap. For AGC, the image current in the orbitrap is used to predict the number of ions in the C-Trap. Since the ion flux can drastically

change across a chromatographic run, this prediction of ion flux to the C-Trap is enabled by implementing rapid microscans to the orbitrap. The occurrence of these microscans is not visible to the user and is automatically triggered in the Q-Exactive when there is a dramatic change in ion flux across time.

By controlling the number of ions sent to the detector using AGC, the signal for trace ions can be increased by increasing trapping times. Likewise, the signal for highly abundant ions can be reduced, by decreasing trapping times. Reducing the number of ions entering the orbitrap limits signal saturation and drift in resolution due to ion coalescence, where ions close in mass lock phases (axial frequency) [66]. A high deflection voltage on the split lens (lens 6) is used as a “binary switch” controlling ion transmission into the C-Trap. Therefore, a precise start and end injection time (IT) into the C-Trap is obtained. Ions are either ejected from the C-Trap to the orbitrap once the AGC is reached, or once the maximum IT is reached.

Measuring high-resolution with orbitrap mass spectrometers

Using an offset Z lens (lens 9) to remove carrier gas, the ions are ejected into the orbitrap with a time of flight (t_{inj}) between the C-Trap and the orbitrap governed by Equation 1-2 [67]**,** where L_{eff} is the effective length of the flight path. Hence the arrival of ion packets is proportional to the square root of the *m/z*, and ions with lower *m/z* arrive before ions with higher *m/z*.

$$t_{inj} = \frac{L_{eff}}{\sqrt{\frac{2eV}{m/z}}} \tag{1-2}$$

As ion packets are injected into the orbitrap, the electrical field of the inner electrode is ramped up, ions are pulled closer to the inner electrode, and the ions begin

to oscillate around the inner electrode. Once the ramping voltage is stabilized, the ions reach equilibrium between their centrifugal force of rotation around the inner electrode (inertia) and the DC potential of the inner electrode (centripetal force). These oscillations are similar to how planets circulate the sun, where the centripetal force in the case of planets is gravity. By balancing centripetal and centrifugal force, a simplified characterization of the radius of the ion motion around the inner electrode (r_m) can be described as in Equation 1-3, where *eV* is the ions' kinetic energy and *eE* is the inward force of the electric field on the ion. [68]

$$r_m = \frac{2eV}{eE} \tag{1-3}$$

Based on Equation 1-3, the radius of the ion path is independent of *m/z* and cannot be used to measure *m/z* directly. Radial ion motion is therefore not important for measuring *m/z*, but it is important for trapping ions in thin circular bands surrounding the inner electrode. With a sector instrument, which is technically a kinetic energy analyzer, the kinetic energy and radial component of the electric field must be matched to control the trajectory of the ion packet through the sector. Similarly, if the kinetic energy and the radial component of the electrical field are not matched in an orbitrap, then ions will not follow narrow circular trajectories around the inner electrode and will either collide with the electrode or form broad discs after ion motions are averaged.

In addition to the radial distance, the radial oscillations cannot be used to measure *m/z* to a high degree of accuracy because the oscillations depend on the velocity and initial radius of the ion trajectories. Because ion packets entering the orbitrap from the C-Trap consists of ions with varying velocities, their frequencies differ,

and the ion packets quickly form a circular distribution of ions around the electrode. Therefore, one of the hallmarks of the orbitrap, is to measure resolution using axial oscillations in the horizontal direction (Figure 1-3).

The orbitrap consists of two outer electrodes, shaped like "cups" facing each other, and an inner spindle-shaped electrode. These electrodes form a quadrologarithmic electric field described in Equation 1-4, where r and z are the cylindrical coordinates, k is the field curvature, R_m is the characteristic radius, and C is a constant. While axial oscillations could theoretically be induced by applying excitation from either of the outer electrodes, ions are injected off axis (Figure 1-3), and due to the shape of the shape of the quadrologarithmic electric field described in Equation 1-4 [68], ion packets automatically begin oscillating axially without any additional force. One of the advantages of off axis injection versus injecting ions to the equator of the orbitrap (center of the orbitrap) and applying excitation, is that in the absence of excitation, detection can start almost immediately, without waiting for the electronics to stabilize.

$$U_{(r,z)} = \frac{k}{2}\left(z^2 - \frac{r^2}{2}\right) + \frac{k}{2}(\mathrm{R}_m)^2 \ln\left(\frac{\mathrm{r}}{\mathrm{R}_m}\right) + C \tag{1-4}$$

When injected off axis, the initial force is from ions moving from areas with smaller distance between the inner and outer electrodes to the equator of the orbitrap, where the greatest distance between the inner and outer electrodes occurs. Thereafter, ions experience a restorative force at the opposite end of the orbitrap as the ions continue their axial trajectory. Hence, without any dampening of oscillations due to the low pressures of the orbitrap (less than 8×10^{-8} mbar), oscillations take the form of a harmonic oscillator, with the frequency of oscillations, ω_z, described in Equation 1-5,

where k is the field curvature.[69] The axial frequency, ω_z, is experimentally determined by measuring an image current in the time domain, which is rapidly converted to the frequency domain using the Fast Fourier Transform (FFT) algorithm.[67]

$$\omega_z = \sqrt{\frac{k}{m/z}} \quad \text{(1-5}$$

Because the orbitrap mass spectrometer directly measures *m/z* using FFT of axial oscillations, which are independent of ion energies and spatial distribution, and the electrical field can be tightly controlled in the orbitrap geometry, masses can be measured with very high accuracy and resolution. The resolving power is a function of the smallest *m/z* difference between two ions in which the peaks can be deconvoluted. Since resolving power is directly correlated to the number of axial oscillations, we can derive Equation 1-6 from Equation 1-5, where t is scanning time. Therefore, resolution is inversely proportional to the square root of *m/z*, meaning that higher mass ions are measured at lower resolution and resolution is directly proportional to the scanning time within the orbitrap. Following the same logic, resolution is inversely proportional to the period of oscillations, and hence shortening the period of oscillations can increase the resolution even at the same acquisition rates (scan time). Therefore, resolutions of a million have been achieved by reducing the size of the orbitrap.[70] However, with a fixed sized orbitrap, methodologies are limited to an optimization between the number of scanning events which can happen over a chromatographic peak and the level of mass resolution desired.

$$R \propto t\sqrt{\frac{1}{m/z}} \quad (1\text{-}6)$$

Scanning functions for obtaining fragmentation data

Fragmentation gives both fatty acid moiety information and lipid class information, as neutral losses or fragment ions are often produced at the linkages between the backbone and constituents, for example at the *sn*1, *sn*2, and *sn*3 ester linkages in Figure 1-1. Systematic querying of this fragmentation data can be used for untargeted studies, increasing identifications of unknown or unexpected species. The scanning functions used for obtaining fragmentation data control both the accuracy and number of lipid identifications. For example, due to resolution being proportional to scanning time, only a limited number of ions can be isolated and fragmented across a chromatographic peak. Hence, a number of ions will not have fragmentation data, and cannot be identified. Therefore, novel strategies are needed for acquisition of fragmentation data.

To obtain fragmentation data, ions can first be isolated to a minimum of a 0.4 amu width using the Q-Exactive quadrupole.[71] Targeted lists or data-dependent acquisition (DDA) modes can be used to program ion isolation. Alternatively, all ions can be simultaneously sent through the quadrupole for fragmentation, which is termed All Ion Fragmentation (AIF). Fragmentation occurs in an octapole trap via higher energy induced collisional dissociation (HCD), which produces fragmentation similar to that in a triple quadrupole instrument, giving low mass ions and full-fragment-mass-range acquisition.[72] Contrary to the name, HCD fragmentation energies are similar to those in collision-induced dissociation (CID). The "higher energy" in HCD refers to the higher

RF field used to trap fragment ions as compared to CID experiments. Users define the normalized collision energy (NCE), which is normalized to the isolation center as the energy required for fragmentation generally scales proportionally with *m/z*.[73] The NCE can be stepped between 3 incrementing values for a single orbitrap scan. Fragmenting ions with stepped collision energies has the advantage of providing a larger number of fragments, which in turn can lead to more confident lipid identification.

In data-dependent top-n (ddMS^2-topn), the top-n most intense peaks in full spectrum acquisition are selected and fragmented. Scanning events of ddMS^2-top10 are shown in Figure 1-4, as well as scanning events for AIF. For ddMS^2-topn, additional parameters include loop count (top-n), underfill ratio, apex trigger, exclude isotopes, and dynamic exclusion. The underfill ratio sets an intensity threshold based on the AGC target. Any ions falling below this threshold will not be selected for fragmentation by tandem mass spectrometry (MS/MS), and therefore in a top-n experiment, often not all top-n ions will be selected and fragmented. The apex trigger sets a lag time between the observation of a peak and isolation for fragmentation in order to isolate ions at the apex of the chromatographic peak and obtain maximum ion intensity. Dynamic exclusion places ions selected for fragmentation on an exclusion list for a given amount of time before they can be selected again for fragmentation, hence increasing the number of ions for which MS/MS scans are obtained.

For AIF, the entire ion population is subjected to fragmentation simultaneously, and therefore the number of precursors with fragmentation information is not limited as in ddMS^2-topn. Identification using AIF is problematic because precursor-fragment relationships are lost due to precursors not being isolated in a single Da window.

Therefore, fragments could come from any precursor, which increases the chance of false positives if only exact mass of fragments are used for identification, and false negatives if dot-product identification is used. Deconvolution using chromatographic peak shape of precursors and fragments can be used to reconstruct the precursor-fragment relationship, reducing the number of false positives. Deconvolution can be difficult in the case of lipidomics where there is a high degree of exact mass overlap in narrow retention time regions both in the precursor and fragmentation space. Therefore, because of the overlap in fragmentation from various precursors when acquiring AIF data, both dot product, reverse dot product, and deconvolution based identification will lead to an increase in false negatives.

Processing of LC-HRMS/MS Lipidomics Data

Data-acquisition often contains the most standardized and reproducible steps of the lipidomics workflow. However, data-processing of large volumes of mass spectra acquired using LC-HRMS/MS is a major bottleneck in the lipidomics workflow for many labs. Currently, there is no consensus on which software tools and algorithms are optimal for data-processing in lipidomics. This work introduces software tools and strategies aimed to cover the entire data-processing workflow. The data-processing workflow consists of 3 major steps, which are: feature finding, lipid annotation, and lipid quantification (Figure 1-2).

The first step of the workflow is to detect features, which are molecules or groups of molecules with similar polarity and adduct masses. To understand feature finding, it is important to understand the LC-HRMS data format. There are 3-dimensions to a full-scan dataset: mass-to-charge (*m/z*), retention time, and intensity. An example 3-dimensional dataset is shown for adipose tissue acquired in positive ion polarity in

Figure 1-5. Feature finding in this work only employed full-scan data using MZmine 2.0, and hence no MS/MS data is used to align features. The first major step of feature finding employed in MZmine 2.0 is chromatogram building. In this step the 3-dimensional data-set is cut into 2-dimensional slices consisting of retention time and intensity (or relative abundance, Figure 1-5). In these slices mass is held constant, and hence these slices are termed reconstructed mass chromatograms. The *m/z* data is binned or otherwise *m/z* values are chosen across the mass range, and a reconstructed mass chromatogram exists for each *m/z* interval.

After chromatogram building, the reconstructed mass chromatograms are deconvoluted (Figure 1-6). This feature identifies and integrates multiple peaks sharing the same or similar *m/z* value within a slice. In this work, a local minimum is used for deconvolution, where a peak is found between two local minima in intensity. Variables such as minimum ratio of the top of a peak to an edge (the local minimum) can be controlled to limit the deconvolution of noise into individual peaks. A list of peaks is then obtained with *m/z* values, retention time values, and intensities (peak areas and peak heights). These peak lists can be further filtered by reducing peaks from blanks, removing isotope peaks, and combining peaks from multiple adducts of the same molecular species. In lipidomics, often the interest is in comparing multiple samples. Therefore, peak lists across multiple samples are aligned using a *m/z* and retention time tolerance. Additional factors such as peak shape can also be accounted for during alignment. After alignment, often there will be missing peaks for some samples, while signal was detected for other samples. Therefore, gap-filling is the final step employed. Gap filling lowers the threshold needed to determine a peak and searches for missing

peaks across respective samples. Gap-filling is an important step, because assigning zeroes to missing peaks can have significant effects on downstream statistical analysis.

After features are detected and aligned, the MS/MS spectra must also be aligned to each feature. As previously described, MS/MS is necessary for lipid identification, because lipids have numerous fatty acyl isomers and these isomers can co-elute (Figure 1-6). The second and third step of data-processing are lipid annotation using the MS/MS data and quantification; these steps are described in detail in Chapter 3 and Chapter 4, respectively. The major difficulty with annotation and quantification is the lack of internal standards. Only a negligible portion of known lipids are covered by the currently available lipid standards due to the diversity of lipid structures. In addition, even with currently available lipid standards, routine analysis is limited by the cost prohibitive nature of purchasing all available and relevant lipid standards.

In annotation, the best practice is implementation of retention time and MS/MS libraries for each lipid analyte based on data acquired for lipid standards on the same instrument and column used for analysis. Because of the lack of internal standards, *in-silico* fragmentation libraries, with limited to no retention time criteria, are often used for annotation of lipids in untargeted studies. For absolute quantification, matrix matched standards which cover the dynamic range of lipid concentrations, would be necessary to account for extraction efficiencies, ionization efficiencies, and any effect of data-processing on quantitative values of each lipid species. While this is commonly employed in metabolomics,[74] in lipidomics, because of the lack and cost of lipid standards, only relative quantification can be performed. In addition, even using long chromatographic separations and high-resolution mass spectrometers, there is still

often overlap between lipid isomers (Figure 1-6). Hence, quantified values are often an average across multiple lipid species. In untargeted lipidomics, class specific lipid internal standards, or multiple lipid internal standards per lipid class, are used to obtain relative quantified values for each lipid species. These standards must be exogenous (often odd carbon numbers) or isotopically labeled to not overlap with endogenous lipid species. For initial clinical applications to determine biomarker candidates or disease etiology, where relative changes are of concern, relative quantification is often sufficient [75]. Further clinical trials or studies for validation, which require higher confidence in lipid identifications and absolute quantification, are likely to be performed as targeted analyses with internal standards.

Biological Interpretation of Lipid Perturbations

After data-processing, the final, and currently one of the most problematic steps, is data interpretation. Currently, the bulk of knowledge on lipid biochemistry has not been compiled into easily searchable pathway databases. Furthermore, pathways containing lipids often only include lipid class, not fatty acid constituents. However, lipids within the same class, containing different fatty acid constituents, can have drastically different biological roles.[76–79] Therefore, the interpretation of shifts in lipid concentrations often relies on experts. Even this interpretation by experts is limited because the specific biological functions for the majority of lipid species detected are unknown. Therefore, in most cases, lipidomics provides a wealth of information at the level of lipid class and fatty acyl constituents, but only the shifts in lipid class and total free fatty acids are used for interpretation. Because the technology allowing lipid researchers to obtain and process data with extensive coverage of the lipidome is still continuing to emerge, it can be expected that our understanding of lipid biology will

continue to grow, and this information will be used to expand lipid pathway databases. This rapid expansion of pathway databases, due to introduction of robust low-cost measurement techniques, is exemplified in the field of genomics and proteomics.[80] A more detailed description of the technical difficulties that must be overcome to achieve this increased coverage and accessibility of lipid pathway databases is described in Chapter 6: Conclusions and Future Perspectives.

Dissertation Overview

Lipidomics requires novel data-acquisition strategies and data-processing algorithms to accurately and comprehensively measure the lipidome. While current lipidomics papers employing LC-HRMS/MS frequently report hundreds of lipids, there are possibly thousands, to even tens of thousands, of lipids within their samples. Increasing coverage of the lipidome increases the probability of finding clinical biomarkers and determining the etiology of disease. In this dissertation, I introduce a number of strategies and software which can improve lipid coverage and the accuracy of measurements. The strategies and software introduced here cover the first three major steps of the workflow (sample preparation, data acquisition, and data processing) as shown in Figure 1-7. Chapter 2 describes a technique termed iterative exclusion, which we have automated and applied to lipidomics applications. Using this data-acquisition technique, we increase the coverage of lipid features which have MS/MS fragmentation data, and hence increase annotation of lipid ions up to 2-fold.

In Chapter 3 I introduce LipidMatch, an open source software which contains *in-silico* fragmentation of over 250,000 lipid species, including oxidized lipids. This software also enabled identification using AIF, a technique where fragmentation is obtained for all ions without selection in a quadrupole. This software both increases the

coverage of the lipidome, through increased coverage of lipid MS/MS spectra in the *in-silico* libraries, and more accurately annotates lipids based on experimental MS/MS data. Often MS/MS data do not provide information regarding fatty acyl constituents, fatty acyl position on the backbone, double bond position, or double bond configuration (*cis* versus *trans*). Therefore, unlike other open source software, LipidMatch uses an annotation style which only denotes structural information known based on the MS/MS information, and indicates if multiple lipids are identified under a single feature. LipidMatch was compared with other open-source software across the same dataset for validation.

After annotation, the next data processing step is quantification. Because standards do not exist to cover even a small portion of the lipidome, class representative internal standards are used. In Chapter 4, a software is introduced for automatically selecting which internal standards to use to quantify each lipid feature. The selection of internal standards is based on limiting ion suppression effects and differences in ionization efficiencies between lipid structures. The software, LipidMatch Quant, was applied to show how different data processing methods can significantly affect the resulting concentration calculated.

The first step in the lipidomics workflow, and often the most overlooked is sample-handling and preparation, where enzymatic activity can drastically change the lipid profile from the native profile. Therefore, we applied the lipidomics workflow discussed in Chapters 2, Chapter 3, and Chapter 4 to determine the differences in lipid profiles of invertebrate environmental sentinel species with and without heat treatment. The invertebrates were small enough to place into cartridges and apply heat treatment

as the mode of euthanization. Evidence is presented that shows that heat treatment drastically or completely reduces enzymatic activity.

This dissertation ends with an outlook of what must be accomplished for lipidomics to continue to grow as an important technique in clinical and other scientific fields. Future directions for my work including additional software tools and integration of all software tools into a single platform are discussed.

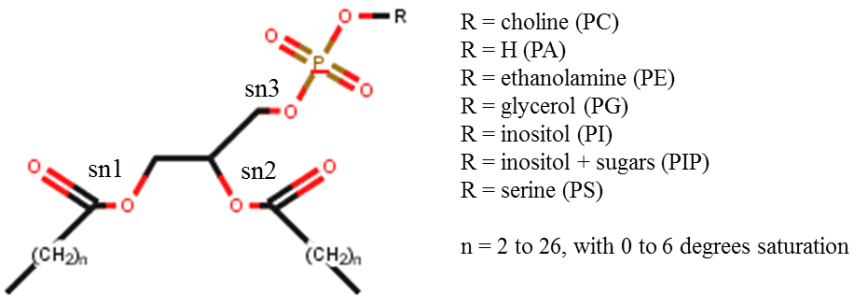


Figure 1-1. General glycerophospholipid structure showing possible head groups (R) attached to the *sn*3 position, and possible fatty acyl constituents attached to the *sn*1 and *sn*2 positions. Replacement of the *sn*3 group with a fatty acid moiety gives the structure of a triglyceride (TG), a type of glycerolipid. Removal of a fatty acid moiety in the *sn*1 *or sn*2 position gives lyso-phospholipids and lyso-glycerolipids. Note that the R groups, chain lengths, and saturations, presented in the figure, are only the most common in mammalian systems. A much larger set of head groups, chain lengths, and saturations exist. [23, 24, 81]

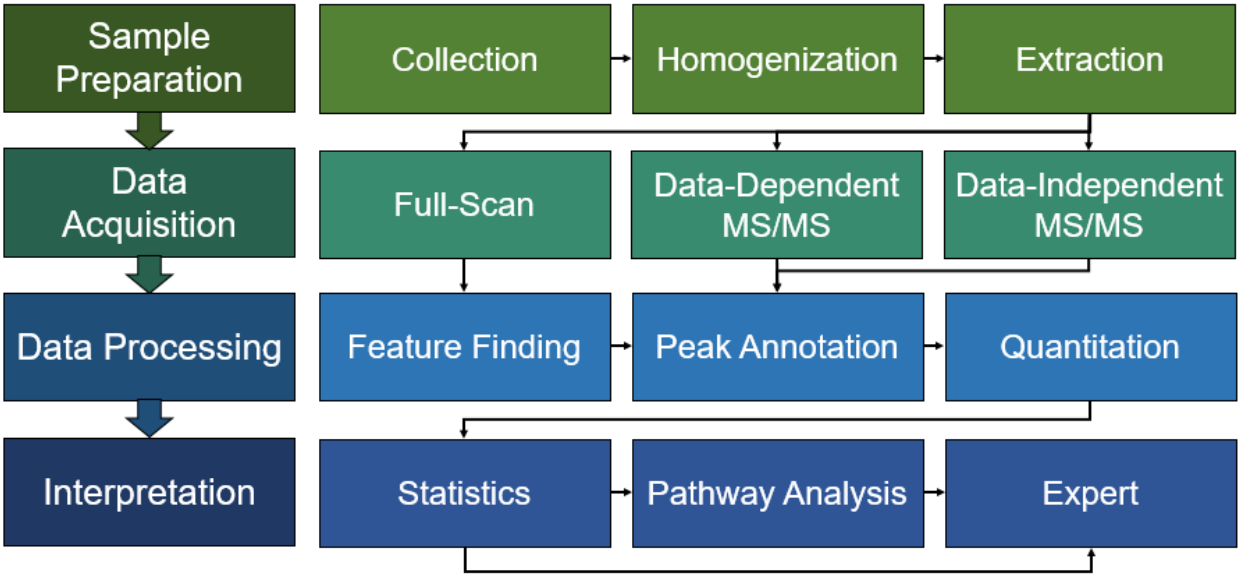


Figure 1-2. Lipidomics workflow for liquid chromatography high resolution tandem mass spectrometry (LC-HRMS/MS)

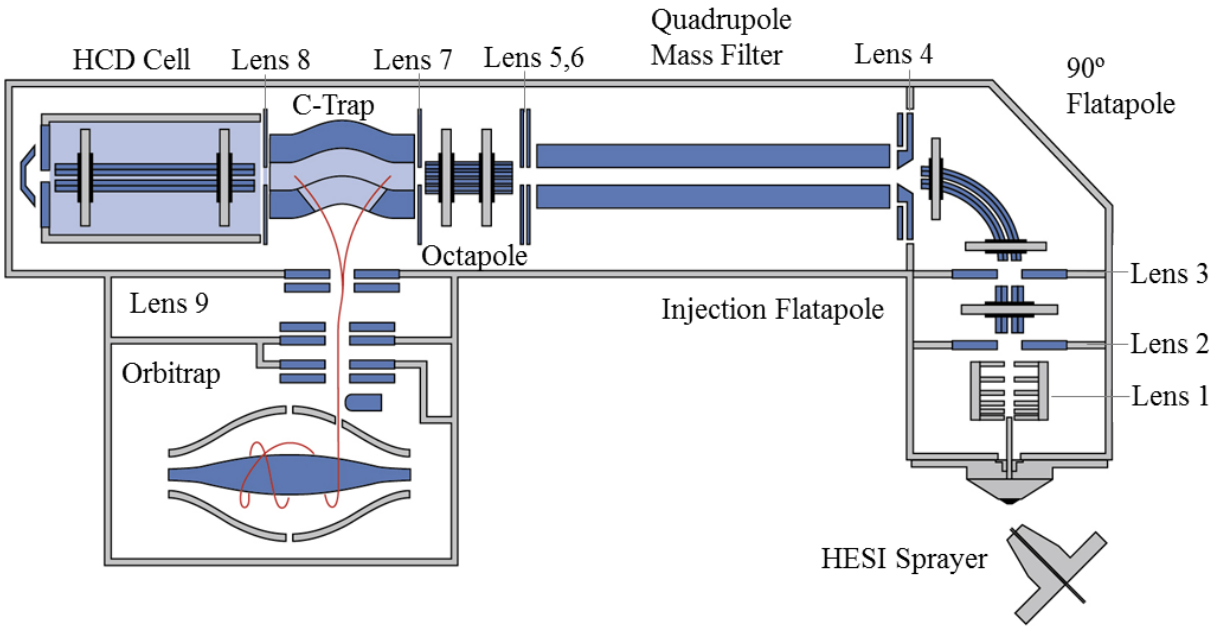


Figure 1-3. Q Exactive schematic adapted from [71]. Lens 1 (S-Lens), lens 2 (S-Lens exit lens), lens 3 (inter-flatapole lens or lens L0), lens 4 (lens L1), lens 5 (quad exit lens), lens 6 (split lens), lens 7 (C-Trap entrance lens), lens 8 (C-Trap exit lens), and lens 9 (Z lens).

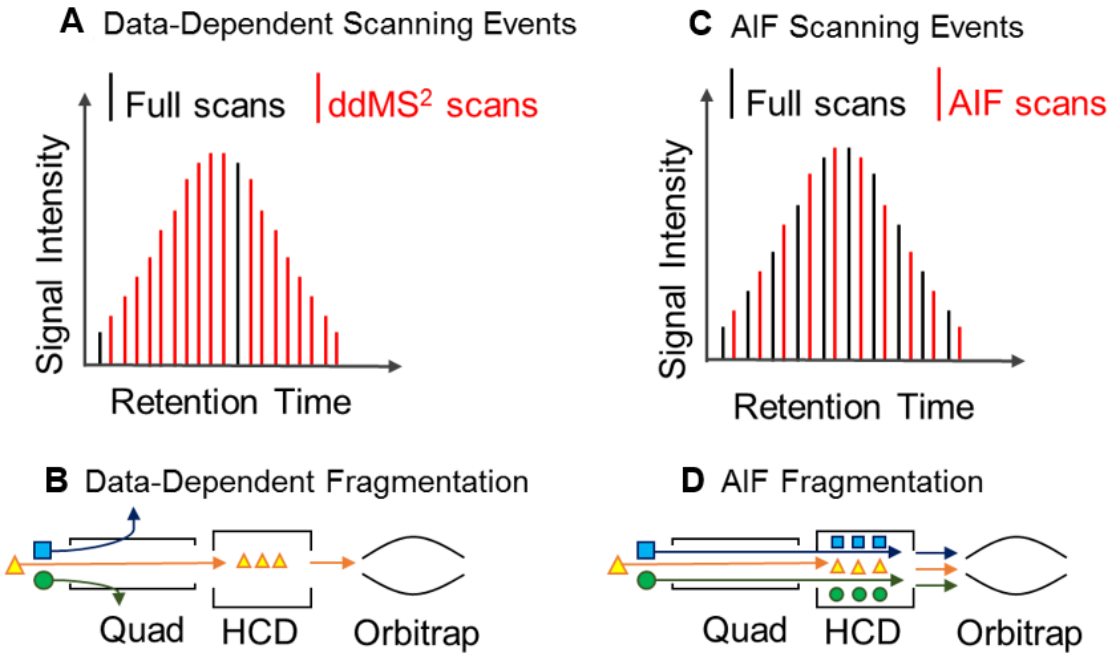


Figure 1-4. Schematic of scanning events for data-dependent acquisition (DDA) and all ion fragmentation acquisition (AIF) mode. A) Scanning sequence for DDA: each full scan is followed by 10 subsequent MS/MS scans, B) for MS/MS in DDA each ion is isolated in the quadrupole (quad) and fragmented in the HCD, before measurement of *m/z* using the orbitrap. C) In all ion fragmentation mode (AIF), scanning events are user defined. In this case, a sequence alternating between full scans and AIF scans is shown, and D) in AIF, all ions are fragmented without mass selection in the quadrupole.

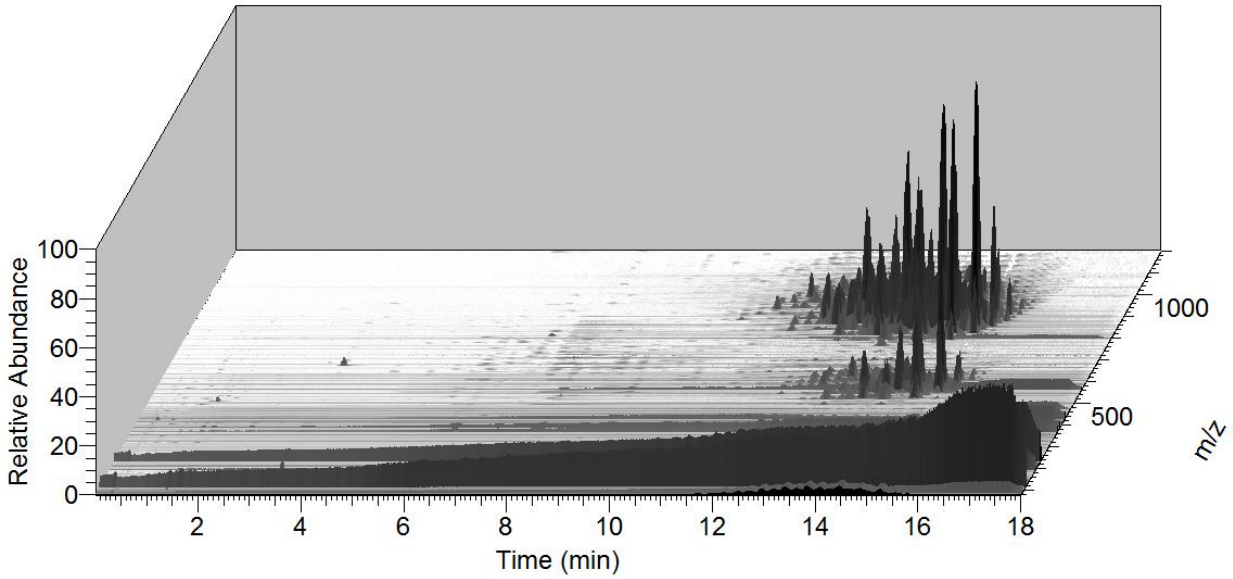


Figure 1-5. Positive ion mode full scan dataset for adipose tissue from Mozambique tilapia species. Data acquired in-house using the workflow described in this dissertation.

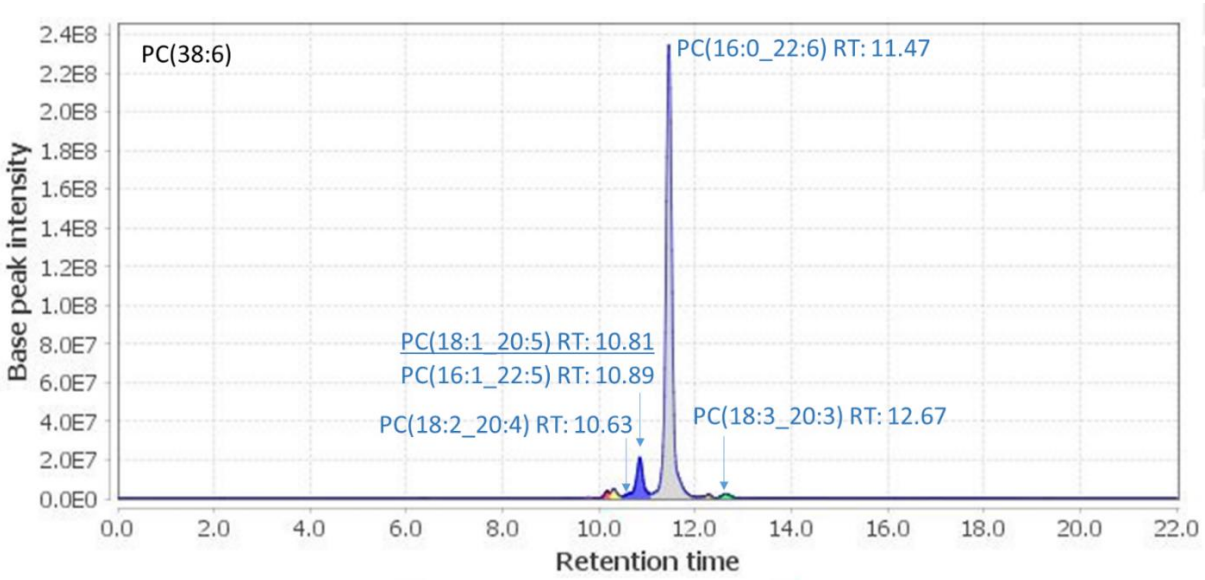


Figure 1-6. Deconvolution of chromatographic data for African sharptooth catfish species plasma using MZmine for the *m/z* corresponding to PC(38:6). Different colors indicate different peaks which are deconvoluted. Blue annotations indicate lipid names based on MS/MS identification for each peak. Data acquired in-house.

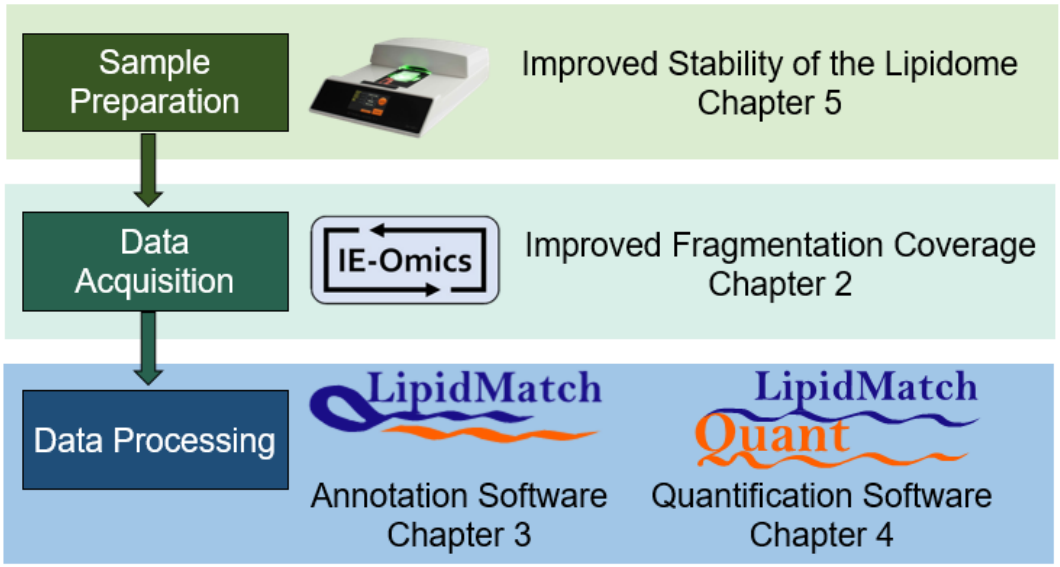


Figure 1-7. Dissertation chapters arranged within the first three major steps of the lipidomics workflow

CHAPTER 2
EXPANDING LIPIDOME COVERAGE USING LC-MS/MS DATA-DEPENDENT ACQUISITION WITH AUTOMATED EXCLUSION LIST GENERATION

The Case for Iterative Exclusion in Lipidomics Experiments

Electrospray ionization mass spectrometry (ESI-MS) is the most widely-employed ionization strategy for lipidomics [82], due to its ability to ionize the diverse range of structures and concentrations. However, isomeric species, for example, lipid species containing different fatty acid constituents with the same total number of carbons and degrees of unsaturation (e.g. PC(16:0_20:1) and PC(18:0_18:1)), cannot be separated using ESI-MS alone. If this structural detail is desired, one solution is to employ liquid chromatography to separate the isomers for quantification based on polarity using reverse phase chromatography, in combination with tandem mass spectrometry (MS/MS) to identify the fatty acid constituents based on fragmentation patterns. However, this strategy is problematic because in order to deconvolute more lipid features for quantification narrower chromatographic peaks are required, thus limiting the number of MS/MS scans which can be obtained across the chromatographic peaks [83]. Therefore, within lipid-rich retention time and *m/z* regimes, numerous ions of different mass-to-charge ratios will be ionized at the same elution time, but only a few can be selected for fragmentation in a single injection. For lipidomic experiments where the lipids of interest are unknown, heuristic rules have been developed to fragment ions. One approach is to select ions with the highest intensity for fragmentation, commonly referred to as data-dependent topN (ddMS^2-topN). Due to concentration bias, this strategy could miss important less abundant lipid species, such as diglycerides and

Published by: Journal of the American Society for Mass Spectrometry

phosphatidylinositols in plasma, which are both important signal molecule classes [84, 85].

A strategy which overcomes the drawback of traditional ddMS^2-topN, in terms of the limited number of MS/MS spectra acquired at any given retention time, is to continuously repeat ddMS^2-topN analysis on the same sample, excluding previously selected precursors ions in each sequential analysis. Theoretically, iterative repeat injections can be used to acquire MS/MS of all precursor ions above background signal, providing a substantial wealth of information for identification. A schematic of this technique, iterative exclusion (IE), is shown in Figure 2-1. While still uncommon, this technique has been applied in the proteomics community. In 2009, Bendall *et al.* designed a proteomics software approach for their strategy termed iterative exclusion – mass spectrometry (IE-MS) [86], which excluded all ions in previous runs regardless of assignment. Using this technique, Bendall *et al.* identified 30% more proteins after 5 IE-MS acquisitions, compared to 5 repetitive traditional data-dependent scans. Rudomin *et al.* [87] were also able to identify 49% more proteins using this strategy. IE has been applied to LC/MS approaches with both LTQ-Orbitrap [88] and qTOF platforms [86] and has been shown to be advantageous for proteomic applications. Examples employing this approach include using IE to study post-translational modifications of proteins [89], discover previously unknown human embryonic growth promoters [86], identify genital track markers [88], characterize Matrigel marketed as a basement membrane matrix for stem cell growth [90], and track pH induced protein changes [91]. In all these applications, IE enabled the ability to characterize a greater variety of proteins, including trace proteins.

While numerous applications have shown the benefit of using IE across platforms for proteomics, most omics analyses do not take full advantage of IE. In part, this may be due to a lack of a simple software program capable of generating exclusion lists automatically as traditionally this is achieved manually. Furthermore, using available software, iterative acquisitions cost time and money, thus putting heavy emphasis on determining whether additional sample injections for IE are worth added instrument time. To this end, we have developed an R script named “IE-omics” for generating exclusion lists from open-source formatted data easily converted from various vendor formats; as a demonstration, we have applied it to lipidomics. IE-Omics is advantageous over the IE-MS script in providing multiple user parameters in a relatively simple interface and providing the ability to directly import multiple vendor formats.

Recently, IE has been adapted to other omics fields, such as in lipidomics and metabolomics, as noted in Sandra *et al.* [92] and Edmands *et al.* [93], respectively. For lipidomics, IE type analyses have been used in direct infusion approaches by increasing the duration of dynamic exclusion to the length of the analysis. For example, Nazari and Muddiman [94] used gas phase fractionation and dynamic exclusion to increase coverage of the lipidome, especially of low abundance species. Schwudke et al. [95] emulated precursor and neutral loss scanning using data-dependent analysis with a dynamic exclusion and inclusion list based workflow for increasing the lipidome coverage. The success of IE type approaches for direct infusion supports the utility of IE for LC-MS/MS. While direct infusion allows rapid biomarker discovery, LC-MS/MS has been found to be more comprehensive [96].

To our knowledge, no research has shown the benefit of applying IE approaches to lipidomics versus traditional LC-MS/MS. In addition, no omics studies have compared results across different matrices with varying amounts of features. Herein we report the use of our user customizable R script for IE to lipid extracts of both Red Cross plasma and *substantia nigra* brain tissue in both positive and negative polarity. The results show that due to the spectral density of lipid species in a chromatographic run, especially in positive ion mode, a substantial benefit is obtained using IE for LC-MS/MS based lipidomics.

Methods: Lipidomics Workflow Implementing Iterative Exclusion Software

Chemicals and Materials

Ammonium acetate and all analytical grade solvents (formic acid, chloroform, and methanol) were purchased from Fisher Scientific (Waltham, MA). All mobile phase solvents were Fisher Optima LC/MS grade (acetonitrile, isopropanol, and water). For Red Cross plasma the following lipid standards were used: triglyceride (TG(15:0/15:0/15:0) and TG(17:0/17:0/17:0)) purchased from Sigma-Aldrich (St. Louis, MO) and lysophosphatidylcholine (LPC(17:0) and LPC(19:0)), phosphatidylcholine (PC(17:0/17:0) and PC(19:0/19:0)), phosphatidylethanolamine (PE(15:0/15:0) and PE(17:0/17:0)), phosphatidylserine (PS(14:0/14:0) and PS(17:0/17:0)), and phosphatidylglycerol (PG(14:0/14:0) and PG(17:0/17:0)) purchased from Avanti Polar Lipids (Alabaster, Alabama). For substantia nigra samples the following standards were used: TG(15:0/15:0/15:0) from Sigma-Aldrich, and PC(19:0/19:0), DG(14:0/14:0), SM(d18:1/17:0), Cer(d18:1/17:0), $^{13}C_2$-cholesterol, PE(15:0/15:0), LPC(19:0), PG(14:0/14:0), and PS(14:0/14:0) purchased from Avanti Polar Lipids. All lipid standards were diluted prior to analysis in 1:2 (v:v) chloroform:methanol and a working

100 mg/L standard mix was then prepared by diluting the stock solution with the same solvent mixture.

Sample Preparation

Pooled Red Cross human EDTA plasma was purchased from the American Red Cross National Testing Laboratories (Detroit, MI) samples were stored at -80˚C. All plasma aliquots (40 µL) were thawed on ice prior to extraction.

Substantia nigra samples were obtained from C57BL/6 mice. The Institutional Animal Care and Use Committee (IACUC # 20148382) at the University of Florida approved the use of all mice and procedures. The mice were housed with a 12 h light-12 h dark schedule and were provided food and water *ad libitum*. Five month old mice were anesthetized using isoflurane vapors. The mice were sacrificed and whole brains were harvested immediately from the skull and placed on a glass petri dish. A scalpel was used to carefully remove the substantia nigra region of the brain. Upon receipt, the tissue was placed in a freezer maintained at -80°C for storage. A Bel-Art Mortar (Bel-Art Scienceware, Wayne, NJ) was used to pulverize the frozen tissue samples under liquid nitrogen. The frozen tissue powder (10-20 mg, in triplicate) was weighed in homogenization tubes containing zirconium beads (0.7 mm diameter, BioSpec Products, Bartlesville, OK).

Both Red Cross plasma and frozen substantia nigra tissue powder were extracted using the Folch method (2:1, v:v, chloroform:methanol) [39]. Briefly, 5 µL of internal standard (IS) mix (100 mg/L) was spiked into the Red Cross blood plasma (40 µL) on ice (IS info in chemicals and materials). For Red Cross plasma, 160 µL of methanol and 320 µL of chloroform was added to all samples. Samples were incubated on ice for 30 minutes and centrifuged at 4 ˚C for 5 min at 15000 rpm. To induce phase

separation, 150 µL of water was added and samples were incubated on ice for an additional 10 min. The organic layer was removed and the aqueous layer was re-extracted with 250 µL of chloroform:methanol (2:1, v:v). The organic layers were combined, evaporated under nitrogen, and reconstituted in 100 µL of isopropanol (for lipidomics). For substantia nigra tissue, the ground tissue was homogenized for 120 seconds, with 100 µL of methanol and 200 µL of chloroform for every 15 mg of tissue. 5 µL of internal standard (IS) mix (100 mg/L) was spiked into the chloroform:methanol (2:1, v:v) mixture before homogenization. Samples were incubated as for plasma, and water was added at a volume of one fourth of the Folch solvent. Substantia nigra was re-extracted with 50 µL of methanol and 100 µL of chloroform for every 15 mg of tissue and dried down. Substantia nigra samples were reconstituted in 200 µL of isopropanol.

Data Acquisition

For both lipidomics analyses, a Dionex Ultimate 3000 RS UHLPC system (Thermo Scientific, San Jose, CA) was employed. Ionization was performed with heated electrospray ionization probe (HESI II) and mass spectra acquired using a Q-Exactive Orbitrap (Thermo Scientific). Source parameters for lipidomics, in positive and negative polarity are provide in Table S-3. Samples were maintained at 4ºC in the autosampler. 2 µL of sample was injected onto a Waters Acquity BEH C18 column (50 mm x 2.1 mm, 1.7 µm, Waters, Milford, MA) maintained at 30ºC. For negative ion mode, 5 µL of sample was injected onto the column and analyzed with the same mass spectral parameters (Table S-2). A gradient ramp (Table S-1) was employed consisting of mobile phase C (60:40 acetonitrile:water, volume fraction) and mobile phase D (90:8:2 isopropanol:acetonitrile:water, volume fraction), both with 10 mmol/L ammonium formate and 0.1% formic acid.

Mass spectra were acquired in full scan mode using data-dependent top 5 analysis (ddMS^2-top5) in both positive and negative polarity with a mass resolution of 70,000. Full scan and ddMS^2-top5 scan parameters are shown in Table S-2. Before each analysis, the instrument was externally calibrated and at least 3 blanks were analyzed. Internal mass calibrants (lock masses) were used in positive ion mode and consisted of diisooctyl phthalate (*m/z* 391.2842) and polysiloxanes (*m/z* 371.1012 and 445.1200). No stable lock mass was observed to be used in negative ion mode. To compare iterative exclusion (IE) with traditional ddMS^2-top5 for lipidomics, a minimum of 4 sequential injections were analyzed by ddMS^2-top5 with IE and 4 without IE, for both negative and positive polarity analysis of Red Cross and substantia nigra lipid extracts. For excluding ions previously selected for fragmentation and placed on an exclusion list, a 10 ppm exclusion tolerance was used.

Software Platform for IE

A software program “IE-Omics” was written using R [97] to directly process a .ms2 file (converted using MSConvert [98]) and output an exclusion list in a format (.csv) which can be directly imported by Q-Exactive instruments. User-defined parameters include the retention time and *m/z* window for combining selected precursors to reduce the size of the exclusion list. A 0.02 *m/z* window and 0.3 min retention time window was used in this experiment. In this case, ions selected at *m/z* values of 400.02, 400.01, and 400.01, and respective retention times of 5.10, 5.15, and 5.30 min, would be combined in one row as *m/z* 400.02 excluded between 4.95 and 5.45 min. In addition, users can denote the number of times ions with the same *m/z* are selected before being considered background ions and excluded for the entire duration of the chromatographic run. In this experiment, a minimum of 15 instances of ions

selected for fragmentation with the same *m/z* was used, excluding these ions from 0 to 18 minutes. The IE-Omics script can be found in the supplementary information and the most up to date version on the Southeast Center for Integrated Metabolomics (SECIM) webpage (http://secim.ufl.edu/secim-tools/).

Feature Detection and Identification

Lipids were identified using both an in-house workflow, LipidMatch, and LipidSearch (Thermo Scientific, San Jose, CA) [99]. LipidMatch consists of R-scripts which identify lipids by matching MS/MS fragments indicative of class and fatty-acid constituents from experimental fragmentation to *in-silico* fragmentation libraries (covering over 250,000 lipid species across 56 lipid types). Only exact mass of the MS/MS fragments (not intensity) is used for matching. Before LipidMatch was applied, features were determined using MZmine 2.0 [100], with the batch mode file containing all the parameters. Both the MZmine batch file and LipidMatch software can be found at <http://secim.ufl.edu/secim-tools/> in the LipidMatch zip file. LipidMatch used the features exact mass determined by MZmine with a 10 ppm *m/z* window for precursor ion matching. Both for LipidMatch and LipidSearch, a 10 ppm *m/z* window was used for fragment matching. In LipidMatch, fragments were only considered confirmed if they were above 1000 intensity units and found in at least one scan within a 0.3 min window of the feature being identified; in LipidSearch, only lipids classified with grade A were kept. The 0.3 min retention time used for finding MS/MS scans and excluding precursors in sequential injections employing the IE-Omics script was close to the median of the full width at half maximum (FWHM) for all features (Figure S2-1). After lipid ions were annotated, redundant annotations, for example, different lipid ion adducts

of the same molecular species, were combined separately for positive and negative analysis.

Results and Discussion: Iterative Exclusion and Lipidome Coverage

By only fragmenting ions not selected in the previous injections, applying IE increased the coverage of both analyte and background ions for lipids (Figure S2-2 and Figure 2-3). As can be seen, for example for the background ion *m/z* 300.2253 in Figure 2-2, some ions selected in the first injection and placed on an exclusion list, were unexpectedly selected in the second injection. The reason for these ions not being excluded is that the mass trigger used to select ions for fragmentation are stored with a *m/z* value with two decimal digits in the Thermo .raw file. Therefore, the *m/z* 300.23 was placed on an exclusion list, and using a 10 ppm exclusion tolerance ions from 300.2270 to 300.2330 were excluded. In the second injection the ion was measured at 300.2253 and therefore was not excluded, and again was placed on an exclusion list at 300.23. This problem can be overcome by either changes in Thermo .raw data storage, in order to store the mass trigger for obtaining MS/MS past the 3rd decimal point (which has been implemented in the Q-Exactive Plus and HF), or by the user increasing the exclusion tolerance, such as to 100 ppm. Since different ions with *m/z* values within 100 ppm will all be isolated and fragmented using a 1 Da isolation window, this solution would be sufficient. Either modification would ensure that ions isolated and fragmented once are never isolated and fragmented again, thereby decreasing the number of injections needed to fragment all ions of interest.

It is a well-known that background ions compete with analytes of interest for selection and fragmentation. Therefore, by discerning the background ion patterns and automatically placing those ions on an exclusion list, analyte coverage can be

increased. A background ion from the mobile phase, ESI source, or from column bleed (to name a few sources), can be discerned by a single *m/z* covering a large portion of the retention time region, as in 300.2253 discussed above. An exclusion list for background ions can often be generated by running several blank injections, but this can be an inefficient process; when the column or mobile phase changes, a new list would need to be created. From this IE type of run, it can be seen that a large portion of ions selected and fragmented are background ions as depicted by a horizontal pattern across the chromatographic run (Figure 2-2). Therefore, this pattern can be readily used to exclude background ions in the IE-Omics software.

Using the default IE-Omics parameters, if the same *m/z* is selected more than 15 times, with each instance being at least 0.15 minutes apart, that *m/z* is annotated as background and placed on an exclusion list across the entire analysis time. For example, *m/z* 391.28 was excluded across the entire retention time in the second injection, after being selected in the first injection of Red Cross plasma 134 times (Figure 2-2). Annotated background ions can also be excluded in future experiments. After the first injection, 17 ions were automatically annotated as background ions by IE-Omics software in positive ion mode of Red Cross plasma. After 6 injections, 54 *m/z* values were annotated as background ions according to this algorithm.

In addition to generation of exclusion lists of background ions, IE also enhanced coverage of the lipidome. When comparing the second to first injection after applying IE, it can be seen that many additional unique precursors are selected in both the glycerophospholipid (GPL) (about *m/z* 700-900 at 5-10 min) and triglyceride (TG) regions (about *m/z* 700-1100 at 11-16 min) (Figure 2-2). Figure 2-3 compares unique

ions selected for fragmentation in positive mode analysis of Red Cross plasma lipid extracts in 6 sequential injections without IE applied (Figure 2-3a) and with IE applied (Figure 2-3B). The IE-omics approach shows that after 6 sequential injections, the number of unique ions fragments is substantially higher (Figure 2-3B). As discussed previously, both new background ion signatures and lipid ions (as can be seen in the GPL and TG region) are selected using IE (Figure 2-3). An analogous figure for substantia nigra is shown in the supplementary information (Figure S2-2), with a zoom in of the glycerophospholipid (GPL) region overlaid with unique molecular species annotated by LipidMatch (Figure S2-3).

Figure 2-4 displays the cumulative number of features with lipid identifications across injection number using LipidMatch from both positive and negative mode analysis of plasma and substantia nigra lipid extracts. In all cases, a greater number of features were identified as lipids using the IE approach compared to a traditional $ddMS^2$-top5 approach. The application of IE was most advantageous in positive ion mode analysis of plasma and substantia nigra lipid extracts. For plasma extracts in positive mode, applying IE and using 6 sequential injections increased the coverage of features annotated with unique molecular species by 69 % compared to the traditional $ddMS^2$-topN approach across 6 sequential injections. A total of 728 unique lipid molecular species were identified with IE, compared to 431 without IE (Figure 2-4A). In negative mode analysis of Red Cross plasma, only 10 % more identifications were obtained using IE. In positive mode analysis of substantia nigra, after the 5 sequential injections, 40 % more features were identified using IE compared to without IE, while in negative mode analysis, 18 % more identifications were obtained when applying IE.

Applying a different identification software, LipidSearch, provided the same general trend, with IE providing the most advantage in positive mode analysis of Red Cross plasma and substantia nigra (69 % and 34 % more identifications, respectively) and least advantage in negative mode analysis of Red Cross plasma and substantia nigra (18 % and 4 %, respectively) (Figure S2-4). Unique annotations of lipid molecular species with retention time information, exact *m/z* from full scan data, and average peak intensity compiled across all sequential injections for positive and negative polarity analysis of Red Cross plasma and substantia nigra can be found in Table S-5 and S-4, for LipidSearch and LipidMatch, respectively. Fragments observed for identification by LipidSearch are also included in Table S-5, and fragmentation criteria for LipidMatch is included in Table S-6.

Based on these results, it is clear that the number of additional identifications obtained when applying IE depends on sample type and the polarity measured by the mass spectrometer. It is expected that if mass spectrum is sparse, a traditional ddMS^2 approach will likely select the majority of ions above an MS^2 threshold limit. In the lipidomic analyses, where applying IE was less advantageous in negative ion mode, negative ion spectra showed fewer ions than positive ion spectra. For example, the number of features (which is related to spectral sparseness), was drastically lower in negative mode than positive mode, with only 4258 features in Red Cross negative mode data versus 19,231 features in Red Cross positive mode data. Therefore, after applying exclusion lists generated by IE, fewer precursors remain above the threshold to be selected for fragmentation in negative ion mode. For example, in negative polarity analysis of plasma, MS/MS scans drastically declined from the first to the fifth

sequential injection (from 2491 to 414 scans), showing depletion of precursors for selection, while in positive polarity there was less of a decline in MS/MS scans (from 2746 to 2581 scans) (Figure S2-5). Therefore, the number of sequential injections required may vary depending on spectral density, and spectral density will be a major factor in determining the additional benefit of IE. For example, increasing the chromatographic gradient time would increase separation of lipids while decreasing spectral density at a given time point, and hence potentially decreasing the advantage of applying IE versus traditional ddMS^2-topN approaches.

It should be noted that additional identifications using IE are only useful if they provide unique information. After excluding previously selected high abundance lipids for fragmentation, sequential injections should provide fragmentation of lower abundance species when applying IE. Often less abundant or trace species serve as critical biomarkers, such as phosphatidylinositol (PI), which is an important signaling molecule class. Phosphatidylcholine (PC) concentrations in plasma, for example, are about 20-fold higher than concentrations of phosphatidylinositol (PI), phosphatidylserine (PS), and phosphatidic acid (PA) combined [101]. After applying IE in the second injection, peak heights of identified lipids were significantly lower than the initial injection for positive and negative polarity analysis of both plasma and substantia nigra lipid extracts (p-value < 0.05) (Figure 2-5). For plasma in positive ion mode, the average intensity of selected precursors seemed to continue to decrease using IE up to the fourth injection, although not significantly (Figure S2-6; p-value > 0.05). Exclusion of trace ions close to the threshold intensity for fragmentation in certain chromatographic regions, while high intensity species, such as in the TG region where spectra are dense,

continue to be selected, would explain why the average intensity of ions does not continue to decrease after a certain number of sequential injections. This is supported by the fact that the number of precursors selected declines across sequential injections when applying IE, and therefore in certain regions ions are no longer being selected for fragmentation (Figure S2-5). For example, the TG region contained 4819 features above 5×10^4 in 4 minutes (11 to 15 minutes) in Red Cross plasma, while the lysophospholipid region contained 3859 features in 4 minutes (0.5 to 4.5 minutes).

The reduced intensity of precursor ions selected after applying IE suggests lower MS/MS spectral quality. This is especially true for positive ion mode fragmentation of most glycerophospholipids, where fatty acyl indicative fragments are of low abundance. To determine the quality of MS/MS spectra across sequential injections, the percentage of lipids identified with grade A, calculated by (A / (A + B + C)), was determined using LipidSearch. These grades are based on the number of fragments identified which contain species specific structural information, with lipid identifications graded A having the most structural information in MS/MS spectra (for example fragments indication both fatty acyls and the head group of glycerophospholipid species). Following a similar trend to the selected ion signal, the sequential injections after applying IE had a general drop in percent A, and hence decrease in MS/MS spectral quality (Figure S2-7). Injections in which IE was applied had significantly lower average percentages of A compared to sequential injections without IE, for negative and positive polarity analysis of substantia nigra tissue lipid extracts (p-value < 0.05) and for positive polarity analysis of Red Cross plasma (p-value < 0.005). No significant difference was observed for negative polarity

analysis of Red Cross plasma. Therefore, unique identifications provided by IE of low abundance species often provide less structural information and are more tentative.

Diglycerides (DG) are often present at low abundance and have been noted as important signaling molecules. In plasma lipid extracts, lower abundance DG species were identified after sequential injections applying IE (Figure S2-8). In the initial injection, all DG species identified except one had extracted mass chromatographic peak heights of 10^6 or 10^7, while after the sixth injection applying IE, all species identified had peak heights of 10^5 (Table 2-1). All DGs at the level of carbons and double bonds in Table 2-1 have been confirmed previously in human plasma [102], except for DG(30:3), identified as DG(12:0_18:3). In addition, all fatty acids constituents contained in DGs have been confirmed in plasma using fatty acyl profiling [102] or have been found in DGs using derivatization [103]. The lower abundance DG species identified after applying IE contained both odd-chain (15:0, 17:1, and 17:2), and shorter chain (12:0, 14:0, and 14:1) species, which were not identified without IE (Table 2-1). These fatty acids are in lower abundance in human plasma [102] and odd-chain species could represent exogenous fatty acid species or those produced by gut microbiota [104]. In this case, these species represent additional biological information otherwise not obtained.

Coverage was improved for certain lipid classes using the IE approach. The majority of unique lipid molecules identified by IE in positive analysis of Red Cross plasma, but not by the traditional data-dependent approach, were mainly glycerolipids, specifically, TGs, oxidized TGs, ether-linked TGs and DGs (Figure 2-6A and Figure 2-6C). These molecular species were of low intensity and were present in

chromatographic regions where mass spectra were dense. Hence, using traditional approaches, these lower intensity ions generally never make it on the list of the top-5 most intense ions to be included for fragmentation. In addition, there was minimal identification of ether-linked TGs, oxidized lyso-phosphatidylcholines (OxLPC), and acyl-carnitines using the traditional ddMS^2-top5 approach. Applying IE significantly increased the coverage of these lipid classes. Ether-linked TGs are a trace fraction of the total TGs in blood, for example, only comprising of 0.1 % of chylomicrons in human blood plasma, where they have been noted to concentrate [105]. By applying IE, these low abundant ions (making up less than 0.1% of TG peak area signal) were selected for fragmentation and tentatively identified by exact mass of the precursor and exact masses of the neutral losses of the two non-ether fatty acyl constituents.

In substantia nigra positive mode analysis (Figure 2-6B and Figure 2-6D), IE improved the coverage of phosphatidylserine (PS), oxidized phosphatidylcholine (OxPC), phosphatidylglycerol (PG) and sulfatide species, which were minimally covered by the traditional approach. In negative ion mode analysis of substantia nigra tissue, using the traditional ddMS^2-top5 approach, there was no coverage of sulfatides and minimal coverage of phosphatidic acid (PA). Applying IE significantly increased coverage of both of these species (Figure S2-9). These findings highlight that IE not only increases the total number of lipid identifications, but increases identifications of trace lipid species of potential interest, which are minimally covered by traditional approaches.

Future developments will continue to increase the advantages of applying IE-Omics. Currently, the script is not integrated in Xcalibur software, and therefore

exclusion lists are not generated in real time and must be uploaded into new method files before each iterative injection. In our lipidomics workflow, we suggest using 3 to 4 iterative injections on pooled samples, which can be used to identify features of a given sample group. Therefore, this method is sufficient for lipid identification in large quantitative studies to determine biomarkers where thousands of samples are required, as only a few additional injections are used for IE, and hence there is minimal addition of acquisition time. Fewer injections may be required if the exclusion window is increased from 10 ppm, for example, to 100 ppm. By increasing the exclusion window, isobaric ions will only be selected in one injection, reducing the number of injections needed to select all ions above a certain threshold. In the future fully automated exclusion list generation may be developed.

Conclusion: Iterative Exclusion Increases Lipidome Coverage

We have semi-automated the IE approach using a simple open source R script. The script uses open source formatted files which can be converted from various vendor formats and produces an exclusion list in a vendor neutral format required for importing into Thermo-Scientific instruments (csv). Features include smart exclusion list generation, which combines ions selected in similar *m/z* and retention time windows to generate a shorter exclusion list, and automatic annotation of background ions. After applying the software, IE-Omics, to lipidomic datasets in Red Cross plasma and substantia nigra brain tissue lipid extracts, IE was shown to be most advantageous in complex matrices with a high number of analyte species. Applying IE to lipidomics analyses in certain cases increased identifications by over 50 %. The greatest advantage using IE was shown in positive ion mode and in Red Cross plasma versus substantia nigra lipid extracts, where spectra were most dense. In lipidomics, trace

species, such as odd-chained and short-chained DGs, were identified only after applying the IE technique.

Future data acquisition strategies, for example only including precursor ions for fragmentation which match lipid masses and identifying polymer patterns for exclusion, could prove advantageous. In most cases, however, such as in negative mode, after only using a few sequential injections, all ions above the threshold limit for fragmentation were selected, and therefore new data-acquisition methods would not provide additional advantage in terms of MS/MS spectral coverage. New data-acquisition methods might be able to reduce the number of injections needed for coverage of the majority of lipid ions and notify the user when additional injections are no longer required. In conclusion, applying IE expands the scope of the lipidome covered, both increasing the total number and diversity of lipids identified.

Table 2-1. Comparison of diglyceride (DG) peak heights and fatty acid compositions. Data are from Red Cross plasma acquired in positive polarity. A) The first ddMS^2-top5 acquisition using LipidMatch (IE 1) and B) the 6th injection after applying an exclusion list using the algorithm described in this paper (IE 6).

a)	DG(C:DB)	Peak Height	DG fatty acid chains
	DG(32:1)	1.7 x 10^7	DG(16:0_18:1)
			DG(16:1_18:0)
	DG(32:2)	2.0 x 10^7	DG(16:0_18:2)
			DG(16:1_18:1)
	DG(36:1)	2.4 x 10^6	DG(16:0_20:1)
			DG(18:0_18:1)
	DG(36:2)	1.9 x 10^7	DG(16:0_20:2)
			DG(16:1_20:1)
			DG(18:0_18:2)
			DG(18:1_18:1)
	DG(38:5)	5.1 x 10^6	DG(16:0_22:5)
			DG(18:1_20:4)
			DG(18:2_20:3)
	DG(36:3)	3.5 x 10^7	DG(16:1_20:2)
			DG(18:0_18:3)
			DG(18:1_18:2)
	DG(36:4)	1.0 x 10^7	DG(16:1_20:3)
			DG(18:1_18:3)
			DG(18:2_18:2)
	DG(36:3)	4.7 x 10^4	DG(18:1_18:2)

b)	DG(C:DB)	Peak Height	DG fatty acid chains
	DG(30:0)	4.2 x 10^5	DG(12:0_18:0)
			DG(14:0_16:0)
			DG(15:0_15:0)
	DG(30:1)	3.5 x 10^5	DG(12:0_18:1)
			DG(14:0_16:1)
			DG(14:1_16:0)
	DG(30:3)	2.6 x 10^5	DG(12:0_18:3)
	DG(32:3)	2.2 x 10^5	DG(14:0_18:3)
			DG(14:1_18:2)
	DG(34:4)	3.6 x 10^5	DG(14:0_20:4)
			DG(16:0_18:4)
	DG(36:5)	3.3 x 10^5	DG(16:0_20:5)

Table 2-1. Continued

b)	DG(C:DB)	Peak Height	DG fatty acid chains
	DG(38:4)	3.9×10^5	DG(16:0_22:4)
			DG(18:1_20:3)
	DG(35:3)	3.0×10^5	DG(17:1_18:2)
			DG(17:2_18:1)
	DG(38:2)	3.4×10^5	DG(18:1_20:1)
	DG(38:4)	5.8×10^5	DG(18:1_20:3)
	DG(38:5)	6.9×10^5	DG(18:1_20:4)
			DG(18:2_20:3)
	DG(40:7)	9.1×10^5	DG(18:1_22:6)
	DG(40:7)	4.1×10^5	DG(18:2_22:5)

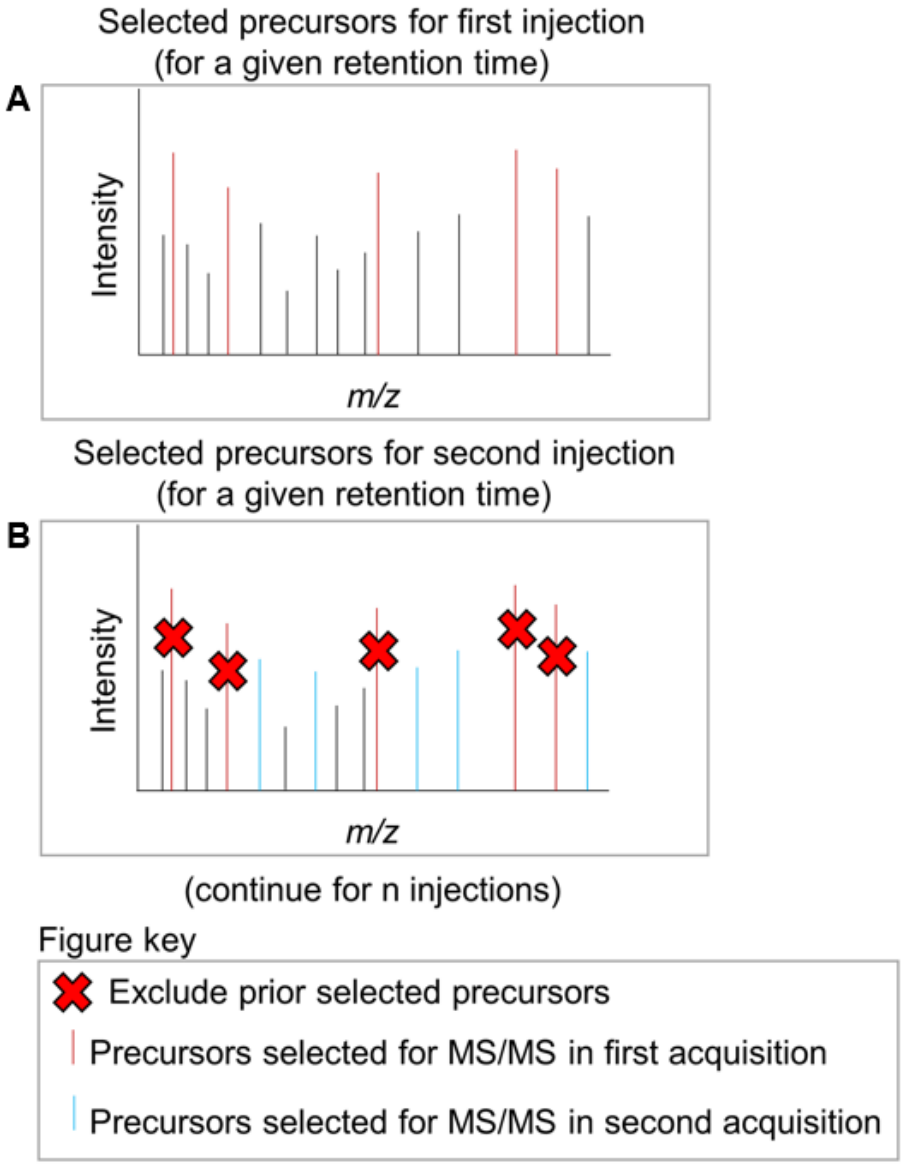


Figure 2-1. Strategy for iterative exclusion based data-dependent topN analysis (IE-ddMS^2-topN). Multiple injections of a sample are analyzed. A) The first injection is used to create an exclusion list and B) this exclusion list is applied to the second injection. Hence, the next top-n most abundant ions are selected and this process can be continued for n injections.

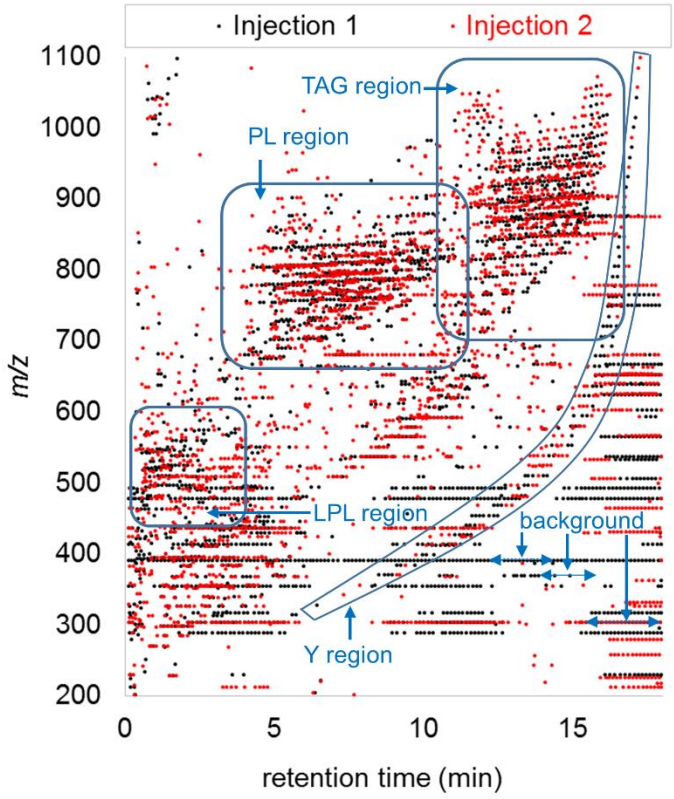


Figure 2-2. Selected precursor ions retention time and *m/z* for Red Cross plasma compared between the first injection (black dots) and second injection (red dots) with iterative exclusion-based (IE) ddMS^2 applied. The Y region is an unknown region with molecules separated by 14 Da corresponding to CH_2 repeating units, likely representing polymer species. Three background ions are indicated with arrows, which were selected at *m/z* 391.28, 354.29, and 303.23 from highest to lowest mass, respectively.

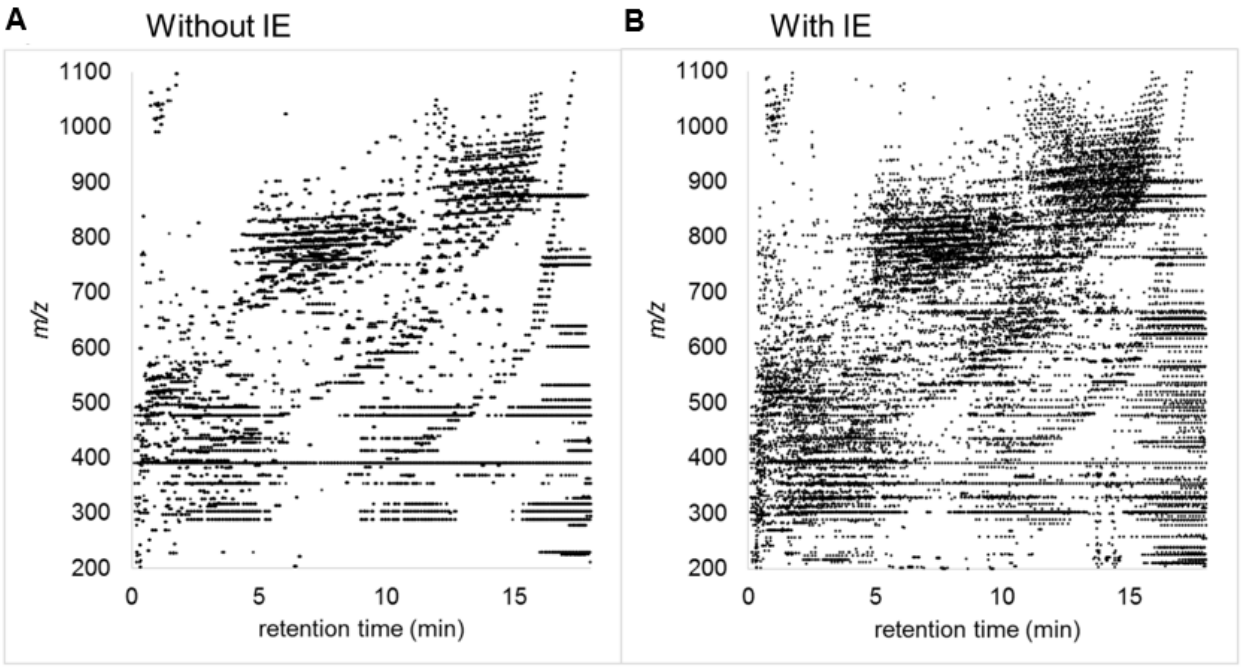


Figure 2-3. The use of iterative exclusion increased fragmentation coverage. Selected precursor ions *m/z* and retention times for: A) 6 repetitive injections using the traditional ddMS^2 approach and B) iterative based-exclusion ddMS^2 (IE-ddMS^2) for Red Cross plasma lipid extracts analyzed in positive mode.

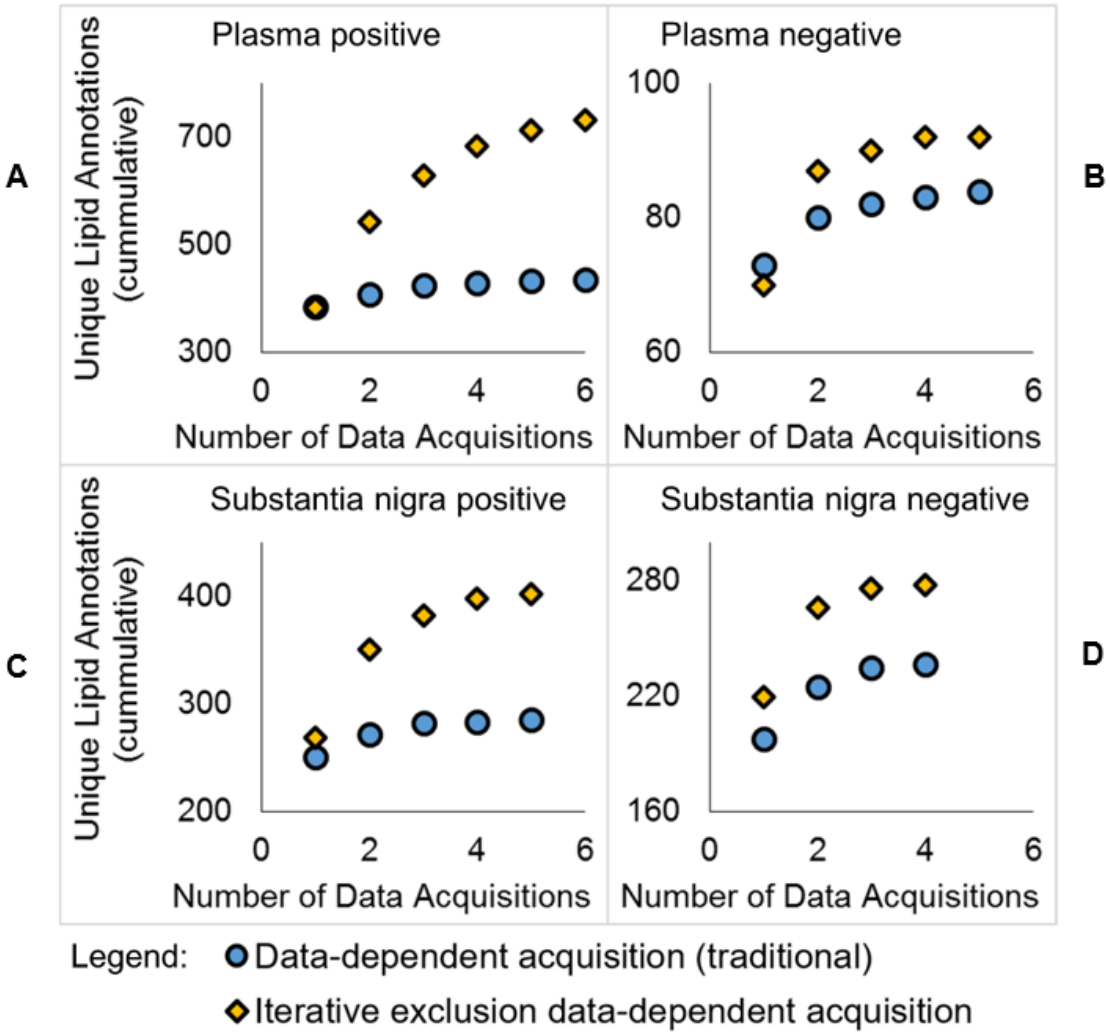


Figure 2-4. Cumulative unique lipid molecular identifications using LipidMatch software across multiple data acquisitions are shown. Iterative exclusion-based data-dependent top5 (IE-ddMS^2-top5) described in this paper is compared with traditional ddMS^2 top5 for extracts of: A) Red Cross plasma in positive mode and B) negative mode, and C) extracts of substantia nigra in positive mode and D) negative mode.

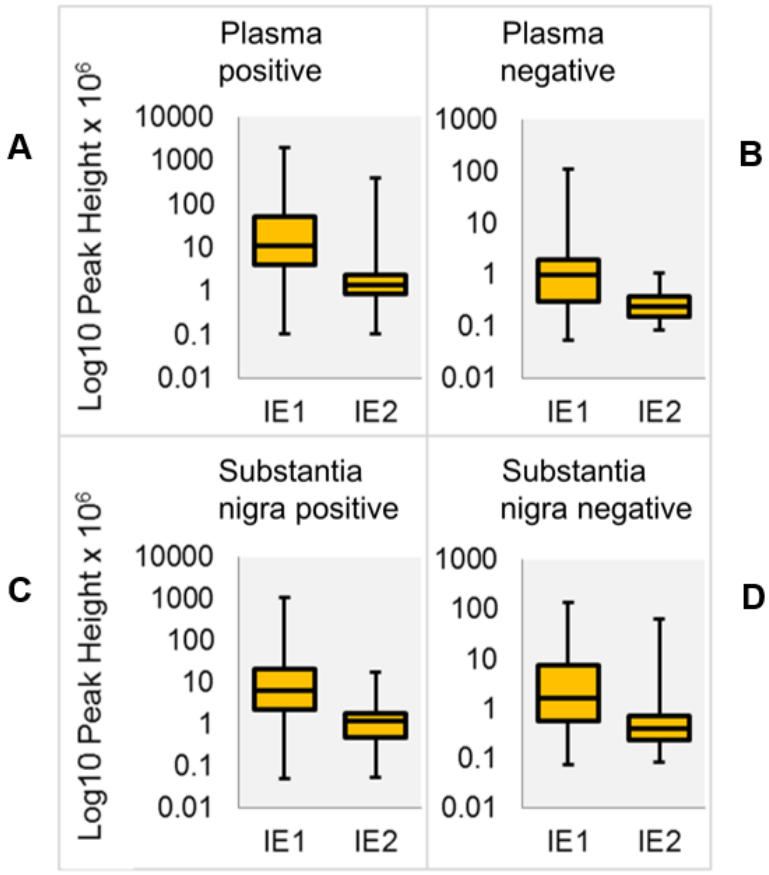


Figure 2-5. Boxplots of log transformed peak heights (base 10) from MZmine for unique lipid molecules identified in the first ddMS^2-top5 acquisition using LipidMatch (IE1) and after applying an exclusion list using the algorithm described in this paper (IE2). All differences where highly significant with a p-value for a Student t-test less than 0.001. Comparisons are made for: A) extracts of Red Cross plasma in positive mode and B) negative mode, and C) extracts of substantia nigra in positive mode D) and negative mode.

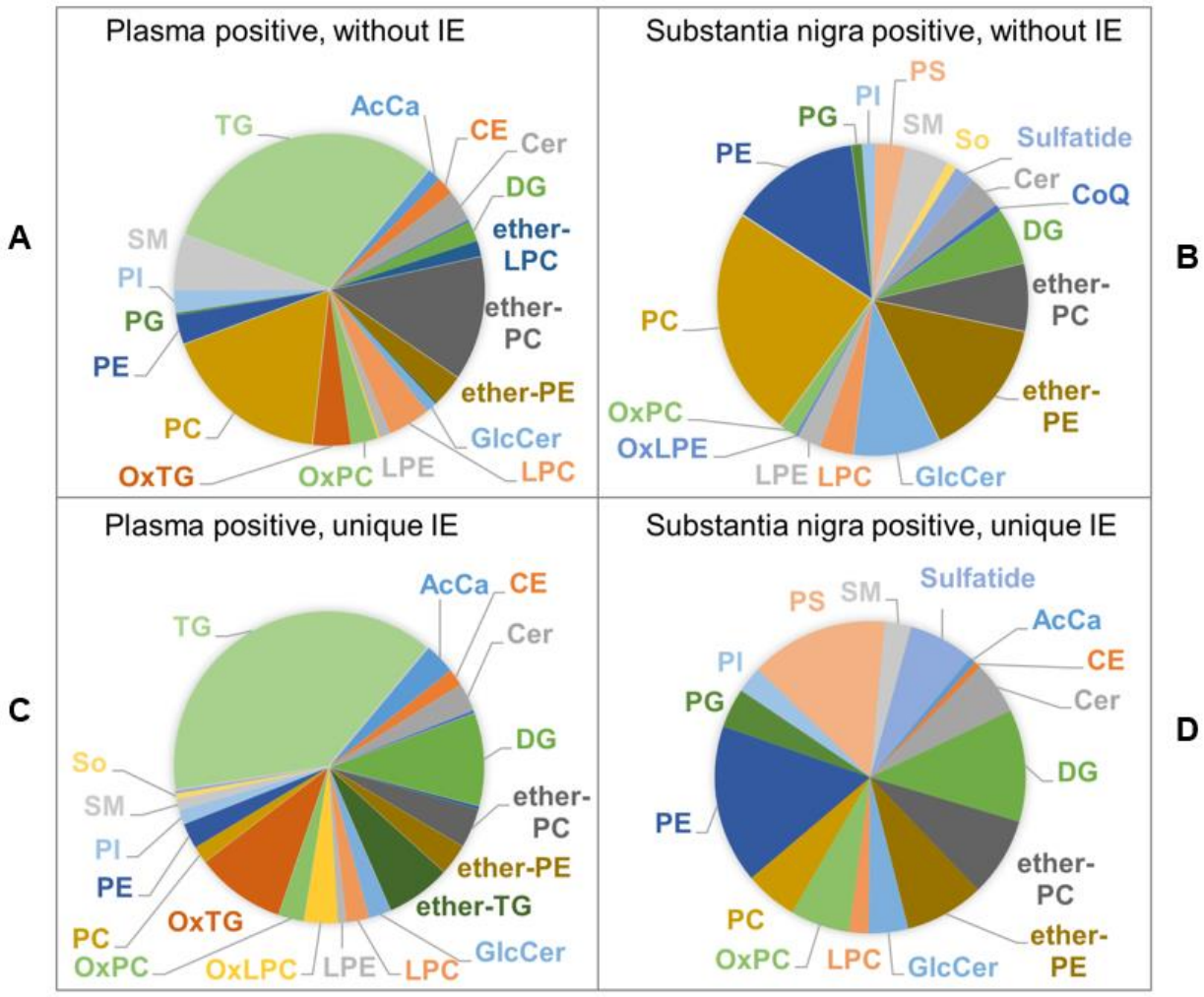


Figure 2-6. Distribution of lipids identified using LipidMatch by lipid class using iterative exclusion-based data-dependent top5 (IE-ddMS^2-top5) acquisitions in positive ion mode. A) The lipid class distribution of all identifications across sequential injections using the traditional ddMS^2-top5 approach is shown for Red Cross plasma and B) substantia nigra tissue lipid extracts; C) The distribution of additional unique lipid molecular identifications after applying iterative exclusion (IE) across lipid classes are shown for Red Cross plasma and D) substantia nigra lipid extracts.

CHAPTER 3
LIPIDMATCH: AN AUTOMATED WORKFLOW FOR RULE-BASED LIPID IDENTIFICATION USING UNTARGETED HIGH-RESOLUTION TANDEM MASS SPECTROMETRY DATA

The Challenges of Lipid Identification

In comparison to proteomics, lipidomics is an emerging technique which currently lacks community-wide agreement concerning the best software choice for the comprehensive and accurate identification of lipids based on chromatographic and tandem mass spectrometric data. A major challenge is the limited number of synthesized standards available, making it difficult to cover the much larger variety of lipid structures for MS/MS spectral matching. In the absence of authentic standards, this challenge has been partially ameliorated by developing *in-silico* libraries for acyl-containing lipids. For example, in 2013, Kind *et al.* released LipidBlast [106], developing a computer generated MS/MS library of 119,200 lipids across 26 lipid classes, which includes predicted mass/intensity pairs.

A second major challenge is the accurate annotation of lipid identifications based on the fragmentation observed [107]. The annotation depends on the structural resolution, which is the structural detail inferred by experimental data, specifically the MS/MS spectra. Structural resolution for lipids is dependent on specific structural characteristics known, such as double bond location, geometric isomerism (*cis* versus *trans*), and the position, lengths and degrees of unsaturation of fatty acyl constituents. For example, if only the exact mass of the precursor and choline head group of a phosphatidylcholine species is observed, the species can only be annotated by total

Published by: BMC Bioinformatics
Published by: Biochimica et Biophysica Acta (BBA)-Molecular and Cell Biology of Lipids

carbons and degrees of unsaturation (e.g. annotated as PC(32:1)) (assuming no overlap from fragmentation of other choline containing species, such as the ^{13}C isotopic peaks of SM). If the precursor mass and fatty acyl fragments are observed, then the lipid can be identified by acyl-constituents (eg. PC(16:0_18:1)), with an underscore denoting that the position of the fatty acyl constituent on the backbone is unknown. For most lipid types, this is the limit of structural resolution that can be accurately annotated using UHPLC-HRMS/MS without specialized or additional approaches. Currently, most lipidomics software over-report structural resolution, which can lead to incorrect biological interpretation of the data.[108]

A third challenge for lipid identification is the fact that features (peaks defined by a mass-to-charge ratio (*m/z*) and retention time) often contain multiple co-eluting molecules with similar *m/z* values. One common case is lipids sharing the same class, total carbons and degrees of unsaturation, but different acyl constituents, for example PC(18:0_18:1) and PC(16:0_20:1). This overlap reduces spectral similarity scores, which are used for identification by most software.

To overcome these challenges, we have developed LipidMatch. LipidMatch currently contains the most comprehensive lipid fragmentation libraries of freely available software, when ranked by the number of lipid types. LipidMatch includes *in-silico* libraries with over 250,000 lipid species across 56 lipid types, including oxidized lipids. LipidMatch incorporates user-modifiable, rule-based lipid identification, which allows for accurate lipid annotation in regards to structural resolution. In addition, if multiple identifications exist for one feature, LipidMatch outputs include all possible identifications ranked by summed fragment intensities.

Lipid Annotation Guidelines for Correctly Reporting Structural Resolution

While the lipidome - the entire collection of individual lipid species in cells, tissues or biofluids - has been estimated to be composed of 1,000 to more than 180,000 molecular lipid species [25, 26], these estimations do not consider isomeric lipid species with different fatty acyl double-bond positions and configurations (*cis* or *trans*), positional isomers (e.g., *sn*1, *sn*2), and stereoisomers (*R* or *S*). Ekroos et al. determined that the number of phosphatidylcholine (PC) positional isomers in Madin-Darby canine kidney II cells nearly doubled the total number of individual lipid species [109], which highlights the substantial presence of lipid isomers in nature. Furthermore, these lipid isomers can also exhibit a variety of specific biological roles. For example, the acyl position of membrane lipids can impact the enzymatic activity that occurs within cellular membranes [110]. Shinzawa-Itoh et al. [79] found biological specificity of acyl chain double bond configurations; though the mitochondrial inner membrane where bovine cytochrome c oxidase (CcO) acquires its phospholipids contains *trans*-vaccenate, only *cis*-vaccenate is incorporated into subunit III of CcO [79]. Researchers have shown differing roles of individual conjugated linoleic acid (CLA) isomers; while the *cis*-9,*trans*-11 isomer has been shown to more broadly inhibit tumorigenesis *in vitro*, the *trans*-10,*cis*-12 isomer has been shown to increase concentrations of human blood lipids, such as triglycerides (TG) and the ratio of LDL to HDL cholesterol, when compared to the *cis*-9,*trans*-11 isomer [76–78].

Structural elucidation is vital in ensuring that biological properties are properly associated with the correct lipid species. Therefore in this section we provide guidelines for annotating lipids and discuss the limitations in biological interpretation of lipid

species. Software, such as LipidMatch, which applies these guidelines is essential to implementation and harmonization by the wider lipidomics community.

Community - accepted guidelines for lipid annotations [111–113] generated/accepted by the International Lipids Classification and Nomenclature Committee have been implemented and promoted by the LIPID Metabolites And Pathways Strategy (LIPID MAPS) consortium [23, 81, 114], and are meant to completely characterize the lipid molecule as shown in Figure 3-1. However, conventional tandem mass spectrometric experiments cannot be used to generate all structural information of a given lipid molecule. Therefore, shorthand notation has been proposed to only confer the level of structural detail known based on experimental data [107]. We will define this structural detail for a given lipid species, as the *structural resolution*. Moreover, we summarize existing guidelines supplemented by new recommendations to prevent over-reporting of lipid structural resolution and to further encourage the use of a common nomenclature system for lipidomics.

Do not annotate lipids using only exact mass. Often researchers entering the lipidomics field will annotate peaks and features based on exact mass only, for example as in Gerspach *et al.* [115]. A feature is a peak, or group of peaks across numerous samples, represented by a specific *m/z* and any other measurements, such as a specific retention time if chromatography is used, or drift time if ion mobility is used [116]. Since the lipidome is diverse, with enormous overlap in exact mass, we strongly warn against annotating features using only exact mass, especially for previously uncharacterized sample types. It is important to note that exact mass search engines,

such as Metlin and LIPID MAPS, provide lipid matches annotated as fully characterized molecular species, which cannot be elucidated from exact mass alone.

Annotate by sum composition when class specific fragmentation is observed. The most basic annotation of lipids is by lipid class and the sum composition of carbons and double bonds in the lipid fatty acyl constituents (Table 3-1). The sum composition annotation is useful in cases where the majority of fragmentation intensity is in class specific fragments. Examples include phosphatidylethanolamine (PE) $[M+H]^+$ (neutral loss (NL) of *m/z* 141.0191) [117], phosphatidylinositol (PI) $[M + NH_4]^+$ (NL of *m/z* 277.0562) [95], and sulfatide $[M-H]^-$ (*m/z* 96.9601) [118], with these base peaks in the fragmentation specific to the lipid head group. Annotation by lipid class can often lead to false positives if fragments are not specific to only that lipid class. A common case is the incorrect annotation of protonated adducts of sphingomyelin (SM) and phosphatidylcholine (PC) and their lysolipid, oxidized lipid, and ether-linked lipid corollaries using *m/z* 184.0733, for example as in Jin *et al.* [119]. Isobaric isotopic peaks of co-eluting SM and PC species will be co-isolated for fragmentation and hence the lipid class represented by *m/z* 184.0733 is ambiguous. In this case, identifications should be noted as tentative unless reconstructed ion chromatograms of the PC and SM species within 3 daltons do not overlap or fatty acyl constituents are observed.

Denote fatty acyl constituents only when fatty acyl fragment(s) are observed. Lipids can often be annotated based on fatty acyl constituents (Table 3-1). Technically, in lipids with two fatty acyl constituents, only one fatty acyl constituent is needed for identification, as the other can be deduced using the exact mass of the precursor. This can be a helpful strategy when *sn1*- and *sn*2-linked fatty acyl

constituents have different fragment efficiencies, as in PC $[M+H]^+$ adducts [120]. Without assumptions based on biology or specific approaches, identification by fatty acyl constituent constituents is often the limit of structural resolution that can be obtained.

Use the underscore to annotate lipid species with unknown positional isomers. Traditional UHPLC-HRMS platforms with tandem MS do not provide information about the double bond position or orientation, stereochemistry, and in many instances, the position of the fatty acyl constituent on the glycerol backbone (*sn*1 or *sn*2). Lipid identifications where the positional isomeric level of the fatty acyl constituents is known is indicated by a slash "/". The underscore "_" was proposed by Liebisch et al. (2013) [121] for instances where there is certainty in the composition of the fatty acyl constituents, but not their placement on the glycerol backbone. Despite the proposed shorthand notation, there has been a mix of annotation styles present in literature. For example, of the lipidomics research articles published in 2017 determined on Science Direct (accessed 02/09/2017), 3 articles [119, 122, 123] incorrectly used "/", 3 articles [124–126] used "-", and 1 article [127] used "_" between fatty acyl constituents, when positional isomers were not identified. One potential source of confusion in annotating lipids is that current lipid identification software, for example LipidSearch, MS-DIAL[128], LipidBlast[106], and Greazy[129] all employ "/" when fatty acyl positions are not known. However, to further advance the lipidomics community, all lipid identification software should improve lipid annotation by incorporating the slash "/" or "_" correctly based on the MS/MS data. Otherwise, an incorrect level of structural detail is assigned to the lipid annotation, providing the user with a level of certainty,

which is misleading for biomarker discovery, disease etiology studies, and translational science with other omics areas.

Report plasmanyl species using O- and plasmenyl species using P-. Some of the most problematic cases for lipid annotation include plasmenyl and plasmanyl ether-linked species, which are depicted in Figure 3-2. One problem is the use of varying annotation style. For plasmanyl lipid species, lipids are often annotated using an "e" or an "O-", while plasmenyl lipids are often annotated using a lowercase "p" or a capital "P-". We suggest using "O-" and "P-", the annotation style used by LIPID MAPS [23, 81, 121]. Another problem arises because the vinyl ether linkage in plasmenyl species and the ether linkage in plasmanyl species only differ by a degree of unsaturation, leading to differing structures with the same molecular formula. For example, plasmanyl PE(O-16:0/22:6) will have the exact same mass as plasmenyl PE(P-16:0/22:5) and cannot be distinguished based on class specific fragments. In this case we suggest including both annotation by sum composition, for example PE(P-38:5) and PE(O-38:6). In the case of ether-linked PC, the formate adduct will yield an abundant *sn*2 fatty acyl fragment when fragmented in negative ion mode; the ether-linked PE species can also be distinguished using fragmentation [130]. Hence, the vinyl ether- and ether-linked lipids can be distinguished using fragmentation, although co-elution of plasmenyl and plasmanyl species often occurs, in which case both species should be reported.

Report all possible lipid candidates for a feature separated by a pipe "|", not just the top few lipid candidates. TGs are the most common case where co-elution of isomeric species occurs. For example, our laboratory tentatively identified

2,607 TG ions ($[M+Na]^+$ and $[M+NH_4]^+$) in human plasma across 370 features, meaning that, on average, each feature had 7 co-eluting TGs identified (unpublished data). For one feature at *m/z* 920.8635 in human plasma, 49 TGs were tentatively identified. TGs are just one example of co-eluting molecules, for the same human plasma analysis in positive ion mode we found that 40% of features with lipid annotations have at least two co-eluting lipids identified. It is important to note that most software only include one lipid identification for a given feature in the final report, which is based on the false assumption that there are few instances of co-eluting lipids. Examples of annotated lipids using pipes can be found in Supplementary Table S-4 of Koelmel et al. [131], for example for *m/z* 766.5391 at retention time 7.06, the feature was annotated as PE(18:0_20:5)+H | PE(18:1_20:4)+H | PE(16:0_22:5)+H, with annotations ranked by a score based on the MS/MS spectra.

Use comprehensive MS/MS libraries whenever possible. Even when annotations include all lipids identified for a respective feature, co-eluting lipids not contained in that software's libraries may still exist. In this case, biological interpretation will be confounded by multiple uncharacterized lipids or other molecules contributing to a feature’s intensity. One potential example is oxidized species, which can overlap with non-oxidized species, but are not contained in most lipidomics software.

Use pre-analytical steps to prevent degradation and interconversion. Pre-analytical steps can also influence the correct annotation of features by affecting the stability and intensity of lipids or leading to interconversion of the lipid species observed. Sample handling and preparation techniques, involving homogenization, freeze-thawing, and/or exposure to air or light, can result in lipid oxidation or (non)enzymatic

degradation or interconversion [132, 133]. For example, our studies have shown that by not quenching enzymatic activity during sample preparation leads to increased lysophosphatidylcholines (LPCs) (+19.3±1.8%) and decreased phosphatidylcholine precursors (−13.4±2.6%) (unpublished data), likely caused by phospholipase A activity. In this case, stabilization techniques (e.g., heat treatment, additives such as antioxidants, and freeze drying) can be employed, and common byproducts of degradation and interconversion can be measured. For certain lyso-lipids, such as LPCs, acyl migration during sample preparation exists between the *sn*1 and *sn*2 isomer, complicating annotation and quantification [134]. Therefore, *sn*1 and *sn*2 isomers of lyso-species should be combined and reported as sum composition.

For MS/MS based identification, LipidMatch is the only lipid identification software to date which employs all the annotation guidelines presented here, including using pipes "|" for multiple identifications, "_" when fatty acyl position on the glycerol backbone is unknown, and annotates lipids by total carbons and degrees of unsaturation when only class specific fragments are observed. As annotation of lipid species becomes more accurate, we will continue to advance our understanding of the precise roles of individual lipids species in biological systems, advancing the utility of lipidomics.

Implementation of LipidMatch Software

LipidMatch was written in R [135]. The user interface for LipidMatch consists of a series of dialogue boxes developed using gWidgets API and the tcltk R package. Users can access LipidMatch as a file in the supplementary material, with the latest version available at <http://secim.ufl.edu/secim-tools/>. A manual and video tutorials are provided to walk users through the entire lipidomics workflow, including vendor file

conversion to open source format, feature processing, LipidMatch identification, *in-silico* lipid library development, and the ability to append identifications from other software (e.g. MS-DIAL or Greazy).

Generation and Validation of LipidMatch *in-silico* Libraries

In-silico libraries were developed in Excel as described in video tutorial 6 in the supplemental information. Briefly, an R script was used to generate a list of possible fatty acid combinations for acyl containing lipids with 2 or 3 fatty acids. A list of 39 possible endogenous fatty acids and 214 potential oxidized fatty acids were incorporated (contained in the LipidMatch zip file). Combinations excluded redundant possibilities such as 18:0_20:0 and 20:0_18:0. For oxidized lipids, a list of 126 potential long chain oxidized fatty acids was generated by the addition of one or more (depending on the degrees of unsaturation) O (as a ketone or epoxy), OH (as a hydroxyl radical), and OOH (as a perhydroxyl radical) to unsaturated fatty acids within the list of 39 endogenous fatty acids. A list of 88 potential short chain oxidized fatty acids were generated by cleavage of unsaturated fatty acids contained in LIPID MAPS and addition of a terminal CHO (aldehyde) or COOH (carboxylic acid). Oxidized fatty acyl constituents were combined with the original list of fatty acyl constituents to generate possible fatty acyl combinations for oxidized lipids.

For each lipid class, structurally indicative fragments were compiled using other MS/MS databases (LIPID MAPS [136], LipidBlast [137], and MS-DIAL [128]), literature, and/or experimentally derived fragmentation. A summary of all lipid libraries contained as of 10/01/2016 is presented in Table S3-1. Note that this list is constantly growing, and a complete list can be found in the LIPID_ID_CRITERIA.csv file found in the most up to date LipidMatch zip file at <http://secim.ufl.edu/secim-tools/>. Using multiple

sources to obtain fragmentation allowed for cross-validation of fragments and generation of lipid class-specific fragmentation rules (see video tutorial 6 of the supplementary information for details). Fragment masses calculated were validated with MS/MS of internal standards obtained using HCD fragmentation [131] on a high-resolution orbitrap mass spectrometer, or literature searches. The following internal standards were used for verification (acronyms are defined in Table S3-2): CE(17:0), CE(19:0), CE(2:0), Cer(d18:1/17:0), Cer(d18:1/25:0), MAG(17:0), DG(14:0/14:0), DG(19:2/19:2), DG(20:0/20:0), GlcCer(d18:1/12:0), LPA(17:0), LPC(17:0), LPC(19:0), LPE(14:0), MG(17:0), OxPC(16:0/9:0(CHO)), PA(14:0/14:0), PC(14:1/14:1), PC(17:0/17:0), PC(19:0/19:0), PE(15:0/15:0), PE(17:0/17:0), PG(14:0/14:0), PG(15:0/15:0), PG(17:0/17:0), PI(8:0/8:0), PS(14:0/14:0), PS(17:0/17:0), SM(d18:1/17:0), SM(d18:1/6:0), TG(13:0/13:0/13:0), TG(15:0/15:0/15:0), TG(17:0/17:0/17:0), TG(17:1/17:1/17:1) and TG(19:0/19:0/19:0). All internal standards were obtained from Avanti Polar Lipids (Alabaster, Alabama), except TG species, which were purchased from Sigma-Aldrich (St. Louis, MO), and cholesterol esters, which were obtained from Nu-Chek Prep (Elysian, MN).

Lipidomics Workflow with LipidMatch

LipidMatch is designed to be integrated with other open-source software to streamline the lipidomics workflow as described in Figure 3-3. LipidMatch was developed and tested using data acquired from a Q-Exactive orbitrap mass spectrometer (Thermo Scientific, San Jose, CA). LipidMatch has also been tested using data acquired on an Agilent 6540 Q-TOF (Agilent Technologies, Santa Clara, CA). LipidMatch can be used with a variety of other vendors and data formats. Ion selection techniques used to acquire fragmentation, including all-ion-fragmentation (AIF),

inclusion list-based targeted approaches, and data-dependent topN (ddMS^2-topN) approaches can be used with LipidMatch to annotate lipids acquired using liquid chromatography, direct injection, or imaging approaches. LipidMatch is not recommended for most applications using low resolution mass spectrometers. For brevity, we will focus on UHPLC MS/MS methods using the data-dependent topN approach, although video tutorials for imaging approaches and AIF approaches are included in the supplemental materials.

In the workflow recommended for LipidMatch, users acquire full scan data for all the samples in negative and/or positive polarity. In addition, users acquire ddMS^2-topN spectra from pooled samples or from other representative samples. Using iterative exclusion (IE) [131] on pooled or representative samples can increase the number of ions with respective fragmentation spectra. This is highly recommended if spectra are dense (many overlapping lipid signals).

Following data-acquisition, full scan data (either centroid or profile) can be processed to determine features, defined as an ion or ions sharing the same *m/z* and retention time. Features can be determined from various peak picking software such as MZmine [100], XCMS [138], or MS-DIAL [128]. The feature table can have nearly any format, allowing flexibility in choosing feature processing workflows. Video tutorial 2 explains how users can process data using MSConvert [98] and MZmine 2.20 using a batch file for MZmine. The batch file was optimized for lipids using the chromatographic methods in Table S2-1 and is included with the tutorial videos in the LipidMatch file.

Once feature tables are created for each biological substrate and each polarity, features can be directly annotated using LipidMatch and the previously generated

MS/MS data. Peak picking of MS/MS data and conversion to .ms2 file format should be done using MSConvert [98]. Feature table(s) and MS/MS data are placed into a directory as shown in Figure 3-4. Often researchers may have multiple feature tables, one for each polarity type and feature tables for each substrate analyzed. Users can include a subfolder for each sample type, and feature tables should end in “n.csv” or “neg.csv” (not case sensitive) for negative mode, and “p.csv” or “pos.csv” for positive mode. Each folder should contain respective MS/MS data for that substrate in .ms2 format also ending in “neg.csv” or “pos.csv”, depending on polarity. The file should have “dd” in the name if it is data-dependent (DDA) data or targeted data, and “AIF” if it is all-ion-fragmentation data. For example, the user could create a folder for a lipidomics experiment on cancer, with two sub-folders, one for plasma from cancer and non-cancer patients and one for healthy tissue and tumor tissue. Each sub-folder could contain, for example, 2 DDA .ms2 files in positive mode and 2 DDA files in negative mode, one pooled for participants with cancer and one pooled for non-cancer participants, as well as the corresponding feature table in negative and positive polarity. Once the user runs LipidMatch and enters user parameters, LipidMatch will automatically append identifications to each feature table using MS/MS files contained in that feature table’s subfolder.

Once lipid identifications are obtained using LipidMatch, identifications from any other software such as Greazy [129], LipidSearch (Thermo Scientific, San Jose, CA), and MS-DIAL can be appended in additional columns to the feature table (Figure 3-3). The annotations are appended from one file to another if the retention time and *m/z* of a feature in one table matches the retention time and *m/z* of a feature from a second table

within a user defined mass tolerance and retention time tolerance. For example, if a retention time tolerance of 0.1 minutes and mass tolerance of 10 ppm is used, a feature annotated PE(36:2)+H with a retention time of 6.72 and *m/z* of 744.5536 will be appended to a feature generated by a different software with a retention time of 6.68 and *m/z* of 744.5540. Lipidome coverage and confidence in identifications can be increased by appending identifications from multiple software onto one feature table. In addition, metabolite, xenobiotic, or other identifications from software such as Compound Discoverer (Thermo Scientific, San Jose, CA) or MS-DIAL can be appended for a more global approach. Furthermore, lipidome coverage can be increased by the user community by adding new *in-silico* fragmentation libraries. Libraries for LipidMatch can be developed using LipidBlast Templates [137] or as explained in video tutorial 6 found in the supplementary information. Each library should be developed with the correct annotation based on the structural resolution that can be inferred by fragments chosen for the identification criteria.

LipidMatch Inputs and Operations

LipidMatch user inputs and respective operations are exemplified in Figure 3-5 using experimental data for PC(38:6) $[M+HCO_2]^-$. A similar schematic to Figure 3-5, which includes user inputs and modifiable parameters, is provided in the supplementary information (Figure S3-1). The user first chooses directories containing feature table(s), for example those generated by MZmine (Figure 3-4). Then, LipidMatch performs exact mass matching at the MS1 level between *in-silico* precursor ions and each features *m/z* using a user defined *m/z* tolerance (Da) (Step 1; Figure 3-5). Precursor ions include all adducts contained in the *in-silico* libraries for the respective polarity, but do not include dimers, multimers or in-source fragments. Each feature and lipid match will be termed a

“feature-lipid pair”. MS/MS scans from .ms2 files within a user defined retention time and *m/z* tolerance of each feature is determined (Step 2; Figure 3-5). The *m/z* tolerance is the same as the isolation window used for selecting ions.

For each MS/MS scan of each feature, experimental fragments are matched against *in-silico* lipid fragments *m/z* using a tolerance window (ppm). The total number of scans across a feature containing that fragment is calculated. In addition, the fragments average *m/z*, maximum intensity, and retention time at maximum intensity across all scans are calculated for a feature (Step 3; Figure 3-5). This information on fragments for each feature-lipid pair is saved as a table in .csv format for each lipid class. Each fragment is assigned 1 if it is above the user defined minimum intensity and scans threshold and 0 if the fragment does not meet these criteria or was not found within the *m/z* tolerance (Step 4; Figure 3-5). The default number of scans required is 1 based on orbitrap mass spectrometers, but can be increased for other applications. The user modifiable intensity threshold for fragment ions to be considered real is dependent on the mass analyzer, the type of detector and the noise level.

In Step 4, fragments assigned a 1 are considered observed based on the threshold criteria discussed above. Lipids are identified if they contain the necessary observed fragments. For example, for PCs measured as formate adducts, both negative ions of the fatty acyl constituents must be observed (Step 5 of Figure 3-5), while for protonated PCs the PC head group ion 184.0733 must be observed, along with at least one fatty acyl indicative fragment if the lipid is to be characterized at the level of fatty acyl constituents. Default fragments which must be observed for each lipid class were determined using high collisional induced dissociation (HCD) on a Q-Exactive orbitrap

mass spectrometer of internal standards, or endogenous lipids verified in literature. Users can modify which fragment ions for each lipid class must be observed for identification using a simple Excel sheet as outlined in the 6th video tutorial. In certain cases it may be important to optimize fragment criteria for applications not employing HCD fragmentation with an orbitrap analyzer. Experimental protocols including mobile phase (adducts observed), low and high mass cutoff, resolution, and type of fragmentation (e.g. HCD, CID, or UV) will determine what fragment ions are necessary for each lipid type to be identified. Therefore, for applications other than those using HCD fragmentation and orbitrap detection, we strongly recommend checking the existing fragmentation rules against MS/MS obtained in-house. Fragments chosen for confirmation should be of relative high intensity and distinguish the lipid structure from other lipids with similar fragmentation. It is important to note that while fragmentation measured on other high resolution instruments, such as qTOF platforms, can result in significant changes in the relative fragment intensities, in most cases the fragment masses observed are the same. Therefore, since LipidMatch does not include intensity in *in-silico* fragmentation libraries and does not include relative intensities in identification, criteria for identification will often be similar between instruments.

After lipids are identified, they are assigned a number based on whether they are identified by class and fatty acyl constituents (1), by data-independent analysis (2), only by class (3), or only by precursor *m/z* without fragment matching (4) (Step 6, Figure 3-5). If multiple lipids are identified for a single feature, the lipids are ranked by the summed intensity of all their fragments with *in-silico* fragment exact mass matches, including those not used for confirmation (Step 7, Figure 3-5). The final ranked lipid

identifications are appended onto the feature table, along with the lipid class and adduct of the top ranked lipid and summed fragment intensities for each identification.

Benchmarking LipidMatch against other Open-Source Software

Comparison of Lipid Software Features

Table 3-2 compares features in LipidMatch, MS-DIAL, Greazy, and LipidSearch which can all be used to analyze UHPLC-HRMS/MS data. LipidMatch, MS-DIAL, and Greazy are open source, while a license must be purchased for LipidSearch.

Currently, MS-DIAL and LipidSearch provide the most user-friendly interfaces and ease of use. In contrast to other UHPLC-HRMS/MS identification software, LipidMatch is completely written in R. Compared to the other lipid identification software written in middle level languages, such as C++, LipidMatch can take longer to run, especially for high resolution data. This is due to the slow speed of imbedded for-loops in R and the extensive LipidMatch libraries and hence large search space. While run time can be longer, LipidMatch can readily be integrated with diverse R tools and statistical packages available for mass spectrometry and omics-based studies

Databases for lipid identification differ both in coverage and information type. For example, LipidMatch and Greazy databases contain only the exact *m/z* of precursor ions and fragment ions, while MS-DIAL and LipidSearch include simulated intensities. In addition, software such as MS-DIAL and LipidMatch contain static *in-silico* libraries, while libraries in Greazy are generated as the program is executed, based on the types of lipids and fatty acyl constituents the user specifies. While LipidMatch libraries are static excel files, as with all four software previously mentioned, the user can select which lipid types to query using LipidMatch, hence limiting searches only to biologically relevant or expected lipid types and reducing run time. LipidMatch libraries contain only

exact *m/z* values of precursors and fragment ions, making it relatively trivial for users to generate *in-silico* libraries and/or convert other databases to the LipidMatch library format. LipidMatch contains all lipid types in MS-DIAL 2.24, as well as LipidBlast release 3 development libraries. With 56 lipid types, the LipidMatch *in-silico* libraries cover the greatest number of lipid types of any open source software to date, with MS-DIAL containing 34 lipid types, and Greazy containing 24 lipid types (Table 3-2).

All four programs use different identification strategies. MS-DIAL and LipidSearch include intensity to rank lipid identification by a similarity score. Greazy includes a similarity score as well as a false discovery probability based on the total number of fragments observed, thus solely relying on *m/z*. Both LipidMatch and LipidSearch include rule-based identification, which allows correct annotation of lipid structure based on fragments observed (correct structural resolution). While all other open-source software sort identifications by similarity score, LipidMatch sorts lipid identifications by summed fragment intensity. For each lipid species identified, all expected fragment ions are summed (using the scan with the highest intensity for each fragment). Fragment ions to sum are determined from the *in-silico* fragment *m/z* values for that species and include fragments not necessary for lipid identification (for example the loss of the PC head group for PCs when the 184.0733 *m/z* PC fragment is observed). For each feature, the lipid ions are ranked from maximum to minimum summed intensity.

LipidMatch ranking is based on the assumption that a feature often represents multiple lipid ions and that ranking is meant to determine the relative signal contribution of each lipid to the feature. In other software, by using similarity score, ranking is based on which lipid identification is most confident. While both ranking algorithms produce

similar results in many cases (see results and discussion), LipidMatch algorithm is designed based on a more accurate assumption of multiple co-eluting lipids sharing *m/z* values within the same accurate mass. In simple dot product matching, the algorithm is based on the assumption that the fragmentation spectra are solely based on the ion of interest. Any deviation from the predicted fragmentation spectra, such as additional high intensity fragment peaks from co-eluting isobaric species, will reduce the dot product score. Many lipids will not be identified due to co-eluting isobaric species adding more fragments to the spectra and hence reducing the dot product score. MS-DIAL has approached this issue by reducing the impact of peaks not contained in the *in-silico* fragmentation library on the modified dot-product score. Fragments from different species which overlap in exact mass, for example fatty acyl fragments from 18:0 in TG(18:0/18:0/18:0) and TG(16:0/18:0/20:0), will still decrease the modified dot-product score in MS-DIAL, and hence lead to false negatives.

Ranking lipid identifications for a given feature is complicated by overlapping mass spectral fragments in LipidMatch as well. A number of problematic cases can arise. For example, for a given lipid type with high intensity fragments below the *m/z* cutoff, the ions-summed fragment intensity will be reduced compared to lipid species with the bulk intensity of fragments within the *m/z* range. Similarly, if high intensity fragments are missing from the *in-silico* library for a lipid type, these lipids will be artificially lowered in their ranking in terms of contribution to feature signal. In addition, shared fragment ions for some lipids will artificially inflate summed fragment intensity (Figure 3-6B) and fragment intensity will depend on the MS/MS scans proximity to a

given ions apex (Figure 3-6C). Similarity score matching, such as that used by MS-DIAL, suffers similar problems.

To determine the accuracy of lipid rankings and identifications using LipidMatch, identification of lipids in Red Cross plasma using LipidMatch was compared toMS-DIAL, and Greazy. Lipid software excluded for comparison included LipidSearch (Thermo Scientific), Lipidyzer (SCIEX), and SimLipid (PREMIER Biosoft), which are not open source software, and Alex [139], LipidXplorer [140], MS-LAMP [141], LIMSA [142], LOBSTAHS [143], Lipid Data Analyzer [144], LipidQA [145], and Lipid-Pro [146], which were not designed for UHPLC-HRMS/MS untargeted experiments. As stated previously, LipidMatch, MS-DIAL, and Greazy differ in lipid identification strategy; hence, the amount of features with the same identifications between LipidMatch and the other software platforms was used to assess the accuracy of the LipidMatch ranking algorithm. Further work, with spiked co-eluting standards sharing the same exact mass at varying concentrations would be helpful to further assess the ranking algorithm accuracy.

A Case Study: Identification of Lipids in Red Cross Plasma

LipidMatch, Greazy, and MS-DIAL were applied to five replicate injections of Red Cross blood plasma. Data was acquired in positive and negative polarity, using iterative exclusion [131] and data-dependent top 5 (ddMS^2-top5) to acquire MS/MS fragmentation. Liquid chromatography and mass spectrometer parameters are shown in supplementary Table S2-1, Table S2-2, and Table S2-3. Identifications from all software were appended to the MZmine feature tables using the CombineSoftwareIDs.R script. Both the script, MZmine parameters (batch file), and an excel sheet with the resulting annotations of features across all 3 software (Table S3-3) are included in the

supplementary information. The script aligns features with similar *m/z* (10 ppm window used) and retention times (0.2 min window used) from two different peak picking or identification software.

Compared to the other major open-source software platforms, such as MS-DIAL and Greazy, LipidMatch annotated more lipid ions. LipidMatch was used to identify 210 lipid ions across 159 features and 15 lipid types in negative polarity. In positive ion mode, LipidMatch was used to annotate 5159 unique lipid ions across 1401 features and 26 lipid types. The large number of unique lipid ions in comparison to a smaller amount of identified features is due to overlap of co-eluting lipids sharing the same exact mass, allowing for multiple lipids identified for a given feature. It is important to note that annotations of class-specific fragments (as indicated by "3_" in Table S3-3), are significantly more tentative than identifications using fatty acyl fragments. This is especially true for choline containing lipid classes such as SM and PC, which share common fragments. For positive ion mode, 987 features were annotated with fatty acyl information. It is also important to note that in this study, we look at the number of lipid ions annotated, including multiple adducts for a given lipid species. When only unique lipid molecules were taken into account by manually removing redundant adducts and features, and identifications using only choline specific fragmentation were removed, a total of 728 features with unique lipid molecular annotations were identified by LipidMatch for this dataset, as has been published previously [131]. The curated 728 lipid molecular identifications using LipidMatch is still significantly greater than the total lipid ions identified by MS-DIAL and Greazy combined. Table S3-3 includes all features identified in Red Cross plasma, with LipidMatch, MS-DIAL, and Greazy annotations.

MS-DIAL and Greazy identified 143 and 94 features in negative mode, respectively, and 411 and 180 features in positive mode, respectively. Lipid types identified, which were unique to LipidMatch, included oxidized species (151 across TG, PC, and LPC in positive polarity), plasmenyl and plasmanyl TGs (19 species in positive mode), sphingosines (2), sulfatides (1), and PI species in positive mode as ammonium adducts (18). It is important to note that many additional unique *in-silico* libraries exist in LipidMatch, for example cardiolipin as ammonium adducts, but these species are not observed in plasma samples. Bar graphs displaying the number of lipid species in each lipid type identified by LipidMatch, MS-DIAL, and Greazy, and overlapping identifications between software are shown in supplementary Figure S3-2 (negative polarity) and Figure S3-3 (positive polarity). In addition, pie charts showing the lipid types covered by LipidMatch are shown in Figure S3-4 (negative polarity) and Figure S3-5 (positive polarity).

Since Greazy is limited to glycerophospholipid species, only 65 features in negative polarity and 68 features in positive polarity had identifications across all software. In negative polarity, 97% of these features had the same identification at the structural resolution of fatty acyl constituents across all 3 software platforms. In positive polarity, 71% of features with identifications across all software tested were the same. Note that plasmenyl and plasmanyl species with differences in one degree of unsaturation were considered the same identification due to minimal difference in MS/MS spectra. The greater discrepancy in identifications in positive mode is most likely to do to the low abundance of acyl chain fragments for glycerophospholipids in positive mode, thus making identification by fatty acyl constituents difficult. At the

structural resolution of lipid class and total carbons and double bonds, 94% of features contained the same identifications across all 3 software platforms in positive polarity, and 100% of features were identified the same in negative polarity.

Of all lipid types identified by both MS-DIAL and LipidMatch, TGs had the most discrepancy. Of the 136 features identified as TGs by both LipidMatch and MS-DIAL (both sodiated and ammoniated forms), 100% of the top hits were the same at the structural resolution of total carbons and degrees of unsaturation, but only 61% of the top hits were the same at the structural resolution of fatty acyl constituents. TG identification is complicated by the number of co-eluting isomers, for example, LipidMatch identified over 20 co-eluting TG isomers for a number of features. These co-eluting isomers can share one or more fatty acyl constituents, and therefore share common fragments, further complicating identification.

LipidMatch had a significant number of lipid identifications by fatty acyl constituents corroborated by at least one other software, suggesting that LipidMatch identification and the ranking strategy results in similar identifications for glycerophospholipid species compared to other identification algorithms. For the 68 features identified by all software in positive polarity, 92% of identifications by LipidMatch were corroborated by at least one other software. MS-DIAL and Greazy had 86% and 84% of identifications corroborated for these features by at least one other software, respectively. In negative polarity, 98% of LipidMatch identifications (all except one) were corroborated by at least one other software, with MS-DIAL having 98% identifications corroborated and Greazy having 100% of identifications corroborated.

Conclusion: LipidMatch is a Flexible, Comprehensive, and Accurate Annotation Software

LipidMatch is a freely available tool with the potential to be incorporated into a diverse range of lipidomics workflows, including imaging, direct-infusion, and LC-MS/MS experiments with both low and high mass resolution. For LC-MS/MS workflows, LipidMatch can be used with any feature processing software, such as MZMine, XCMS, or MS-DIAL. LipidMatch contains the greatest diversity in lipid types of any current open-source software platform and a unique rule-based strategy for identification and summed fragment intensity based strategy for ranking top hits. Compared to other software, LipidMatch is highly customizable. For example, users can select which fragments are necessary for confirmation and develop their own fragmentation libraries in Excel. Additional tools allow the user to pool results from multiple identification software platforms into one feature table. Compared to MS-DIAL and Greazy, LipidMatch was found to provide the most lipid identifications for Red Cross plasma. For features with identifications using all 3 software platforms, identifications were comparable at the level of fatty acid constituents. 92% and 98% of LipidMatch identifications were corroborated by at least one of the other software platforms in positive and negative mode, respectively.

Table 3-1: Structural resolution and annotation of lipids using mass spectrometry

Structural Resolution	Annotation
Carbons and Double Bonds	PC(34:2)
Fatty Acyl Constituents	PC(16:0_18:2)
Positional Isomers	PC(16:0/18:2)
Double Bond Position	PC(16:0/18:2(10,12))
Double Bond Cis/Trans	PC(16:0/18:2(10E,12Z))
Stereochemistry	PC(16:0/18:2(10E,12Z)[R])

Table 3-2. Comparison of lipid identification software.

	LipidMatch	MS-DIAL	GREAZY	LipidSearch 4.1
Identification (ID) Strategy*	Rules	Similarity	Similarity	Rules and Similarity
Fragment Intensity for ID*	Yes (ranking)	Yes	No	Yes
in-silico Library (Types)	56	34	24	59
User Developed Libraries	Yes	Difficult	Difficult	Difficult
Programming Language	R	C#	C++	Java
Restrictions	None	None	None	License
Multiple Vendor Formats	Yes (.ms2)	Yes (.abf)	Yes (.mzML)	Yes
Data Independent Analysis**	Yes	Yes	No	No
MS^3 analysis	No	No	No	Yes
Multiple Hits in Final Report	Yes (ranked)	No	No	Yes (ranked)
Structural Resolution	Correct	Over Reports	Over Reports	Correct
Identifiers (eg. LipidMaps)	No	Yes	No	No
Computational time (HR data)	Slow	Medium	Fast	Fast
Employs False Discovery	No	No	Yes	No

Note that in determining total types of lipids contained in each software's *in-silico* library all ether linked lipids contained were considered two types (plasmenyl and plasmanyl) and all oxidized lipids contained across numerous classes were considered one lipid type
*Please read text for further information
**Not discussed in-depth in this manuscript. LipidMatch can be applied to AIF data independent analysis (currently only supports Thermo files), while MS-DIAL can be applied to AIF and SWATH approaches
***Correct reporting of structural resolution means that lipids are annotated only at the level of structure known based on fragmentation

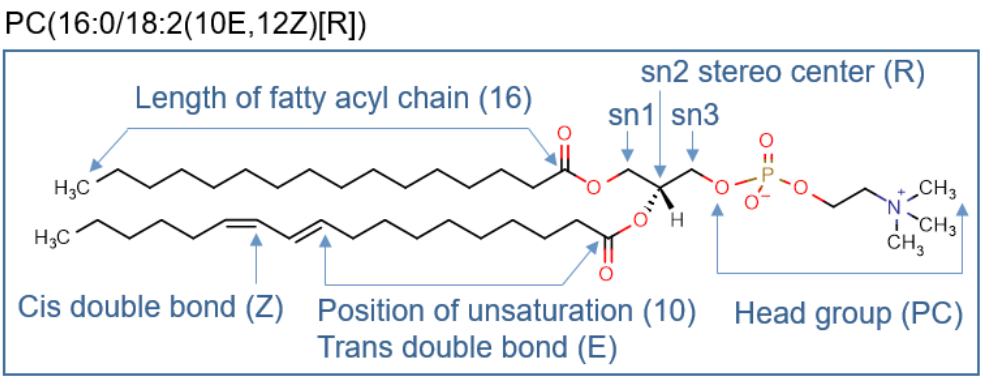


Figure 3-1. Annotation of a phosphatidylcholine (PC) species outlining how to annotate each structural detail of glycerophospholipids. The lipid is annotated using Lipid Maps nomenclature based off of International Union of Pure and Applied Chemists and the International Union of Biochemistry and Molecular Biology (IUPAC-IUBMB) Commission on Biochemical Nomenclature. The [R] conformation is often not indicated in annotation, while [S], being the less common form, is specifically referred to.

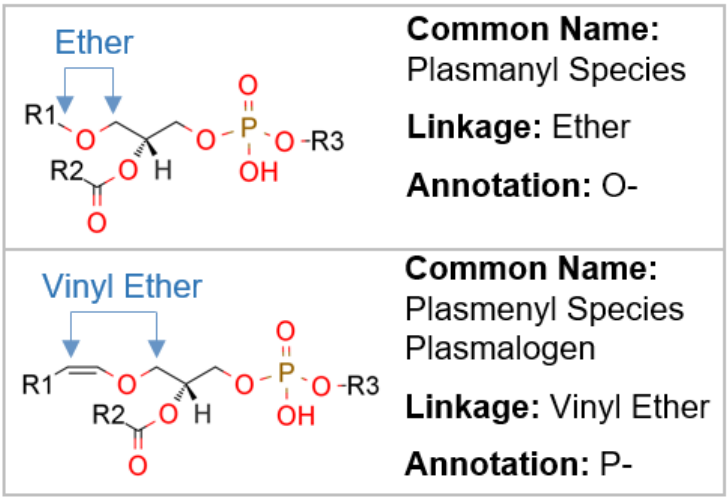


Figure 3-2. General structure for plasmanyl and plasmenyl phospholipid species containing a glycerol backbone

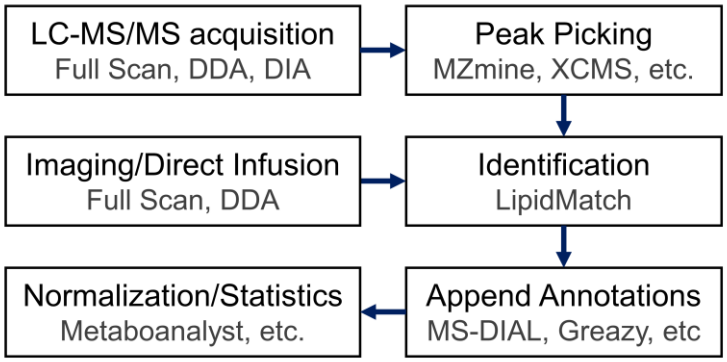


Figure 3-3. Options for open source software integration with LipidMatch in a lipidomics data processing workflow. Acquisition modes for fragmentation which can be used to annotate lipids with LipidMatch include data-dependent analysis (DDA) and data-independent analysis (DIA) for both direct infusion and liquid chromatography (LC) tandem mass spectrometry (MS/MS) approaches.

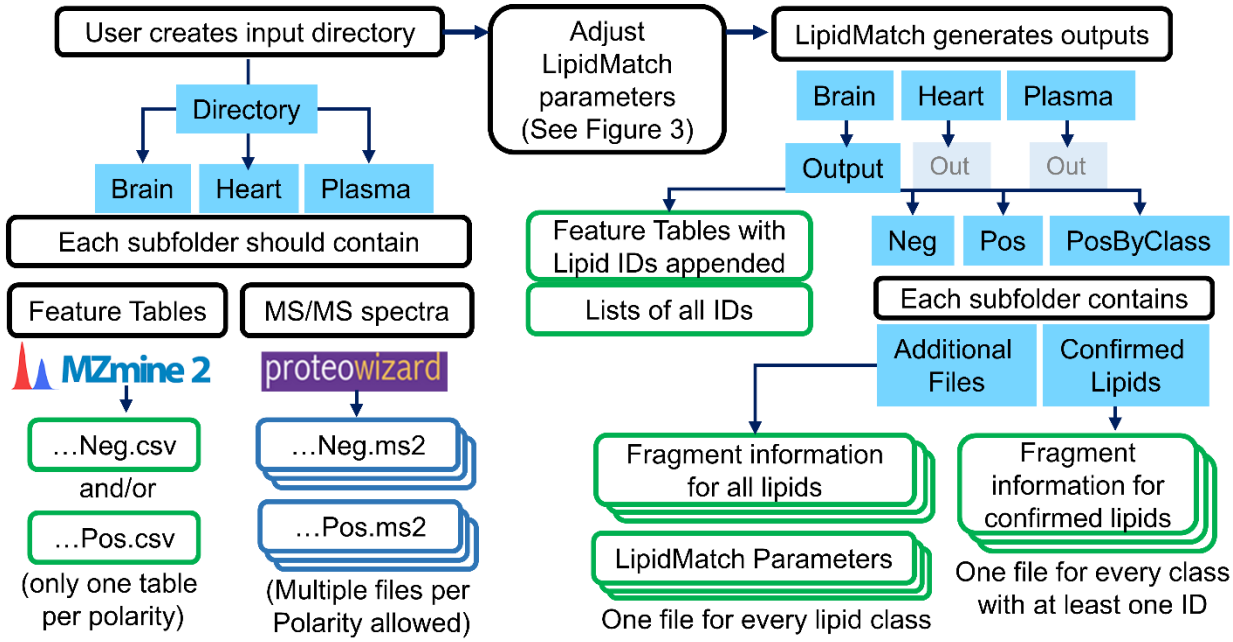


Figure 3-4. Workflow for using LipidMatch, with input and output folder structure and files. Green boxes represent .csv files, dark blue boxes represent open source MS^2 files (.ms2), and filled light blue boxes represent folders. Three stacked boxes represent that multiple files are allowed or generated. The subfolders (brain, heart, and plasma) are examples, these folders can be for any biological substrate. In addition if only one biological substrate is analyzed, only the main directory folder is needed. In the outputs generated by LipidMatch each subfolder contains an output folder as depicted above.

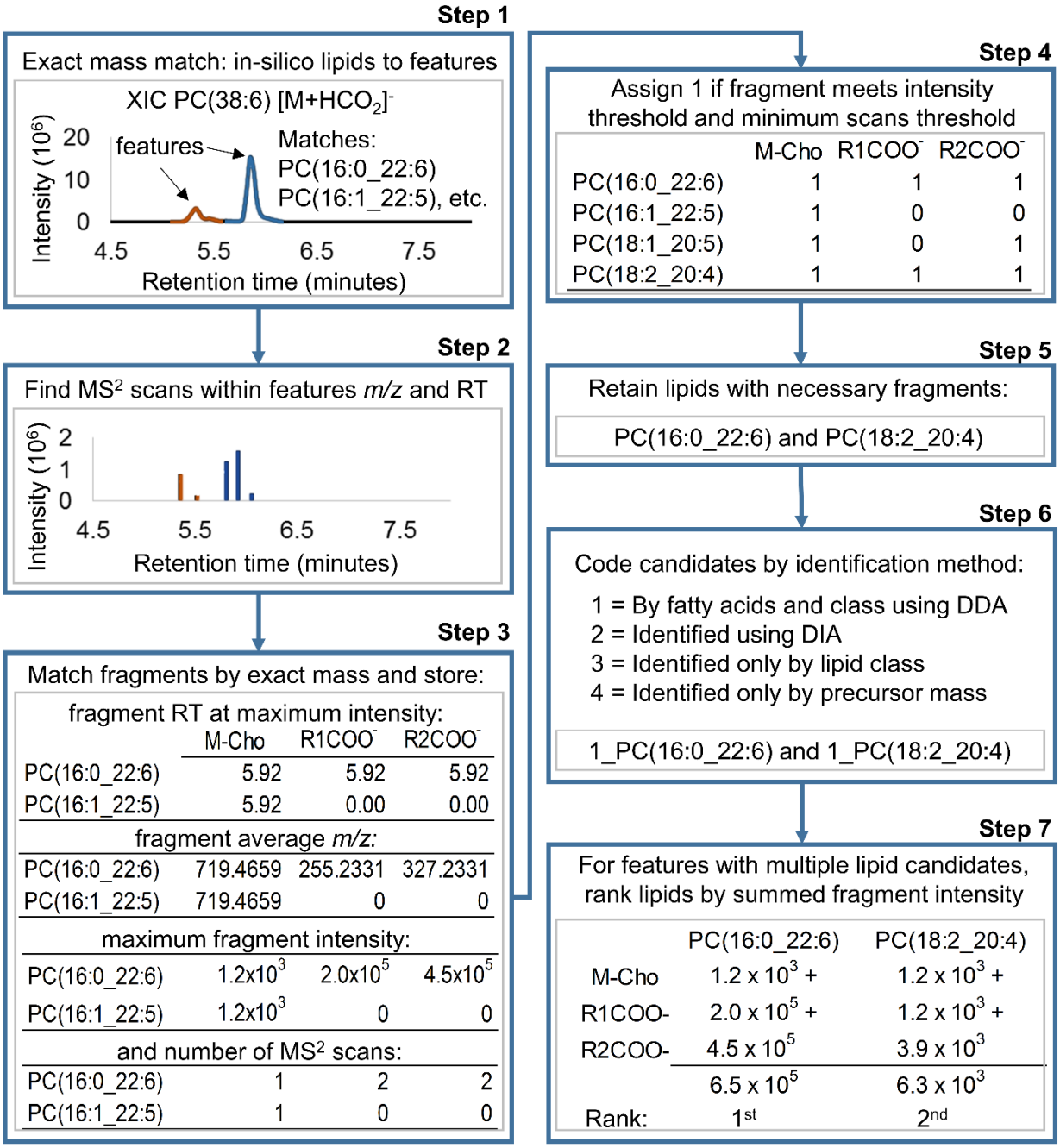


Figure 3-5. Simplified flow diagram of LipidMatch operations. The steps for identification of the feature at *m/z* 850.5604 and retention time (RT) 5.92 as formate adducts of PC(16:0_22:6) and PC(18:2_20:4) are shown as an example in grey boxes for each step. Note that the number of lipid identifications and fragments queried in the example are reduced significantly for illustration purposes. For Step 5, $R1COO^-$ and $R2COO^-$ were required for identification above an intensity threshold of 1000 in at least one scan across the peak.

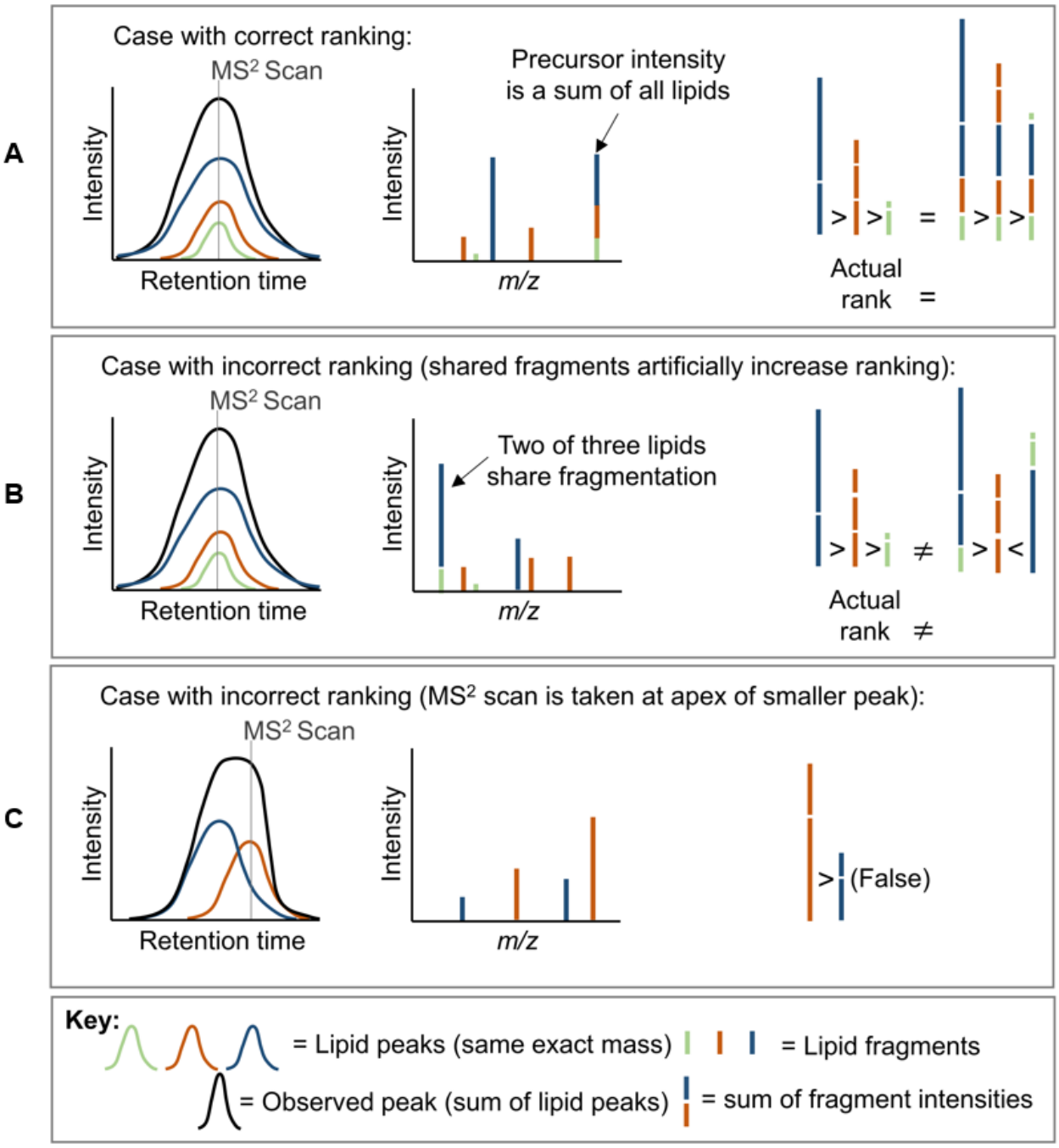


Figure 3-6. Problematic cases which can arise when ranking lipids by the sum of fragment intensities. The first panel a) represents a case were lipids are accurately ranked (far right) based on the areas under the peak (far left). It also show that even in this case, the precursor intensity doesn't reflect a single intensity, but a sum of the intensity of all precursor isomers (middle). In panel b) two lipids (blue and light green) share a high intensity fragment with the same *m/z* (middle), inflating their intensity values leading to false ranking (far right). In panel c) the MS/MS scan misses the apex of the lipid with a blue trace, and hence the summed intensity for the blue trace is reduced.

CHAPTER 4
ANNOTATION AND QUANTIFICATION OF LIPIDS USING AN OPEN SOURCE LC-HRMS/MS WORKFLOW AND LIPIDMATCH QUANT

Relative Quantification in Lipidomics

Lipids diverse biological roles are achieved through the vast heterogeneity and complexity in lipid structure, distribution, and concentration. For example, individual lipids can differ by over six orders of magnitude in concentration [147], while chemical and physical properties can vary in polarity, structural orientation, and charge state (e.g., charged, zwitterionic, and neutral lipid species). Advances in mass spectrometry and the advent of electrospray ionization (ESI) has enabled researchers to begin to detect this wide diversity of lipids; however, quantification of these detected lipids is challenging due to their dynamic range and breadth of chemical properties.

For quantification in lipidomics, either relative or absolute quantification can be performed. Absolute quantification typically employs matrix-matched external calibration curves and/or isotopically labeled internal standards for each lipid quantified. This quantitative approach has limited application to untargeted lipidomics analyses due to the enormous diversity of the lipidome, limited availability of appropriate standards to cover this diversity, and the cost associated with purchasing hundreds of standards. Relative quantification is often sufficient where relative changes are of concern, for example between diseased and control populations [75]. Relative quantification, which does not employ a calibration curve, and involves the addition of a smaller set of internal standards representative of the classes of lipids analyzed, is the most commonly used approach for quantification in untargeted lipidomics experiments.

Submitted to Journal of the American Society for Mass Spectrometry

The selection of the most appropriate internal standard to best represent a lipid feature can be challenging. The dynamic range, ionization efficiency, and specificity, which are all important for quantification, can differ depending on the lipid molecule's structure, more specifically lipid class, degrees of unsaturation, and number of carbons in fatty acyl constituents. Lipid class generally has the greatest effect on ionization efficiency. Previous reports have shown that lipid internal standards spiked into samples at the same concentration have orders of magnitude differences in intensities across different classes [148]. Therefore, lipids should generally be quantified using standards from the same lipid class. To account for the number of carbons and degrees of unsaturation in fatty acyl constituents, which both lead to an increase in ionization efficiency [148], two or more lipid standards per class, each with different carbons and degrees of unsaturation is suggested for polar lipids [149]. For neutral lipids, where fatty acids play a greater role in ionization efficiencies, response curves based on a wide range of internal standards is often employed. The differences in carbons are often a more significant contributor to ionization efficiency than that of unsaturation at low concentrations, while at high lipid concentrations the effect of unsaturation on ionization efficiency becomes more pronounced [148].

In addition to lipid structure, overlapping chromatograms, ion suppression, large dynamic ranges in lipid concentration, extraction procedure [149], and other methodological and instrumental factors can affect the amount of lipid signal observed. Ultra-high performance liquid chromatography (UHPLC) and high-resolution mass spectrometry (HRMS) can be employed to increase specificity. HRMS reduces the overlap of mass spectral peaks from isobars, resulting in a decrease in residual

standard deviations of measurements and more accurate peak integrations, which are used for more accurate quantification [150]. Chromatography also reduces the possibility of peak overlap by adding an orthogonal dimension of separation, and can reduce ion suppression by separating lipid classes and species, reducing the probability of high abundant lipid classes suppressing low abundant lipid classes [149].

In summary, the best choice of lipid internal standards are those that are lipid class representative and elute at similar retention times to the analytes of interest. Manually selecting representative spiked internal standards and the associated lipid analytes to quantify and applying the algorithm for quantification can be a tedious process prone to human error, especially with lists containing hundreds of lipid species. Automation of the quantification process can lead to increased throughput, a reduction in errors, and harmonization of quantification methods within the lipidomics community. Therefore, we developed LipidMatch Quant (LMQ), which can be integrated in an open source workflow to select the most appropriate internal standards for quantification within acquired LC-HRMS data. While numerous open source quantification software for direct infusion based lipidomics currently exists [139, 140, 145, 151], to our knowledge, Lipid Data Analyzer (LDA) [144, 152] is the only open source quantification software for LC-based lipidomics using class representative lipid standards to return units of concentration. LMQ differs from LDA and commercial lipid quantification software such as LipidSearch (Thermo Scientific), SimLipid (PREMIER Biosoft), and Lipidyzer (SCIEX), in that it was built to be integrated into workflows using any combination of peak picking and peak annotation software. For example lipids can be quantified using LMQ and outputs from MS-DIAL [128], LipidSearch, or LipidMatch [153], with little to no

modification. In addition, the LMQ algorithm for selecting internal standards for feature quantification is unique; accounting for both ion suppression effects by matching individual lipid species to lipid internal standards with the closest retention time and ionization efficiencies by matching lipids to internal standards by lipid class and adduct.

The effect of lipid structure on quantification has been investigated previously [56, 149, 150, 154], while to our knowledge the effect of different data processing strategies and adducts utilized on final concentrations has not been examined thoroughly in UHPLC-HRMS experiments. Therefore, we investigated different data processing methods (peak area versus peak height, smoothing versus not smoothing) and utilization of different ions and polarities for lipid quantification using LMQ. Investigating the effect of various aspects of the lipidomics workflows on quantification using open source tools available to the wider community is an important step in validating the utility and establishing community wide protocols for relative quantification in lipidomics.

Methods: Lipidomics Workflow and LipidMatch Quant Implementation

Lipid Extraction and Data Acquisition

Lipids were isolated from 40 µL of National Institute for Standards and Technology (NIST) standard reference material (SRM 1950) Metabolites in Frozen Human Plasma [155]. Lipid internal standards were purchased from Avanti Lipids (Alabaster, AL), which included lysophosphatidylcholine (LPC (19:0)), phosphatidylcholine (PC (17:0/17:0)), phosphatidylglycerol (PG (17:0/17:0)), phosphatidylethanolamine (PE (17:0/17:0)), phosphatidylserine (PS (17:0/17:0)), triglyceride (TG (15:0/15:0/15:0)), ceramide (Cer (d18:1/17:0)), and sphingomyelin (SM (d18:1/17:0)), were spiked into the plasma at 1.4 nmol, 0.92 nmol, 0.93 nmol, 0.97

nmol, 0.92 nmol, 0.26 nmol, 1.3 nmol, and 0.98 nmol, respectively. The extraction was performed using the Matyash method [156] and samples were reconstituted in 40 µL of isopropanol.

Samples were injected onto a Waters (Milford, MA) BEH C18 UHPLC column (50 x 2.1 mm, 1.7 µm) held at 50 °C with mobile phase A consisting of acetonitrile:water (60:40, v:v) with 10 mM ammonium formate and 0.1% formic acid and mobile phase B consisting of isopropanol:acetonitrile:water (90:8:2) with 10 mM ammonium formate and 0.1% formic acid at a flow rate of 0.5 mL/min. A Dionex Ultimate 3000 RS UHLPC system (Thermo Scientific, San Jose, CA) coupled to a Thermo Q-Exactive mass spectrometer (San Jose, CA) was employed for data acquisition. Three replicate injections were run in both positive and negative polarity employing alternating full and all-ion fragmentation (AIF) scans at a resolution of 70,000. In addition, 15 injections employing targeted MS/MS data acquisition were performed, with the inclusion list for fragmentation containing lipids identified by the LIPID MAPS consortium [102]. The ion optic settings for the mass spectrometer included: analyzer temperature of 30 °C and S-Lens radio frequency level of 35 V. The ionization conditions included a sheath gas flow of 30, auxiliary gas flow of 5, and sweep gas flow of 1 arbitrary units, and a spray voltage of 3.5 kV, and capillary temperature of 250 °C. For positive ion mode, lock masses of diisooctyl phthalate (*m/z* 391.2842) and polysiloxanes (*m/z* 371.1012 and 445.1200) were used, while no lock masses were used in negative ion mode. Both AIF and targeted MS/MS injections were used for lipid identification, while full scan data were used for feature finding and quantification. The UHPLC gradient use in this

experiment is shown in Table S4-1, while the mass spectrometric parameters are shown in Table S4-2.

Data Processing

The open source data processing workflow for lipidomics is shown in Figure 4-1. The first step in the workflow is feature finding using MZmine 2 [100], followed by annotation with LipidMatch [153], blank feature filtering (BFF), and quantification by LipidMatch Quant (LMQ). Note that LMQ can be employed with any feature finding and lipid identification software.

A two-step process was used for feature detection. First features and their respective *m/z*, retention time, and peak heights across samples were detected using an MZmine workflow consisting of mass detection, chromatogram building, chromatogram deconvolution using local minimum, isotopic peak grouping, and alignment and gap filling of features (the batch mode is in the supplemental 2017_9_11_LMQ_Software.zip file; 2017_8_01_MZMine_Batch_Step1.xml). This step is untargeted, in that no information on expected peaks is utilized. In this work, the untargeted step for feature detection included a blank sample and three replicate injections of SRM 1950 in both positive and negative polarity. The resulting feature table from the untargeted step was filtered using a modified blank feature filtering (BFF) approach, where the minimum intensity of the replicate injections had to be at least five fold greater than the blank intensity. The BFF method dramatically reduces the number of peaks which are not from biological origin.

After filtering, the median peak height and peak retention time from the SRM 1950 replicate injections were used to develop a targeted peak list. In the second step, a targeted list of peak *m/z* and retention time values was generated from the previous

step, and the internal standard *m/z* values and retention time values were appended to this list. An MZmine workflow consisting of mass detection, targeted peak detection, chromatogram deconvolution, alignment, and gap filling was used (the batch mode is in the supplemental 2017_9_11_LMQ_Software.zip file;2017_8_01_MZMine_Batch_Step1.xml).

Reprocessing the data using a targeted peak list determined from a smaller sample set has two advantages, especially for application to larger datasets. One advantage is that this workflow significantly reduces data processing time for large datasets, while the other advantage is that peak picking and integration using a targeted peak list is more consistent across samples than aligning features from an untargeted workflow. For example, if there are six pooled samples which should be representative of the features present in 100 samples, these pooled samples are the only samples that need to be run through the initial MZmine workflow. Then a target list can be generated after blank filtration and subsequently used to target features across all 100 samples. Note that in this study only three samples were analyzed, and hence all samples were used to determine the targeted peak list. The median of retention time and *m/z* values across all samples was used rather than the average, as often overlapping peaks lead to average *m/z* and retention time values which are actually between the two peaks and neither represents the first or second peak. For cases in which there are odd sample numbers, the median will always represent the location of a true peak. For cases in which there are even sample numbers, the median will represent the average of two peaks and therefore the value at the i^{th} position of the ranked values can be used, where i = n/2, and n is the total number of samples.

Once the final peak list with retention time, *m/z*, peak area, and peak heights were obtained, the data were annotated using LipidMatch (Figure 4-1). LipidMatch [153] identification was performed using all ion fragmentation data (AIF) and targeted MS/MS data acquisition using precursor ions from lipids determined by the LIPID MAPS consortium [102]. If multiple lipids are annotated for a single feature, the lipid annotations are then ranked by the sum of fragment intensity. The annotated feature table was further reduced to molecular species in positive and negative ion mode by selecting the top most abundant ions for features with the same molecular lipid (this was performed using the Excel "highlight duplicate" and sort functions).

The finalized feature tables with unique lipid molecular species was uploaded into LipidMatch Quant (LMQ) for quantification. A table containing the internal standard name and corresponding concentrations (nmol lipid per mL plasma) was created and uploaded.

LMQ User Workflow

The LMQ software requires two comma separated values (.csv) files as input for proper operation. The first required file is a feature table with the following content for each feature: (1) peak height or peak area, (2) lipid annotation, (3) lipid class, (4) lipid adduct, (5) retention time, and (6) *m/z*. The second required file is an internal standard sheet, which lists the names of all internal standards added, their concentrations, retention time, and *m/z* for each adduct. The names of the internal standards can be in any format familiar to the user. Examples and templates of the two input tables can be found in the LipidMatch Quant zip file available at <http://secim.ufl.edu/secim-tools/> and in the supplemental information (2017_9_11_LMQ_Software.zip).

The user can easily generate the *m/z* of the adducts expected for each lipid internal standard using only the internal standard name, with a separate tool, LipidPioneer [157]. The user then specifies which internal standard will be used for each lipid class in the internal standard sheet. Note that multiple lipid classes can be represented by a single internal standard in the internal standard sheet. For example in this work, we included the following lipid classes to be quantified by PC(17:0/17:0): PC, Plasmanyl-PC, Plasmenyl-PC, and OxPC (oxidized phosphatidylcholine). We chose to represent ether-linked species using a non-ether-linked internal standard, as it has been shown that ether linked glycerophospholipids have the same response factor to their non-ether linked counterparts [154]. This internal standards sheet can be used for later experiments if the same internal standards and chromatographic conditions are employed (and there is no retention time drift).

After opening and running the R script in the LipidMatch zip file, popup boxes prompt the user to select the working directory folder for all files (feature table and internal standards sheet). The user is then instructed to select the feature table and the internal standard sheet. The user completes a series of input boxes, inputting the location of the columns for m/z, retention time, lipid class, lipid adduct in the feature table, and the row in which data starts. By not predefining the format of the feature table, users can utilize various peak picking and lipid annotation software and directly, or with minor modification, apply LMQ. Other user inputs include retention time and m/z tolerances, which are used for locating features representing the internal standards in the feature table using the retention time and m/z values supplied in the internal standard sheet.

The software outputs a ‘standardsfound.csv’ (all identified internal standards) and ‘[input_sheet_name]_Quant.csv’ (feature table with concentrations) file in .csv format. LMQ outputs a list of all internal standards that were identified in the feature table using the internal standards sheet. In addition, the feature table, with quantified values and appended columns containing internal standard information, is created. Each feature includes columns containing information regarding which internal standard ion (molecular species and adduct) was used and a scoring column which allows the user to see how well the internal standard represents the analyte. Lipids quantified using a score of 2 or 3, should be used only with great caution, as internal standards which match the lipid class of the feature were not found. Since lipid class significantly affects ionization efficiencies, these standards only take into account ion suppression, but not ionization efficiencies. An output table for LMQ can be found in the LMQ supplementary zip file.

LMQ Algorithm

A schematic of the LMQ algorithm is shown in Figure 4-2. The LMQ algorithm incorporates a scoring based approach to classify internal standards selected for each feature. A lower score indicates better representation of the feature by the internal standard while a higher score indicates poorer representation (with scores of 1, 2, and 3). For each feature, the LMQ algorithm associates the appropriate internal standard detected. If the feature and internal standard adduct and class match, the feature is scored as a 1. If the current feature class does not match any of the internal standard lipid classes, but the same adduct is found for an internal standard representing a different lipid class, a score of 2 is given. If no internal standard is found for a feature with a matching adduct or class, a score of 3 is given (Figure 4-2).

It is important to note that multiple internal standards can be provided for a single lipid class. In this case, the internal standard with the closest retention time is used for each feature of the respective lipid class. Since retention time correlates with saturation and carbons in the lipid fatty acyl constituents, this will in part account for different ionization efficiencies due to these structural differences. More importantly, ion suppression can vary across retention time, and therefore using multiple internal standards can better account for these differences in ion suppression. If multiple standards are found using score 2 or 3, the one with the closest retention time to the average retention time for the entire lipid class and specific adduct is used to quantify all lipids with the class and adduct.

Comparison of Quantification Using Different Data Processing Methods and Different Ions

Different data processing methods and ions were used for quantification to determine which methods had the greatest effect on the final quantitative values. The comparisons were: smoothing versus no smoothing (smoothing set to 15 in MZmine), peak height versus peak area, quantification with negative versus positive ions, and quantification on $[M+Na]^+$ adducts versus the major precursor ion. The $[M+Na]^+$ adducts were chosen because for the majority of lipids in positive ion mode an $[M+Na]^+$ peak is present, and hence may affect quantification through competitive ionization. For comparison of similarity, the slope and r^2 of linear correlations on the log10 value obtained between the two comparative methods were used. In addition, Bland-Altman type plots [158] were used to determine the relative percent difference in concentrations using two different methods or ions for quantification. A distinction was that instead of normalizing to the average, as is traditionally done for calculating percent difference to

be visualized in Bland-Altman plots [159], the differences were normalized to the minimum values (hence giving a percent increase from the minimum value). When differences are normalized to the average, the absolute relative percent difference plotted against the fold change (fold changes greater than 1) is non-linear and asymptotic to 200 %, while the relative percent difference, calculated by normalization to the minimum, is linear as compared to fold change and hence is easier to interpret (Supplemental Figure S4-1). The equation used to calculate relative percent difference is shown below:

$$Relative\ percent\ difference = \frac{x - y}{\min(x, y)} \times 100 \quad (4\text{-}1)$$

Where x and y represent concentrations calculated using different methods or ions

For comparison of overall deviation between measurements, the absolute value of x-y was taken in the formula above. In this case, if relative percent differences were at or below 50 % using modified Equation 4-1, the results were considered similar (for example, 0.5 nmol/mL and 0.75 nmol/mL), while a relative percent difference above 50 % was not considered similar. A sign test was used to determine whether the quantitative values using different methods or ions provided significantly similar results (less than or equal to 50 % difference) across the majority of features or significantly different results (greater than 50 % difference). For example, if 90 out of 100 concentrations calculated using two different methods had equal to or less than 50 % difference, they would be considered to generally provide similar results as corroborated by the sign test p-value of 0.01; while if 90 out of 100 calculated concentrations were

different by over 50 % they would be considered to generally provide different results as corroborated by the sign test p-value of 0.01.

Precision of quantification using different methods or ions for replicate injections was determined using relative standard deviation (RSD). A sign test was used to determine whether features tended to have higher RSDs in one methodology compared to another.

Results and Discussion: Coverage by AIF and Comparison of Data-Processing Methods on Lipid Concentration Calculated

A table showing a comparison of available lipid quantification software used to process data from UHPLC-HRMS/MS workflows is shown in Table 4-1. To our knowledge, LMQ and LDA are the only software programs which are both open-source and can employ class representative quantification using internal standards. While LDA is a full solution, from feature detection to quantification, LMQ can more easily be integrated into workflows, leveraging other open source tools, for example MZmine and LipidMatch, as employed in this manuscript. Peak picking and lipid annotation can be performed with various software, and parameter optimization can be application, instrument, and workflow specific. Therefore, by integrating LMQ into a larger open source or proprietary lipidomics workflow, users do not need to validate and optimize new peak picking and annotation strategies. The only requirements are a separate column in the feature table for lipid retention time, m/z, class, and adduct. This can be obtained using the text to columns function in Excel if the information is not separated in the native output format.

A total of 129 unique lipid molecular species across 16 lipid types were identified in negative ion mode, of which 122 had appropriate internal standards for quantification

(with phosphatidylinositol not having a class specific internal standard). In positive ion mode 225 unique lipid molecular species across 20 lipid types were identified, with 185 quantified using appropriate class representative internal standards. The output tables with concentrations calculated for SRM 1950 data acquired in positive and negative mode using LMQ and peak areas can be found in the supplementary 2017_9_11_LMQ_Software.zip file under Example_Files. These outputs are the .csv files generated via LMQ, and include the LMQ quantification score, and the internal standard species and adduct used for quantification for each feature. Annotations in column 9 of the tables were obtained from LipidMatch, with an annotation beginning with "1_" representing identifications by targeted MS/MS, and "2_" by AIF. The majority of annotations were obtained using all ion fragmentation (AIF), with the remainder identified using targeted MS/MS.

In AIF, the precursor-fragment relationship is lost due to the wide isolation window, which allows all ions within the m/z range of interest to be fragmented. This can lead to a drastic increase in false positives. LipidMatch filters fragments using a correlation cutoff obtained from a linear regression of the elution profile of the precursor against the fragment ions. This AIF algorithm is advantageous in correctly annotating closely eluting peaks, as compared to using data-dependent scans (Supplemental Figure S4-2). Due to the high number of fragmentation scans in AIF for any precursor, the elution profile of the reconstructed mass chromatogram of the fragment specific to one overlapping isomer, but not the other, can be used to annotate the closely eluting peaks (Figure S4-2). Example elution profiles of precursors and respective fragments are shown in Figure 4-3 for PC(16:0_20:4) and PC(18:2_20:4) identified in positive and

negative ion mode. For PC(16:0_20:4) all precursor and fragment peaks elute with a similar profile at 7.6 minutes, while for PC(18:2_20:4) all precursors and fragments elute with a similar profile at 7.1 minutes. Overlapping elution profiles in the AIF reconstructed mass chromatograms are due to numerous lipids of different precursor mass containing the same fatty acyl constituents. For example, note that the shared arachidonic acid (20:4) leads to the same fragmentation elution profile for NL R1COOH, LPC(R2)+H, and R2COO- in Figure 4-3. This indicates why it is important to employ correlation of elution profiles of precursors and fragments in AIF to reconstruct the precursor-fragment relationship, rather than only identifying lipids based on the occurrence of their respective fragments in the retention time region they elute.

Of all lipids identified in negative ion mode, 98 features were uniquely identified by AIF, 20 uniquely identified using target MS/MS, and 11 identified by both AIF and the targeted MS/MS. In positive ion mode, 85 features were uniquely identified by AIF, 88 by targeted MS/MS, and 52 by both. Of the features annotated both by AIF and targeted MS/MS, 100 % had the same annotation (top ranked, considering plasmenyl and plasmanyl species differing by one saturation the same) in negative ion mode, and 87 % had the same annotation in positive ion mode. Of those in positive ion mode with differing annotations between AIF and targeted MS/MS, the annotations only differed by fatty acid composition, not by lipid class and total carbons and degrees of unsaturations. The majority (6 of 7) of the features with differing annotations using AIF versus targeted MS/MS were annotated as TGs, which are known to be difficult to annotate due to significant overlap of isomers in the retention time regime. These results suggest that annotation of AIF data using LipidMatch provides similar results to traditional MS/MS

approaches, and has low false positives, especially at the level of sum composition and lipid class.

A table of features annotated with unique molecular lipids, their quantitative values, and the internal standards used for each feature, can be found in the supplementary zip file containing the LMQ software. Comparisons of the concentrations calculated using different ions and data processing strategies were made for each of the quantified lipids. For the comparison of different ions and polarities, only those lipid molecules which were represented by both ions, or both polarities, were used.

Different data processing methodologies and ions for quantification were compared in terms of final concentrations, as well as each method's precision in measuring three replicate injections. The quantification comparisons were as follows: (1) smoothed versus non-smoothed peak heights, (2) smoothed versus non-smoothed peak areas, (3) peak area versus peak height, (4) negative versus positive polarity (peak areas), and (5) major adducts versus sodium adducts (peak areas). The number of features used for each comparison, percent difference, and log two of the fold change, are summarized in Supplementary Table S4-3. Comparisons of smoothed versus non-smoothed peak heights, peak area versus peak height, and quantification on positive versus negative ions, all had an r^2 above 0.9 and slopes about equal to 1 in log-log plots shown in Figure 4-4. In addition, a significant proportion of relative percent differences were at or lower than 50 % for comparisons (Figure 4-5), with p-values of a two-sided sign test less than $p < 0.05$. Smoothing had the least impact on final concentration, with none of the 185 lipids above 50 % difference, and only two above 25 % difference. Peak height versus peak area also provided relatively similar

concentrations with only about 13 % of the 185 lipids above 50 % difference. Of these three comparisons, polarity had the greatest effect on concentration, with 25 % of lipids having percent differences above 50 % (in this case only lipids common between polarities were utilized). This has major implications for which polarity is chosen as "correct" for a given set of lipids, with the greater concentration not always correct for a host of reasons. For example the greater intensity or concentration for one adduct over another could be due to overlap of similar peaks, not because of better ionization efficiency.

For comparisons of peak area versus height, the greatest percent difference was for triglycerides, with concentrations calculated in peak area much greater than those calculated by peak height. For ten of the triglycerides, the concentrations calculated using peak area were more than 2-fold higher than those calculated by peak height (over 100 % percent difference; Figure 4-5B). A closer look at extracted ion chromatograms (EICs) and integration using MZmine 2 of these peaks showed a common trend (Figure 4-6). For the triglycerides with minimal difference between peak height and peak area (less than 5 % in Figure 4-6B and supplementary Figure S4-3B), the peaks were well defined (Gaussian shaped and baseline resolved) without any visual overlap. For the triglycerides with major differences between peak height and peak area (over 100 % in Figure 4-6A and Supplemental Figure S4-3A), there were overlapping isomers without complete deconvolution. Therefore, the integration of multiple overlapping isomers as one peak (improper deconvolution and/or poor chromatographic separation) was the major cause explaining why concentrations calculated using peak areas were much greater than those using peak height. In

addition, the number of isomers integrated as one peak varied across samples (Figure 4-6A and Supplemental Figure S4-3A). This led to a large variation in concentrations calculated using peak areas in the case of overlapping peaks, and hence using peak height in lipidomics may be advantageous when a large portion of isomeric peaks overlap in retention time.

The majority of lipid concentrations calculated in positive and negative polarity differed by less than 50 %. For those which differed by more than 50 %, there was no clear trend in EICs. For example, the EICs of PC(16:0_20:5) and PC(18:0_20:4) had similar elution profiles between species and as protonated and formate ions (Supplemental Figure S4-4). While EICs looked similar, concentrations calculated in negative and positive polarity for PC(16:0_20:5) differed by over 2-fold (over 100 %), while for PC(18:0_20:4) concentrations differed by less than 10 %. This data suggest that certain species may have very different ionization efficiencies compared to the internal standard and response curves for negative and positive polarity, while others do not. Indeed, Zacarias et al. [154] showed non-linearity in intensity versus concentration in negative ion mode irrespective of instrumental parameters, while lipid intensity versus concentration in positive ion mode was relatively linear in comparison.

While adducts determined in negative ion polarity correlated well and gave similar quantitative values as adducts in positive polarity, sodiated adducts gave very different concentrations (Figure 4-5D) and did not correlate with their corresponding adducts in positive polarity (Figure 4-4D). For comparison of quantification using major ions versus sodium ions, a targeted list for sodium was developed by copying retention times and changing the masses of the $[M+H]^+$ and $[M+NH_4]^+$ ions detected. This

conversion of protonated and ammoniated species to a sodiated m/z was automated by pasting the molecular species into LipidPioneer [157]. The targeted peak list was then uploaded and the data were reprocessed using MZmine as described in the methods section. No trends were observed in EICs between sodiated species and their corresponding adducts ($[M+H]^+$ or $[M+NH_4]^+$). This is potentially due to sodium not being added to solution, and hence concentrations of sodiated species could be impacted by the number of sodium ions dissolved in the mobile phase at the point of elution, the number of competing ions forming sodiated species, co-eluting isomers, and the concentration of the analyte. As shown by lack of correlation to major adducts, the concentration of analyte seems to be a minimal factor in the intensity of sodium adducts of the analyte. It is possible that adding signal intensities of all adducts for the same molecular species and the associated standard could improve quantification. When adding $[M+Na]^+$ to $[M+H]^+$, there was a slight increase in the relative percent difference between the concentrations calculated in positive ion mode compared to negative ion mode for LPCs and PCs and a significant decrease in the percent difference for ceramides. But due to the instability of the sodium adducts intensities across injections, it is not recommended to include them in calculations of concentration, and hence we recommend removal of sodiated features from the dataset for quantitative analysis.

In addition to the overall difference between concentrations calculated using different data-processing methods and ions, the residual standard deviation (RSD) between three replicate injections of SRM 1950 was calculated for each method. For all methods, the average RSD was less than 20 % (Table 4-2). Concentrations calculated using positive polarity, peak area, and non-smoothed data were more reproducible

across multiple injections when compared to concentrations calculated using negative polarity, peak height, and smoothed data, respectively, as indicated by a two-tailed sign test and lower RSDs (Table 4-2). These results may not be generalizable to all datasets and workflows, and further experiments should be done comparing the effect of these parameters on RSD.

Conclusions: LMQ is a Flexible Tool for Applying Current Relative Quantification Methods to Various Worfkflows

LipidMatch Quant (LMQ) employs internal standards to quantify lipids in UHPLC-HRMS/MS open source workflows. The flexibility in the input feature table format allows LMQ to be used as a backend to any lipid annotation software. LMQ utilizes a unique algorithm to select a standard which best represents the lipid being quantified by matching lipid class, adduct, and retention between the feature and the internal standard in order of priority, respectively. LMQ allows for multiple internal standards per lipid class and provides a scoring system allowing for transparency, noting how each internal standard was chosen for each lipid class and adduct.

Applying LMQ to compare quantitative values obtained using various data processing workflows and ions, we found that the ion chosen for quantification had the greatest effect on the resulting concentrations. Negative and positive ions showed slightly different concentrations, while sodium ions provided drastically different concentrations compared to all other ions. We suggest not to utilize sodium adducts in calculating lipid concentration, at least in cases where sodium is not intentionally added to the mobile phase and samples. Data processing had less of a significant effect, with the greatest difference in calculated concentrations being attributed to peak area versus peak height, when the feature consisted of multiple unresolved chromatographic peaks.

Additional features which could be employed for relative quantification, include response factors based on instrument response to lipid structure (carbons and degrees of unsaturation), and dialogue boxes to aid users in selecting internal standards when class representative standards do not exist. Currently, only the relative quantification portion of LMQ is validated, but future work may provide interactive graphing packages to support in-depth visualization and statistical analysis. The incorporation of such packages would allow users the ability to export custom graphs from the LMQ software. Finally, a single interface is under development to incorporate MZmine, LipidMatch, LipidMatch Quant, and MetaboAnalyst [160] into a user-friendly lipidomics workflow.

Table 4-1. Comparison of different lipid quantification software which can be applied to UHPLC-HRMS/MS data

	Output	IS: Class Specific*	Multiple IS per Class**	Response Factors***	Vendor Specific	License
Lipid Data Analyzer	Concentration	Yes	Yes	No	No	Open Source
MZmine 2	Normalized Peak Intensities	No	–	No	No	Open Source
LipidMatch Quant	Concentration	Yes	Yes	No	No	Open Source
SimLipid	Concentration	Yes	Yes	No	No	Purchase
LipidSearch	Concentration	Yes	No	No	No	Purchase

* Can internal standard be matched to features for quantification based on lipid class?
** Can multiple internal standards for a single lipid class be used?
*** Are response factors based on lipid structures and resulting ionization efficiencies employed?

Table 4-2. Comparison of the relative standard deviation (RSD) of concentrations calculated using different methods or ions

Test	RSD (Avg)	RSD (# >)*	Sign Test
$[M+H/NH4]^+$	5 ± 3 %	31	p = 0.057
$[M+Na]^+$	10 ± 10 %	49	
Pos	4 ± 5 %	10	p < 0.0001
Neg	12 ± 15 %	42	
Height	7 ± 5 %	126	p < 0.0001
Area	6 ± 7 %	59	
Smoothed	7 ± 6 %	103	p < 0.0001
Not Smoothed	6 ± 5 %	82	

*The number of species with RSDs greater in the respective method or ion.
Note that comparison for ions were made using peak areas, while those for smooth versus not smoothed utilized peak heights.

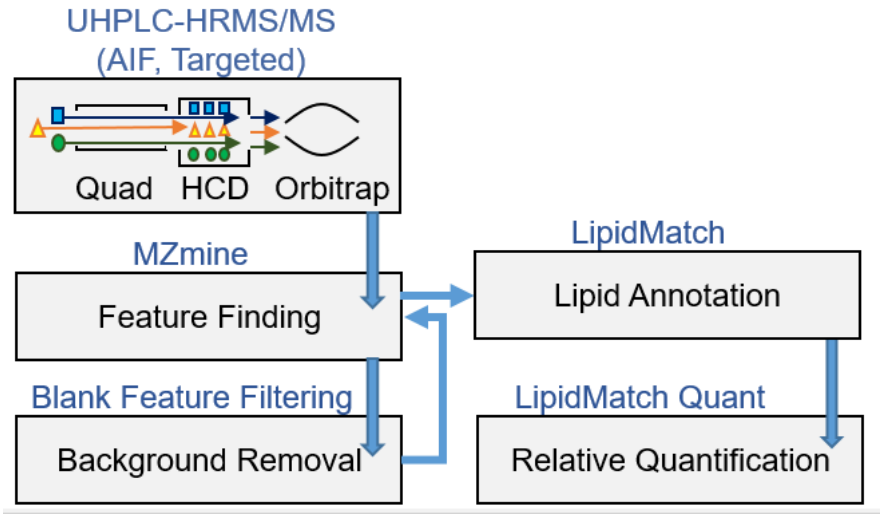


Figure 4-1. Open source lipidomics workflow employed in this study

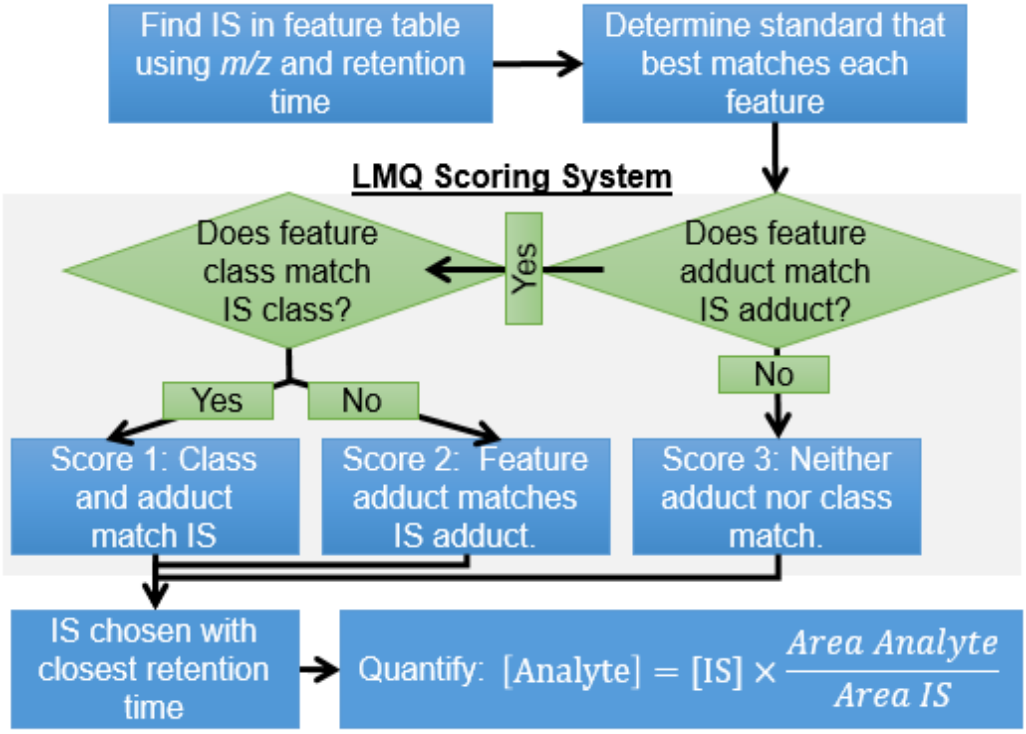


Figure 4-2. Simplified schematic of LipidMatch Quant (LMQ) algorithm.

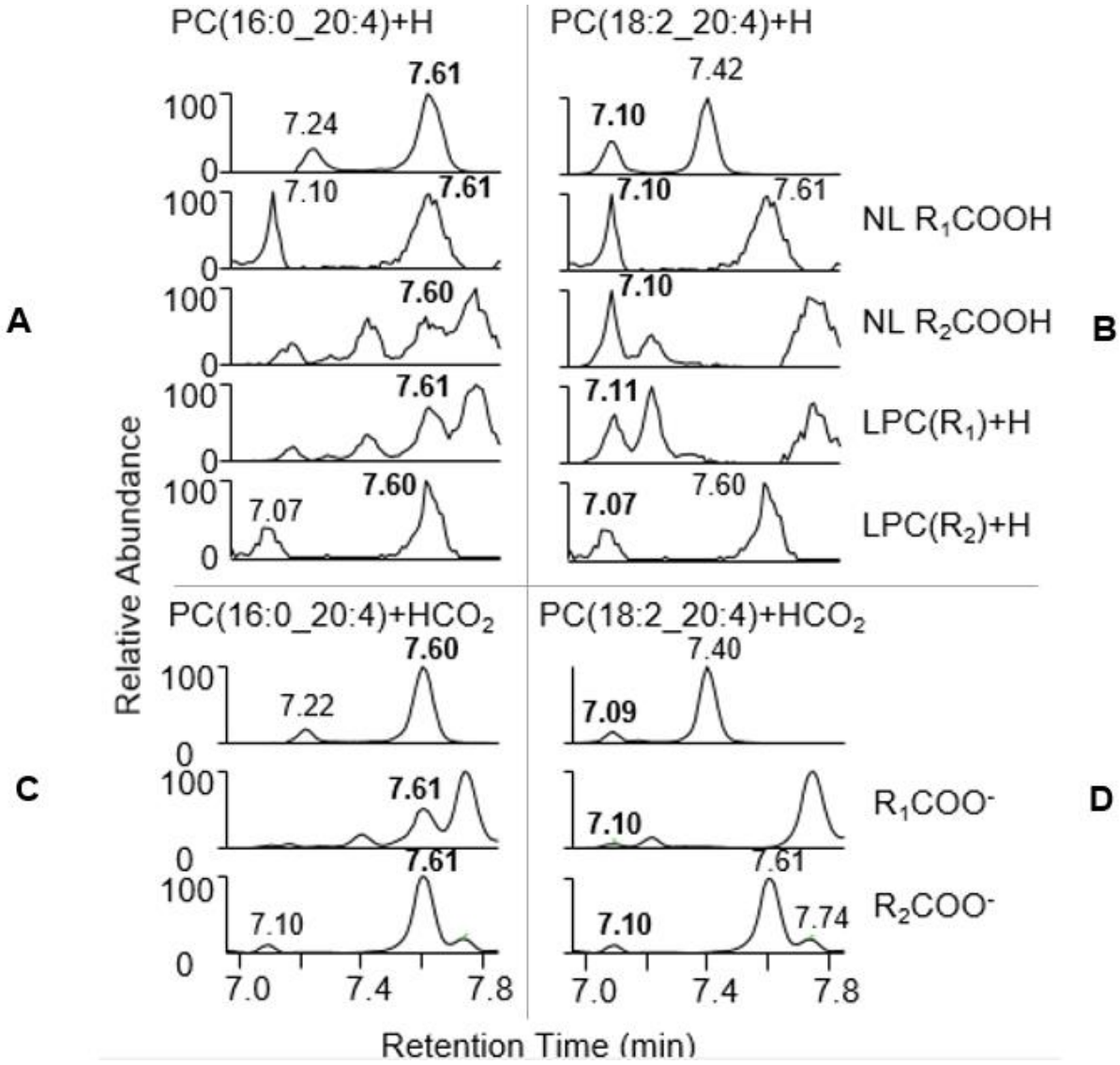


Figure 4-3. Examples of extracted mass chromatograms for the precursors and fragments of PC(16:0_20:4) and PC(18:2_20:4) as protonated and formate adducts in positive and negative mode, respectively. Fragmentation was obtained by AIF and shows correlation of the elution profile of precursor and fragments for both species. The bold retention time values represent the precursor or precursor fragments, while non-bold represent fragments from the precursor in the adjacent panel. A) protonated PC(16:0_20:4), B) protonated PC(18:2_20:4), C) PC(16:0_20:4) as a formate adduct, and D) PC(18:2_20:4) as a formate adduct.

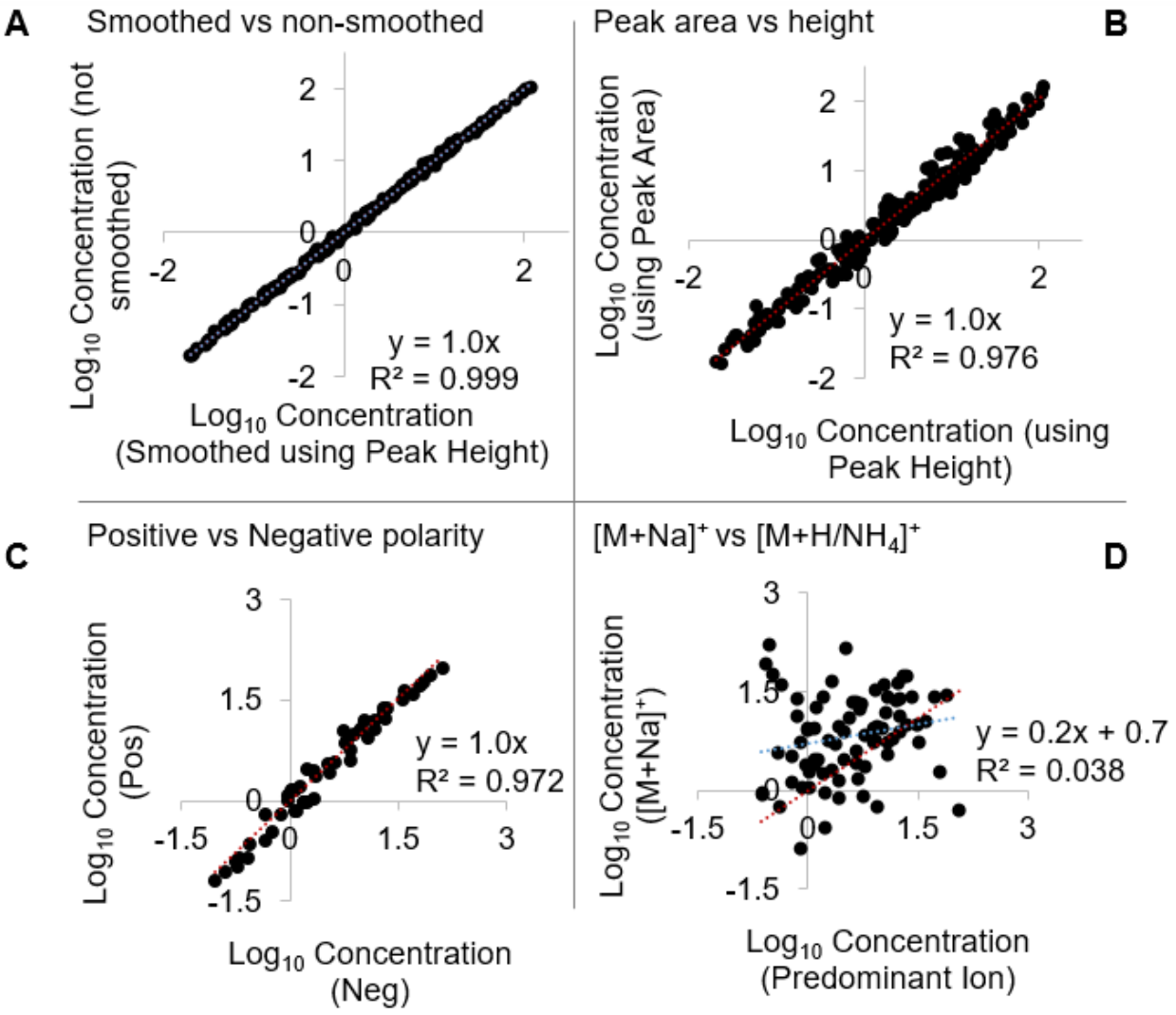


Figure 4-4. Linear regression comparing the $\log_{10}$ of concentrations calculated using different workflows and ions. A slope of 1 and R^2 close to 1 are expected if the methods or ions both result in similar concentrations. The panels show: A) concentrations calculated using smoothed versus non-smoothed peak heights (smoothing was done as the final step in MZmine), B) peak area versus peak heigh, C) positive versus negative polarity using peak area, and D) sodium adducts versus the major adduct observed in positive polarity using peak area. Sodium adducts were compared to protonated adducts except in the case of neutral lipids which formed ammoniated adducts.

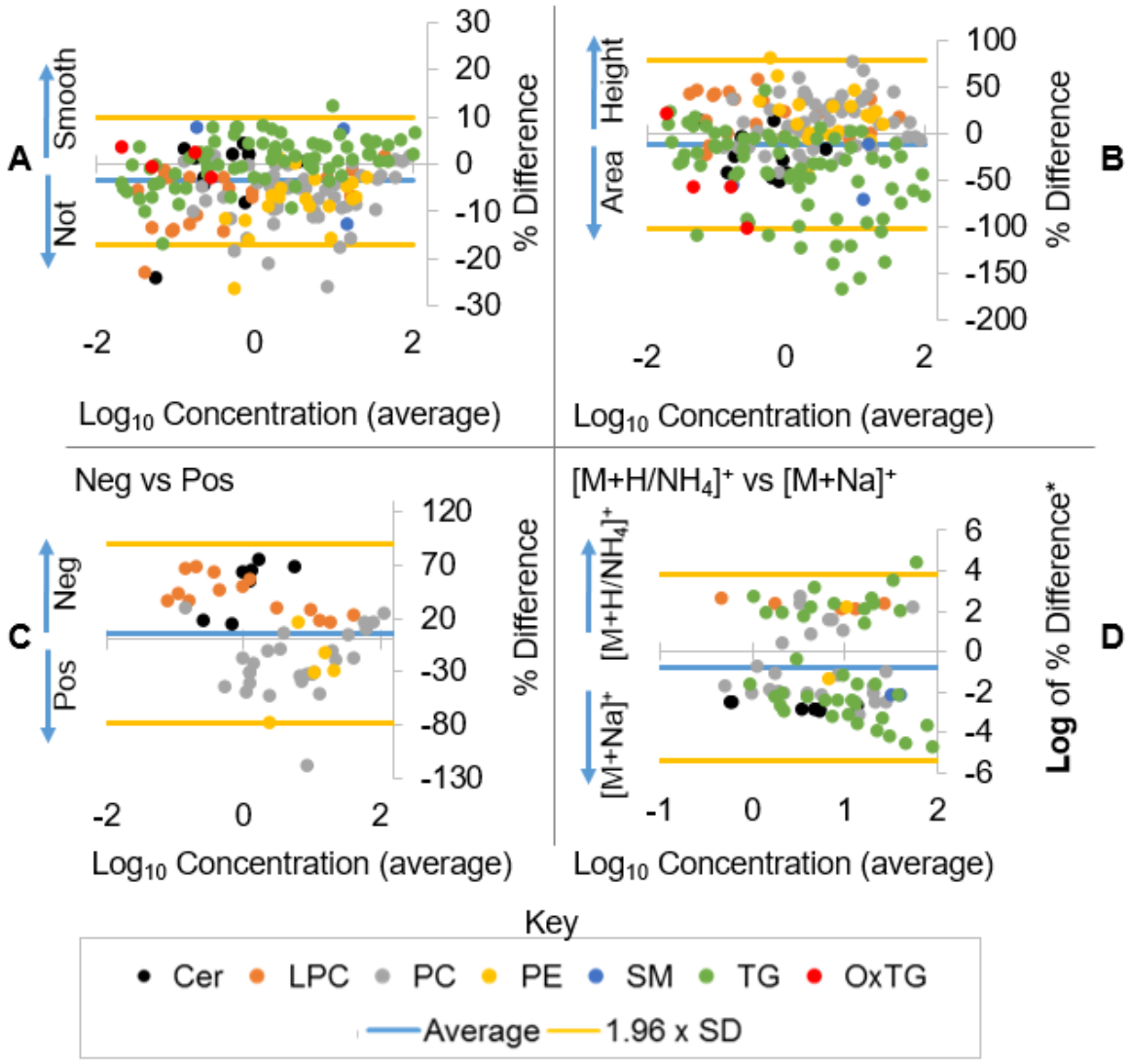


Figure 4-5. Bland-Altman type plots showing differences in concentrations calculated using different methods and ions. See Formula 1 for relative percent difference calculation. Blue arrows delineate the direction of difference.The panels show: A) the percent differences in concentrations calculated using smoothed versus non-smoothed peak heights (smoothing was done as the final step in MZmine), B) peak area versus peak height, C) positive versus negative polarity using peak area, and D) sodium adducts versus the major adduct observed in positive polarity using peak area.

*Note that the differences between major adducts and $[M+Na]^+$ were drastic and ranged over several orders of magnitude. Therefore, the log of the absolute percent difference was used and then multiplied by -1 when the $[M+Na]^+$ concentration was calculated higher than the major ion.

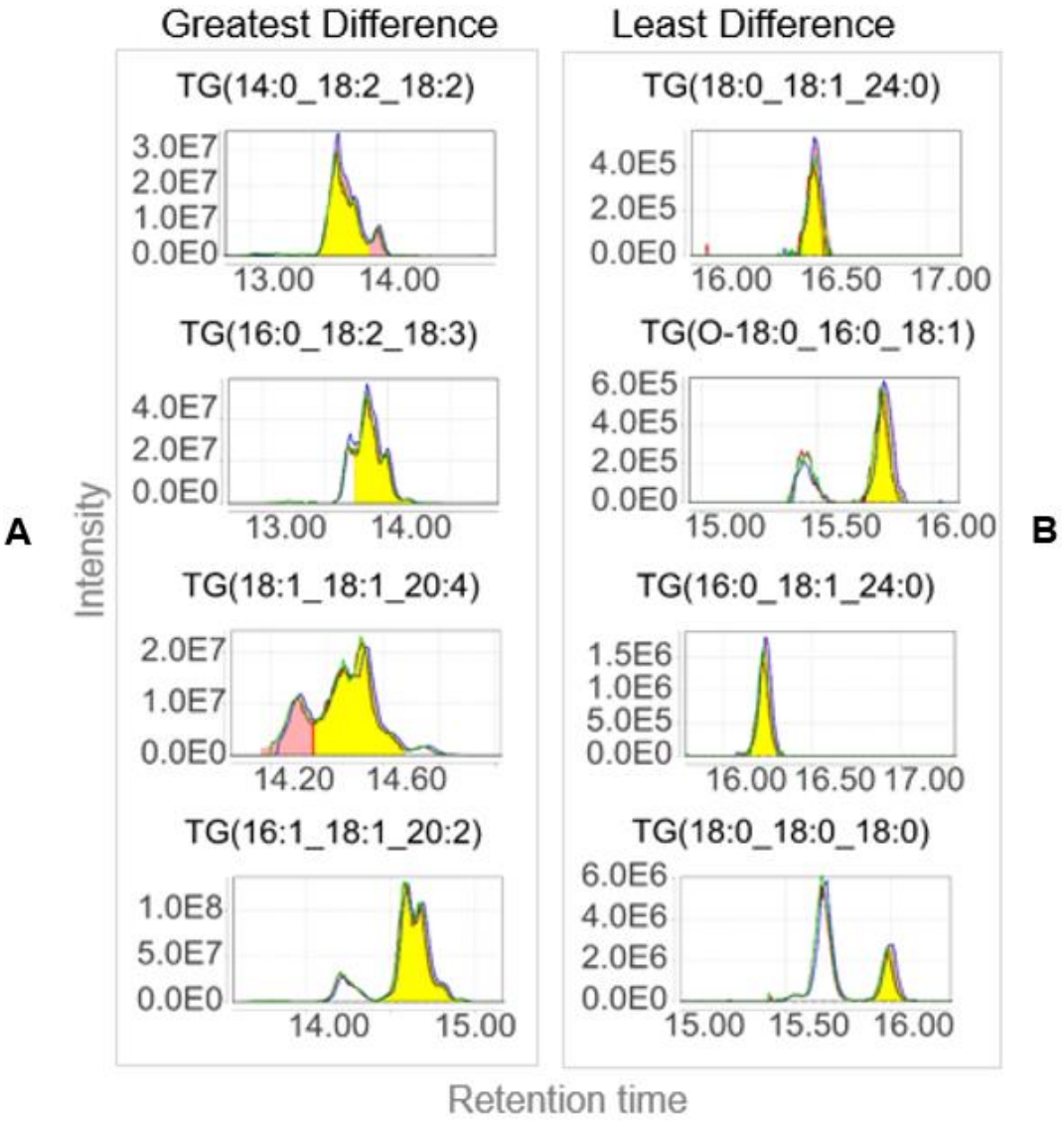


Figure 4-6. Extracted ion chromatograms (EICs) and peak integration by MZmine of the triglycerides (TGs): A) with the most, B) and least percent difference when comparing quantification using peak height versus peak area

CHAPTER 5
EXAMINING HEAT TREATMENT FOR STABILIZATION OF THE LIPIDOME

The Case for Heat Treatment to Improve Lipid Stability in Tissues

Over the last decade, lipidomics has steadily gained status as an established strategy for human health and disease research. Inherently linked to this emergence have been the substantial advances in the lipidomics workflow, specifically advances in lipid extraction,[161] separation, mass spectrometric detection,[27] data processing, and biochemical interpretation/analysis.[162] Success of lipidomics research lies in the understanding of the fundamentals and sources of bias/variability at each step of the lipidomics workflow. Perhaps the most overlooked aspects of the lipidomics workflow are the pre-analytical steps, those steps or decisions made regarding samples prior to analysis. Despite community-wide awareness of these issues[163–165], the current gap between the perceived gravity of pre-analytical influence and the efforts to address these concerns remains large.[166]

The preservation of the native omic profile from the time of collection (t_0) is an ongoing pre-analytical challenge. Flash freezing samples in liquid nitrogen immediately after collection is considered the gold standard for halting metabolic activity. Sample collection in the natural environment (field studies) poses a difficulty in this regard, as the availability of liquid nitrogen and other methods for maintenance of ultra-cold environments may be limited, and thus analyzed samples may not exhibit a true t_0 profile. Laboratory studies are not exempt from issues using flash freezing. The integrity of samples is maintained by remaining in a frozen state until chemical extraction; however, it is often necessary to weigh samples prior to extraction where thawing is undoubtedly occurring. At this point, certain lipids can be susceptible to changes via

enzymatic activity,[30] chemical degradation due to pH, and oxidation,[29] thus complicating the interpretation of the observed lipid profile. A study by Wang et al. analyzed the lipidome in mouse and rat plasma and noted that several lipids significantly increased or decreased after 4 hours on a benchtop.[30] To better preserve the lipidome, an enzyme inhibitor (phenylmethylsulfonyl fluoride) was added to the plasma in the study, thus creating a more native profile. For oxidized lipids, chemical stabilizers, such as butylated hydroxytoluene, are commonly used to prevent oxidation.[31] While chemical stabilizers can be applied to matrices such as plasma, stabilizing solid matrices, such as tissues, at the time of collection requires alternative approaches.

Heat treatment is an alternative approach which has been used as early as the 1940s, for example, to reduce lipid byproducts produced by phospholipase C and D in the presence of alcohols used for extraction.[32] Jerneren et al. examined the application of heat treatment on lipid stability using recent technology that precisely controls temperature and creates a vacuum to reduce oxidation in an effort to study post-mortem effects on free fatty acid composition in brain and liver tissues. It was found that heat was effective in reducing phospholipase activity and *ex vivo* lipolysis in comparison to tissues that were not treated with heat.[167] This idea of heat treatment, as a means to preserve the lipidome in tissues, was expanded upon by Saigusa et al. where it was shown that heat treatment was able to retard metabolic changes in certain sphingolipids.[168] While advantages have been shown for specific lipids, a comprehensive study of the heat treatment effect on the lipidome using state-of-the-art heat treatment equipment, ultra-high pressure liquid chromatography (UHPLC), and

high-resolution mass spectrometry (HRMS), is missing. In addition, to our knowledge, true t_0 studies, where heat treatment is used as the method of metabolic arrest, have not been shown.

In this regard, we aimed to evaluate heat treatment as an approach to improve lipid stability using high-resolution tandem mass spectrometric approaches coupled to a UHPLC system. We examined the specific effects of heat treatment on over 40 lipid types including oxidized lipids using LipidMatch lipid identification software[131] <http://secim.ufl.edu/secim-tools/> in three invertebrate species: the earthworm (*Eisenia fetida*), the house cricket (*Acheta domestica*), and the ghost shrimp (*Palaemonetes paludosus*). The selection of the three invertebrate species was based on the fact that each species has been used previously as bioindicators for environmental monitoring[169–171] and their small size allows them to be directly placed in the heat treatment cartridge for a true t_0 study. Additionally, it has been previously noted that some invertebrate species, like the earthworm, have high enzymatic content.[172] In metabolomics-based studies of earthworms, it was shown that heating whole-body metabolite extracts was beneficial for improving metabolite and total lipid stability,[172, 173] thus making them ideal candidates for this examination. The current study was performed both to highlight the potential lipid changes that occur immediately following animal sacrifice, and to propose heat treatment as a means to reduce this change. We examined whether heat completely stabilizes the lipid profile by allowing heat-treated samples to sit for one hour prior to lipid extraction. Based on trends in lipid profiles, with and without heat treatment, we discuss the underlying enzymatic pathways leading to

the largest lipid transformations, and the benefit of heat treatment to prevent these actions.

Materials and Methods

Materials

The three invertebrate species investigated in this work included earthworms (*Eisenia fetida*) (n = 18) (Haddrell's Point Tackle, Mount Pleasant, SC), house crickets (*Acheta domestica*) (n = 18) (Haddrell's Point Tackle, Mount Pleasant, SC), and ghost shrimp (*Palaemonetes paludosus*) (n = 10) (PetSmart, North Charleston, SC). Female crickets were excluded from analysis based on identification of the ovipositor, and shrimp containing eggs were excluded. Worms, crickets, and shrimp were selected at similar sizes and weights. Optima methanol and HPLC grade chloroform were purchased from Thermo Fisher Scientific (Waltham, MA, USA). HPLC grade 2-propanol was acquired from Alfa Aesar (Haverhill, MA, USA). Ultrapure water with 18 MΩ cm resistivity (Millipore Milli-Q Gradient A10; EMD Millipore, Billerica, MA, USA) was used for sample preparation. Ammonium acetate and analytical grade formic acid were purchased from Fisher-Scientific. All mobile phase solvents were Fisher Optima LC/MS grade (acetonitrile, isopropanol, and water).

All internal standards used in the mixtures were purchased from either Avanti Polar Lipids (Alabaster, Alabama) or Nu-Chek Prep (Elysian, MN). The mixture consisted of the following gravimetrically weighed exogenous internal standards in µg lipid per g solvent: CE(17:0) (45.1 µg/g), Cer(d18:1/17:0) (5.8 µg/g), DG(20:0/20:0) (43.1 µg/g), FA(18:0-d_{35}) (34.1 µg/g), GalCer(d18:1/12:0) (7.8 µg/g), LPC(17:0) (11.4 µg/g), LPE(14:0) (15.5 µg/g), PA(14:0/14:0) (45.0 µg/g), PC(14:1/14:1) (46.0 µg/g), PE(15:0/15:0) (17.1 µg/g), PG(15:0/15:0) (20.4 µg/g), PS(14:0/14:0) (6.6 µg/g),

SM(d18:1/17:0) (15.2 µg/g), ST(16:0) (3.4 µg/g), and TG(17:1/17:1/17:1) (257.5 µg/g). About 250 µL of the internal standard mixture was gravimetrically weighed and added to each cricket and worm sample prior to extraction, while 150 µL of the internal standard cocktail was added to each shrimp sample.

Heat Treatment

A schematic of the experimental design is shown in Figure 5-1. Samples were immediately euthanized for heat treatment by placing the whole invertebrate into a Denator Maintainor® cartridge (Supplemental Figure S5-1) which was placed in the Stabilizor™ T1 (Denator, Uppsala Science Park, Sweden). Of the samples analyzed, nine worms, nine crickets, and five shrimp were euthanized by heat treatment followed by immediate freezing until extraction. Based on internal measurements employed by the Stabilizor™ T1, the actual heating parameters were as follows: treatment time, 23.9 ± 2.9 seconds; maximum temperature, 95.1 ± 0.2 °C; minimum temperature, 94.0 ± 0.2 °C; and pressure, 4.5 ± 0.9 mbar. Treatment parameters were significantly less variable for shrimp, likely owing to smaller body mass than worms and crickets.

Alternatively, samples were euthanized by rapidly freezing in aluminum foil in a cryocart cooled by liquid nitrogen to below −140 °C. Of the samples analyzed, nine worms, nine crickets, and five shrimp were euthanized by rapid flash freezing.

All common house crickets and earthworms (heat-treated and flash-frozen) were then cryo-pulverized using a Retsch cryomill (Retsch, Haan, Germany), and the resulting powder was then transferred to a 2-mL cryovial and subsequently placed in a −80 °C freezer overnight. Whole shrimp were placed in 2-mL cryovials without cryo-pulverization because of their small size. All samples were stored for less than 12 h in a −80 °C freezer before extraction.

Lipid Extraction

Shrimp. Each whole ghost shrimp was removed from the −80 °C freezer and immediately weighed (average mass of 0.11 g, n = 10) and homogenized in a Bullet Blender® (Next Advance, Averill Park, NY, USA) with 1 mL cold methanol using 0.5 mm zirconium beads (Next Advance, Averill Park, NY) for 4 min cycles at speed 10. The shrimp homogenate was then transferred to a new vial for lipid extraction.

Cricket and earthworm. The cricket and earthworm powdered samples were removed from the −80 °C freezer and a portion of the powder was weighed for extraction. The tissue powder (cricket: average mass, 0.089 g (n = 18); earthworm: average mass, 0.10 g (n = 18)) were then transferred into new vials. To evaluate lipid stabilization as a result of the euthanization method (heat-treated or flash-frozen), 10 cricket and 10 earthworm samples (5 heat-treated and 5 flash-frozen per organism) were incubated on ice for 1 h prior to lipid extraction.

Lipid extraction followed the Bligh–Dyer extraction protocol;[38] three mL of a methanol/chloroform mixture (2:1, v:v) were added to all of the samples. Then, 250 µL of the internal standard mixture was added to the cricket and earthworm homogenates, while 150 µL was added to the shrimp homogenates. Following the addition of the internal standard, 1.8 mL of water was added to the cricket and worm homogenates, while 1.7 mL was added to the ghost shrimp homogenates. One milliliter of chloroform was then added to all of the homogenates and vortexed. The samples were subsequently centrifuged (IEC Centra CL3, Thermo Fisher Scientific, Waltham, MA, USA) at 2000 rpm for 10 min, and the chloroform layers were transferred to new vials. An additional 2 mL of chloroform were added to the homogenate samples for re-extraction and samples were vortexed and centrifuged again. The chloroform layers

were combined with the previous chloroform layers. Samples were dried (Biotage TurboVap® LV, Charlotte, NC, USA) and reconstituted in 1 mL of 2-propanol. Then, samples were transferred to autosampler vials, dried, and reconstituted in 200 µL of 2-propanol to concentrate them for mass spectrometric analysis.

Lipidomic Analysis via Mass Spectrometry

Cricket, earthworm, and ghost shrimp reconstituted extracts were kept at 4 °C in the autosampler. Five µL was analyzed using reversed-phase liquid chromatography with a Waters Acquity BEH C18 column (50 mm × 2.1 mm, 1.7 um, Waters, Milford, MA) maintained at 30 °C on a Dionex Ultimate 3000 RS UHPLC+ system (Thermo Scientific, San Jose, CA). The gradient ramp was performed using mobile phase A (60:40 acetonitrile:water, v:v) and mobile phase B (90:8:2 isopropanol:acetonitrile:water, v:v:v), both with 10 mmol/L ammonium formate and 0.1 % formic acid (Supplemental Table S5-1).

Mass spectra were acquired using a Q-Exactive Orbitrap (Thermo Scientific, San Jose, CA) equipped with a heated electrospray ionization (HESI II) probe, in polarity switching and data-dependent top 10 (ddMS^2-top10) modes. In positive-ion mode, diisooctyl phthalate (*m/z* 391.2842) and polysiloxanes (*m/z* 371.1012 and *m/z* 445.1200) were used as lock masses. No lock masses were used in negative-ion mode. The Q-Exactive was externally calibrated before data-acquisition. The S-lens voltage, skimmer voltage, inject flatapole offset voltage, and bent flatapole DC voltage were 35 V, 25 V, 15 V, 8 V, and 6 V, respectively. All samples were analyzed in batches (each invertebrate constituted a batch) with at least three solvent blanks analyzed per batch. Samples were randomized within batches. Four ddMS^2-top10 analyses were conducted

for each invertebrate for lipid identification (two in positive polarity and two in negative polarity). Scanning parameters are provided in Supplemental Table S5-2.

Data Processing

A list of features (*m/z* and retention time) and peak areas were obtained using MZmine 2.2.[100] Chromatograms were smoothed and deconvoluted using a local minimum search; isotopic peaks were grouped, features were aligned, and missing features were gap-filled. Peaks with peak heights below 1×10^5 intensity were removed. An MZmine batch file containing exact parameters can be found in the Supplementary Information. Features were then identified using a proprietary software LipidSearch (Thermo Scientific, San Jose, CA)[99] and an open source software LipidMatch[131], which can be accessed at <http://secim.ufl.edu/secim-tools/>. These software have different confirmation criteria and fragmentation libraries, thereby adding additional confidence if features were identified as the same lipid by both software and expanding the total number of identified lipids. Resulting identifications from both LipidSearch and LipidMatch were aligned with code available in LipidMatch to the MZmine outputted feature list. A script was also used to combine duplicate features using a retention time window of 0.1 and an *m/z* window of 0.006, retaining the maximum value of all the replicates.

Identified lipids were semi-quantified using class representative standards, and if no standards were analyzed for the lipid class in question, the standard with the greatest similarity in structure and retention time was used (Supplemental Table S5-3). Quantified lipids were normalized to tissue weight (µg lipid per g of wet tissue). A list of defined acronyms and the observed adducts in LipidMatch are provided in Supplemental Table S5-4.

After quantification, features in negative and positive polarity were combined, using negative mode data if molecular species were represented in both polarities. Lipids quantified in negative mode were prioritized due to lower background noise and more accurate annotation due to high intensity fatty acyl fragments.[174] Molecular species represented by multiple adducts were quantified using the adduct with the highest calculated concentration. Tentative identifications, such as identifications by exact mass only or identification of sphingomyelin (SM) or phosphatidylcholine (PC) solely by the *m/z* 184 fragment ion, were removed. Sodium adducts were also removed from the final dataset, as quantification on the sodium ion without excess sodium is problematic. For oxidized lipids, which are both present in LipidMatch and LipidSearch databases, only lipids identified by MS/MS in both negative and positive polarity were kept to reduce the probability of false positives. Therefore, oxidized lipids that only occur in positive polarity, for example TGs, were excluded from analysis; oxidized lipidomics is an emerging field without proper guidelines for proper annotation, and therefore only the most confident identifications were retained.

Statistical Analysis

After quantification and annotation of features, both univariate and multivariate statistical methods were applied. Mean centering and a generalized log transformation (Glog) was used prior to principal component analysis (PCA) using Metaboanalyst 3.0.[160]

For univariate methods, analysis of variance (ANOVA) was used to investigate the following questions: (1) euthanization method: heat-treated versus flash-frozen (D vs N), (2) heat stabilization: direct extraction versus 1 h ice incubation then extraction (D vs D1hr), (3) flash-frozen stabilization: direct extraction versus 1 h ice incubation then

extraction (N vs N1hr), and (4) compare stabilizations: heat-treated versus flash-frozen samples (D1hr vs N1hr). For shrimp, only the comparison between the euthanization method (D vs. N) was made. The p-values from ANOVA across all comparisons were adjusted using Benjamini and Hochberg's false discovery rate (FDR) method.[175] Analyses across different organisms, were treated as separate experiments, and hence FDR was calculated separately for each of these instances. To determine individual lipids that were significant, an FDR adjusted p-value of 0.05 was used as the cutoff. To determine if there were trends by lipid class, a Fisher's exact test was used with a p-value cutoff of 0.05 using the Fisher.test R function. A one-sided Fisher's exact test was used to determine if a significant number of lipids within a class were upregulated or down regulated compared to lipids across all classes. A list of potentially upregulated or downregulated lipids for Fisher's exact test was determined using an FDR adjusted p-value of 0.25. Note that an FDR adjusted p-value of 0.25, means that about one of four significant compounds will be a false positive. Because the Fisher's exact test was used to determine general trends across lipid classes, we are not interested in whether a particular lipid is significant at this FDR rate, but whether lipids increase or decrease within a class compared to across all classes using this FDR rate.

Results and Discussion: Evidence for Deactivation of Lipases

Lipid identifications from extracted common house crickets, earthworms, and ghost shrimp, using LipidSearch and LipidMatch generally agreed (Supplemental Table S5-5), especially at the level of lipid class, total carbons, and degrees of unsaturation. There was better agreement between software in negative ion mode than positive ion mode, and therefore when combining polarities, if a molecular species was represented by both polarities, negative ion mode data was retained. Using these software, a total of

426 (common house cricket), 593 (earthworm), and 429 (ghost shrimp) unique lipids were annotated by MS/MS (Supplemental Table S5-5). A total of 40 unique lipid types were identified when combining lipids from all sampled organisms. For an overview of the lipidome coverage for each organism, distributions of lipid concentrations and the number of species annotated per class are presented in pie charts in Supplemental Figure S5-2.

Multivariate statistics were used to determine general changes in the lipid profile across samples and treatments. PCA scores plots showed different patterns of enzymatic degradation or enzymatic transformation of lipids depending on organism (Figure 5-2). All heat-treated (D) and flash-frozen samples (N) were separated along the first principal component (PC1) with *t*-test p-values of the scores less than 0.05, and PC1 explained variances of 25 % (earthworm), 66 % (cricket), and 62 % (shrimp) (Figure 5-2). In crickets, PC2 better separated heat-treated and flash-frozen samples, with a p-value of 0.01 compared to 0.03 for PC1 scores of all heat-treated versus frozen samples.

To determine whether lipid degradation continued following euthanization, select samples were incubated on ice for 1 h (D1hr and N1hr) (Figure 5-1). For crickets and worms, the PCA scores plot suggests the heat-treated samples (D) and flash-frozen samples (N) had the least amount of variance, while samples after 1 h incubation on ice (D1hr and N1hr) showed higher variance (Figure 5-2A and Figure 5-2B). This suggests lipid changes occurred after an hour on ice, and that heat treatment did not completely stabilize the lipids. The earthworms had distinct separation between flash-frozen (N and N1hr) and heat-treatment (D and D1hr) (Figure 5-2B), unlike the crickets (Figure 5-2A)

where the variability of the ice incubated samples (D1hr and N1hr) resulted in less distinct groupings.

The top 30 features indicative of the lipids differentiating between flash-frozen (N) and heat-treated samples (D) in PCA and their loadings are shown in Supplemental Table S5-6. Of the top 15 features influencing heat-treated samples, for crickets, worms, and shrimp, 100 % consisted of intact glycerophospholipids and oxidized glycerophospholipids, suggesting minimal enzymatic degradation compared to flash-frozen samples. Of the top 15 features indicative of flash-frozen samples, over 60 % consisted of lysoglycerophospholipids for crickets and 100 % consisted of lysoglycerophospholipids for worms (Supplemental Table S5-6a and Supplemental Table S5-6b), indicating enzymatic degradation of intact glycerophospholipids to lysoglycerophospholipids. For shrimp, 87 % of the top 15 loadings indicative of flash-frozen samples consisted of phosphatidylmethanol (PMe), an enzymatic product produced when using methanol for extraction (Supplemental Table S5-6c).

These enzymatic products are likely generated in the pre-analytical steps of lipid analysis, such as homogenization or freeze/thawing, which can cause lipid membranes to rupture, resulting in stored calcium in the mitochondria and endoplasmic reticulum lumen to be released. Calcium activates proteases, lipases, and kinases, which in turn can lead to degradation of lipid species.[176] For example, glycerophospholipids (GPL) in the presence of phospholipase A, phospholipase C in concert with various kinases, and phospholipase D in concert with PA phospho-hydrolase are cleaved, producing products including diglycerides (DGs), lysophosphatidylcholines (LPCs),

lysophosphatidylethanolamines (LPEs), and LPC and LPE plasmanyl and plasmenyl corollaries, as shown in Figure 5-3.

In a univariate analysis of the euthanization effect on lipids, a Fisher's exact test was used to determine trends across lipid classes. The Fisher's exact test p-values (Table 5-1 and Supplementary Table S5-7) show which lipid classes were the most significantly upregulated or downregulated in comparison of euthanization techniques. One-sided Fisher's exact test on heat-treated (D) and flash-frozen (N) shrimp and worm samples showed an overall increase in LPC and LPE species in flash-frozen samples (Table 5-1b and Table 5-1c), with p-values less than 0.005. The general lower concentration of lyso-glycerophospholipids (lyso-GPL) in heat-treated samples (Table 5-1), provides evidence that phospholipase A_2 activity[177] (Figure 5-3) was reduced. For crickets, LPC and LPE species did not differ in heat-treated (D) and flash-frozen samples (N) that were not incubated on ice.

While no significant difference in LPC and LPE was observed between heat treated (D) versus flash-frozen (N) cricket samples, which were immediately extracted, after samples were incubated on ice for one hour, there was a significant difference between heat-treated (D1hr) and flash-frozen (N1hr) cricket samples (p-value < 0.005; Table 5-1a). This is due to an increase in LPC and LPE concentration only in samples without heat-treatment after one hour incubation (N1hr versus N). On the contrary, one hour incubation on ice (N versus N1hr and D versus D1hr) did not increase enzymatic production of LPC and LPE species in worm samples, except ether-LPC and ether-LPE species for flash-frozen samples (N versus N1hr) (Table 5-1b). Ether-linked lipids are more resistant to enzymatic degradation by phospholipase A_1,[178] and therefore

slower degradation may explain the continued increase of ether-LPC and ether-LPE after one hour in samples without heat-treatment. This data indicates that the amount, and rate of enzymatic activity are both dependent on lipid class and the sample organism.

LPCs incorporating very long chain fatty acids with over 22 carbons were detected in shrimp and worm by LipidSearch using MS/MS (Table S5-5). For LPC species with 22 or less carbons, 14 out of 16 LPC species in worms, and 8 of the 10 LPC species in shrimp, had FDR p-values less than 0.05 for heat-treated (D) versus flash-frozen (N) samples. For species with over 22 carbons, of the 8 LPC species in worms, and 6 LPC species in shrimp, none had FDR corrected p-values less than 0.05 (Table S5-5). These results suggest that due to the unique enzymatic synthesis and degradation of very long chain fatty acids,[179] either heat treatment is ineffective at preventing enzymatic activity, which generates LPC species with over 22 carbons, or significant enzymatic activity generating LPCs for species with over 22 carbons does not occur during sample preparation. Hence, not only is enzymatic activity dependent on the organism and lipid class, but may also be dependent on fatty acyl constituents.

While low concentrations of lyso-GPL in heat-treated organisms (Supplemental Table S5-5) support that heating reduced phospholipase A activity for species with 22 or less carbons, intact glycerophospholipids would be expected to be higher in heat-treated organisms, as these lipids would not be cleaved by phospholipase A (Figure 5-3). However, glycerophospholipids were not significant according to the Fisher's exact test (Supplemental Table S5-7), potentially owing to the small sample size used in this study. In this case, trends in GPL may be less likely to be discerned against biological

variation and other sources of noise because of the relatively small percent change in concentration compared to their enzymatic products, lyso-GPLs. For example, the total LPE concentration in shrimp was 417 % higher (p-value = 0.0000009) in flash-frozen shrimp (11.8 ± 0.8 μg/g) compared to heat-treated shrimp (2.3 ± 1.1 μg/g), while the average total PE concentration was 27 % lower (p-value = 0.032) in flash-frozen shrimp (647.4 ± 44.3 μg/g) compared to heat-treated shrimp (820.8 ± 123.6 μg/g) (Supplemental Figure S5-3). Using the total sum of PC, PE, and respective plasmanyl and plasmenyl species of earthworms (including samples with and without the additional 1h incubation) suggested that heat treatment reduced degradation of PE to LPE, PC to LPC, and ether linked species to their respective lyso-species (Supplemental Figure S5-3 and Supplemental Figure S5-4). Our results agree with the findings of Wang et al. where lyso-GPL were lower and GPL higher in heat-treated rat and mouse plasma samples compared to control samples stored on a benchtop.[30]

Glycerophospholipids can also be cleaved by phospholipase D to form phosphatidic acids (PA), which can further be converted into DGs by PA phosphohydrolase (Figure 5-3). While no PAs were detected in samples, DGs were detected in crickets and shrimp (Supplemental Table S5-5; Supplemental Table S5-7). In shrimp, all DGs detected were an order of magnitude lower in heat-treated samples, with FDR corrected p-values < 0.005, and Fisher's exact test p-values < 0.005 (Supplemental Table S5-5, Table 5-1c). This trend was not observed for crickets, but, when summing total DG concentration including the samples incubated for 1 h, DGs were 122 % greater (p-value = 0.041) in flash-frozen compared to heat-treated crickets. Further evidence that heat treatment denatures, and hence deactivates, phospholipase

D is shown by the presence of phosphatidylmethanol (PMe) in frozen shrimp (Figure 5-4A). It has been shown previously that water can be replaced with alcohols as acceptors in transphosphatidylation via phospholipase D[34] (Figure 5-3), and as shrimp were homogenized in methanol, PMe are expected. In the flash-frozen samples, 14 PMe species were detected as deprotonated ions, while no PMe ions were found in heat-treated samples. It is important to note that PMe was not detected in crickets or earthworms, as these organisms were large enough for cryopulverization, eliminating the need for an additional homogenization step using methanol. Therefore, it is important to note, that in addition to lipid structure and sample organism affecting enzymatic activity, the sample homogenization and extraction procedure employed also affects enzymatic activity.

Phospholipase D has multiple isoforms that hydrolyze PC, PE, phosphatidylglycerol (PG), phosphatidylserine (PS), phosphatidylinositol (PI), LPC, cardiolipin, and plasmalogen head groups. Certain lipid classes and fatty acyl constituent constituents will preferentially be hydrolyzed depending on the organism. [180] In shrimp, PE(20:5_22:6), PS(20:5_22:6), PC(20:5_22:6), and potential enzymatic products PMe(20:5_22:6), DG(20:5_22:6), LPE(20:5), LPC(20:5), LPE(22:6), and LPC(22:6), suggest PE, PS and PC conversion by phospholipase D into PMe and DG at least for GPLs containing 20:5 and 22:6 (Figure 5-4A-I). Potential fatty acid preferences for enzymatic reaction of phospholipase D existed. For example, in all 13 DGs generated by enzymatic reaction in shrimp (Supplemental Table S5-5), each DG contained either a 20:4, 20:5, 22:5, or 22:6 fatty acyl constituent, despite high intensity species such as PC(16:0_18:1) and PE(16:0_18:1).

Phosphatidylserine (PS) lipid species can be substrates for a specific lipase: PS-specific PLA_1. For shrimp, a closer investigation of PS(20:5_22:6) (Figure 5-4B) suggests that the lower concentration in frozen samples is mainly because of phospholipase D and not PS-specific PLA_1 or phospholipase A . The difference in the amount of degradation products in frozen shrimp compared to heat treated shrimp was found to be similar to the difference in the product source assuming phospholipase D activity. While PC(20:5_22:6) and PE(20:5_22:6) are only 11.7 μg/g lower in frozen shrimp compared to heat-treated shrimp, their likely degradation products in the presence of phospholipase D, PMe(20:5_22:6) and DG(20:5_22:6), are 74.8 μg/g higher in frozen shrimp. This suggests another source of these degradation products. When PS(20:5_22:6) is included as a source of DG(20:5_22:6) and PMe(20:5_22:6) via phospholipase D activity, the concentration becomes 66.9 μg/g (Figure 5-4A-C) for sources of the 74.8 μg/g enzymatic products (Figure 5-4D-K). This suggests that in shrimp, phospholipase D degradation of PS species predominate, versus PS-specific PLA_1, which would lead to lysophosphatidylserines (LPS), which were not detected. Other sources of PMe(20:5_22:6) could be PI(20:5_22:6), PG(20:5_22:6), or PA(20:5_22:6), but none of these species were detected. In addition, DG(20:5_22:6) could be a product of triglyceride lipase activity for TG(16:0_20:5_22:6), TG(18:0_20:5_22:6), and TG(18:1_20:5_22:6) lipids, but together these only accounted for less than 5 μg/g in concentration, and hence could not be a major factor in the 45.1 μg/g increase of DG(20:5_22:6) observed in flash-frozen shrimp.

This study supports previous evidence that lipase activity depends on fatty acyl constituents, which have been shown for numerous lipases including neutral

lipases[181, 182] and phospholipase D.[180] Oxidized lipids have potential to also show various trends depending on acyl chain composition, as they can be produced both via enzymatic and chemical routes owing to exposure to catalysts, electromagnetic radiation, and/or oxygen.[183] On average, all 9 oxidized PEs (OxPE) and oxidized PCs (OxPC) were higher in concentration when incubated on ice for 1 h compared to those immediately extracted after weighing for both heat-treated (except 1) and frozen crickets, but were not significantly different based on FDR corrected p-values and Fisher's exact test (Supplemental Table S5-5a and Supplemental Table S5-7a). The only significant change in oxidized lipids was a general increase in OxPE after one hour in the heat treatment samples (Fisher's exact test p-value = 0.016; Supplemental Table S5-7a). Measurement of oxidative markers such as malondialdehyde could be used in the future to determine if heat treatment increases oxidation.

This manuscript investigated heat treatment as a mechanism to prevent enzyme degradation of lipids. Untargeted studies often are not concerned with absolute quantification, but instead investigate profile changes in response to the experimental design. If all samples degrade similarly, and compounds are detectable, the occurrence of degradation (within acceptable experimental design) may not be a hindrance for a biomarker type assessment. Based on this data, it is not conclusive whether heat treatment reduces variance introduced from enzymatic activity. RSDs across all features (Supplemental Table S5-5) did not follow any trend between heat treatment and flash-frozen samples (t-test). Surveying RSDs of total LPC and LPE concentrations across earthworms and crickets showed higher variability in flash-frozen samples compared to heat-treated samples when incubated and non-incubated samples were

combined (Table 5-2). When RSD's for incubated and non-incubated samples were compared without combining euthanization methods (D vs N and D1hr vs N1hr compared separately) there was no clear trend in RSD between heat-treated and flash-frozen samples. This suggests heat treatment reduces variation, when samples are left on a benchtop for varying amounts of time, but may not reduce variance if all samples are left on the benchtop for the same amount of time prior to extraction. If the amount of degradation is consistent across all samples, then for comparative studies where only relative amounts of lipids are of concern, heat treatment may not be necessary. A future study with a larger sample size that minimizes biological variability, such as aliquots of monoclonal cells grown under the same conditions, could provide more conclusive evidence for using heat treatment for stabilization of the lipidome.

Conclusion: Heat Treatment Warrants Further Investigation

Our study provides evidence that heat treatment deactivates enzymes in the early pre-analytical phase, which may provide better qualitative and quantitative results for lipids. Based on higher concentrations of PE and PCs, and lower concentrations of LPE, LPC, DG, and PMe species in heat-treated organisms, heat treatment reduced or prevented enzymatic degradation of important lipid types in common house cricket, earthworm, and ghost shrimp extracts. This evidence supports reduction in the activity of phospholipase A and phospholipase D, specifically. However, non-enzymatic processes such as oxidation may continue to occur in both heated and frozen samples.

We believe that heat treatment could help minimize variance within treatment groups that can occur as a result of freeze-thaw cycles, variable time between sample collection and freezing, or other factors that affect the rate or duration of enzyme activity. To compare and integrate lipidomes characterized across the international

community, sample stabilization is extremely important, and heat treatment may be a simple solution. This is especially noteworthy for solid matrices, where the inclusion of additives to slow lipid degradation is impractical. Additional studies with larger samples sizes and a homogenous replicated sample would be beneficial in determining whether heat treatment stabilizes the lipid profile during sample preparation.

Table 5-1. Lipid enzymatic products observed for worm, cricket, and shrimp, respectively, and the p-values of the Fisher's exact test. Bold p-values are less than 0.05. The Fisher's exact test was used to determine if the lipid species within a particular lipid class significantly increased or decreased as compared to other lipid species. Comparisons are for flash frozen (non-Denator) versus heat-treated (using the Denator device) samples (N vs D), flash frozen samples which sat on ice for one hour prior to extraction versus heat-treated samples which sat on ice for one hour prior to extraction (N1hr vs D1hr), flash frozen versus flash frozen samples which sat on ice for one hour (N vs N1hr), and heat-treated versus heat-treated samples which sat on ice for one hour. Tables are for A) earthworms, B) common house cricket, and C) ghost shrimp.

Table 5-1 a)

Earthworm	LPC	LPE	ether-LPC	ether-LPE
N vs D	0.000*	0.000*	1.000	1.000
N1hr vs D1hr	0.000*	0.000*	0.000*	0.000*
N vs N1hr	0.315	1.000	0.003*	0.000*
D vs D1hr	0.694	1.000	1.000	1.000

Table 5-1 b)

Common House Cricket	LPC	LPE	ether-LPE	DG
N vs D	1.000	0.142	1.000	1.000
N1hr vs D1hr	0.000*	0.000*	0.028*	1.000
N vs N1hr	0.000*	0.000*	0.012*	1.000
D vs D1hr	1.000	1.000	1.000	1.000

Table 5-1 c)

Ghost Shrimp	LPC	LPE	PMe	DG
N vs D	0.006*	0.001*	0.000*	0.000*

Table 5-2. Residual standard deviations (RSD) of lysophosphatidylcholine (LPC) and lysophosphatidylethanolamine (LPE) measured in heat-treated (D All) and flash frozen samples (N All) across all samples, including those which were immediately extracted and those which sat on ice for one hour.

	Cricket		Worm	
	N All	D All	N All	D All
LPC RSD (%)	72	29	23	20
LPE RSD (%)	72	41	44	32

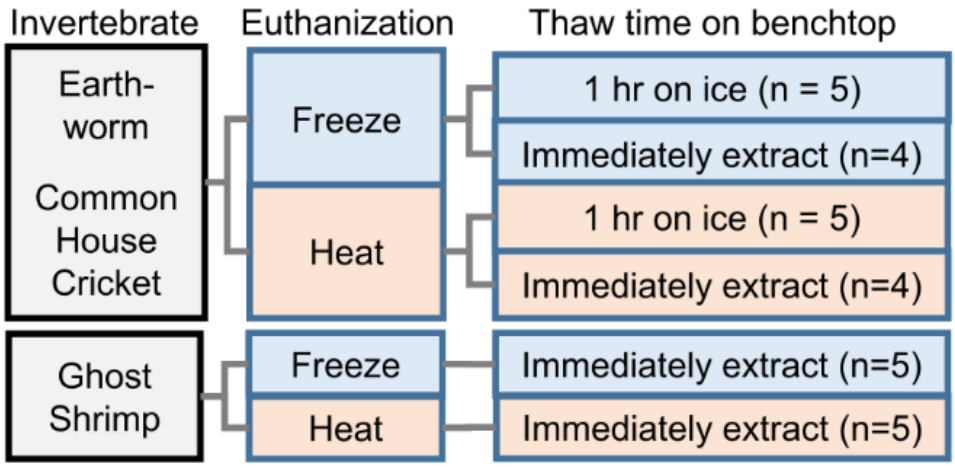
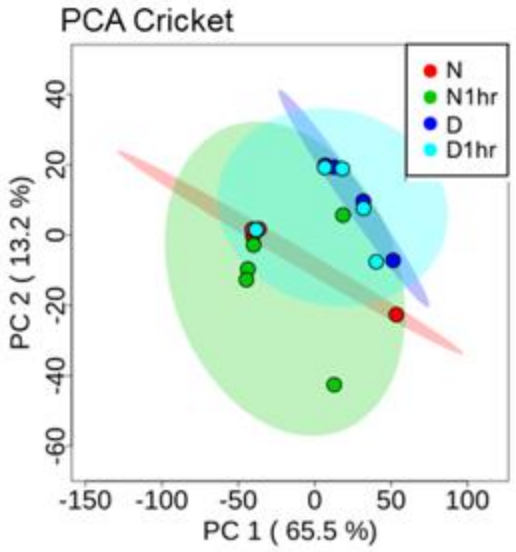


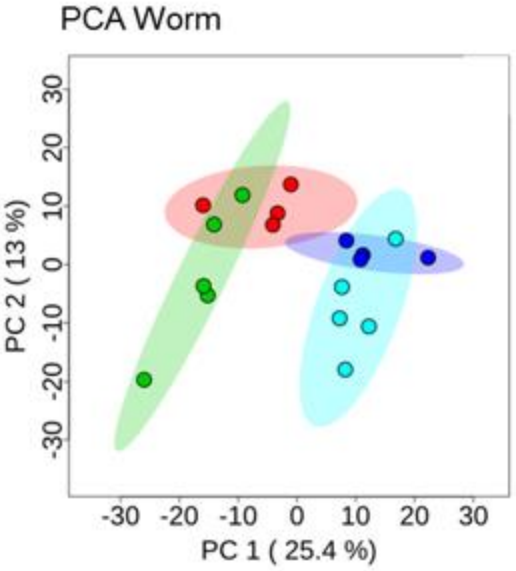
Figure 5-1. Experimental design. Earthworms and crickets were utilized to study the comparison of euthanization by heat treatment or flash freezing, followed by extraction of the homogenized tissue. Additionally, these organisms were used to assess the stabilization of the lipidome by the euthanization method with a 1 h ice incubation. Ghost shrimp were used to test the same euthanization methods but instead incorporated a methanol homogenization step during extraction.

Figure 5-2. Principal components analysis (PCA) of based on (N) flash-frozen, (N1hr) flash-frozen and 1 h ice incubation, (D) heat treatment, and (D1hr) heat treatment followed by 1 h ice incubation (next page). PCA plots are for: A) Crickets, B) Worms, and C) Shrimp

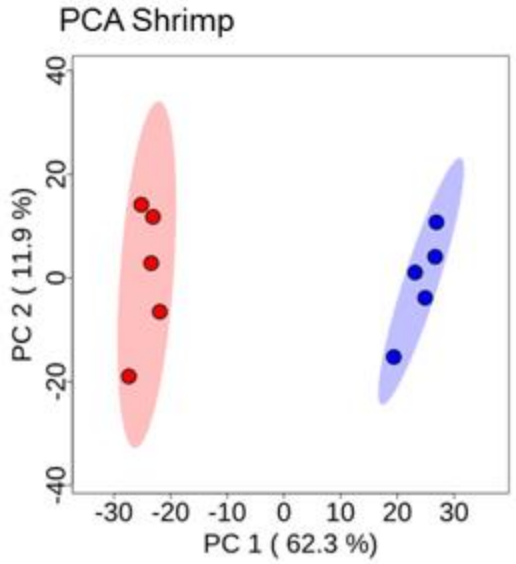
A


PCA Cricket
N
N1hr
D
D1hr
PC 2 (13.2 %)
PC 1 (65.5 %)
40
20
0
-20
-40
-60
-150
-100
-50
0
50
100

B


PCA Worm
PC 2 (13 %)
PC 1 (25.4 %)
30
20
10
0
-10
-20
-30

C


PCA Shrimp
PC 2 (11.9 %)
PC 1 (62.3 %)
40
20
0
-20
-40

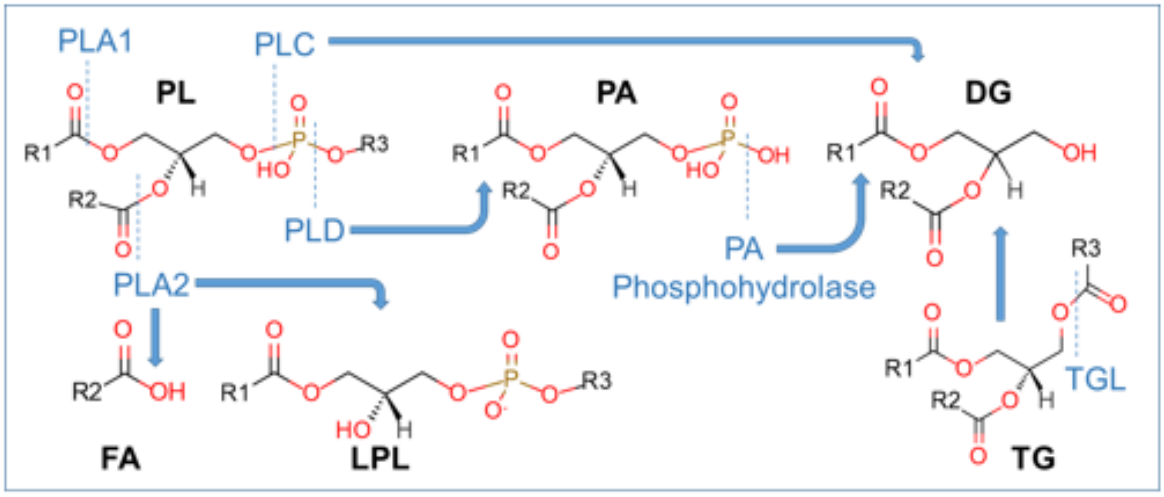


Figure 5-3. Schematic of enzymatic degradation of glycerophospholipids (GPL) and triglycerides (TG). Products and intermediates include fatty acids (FA), lysoglycerophospholipids, phosphatidic acid (PA), and diglycerides (DG). Enzymes involved include phospholipase A1 and A2 (PLA1 and PLA2, respectively), phospholipase C (PLC), phospholipase D (PLD), and triglyceride lipase (TGL). The FA and LPL products from cleavage via PLA1 are not shown.

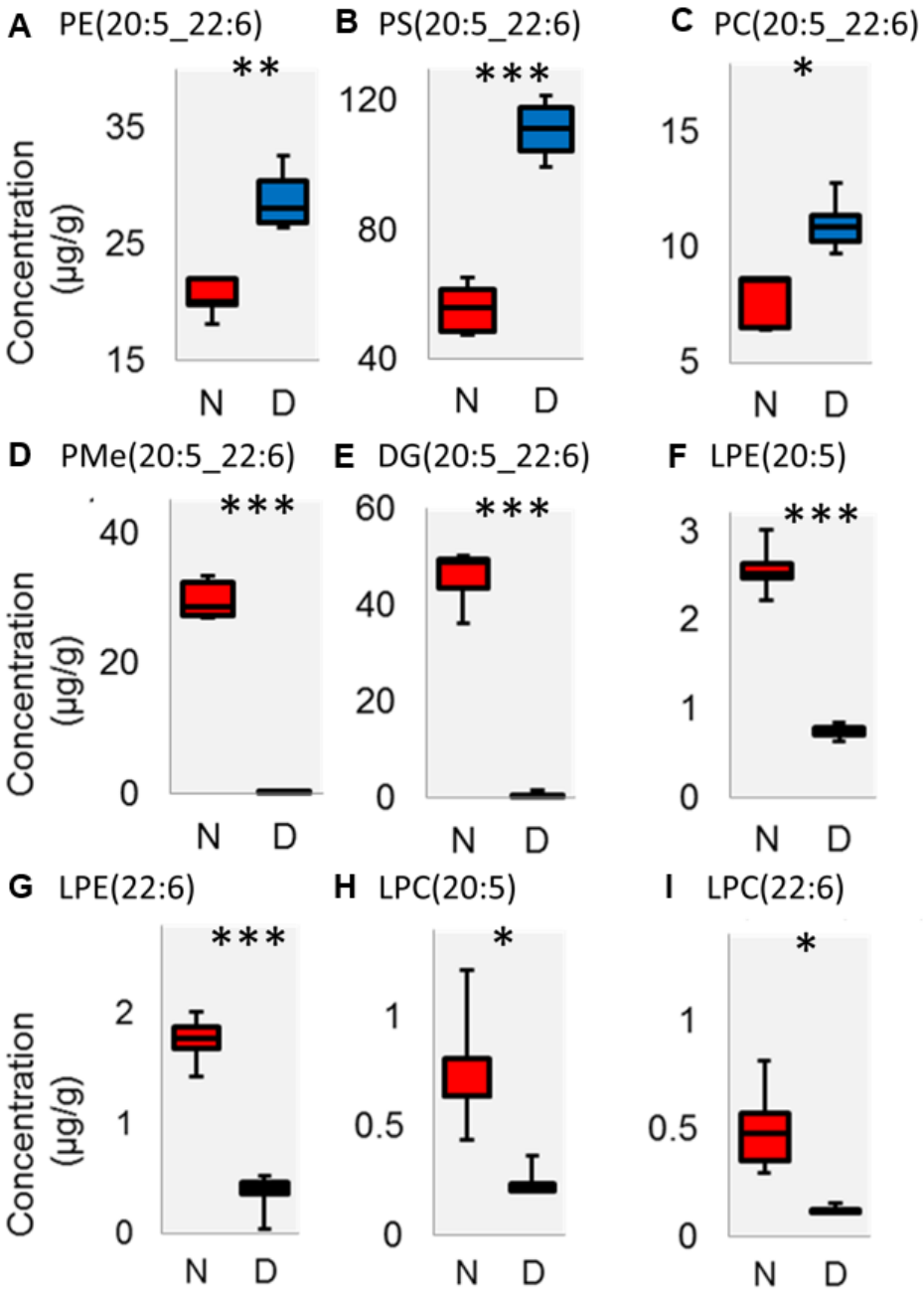


Figure 5-4. Box plots of lipid species concentrations in flash-frozen (N) and heat-treated (D) shrimp. Lyso-lipid (LPC and LPE) concentrations are from protonated ions, glycerophospholipids PE and PMe are deprotonated ions, PC are formate adducts, and DG are ammonium adducts. P-values are Hochberg FDR corrected with three significance levels: $p < 0.05$ (*), $p < 0.005$ (**), and $p < 0.0005$ (***). Panels show: glycerophospholipids containing 20:5 and 22:6 (A - C) and potential enzymatic products (D - I).

CHAPTER 6
CONCLUSION AND FUTURE PROSPECTIVES

While lipidomics has significant potential in aiding biomarker discovery and our understanding of health and disease, the lipidomics workflow is tedious, and there is no community wide consensus on the proper data processing protocols. Without consensus on protocols, untargeted lipidomics is unable to provide reproducible annotations and measurements of lipids that are needed for adaptation to the clinical field. Therefore, we introduce modular lipid data processing tools, which are available to the wider lipidomics community. These tools cover all the major steps of the lipidomics workflow (Figure 1-7) including data conversion to open source format, peak picking and processing (feature finding), blank filtration, annotation, quantification, and removal of redundant adducts. Increasing the throughput and reproducibility of lipidomics data processing, these tools increase the coverage of the lipidome and introduce new protocols for lipid feature finding and quantification.

Specifically, we introduce data-acquisition methods (iterative exclusion) and lipid annotation strategies (LipidMatch), which drastically increase the coverage of the lipidome. The expanded coverage allows for characterization of species that are often un-identified, such as oxidized lipids and ammoniated adducts of cardiolipin. We achieve this by expanding the current high-resolution *in-silico* MS/MS libraries for lipids and developing software to match *in-silico* fragmentation to experimental spectra acquired using data-dependent and data-independent acquisition. In contrast to data-dependent acquisition, data-independent acquisition provides MS/MS of less abundant lipids, which are important in signaling and other important lipid functions. In data-independent analysis, specifically all ion fragmentation, precursor-fragment

relationships are lost due to isolating ions in a window that spans the entire mass range of interest; therefore, fragments could originate from any precursor. The algorithm employed in LipidMatch reconstructs precursor-fragment relationships by correlating precursor chromatographic elution profiles to fragment elution profiles. This correlation dramatically reduces false positives by only using fragments from the precursor in question for identification and excluding fragments from other coeluting ions. However, the correlation of elution profiles drastically increases false negatives, lipids which truly exist but are left unidentified. False negatives occur because overlapping elution profiles for fragments or precursors reduce correlations below the minimum threshold, and as a result, these fragments are excluded from use in identification even though they came from the precursor of interest.

Therefore, to better annotate low abundance lipids, we designed a script to apply iterative-exclusion data-dependent tandem mass spectrometry. The script, which we have called IE-Omics, generates exclusion lists from ions selected in previous acquisitions. After applying the exclusion list to a sequential injection only unique ions unselected in the previous acquisition are fragmented. When applying this method iteratively, over 60 % more lipids can be annotated. This technique provides a drastic increase in coverage as compared to traditional data-dependent analysis and AIF. When all three techniques are used together, the greatest coverage of lipid fragmentation and hence annotations are obtained.

In addition to increasing lipid coverage, it is important to report only the structural details known (referred to as structural resolution) based on experimental data. LipidMatch addresses this issue by using an annotation style which provides users with

only the structural resolution that is known based on their experimental data. While LipidMatch's accurate reporting of structural resolution is an advantage over other open source software, traditional lipidomics experiments employing UHPLC-HRMS/MS do not provide exact lipid molecular structure. However, mass spectrometry techniques have been developed to report more detailed structural resolution of lipids, including fatty acyl positional isomers, double bond position, and double bond *cis*/*trans* isomerism.

Positional isomers such as *sn*1 and *sn*2 isomers can be distinguished in tandem mass spectrometric approaches using the relative ratios of the fatty acyl fragments. However, because these relative ratios can vary between instruments, fragmentation method, and lipid classes, internal standards with varying ratios of *sn*1 and *sn*2 isomers must be used for quantitative approaches. Standards are often impure, and therefore must first be characterized by measuring the ratio of the fatty acid concentrations after treatment with phospholipase A2, which removes fatty acids only from the *sn*2 position [109]. Additionally, the lack of synthetic lipid standards to represent the diversity of lipid structures prohibits absolute certainty in the intensity of lipid fragments derived from the glycerol backbone.[121] However, one promising technique for identification of double bond positions is ozone-induced dissociation (OzID) [184], although specialized equipment for onsite generation of ozone and flow control is needed. In OzID, a traditional tandem mass spectrometric approach to characterize lipids by fatty acyl constituents is followed by the introduction of ozone, which induces fragmentation indicative of double bond positions. Another method for determining double bond location as well as *cis* and *trans* isoforms is the use of silver-ion liquid chromatography, where separation occurs due to weak complexes formed between the π electrons in

double bonds of unsaturated lipids with the silver particles in the stationary phase. This technique often requires chromatographic analyses on the order of hours to separate out all isomers, and hence is impractical for high-throughput studies with large sample sizes [185].

Another promising method for identification of both *sn*1 and *sn*2 fatty acyl positions, double bond positions, and *cis* and *trans* isoforms is to apply ion mobility spectrometry (IMS), a rapid and predictable separation device, in tandem with UHPLC-HRMS/MS studies [110]. Currently, these lipid isomers are rarely baseline-resolved using IMS, but the resolving power of IMS is expected to increase with technological advances such as structures for lossless ion manipulation (SLIM) [186]. Because IMS is easily combined with various liquid chromatographic and mass spectrometric techniques, it could revolutionize molecular characterization of lipids in the near future.

In addition to achieving improved structural resolution of lipid annotations, databases must be designed to incorporate the various levels of structural resolution obtained by mass spectrometry. Otherwise, determination of the biological relevance of down-regulated and up-regulated lipids cannot be accomplished using pathway mapping. This is due to the fact that subtle differences in lipid structure can have dramatic influences on the lipid species biology. Currently, biochemical databases (e.g., Kyoto Encyclopedia of Genes and Genomes (KEGG)), are unable to capture the varying lipid structural resolution conferred by mass spectrometry.

Furthermore, it is important to establish identifiers that query only biological information pertaining to known structural motifs. For example, the ideal case for PC(16:0_18:1) would be an identifier specific to the lipid class and to the fatty acyl

constituents, but not to the *sn*1 and *sn*2 positions or the double bond position. While there is a general KEGG entry for the phosphatidylcholine class, C00157, there is no KEGG entry for specific PCs. In this case, searching KEGG reduces the scope of biological inference to mechanisms general to all PCs. For the Human Metabolome Database (HMDB), identifiers exist for the specific lipid molecule, for example PC(16:0/18:1(9Z)). However, these biological inferences can be too specific (i.e., based on *sn*1 and *sn*2 position) and thus lead to false interpretation of the data. It is important to note that currently, while specific lipid molecules exist in databases such as LIPID MAPS and HMDB, the curated pathways predominantly contain general lipid class biology. Therefore, current biological inference in lipidomics relies either on expert opinion or on lipid class and fatty acid profile-based trends.

Universal chemical identifiers, which can convert a chemical structure into a machine readable string, and vice versa, such as the widely used International Chemical Identifier (InChI) [187], would be extremely useful for electronic record finding of mass spectrometric based lipid annotations. InChI consists of layers, each containing additional information about the molecular structure. Many of these layers can be omitted, hence allowing for some flexibility in structural resolution. For example, layers signifying *cis* versus *trans* double bonds, or chirality of the lipid molecule, can be omitted, in which case any lipid isomers will be found. However, currently InChI requires a minimum of 3 layers, one of which is absolute bond connectivity; therefore the position of a double bond and fatty acid on the backbone cannot be left undetermined limiting application to lipidomics. Chemical query languages, such as SYBYL [188], could be used, with the possibilities of storing Boolean logic, wild cards (unknown atoms and R-

groups), and other functionalities allowing lipid annotations to be stored in a machine readable string which can cover all the different levels of structural resolution provided by mass spectrometry. These identifiers are not widely implemented; in order for them to be useful, they should be implemented in annotation software, databases (such as LIPID MAPS), and by chemical manufacturers.

In addition to accurate lipid annotations, it is important to measure lipid concentrations reproducibly, in order that measurements can be compared across laboratories. Therefore, LipidMatch Quant is introduced to automate the relative quantification process. Lipid quantification is problematic, due to the cost prohibitive nature and unavailability of lipid standards to cover the diverse species within a given lipidome. Therefore, strategies to select the best internal standards for each lipid class are employed, in what is termed relative quantification. LipidMatch Quant can be employed to help take into account ionization efficiencies and ion suppression effects; LipidMatch Quant accomplishes this by selecting internal standards based on lipid class, lipid adduct, and lipid retention time. Following selection of internal standards, a table is generated which includes the concentration of each lipid across each sample and the respective internal standard ion used for each lipid.

While these algorithms are currently of great utility to the lipidomics community, these algorithms are not intended as a final solution to a robust and comprehensive lipidomics workflow. Furthermore, because of the complexity of the lipidome, advances in instrumental and data processing strategies in lipidomics will continue to increase the coverage and accuracy of lipid measurements. Additionally, community-wide efforts to advance portions of lipidomics, metabolomics, and proteomics workflows will continue

to provide valuable tools, which can be integrated with or replace certain tools presented here. By utilizing multiple user input parameters, the tools presented here can readily be integrated with other software outputs. Therefore, the workflows presented in this dissertation can utilize community accepted lipidomics software such as MZmine 2, XCMS, LipidSearch, MS-DIAL, and Metaboanalyst. The flexibility of the workflow also allows for integration with various data acquisition approaches and vendor formats. For example, LipidMatch has been successfully employed with Agilent, Thermo, and ScieX data files for imaging, liquid chromatography, and direct infusion based high-resolution mass spectrometry.

As with data-processing strategies, there are numerous sample preparation protocols employed in the lipidomics community. Many factors in sample preparation reduce the accuracy and precision of lipid measurements, and yet are not fully understood. For example, in this work it is shown that enzyme activity can drastically alter the final lipid concentrations measured. Furthermore, it was determined that enzymatic transformations are highly dependent on lipid class, fatty acyl composition, the organism or substrate studied, and sample storage and preparation protocols. Therefore, for lipidomics to be implemented in a clinical setting, one must account for these factors perturbing lipid measurements. However, most researchers do not sufficiently account for enzymatic degradation in their sample preparation. Adaptation of technologies which employ heat treatment, sample processing entirely at cryogenic temperatures, or other strategies for preserving sample integrity, are needed for robust and accurate lipid measurements.

Much work is needed to further improve accuracy of lipid measurements. Lipid concentrations measured across labs are often drastically different, and annotations using even the most widely accepted software often contain false positives. Therefore, in order to benchmark the progress of lipid measurements, we need quality controls, certified reference materials, and inter-laboratory exercises. As lipidomics continues to show promise in the clinical field, advancement in standardization and development of robust workflows will continue to meet the growing demand.

APPENDIX
SUPPLEMENTAL INFORMATION

Chapter 2

Object 2-1. Additional supplemental information for Chapter 2 (Table S2-4, Table S2-5, and Table S2-6 and all supplemental Figures and Tables in power point) (.xlsx and .pptx files 9 KB)

Table S2-1. Gradient for reverse phase liquid chromatography of lipids. Mobile phase C consisted of 60:40 acetonitrile:water and mobile phase D consisted of 90:8:2 isopropanol:acetonitrile:water, with both containing 0.1% formic acid 10 mM ammonium formate. The flow rate was 500µL/min.

Time (min)	0	1	3	4	6	8	10	13	15	16	16.5	17	21
C (%)	80	80	70	55	40	35	35	20	20	10	10	80	80
D (%)	20	20	30	45	60	65	65	80	80	90	90	20	20

Table S2-2. Mass spectrometric parameters. Abbreviations are: Res – resolution, AGC – automatic gain control, IT – injection time, NCE – normalized collision energy (stepped), $ddMS^2$ – data-dependent tandem-mass spectrometry, Iso – isolation width, Apex – apex trigger, and Dyn Excl – dynamic exclusion.

	Res	AGC	IT (ms)	Range *(m/z)*	NCE
Full Scan	70k	$5x10^6$	256	200-1200	NA
$ddMS^2$	70k	$5x10^6$	256	80-900	20 ± 5*

	Iso *(m/z)*	Intensity Threshold	Appex (s)	Dyn Excl (s)	top N
$ddMS^2$	1	$5x10^4$	5 to 20	6	5

*For Red Cross plasma the stepped NCE was 20 ± 4

Table S2-3. Source parameters (electrospray ionization (ESI))

Omics analysis	Lipidomics		Metabolomics	
Polarity	Neg	Pos	Neg	Pos
Sheath gas flow rate	25	30	45	45
Auxilary gas flow rate	15	5	10	10
Sweep gas flow rate	0	1	1	1
Spray voltage (kV)	3.5	3.5	3	3.5
Capillary temperature (°C)	250	300	325	325
S-lens RF level	35	35	30	30
Aux gas heater temp (°C)	350	300	350	350

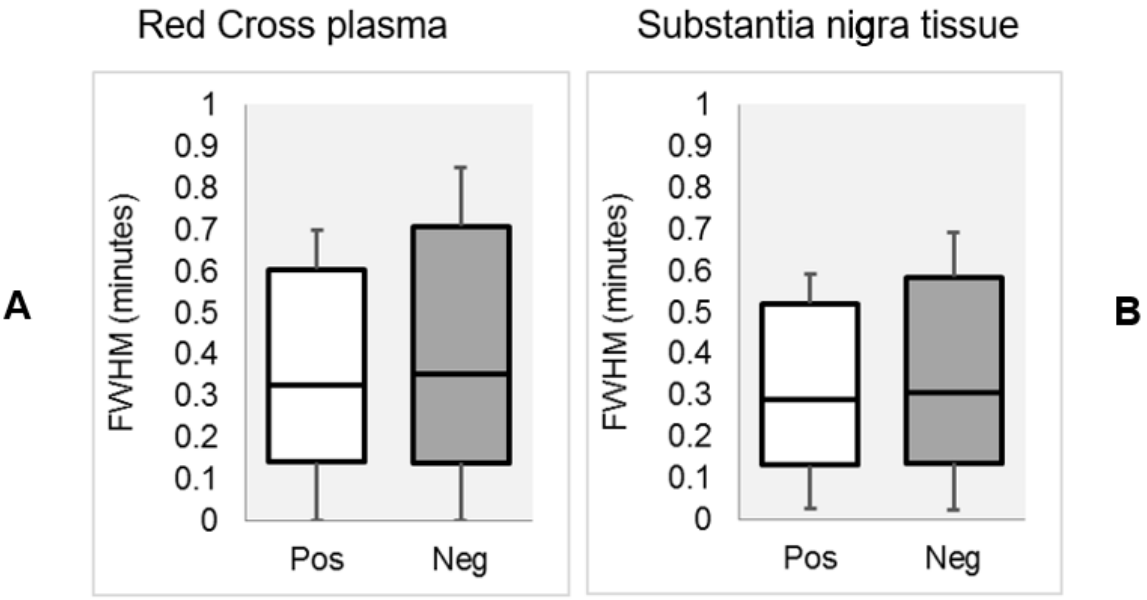


Figure S2-1. The retention time window for both determining the MS/MS scan under a chromatographic feature and for exclusion of ions previously selected, was 0.3 min, which was close to the median of FWHM values. Distribution of chromatographic peaks determined by MZmine full widths at half maximum (FWHM) for: A) Red Cross plasma in positive and negative polarity and B) Substantia nigra in positive and negative polarity.

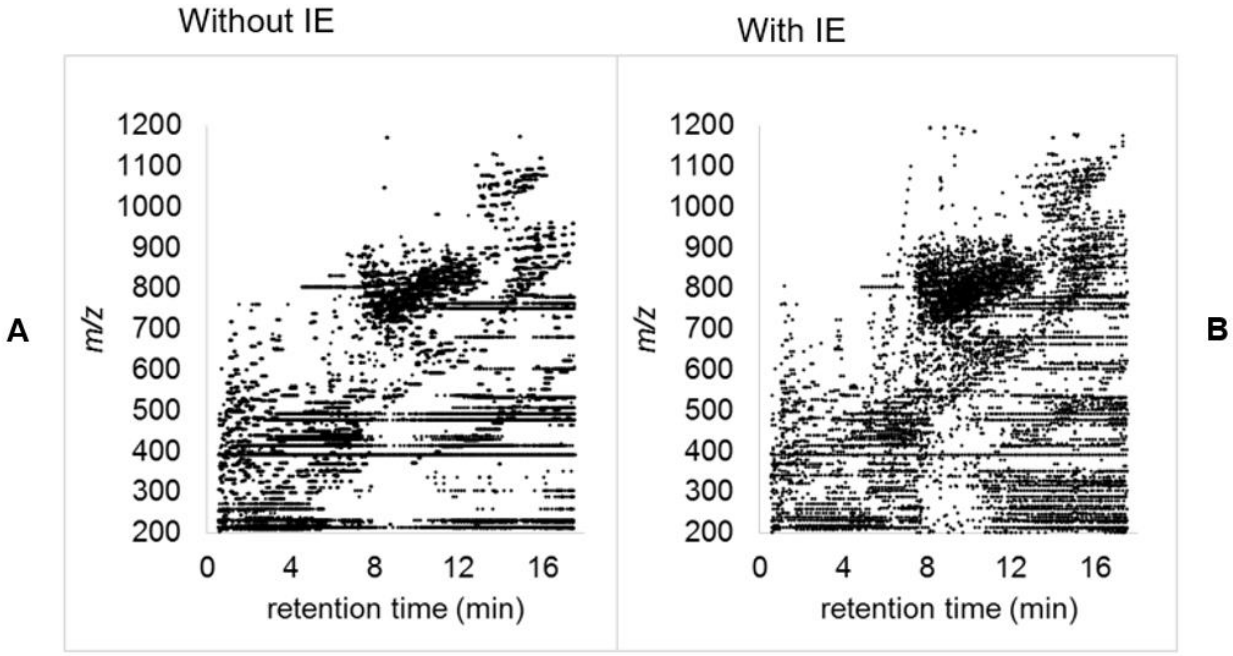


Figure S2-2. A higher density of selected precursor ions in substantia nigra extracts analyzed with IE was observed compared to the traditional ddMS^2 approach. Selected precursor ions *m/z* and retention times for 6 repetitive injections using: A) the traditional ddMS^2 approach and B) iterative based-exclusion ddMS^2 (IE-ddMS^2) for substantia nigra lipid extracts analyzed in positive mode.

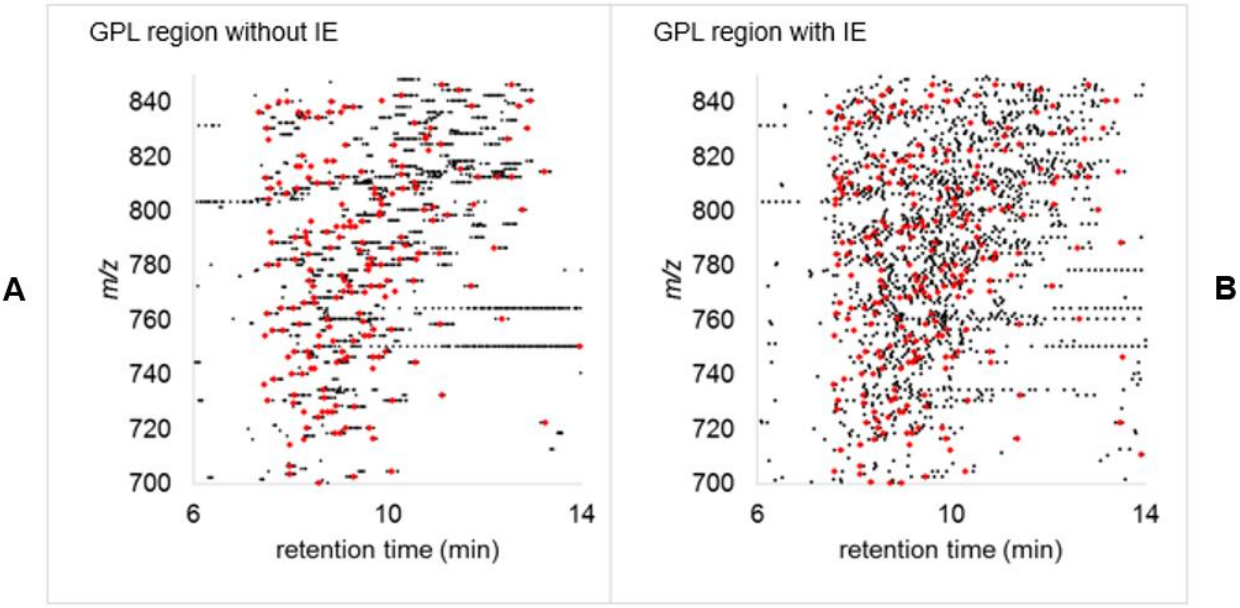


Figure S2-3. Selected precursor ions *m/z* and retention times for: A) 6 repetitive injections using the traditional ddMS^2 approach and B) iterative based-exclusion ddMS^2 (IE-ddMS^2) for substantia nigra lipid extracts analyzed in positive mode, zoomed into the glycerophospholipid (GPL) region. Identified lipid molecules by LipidMatch are also shown in red.

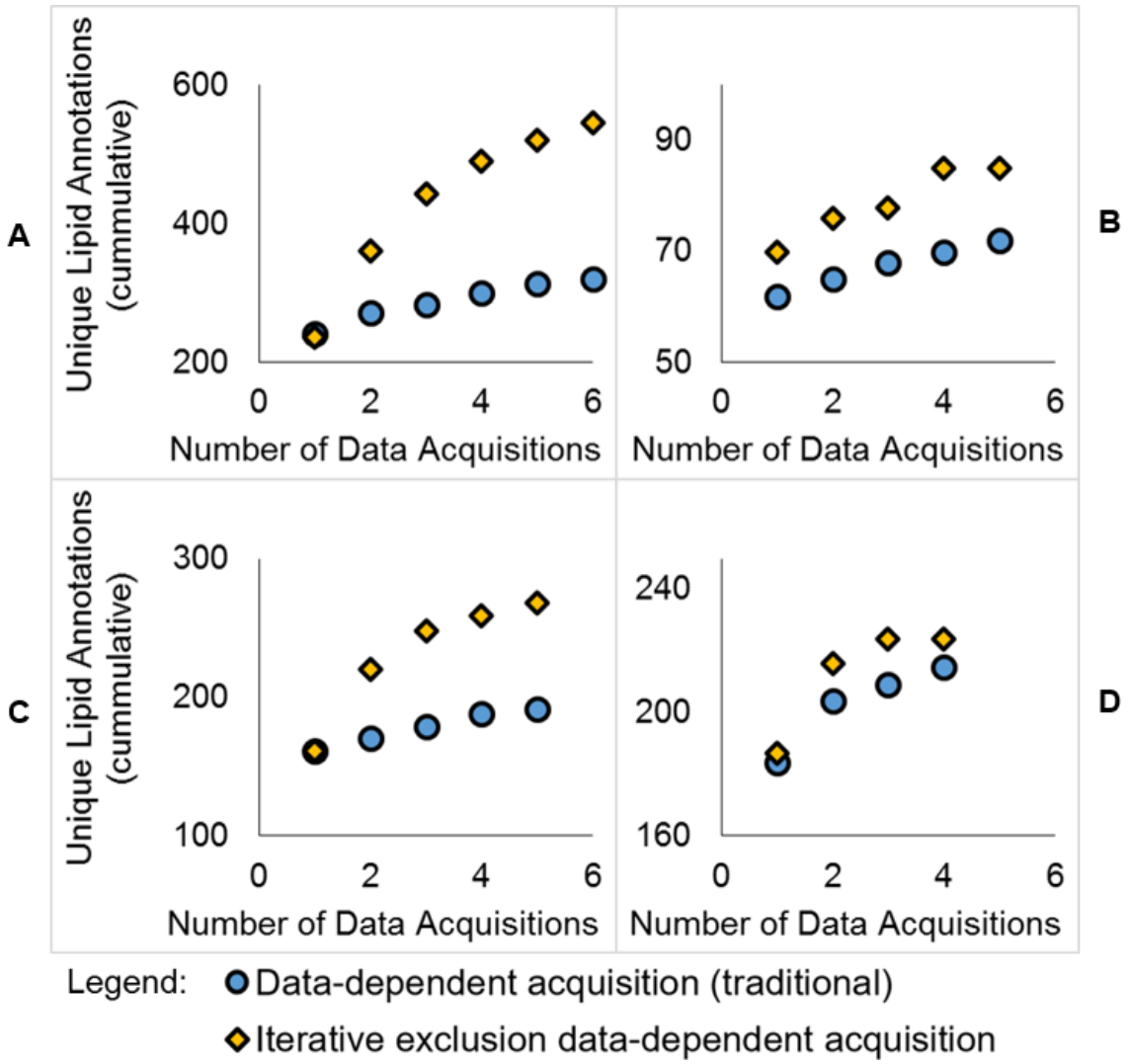


Figure S2-4. Cumulative unique lipid molecular identifications using LipidSearch software across multiple data acquisitions. Iterative exclusion-based data-dependent top5 (IE-ddMS^2-top5) described in this paper is compared with with traditional ddMS^2 top5 for: A) lipid extracts of Red Cross plasma in positive mode and B) negative mode, and C) lipid extracts of substantia nigra in positive mode and D) negative mode.

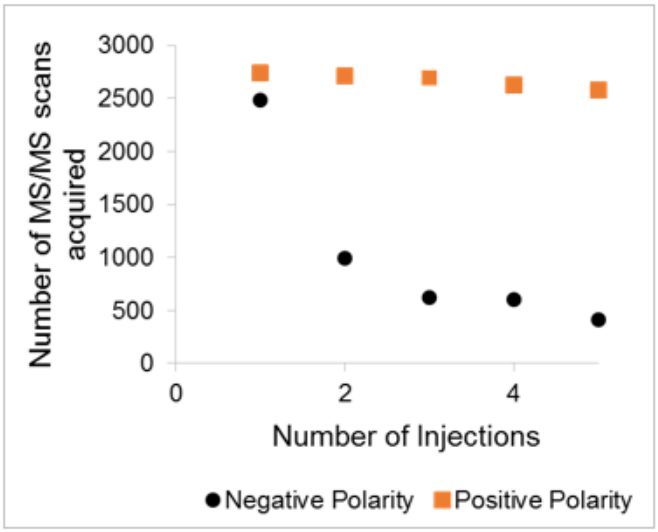


Figure S2-5. Number of precursors selected for fragmentation across sequential injections after applying IE to negative and positive mode analysis of Red Cross plasma lipid extracts.

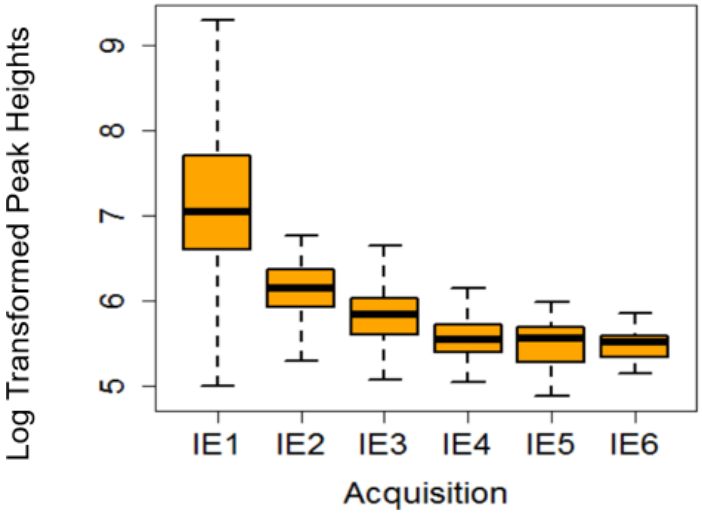


Figure S2-6. Boxplots of log transformed peak heights from MZmine for lipids identified in the first ddMS^2-top5 acquisition using LipidMatch (IE1) and after applying an exclusion list using the algorithm described in this paper (IE2, IE3, IE4, IE5, and IE6). Comparisons are made for lipid extracts of Red Cross plasma in positive mode. Differences between IE1 and IE2, were highly significant with a p-value for a student t-test less than 0.0000005. Differences in the following injections did not have significantly lower intensities (p-value > 0.05).

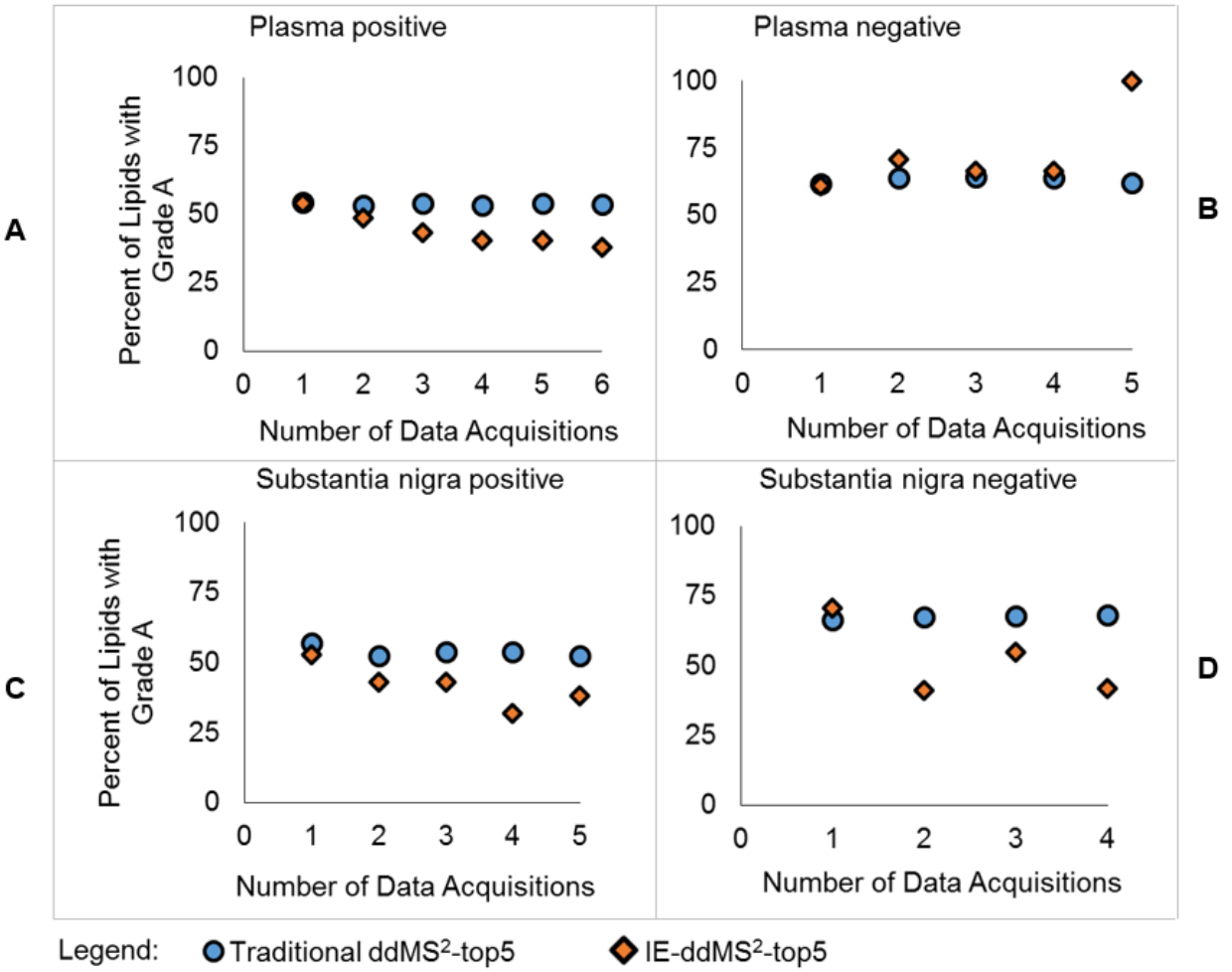


Figure S2-7. Graphs are shown representing MS/MS spectral quality over sequential injections with and without applying IE. The dependent variable is the percent of lipids identified by LipidSearch which were graded A (high confidence and characterization of structural detail) over all grades (A + B + C). B and C grades indicate that there were less fragments identified for those lipids. Iterative exclusion-based data-dependent top5 (IE-ddMS^2-top5) described in this paper is compared with traditional ddMS^2 top5 for lipid extracts of: A) Red Cross plasma lipid extracts in positive mode and B) negative mode, and C) lipid extracts of substantia nigra in positive mode and D) negative mode.

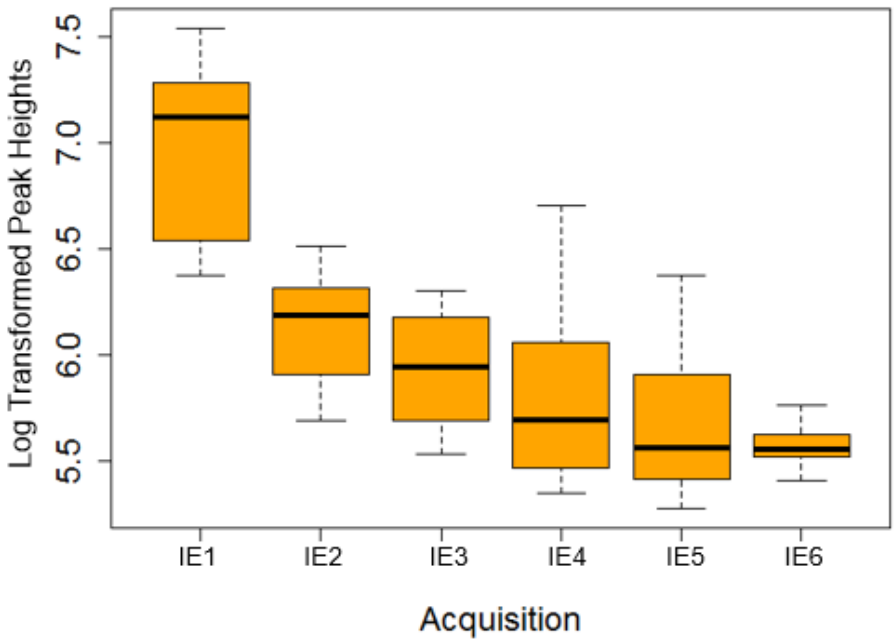


Figure S2-8. Boxplots of log transformed peak heights from MZmine for diglycerides (DGs) identified in the first ddMS^2-top5 acquisition using LipidMatch (IE1) and after applying an exclusion list using the algorithm described in this paper (IE2, IE3, IE4, IE5, and IE6). Comparisons are made for Red Cross plasma lipid extracts in positive mode.

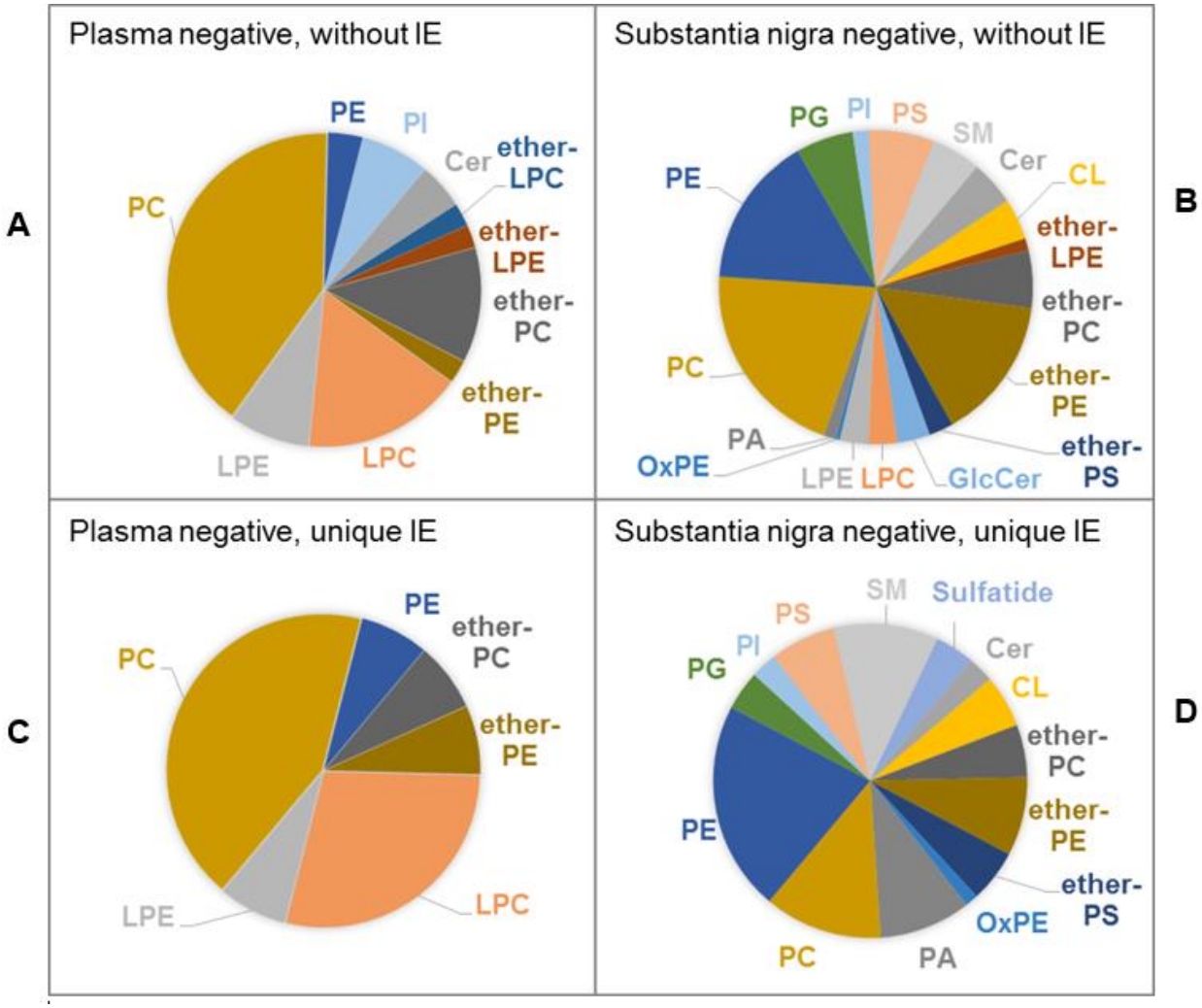
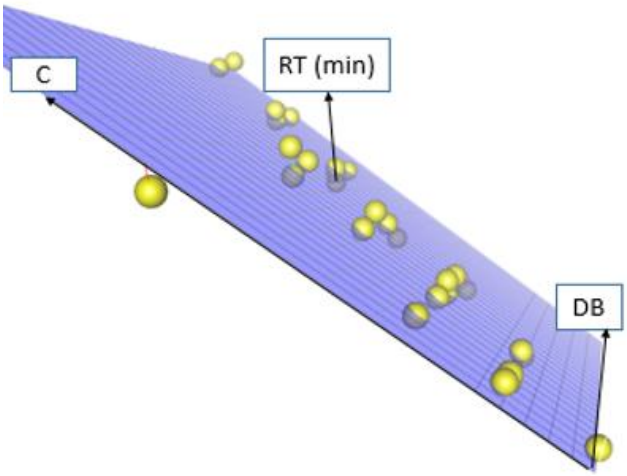


Figure S-9. Distribution of lipids identified using LipidMatch by lipid class using iterative exclusion-based data-dependent top5 (IE-ddMS^2-top5) acquisitions in negative ion mode. The lipid class distribution of all identifications across sequential injections using the traditional ddMS^2-top5 approach is shown for: A) Red Cross plasma and B) substantia nigra tissue lipid extracts. In addition, the distribution of additional unique lipid molecular identifications after applying iterative exclusion (IE) across lipid classes are shown for: C) Red Cross plasma and D) substantia nigra lipid extracts.



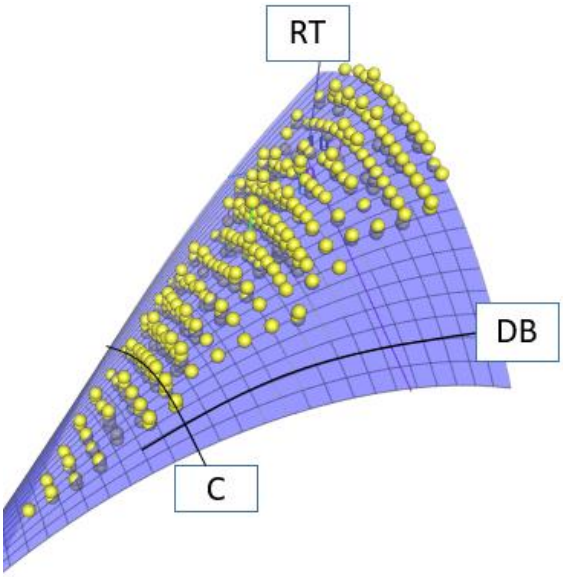
DGs

Min	1Q	Median	3Q	Max
-0.5	-0.1	0	0.2	0.5

$RT = 0.54(C) - 0.80(DB) - 7.36$

Residual standard error: **0.27**
Adjusted R^2: **0.97**
p-value: < $\mathbf{10^{-16}}$

Figure S2-10. Multilinear regression for predicting retention times of diglycerides (DGs) based on DG total carbons and degrees of unsaturation in the fatty acid constituents. DGs fatty acid constituents were determined using tandem mass spectrometry and an *in-house* identification software LipidMatch. Models explain the majority of the variance (97%) as expected. Therefore, the majority of LipidMatch identifications are verified by orthogonal separation, retention time, at least at the level of total carbons and double bonds.



TGs

Min	1Q	Median	3Q	Max
-0.756	-0.125	0.0167	0.135	1.5041

$RT = -0.01(C^2) - 0.02(DB^2) + 1.1(C) - 1.9(DB)$
Residual standard error: **0.23**
Adjusted R^2: **0.97**
p-value: < $\mathbf{10^{-16}}$

Figure S2-11. Multilinear regression for predicting retention times of triglycerides (TGs) based on TG total carbons and degrees of unsaturation in the fatty acid constituents. TGs fatty acid constituents were determined using tandem mass spectrometry and an *in-house* identification software LipidMatch. Models explain the majority of the variance (97%) as expected. Therefore, the majority of LipidMatch identifications are verified by orthogonal separation, retention time, at least at the level of total carbons and double bonds.

Chapter 3

Object 3-1. Additional supplemental information for Chapter 3 (Table S3-4 as .xlsx, all supplemental Figures and Tables in power point, and LipidMatch software, video tutorials, manual, and example files) (.zip, .xlsx and .pptx files 369 MB)

Table S3-1. LipidMatch lipids as of 10/1/2016

Class	Species	Adducts
Ac2PIM1	78	$[M-H]^-$
Ac2PIM2	78	$[M-H]^-$
Ac3PIM2	1728	$[M-H]^-$
Ac4PIM2	20736	$[M-H]^-$
AcCa	53	$[M+H]^+$
BA_Glycine	16	$[M-H]^-$
BA	34	$[M-H]^-$
BA_Taurine	13	$[M-H]^-$
CE	38	$[M+NH_4]^+$
Cer_ADS	784	$[M+HCO_2]^-$;$[M-H]^-$
Cer_AP	2352	$[M+HCO_2]^-$;$[M-H]^-$
Cer_AS	1568	$[M+HCO_2]^-$;$[M-H]^-$
Cer_BDS	784	$[M+HCO_2]^-$;$[M-H]^-$
Cer_BS	2352	$[M+HCO_2]^-$;$[M-H]^-$
Cer_EODS	4368	$[M+HCO_2]^-$;$[M-H]^-$
Cer_EOS	8736	$[M+HCO_2]^-$;$[M-H]^-$
Cer	1444	$[M+H]^+$;$[M+HCO2]^-$
Cer_NDS	784	$[M+H]^+$;$[M+HCO_2]^-$;$[M-H]^-$
Cer_NP	2352	$[M+HCO_2]^-$;$[M-H]^-$
CerP	168	$[M+H]^+$;$[M-H]^-$
CL	1596	$[M-2H+3Na]^+$;$[M-2H]^{2-}$;$[M-H]^-$;$[M+NH_4]^+$
CoQ	5	$[M+NH_4]^+$
DG	741	$[M+NH_4]^+$
DGDG	1176	$[M+HCO_2]^-$;$[M+NH_4]^+$
DGTS	741	$[M+H]^+$
DMPE	741	$[M-H]^-$
Ganglioside	1352	$[M-H]^-$
GlcADG	1176	$[M-H]^-$;$[M+NH_4]^+$
GlcCer_AP	2352	$[M+FA-H]^-$;$[M-H]^-$
GlcCer_AS	1568	$[M+FA-H]^-$;$[M-H]^-$
GlcCer_NDS	784	$[M+H]^+$;$[M+FA-H]^-$;$[M-H]^-$
GlcCer_NP	2352	$[M+FA-H]^-$;$[M-H]^-$
GlcCer	1568	$[M+H]^+$;$[M+HCO_2]^-$
lanosteryl	2	$[M+NH_4]^+$
LDGTS	231	$[M+H]^+$
LipidA-PP	425	$[M-H]^-$;$[M-H]^-$
LPC	38	$[M+H]^+$;$[M+Na]^+$;$[M+HCO_2]^-$
LPE	38	$[M+H]^+$;$[M+Na]^+$;$[M-H]^-$
LPI	38	$[M-H]^-$
MG	49	$[M+NH_4]^+$
MGDG	2304	$[M+H]^+$;$[M+Na]^+$;$[M+HCO_2]^-$;$[M+NH_4]^+$;$[M+NH_4\text{-}CO]^+$

Table S3-1. Continued

Class	Species	Adducts
MMPE	741	$[M-H]^-$
OxLPC	214	$[M+H]^+$;$[M+Na]^+$;$[M+HCO_2]^-$
OxLPE	214	$[M+H]^+$;$[M+Na]^+$;$[M-H]^-$
OxPC	31112	$[M+H]^+$;$[M+Na]^+$;$[M+HCO_2]^-$
OxPE	31112	$[M+H]^+$;$[M+Na]^+$;$[M-H]^-$
OxTG	115520	$[M+NH_4]^+$;$[M+NH_4]^+$
PA	741	$[M+H]^+$;$[M-H]^-$
PC	741	$[M+H]^+$;$[M+H]^+$;M+Na;$[M+HCO_2]^-$
PE	741	$[M+H]^+$;$[M+Na]^+$;$[M-H]^-$
PG	741	$[M+H]^+$;$[M-H]^-$;$[M+NH_4]^+$
PI	741	$[M-H]^-$
Plasmanyl-LPC	11	$[M+H]^+$;$[M+HCO_2]^-$
Plasmanyl-LPE	3	$[M+H]^+$;$[M-H]^-$
Plasmanyl-PC	418	$[M+H]^+$;$[M+HCO_2]^-$
Plasmanyl-PE	114	$[M+H]^+$;$[M-H]^-$
Plasmanyl-PS	114	$[M+H]^+$;$[M-H]^-$
Plasmanyl-TG	4446	$[M+NH_4]$;$[M+NH_4]$
Plasmenyl-LPC	11	$[M+H]^+$;$[M+HCO_2]^-$
Plasmenyl-LPE	11	$[M+H]^+$;$[M-H]^-$
Plasmenyl-PC	418	M+H;$[M+HCO_2]^-$
Plasmenyl-PE	228	$[M+H]^+$;$[M-H]^-$
Plasmenyl-PS	228	$[M+H]^+$;$[M-H]^-$
Plasmenyl-TG	4446	$[M+NH_4]$;$[M+NH_4]$
PS	741	$[M+H]^+$;$[M+Na]^+$;$[M-H]^-$
SM	1444	$[M+H]^+$;$[M+HCO_2]^-$
So	13	$[M+H]^+$
SQDG	1473	$[M-H]^-$
Sulfatide	196	$[M+H]^+$;$[M-H]^-$
TG	9880	$[M+Na]^+$;$[M+NH_4]^+$;$[M+NH_4]^+$
zymosteryl	2	$[M+NH_4]^+$
71	274558	136

Table S3-2. LipidMatch lipid acronyms as of 10/1/2016

Acronym	Definition
Ac2PIM1	Diacylated phosphatidylinositol monomannoside
Ac2PIM2	Diacylated phosphatidylinositol dimannoside
Ac3PIM2	Triacylated phosphatidylinositol dimannoside
Ac4PIM2	Tetraacylated phosphatidylinositol dimannoside
AcCa	Acylcarnitine
BA	Bial Acid: Beta
BA_Glycine	Bial Acid Glycine Conjugate
BA_Taurine	Bial Acid Taurine Conjugate
CE	Cholesterol Ester
Cer_NDS	Backbone: Dihydrosphingosine Fatty acid: non-hydroxy fatty acid
Cer_NP	Backbone: Phytosphingosine Fatty acid: Fatty acid: non-hydroxy fatty acid
Cer	Ceramide (Nonhydroxyacyl sphingosine)
Cer_ADS	Backbone: Dihydrosphingosine Fatty acid: a-hydroxy fatty acid
Cer_AP	Backbone: Phytosphingosine Fatty acid: a-hydroxy fatty acid
Cer_AS	Backbone: Sphingosine Fatty acid: a-hydroxy fatty acid
Cer_BDS	Ceramide Backbone: Dihydrosphingosine
Cer_BS	Ceramide Backbone: Sphingosine
Cer_BS	Ceramide Backbone: Sphingosine
Cer_EODS	Backbone: Dihydrosphingosine Fatty acid: esterified w-hydroxy fatty acid
Cer_EOS	Backbone: Sphingosine Fatty acid: esterified w-hydroxy fatty acid
CerP	Ceramide-1-phosphate
CL	Cardiolipin
CoQ	Coenzyme Q
DG	Diglyceride or Diglyceride
DGDG	Digalactosyldiglyceride
DGTS	Diacylglyceryltrimethylhomo-Ser
DMPE	Dimethyl Phosphatidylethanolamine
Ganglioside	Ganglioside or glycan ceramides
GlcADG	Glucuronosyldiglyceride
GlcCer_NDS	Glucosyl-ceramide Backbone: Dihydrosphingosine Fatty acid: non-hydroxy fatty acid
GlcCer	Glucosyl-Nonhydroxyacyl sphingosine
GlcCer_AP	Glucosyl-ceramide Backbone: Phytosphingosine Fatty acid: a-hydroxy fatty acid
GlcCer_AS	Glucosyl-ceramide Backbone: Sphingosine Fatty acid: a-hydroxy fatty acid
lanosteryl	Lanosteryl
LDGTS	Lysodiacylglyceryltrimethylhomo-Ser
LipidA_PP	Diphosphorylated hexaacyl Lipid A
LPC	Lysophosphatidylcholine
LPE	Lysophosphatidylethanolamine
LPI	Lysophosphatidylinositol

Table S3-2. Continued

Acronym	Definition
MG	Monoacylglycerol
MGDG	Monogalactosyldiglyceride
MMPE	Monomethyl-Phosphatidylethanolamine
OxLPC	Oxidized lysophosphatidylcholine
OxLPE	Oxidized lysophosphatidylethanolamine
OxPC	Oxidized phosphatidylcholine
OxPE	Oxidized phosphatidylethanolamine
OxTG	Oxidized triglyceride
PA	Phosphatidic acid
PC	Phosphatidylcholine
PE	Phosphatidylethanolamine
PG	Phosphatidylglycerol
PI	Phosphatidylinositol
Plasmanyl-LPC	Plasmanyl lysophosphatidylcholine
Plasmanyl-LPE	Plasmanyl lysophosphatidylethanolamine
Plasmanyl-PC	Plasmanyl phosphatidylcholine
Plasmanyl-PE	Plasmanyl phosphatidylethanolamine
Plasmanyl-PS	Plasmanyl phosphatidylserine
Plasmanyl-TG	Plasmanyl triglyceride
Plasmenyl-LPC	Plasmenyl lysophosphatidylcholine
Plasmenyl-LPE	Plasmenyl lysophosphatidylethanolamine
Plasmenyl-PC	Plasmenyl phosphatidylcholine
Plasmenyl-PE	Plasmenyl phosphatidylethanolamine
Plasmenyl-PS	Plasmenyl phosphatidylserine
Plasmenyl-TG	Plasmenyl triglyceride
PS	Phosphatidylserine
SM	Sphingomyelin
So	Sphingosine
SQDG	Sulfoquinovosyldiglyceride
Sulfatide	Sulfatide
TG	Triglyceride
zymosteryl	zymosteryl

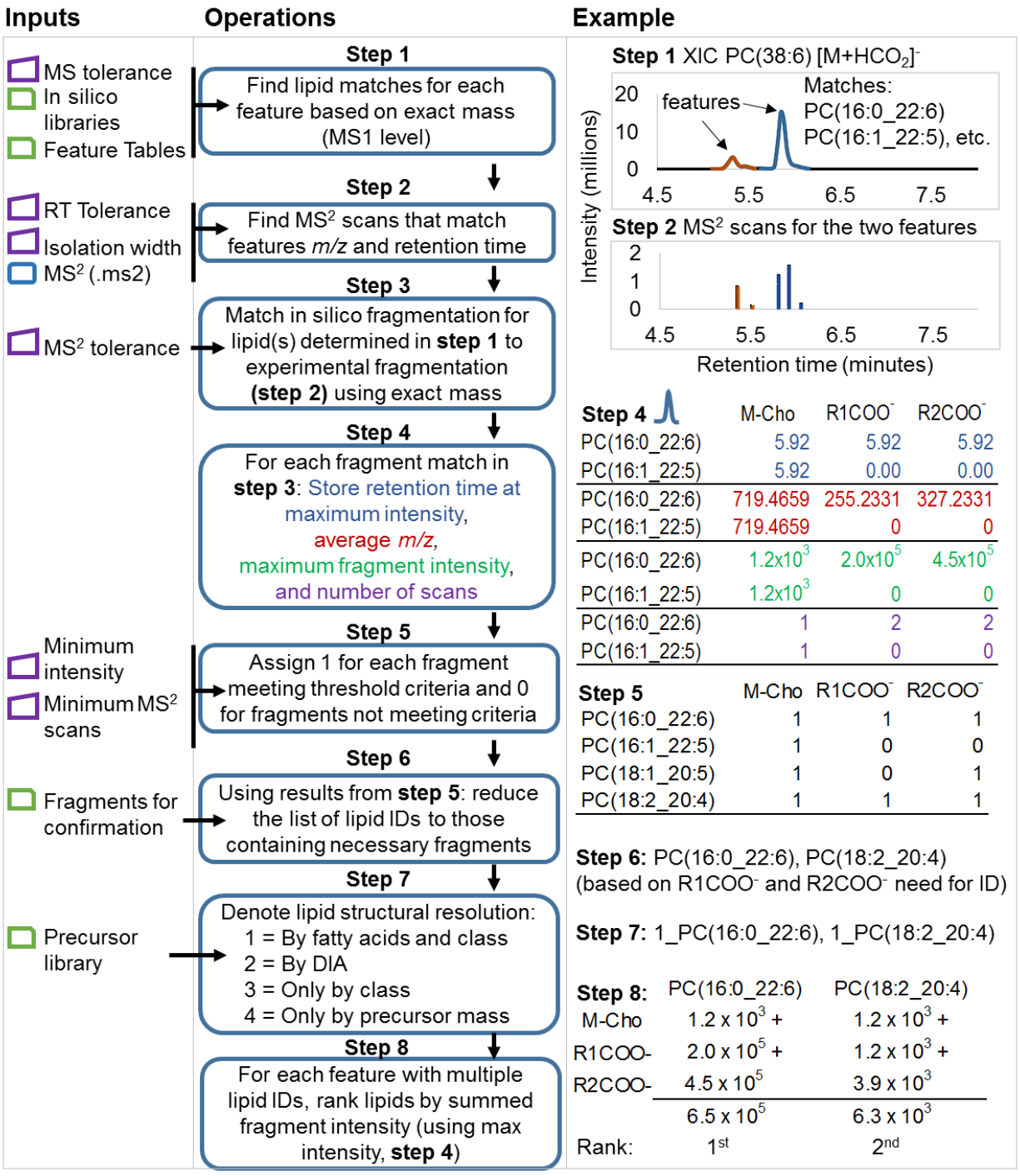


Figure S3-1. Simplified flow diagram of LipidMatch operations. The first panel is input files, the second represents operations performed by LipidMatch, and the third panel illustrates procedures using data from red cross plasma. For the first panel, green boxes with folded top right corners are input csv files. Purple boxes with diagonal tops are input parameters. The third panel uses identification of PC(38:6) $[M+HCO_2]^-$ in negative polarity as an example.

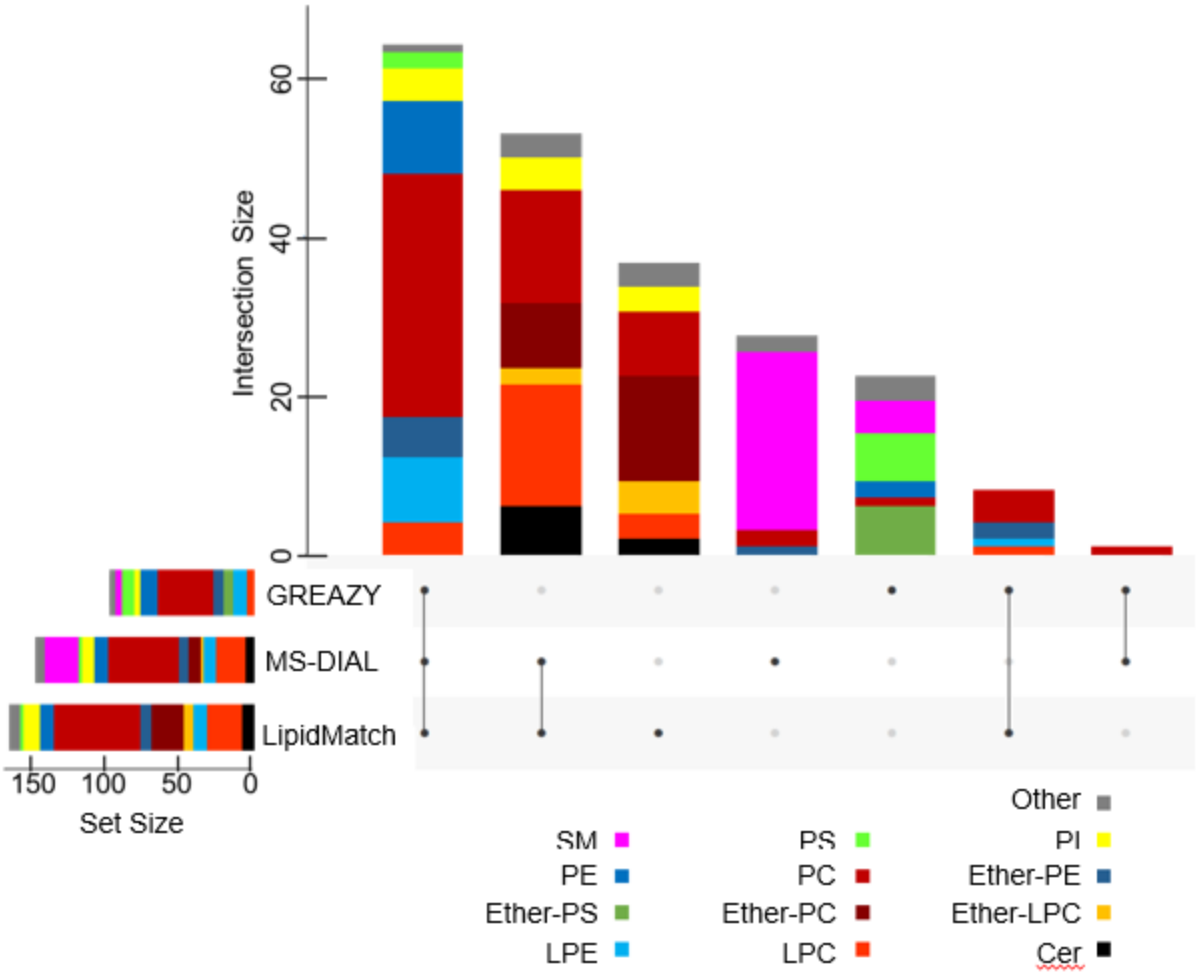


Figure S3-2. Set overlap for LipidMatch, MS-DIAL, and GREAZY in negative polarity analysis of Red Cross plasma. Visualization of sets based on UpSet <http://www.caleydo.org/tools/upset/>. Dots and lines represent which software (sets) overlap, and bars represent the total lipid species contained within each overlap. For example the first vertical bar represents the number of features with the same identification for all 3 software. Color codes show the lipid types making up a specific overlap or set. Sets or sorted by number of lipid species contained within. Species included in other were PIP, PIP2, and GlcCer. Horizontal bars represent the total number of feature identified by each respective software.

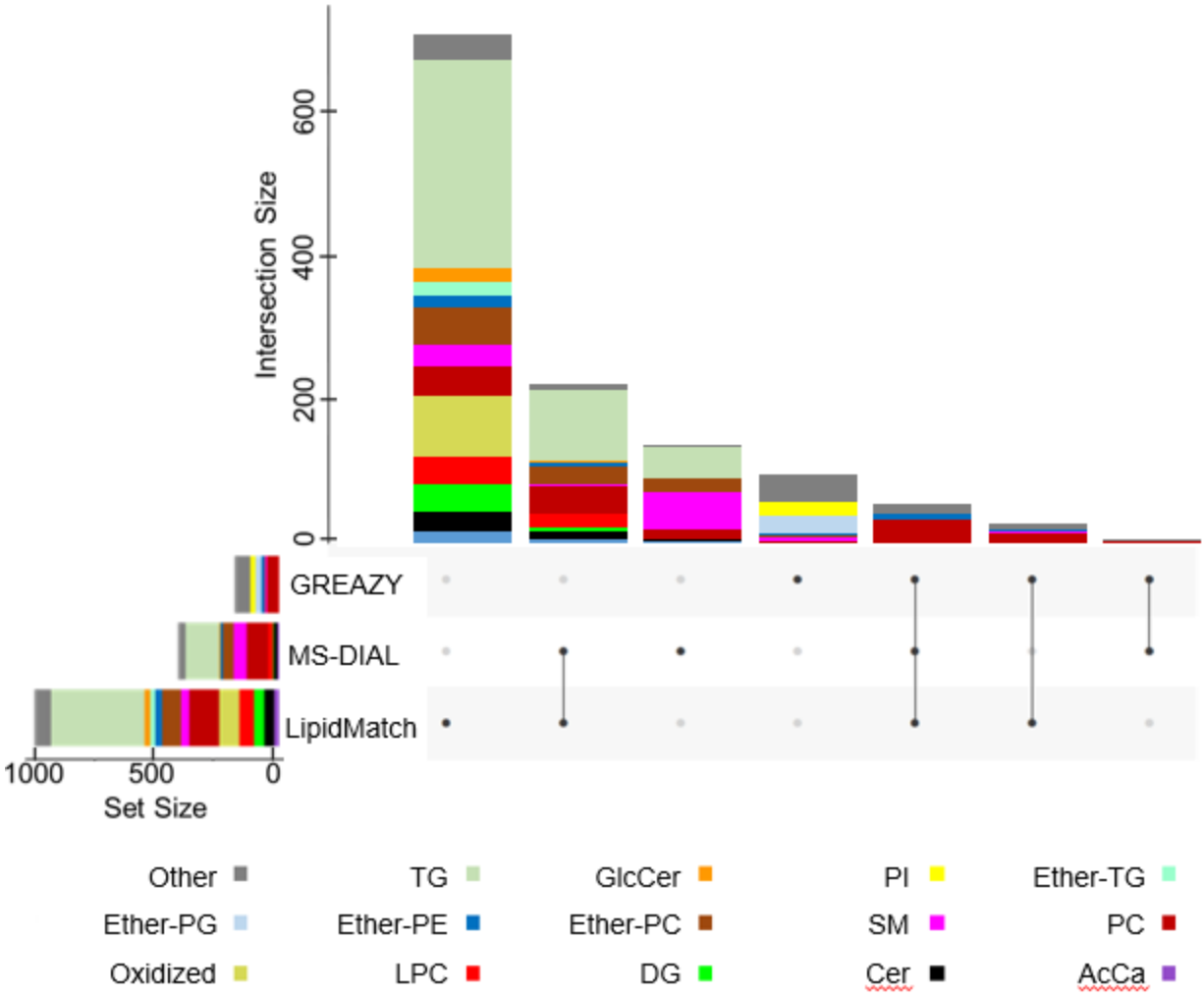


Figure S3-3. Set overlap for LipidMatch, MS-DIAL, and GREAZY in positive polarity analysis of Red Cross plasma. Dots and lines represent which software (sets) overlap, and bars represent the total lipid species contained within each overlap. Color codes show the lipid types making up a specific overlap or set. Species included in other were ether-linked-LPC, Co, So, Sulfatide, PIP3, PE-Cer, PG, PA, PS, ether-linked-PS, PE, CE, and LPE, which all had less than 15 lipids in any given overlap between sets.

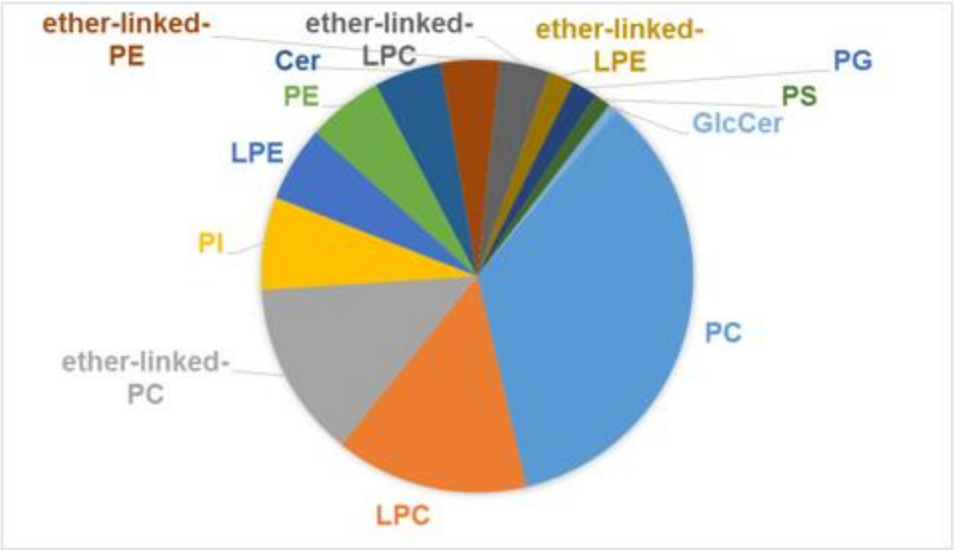


Figure S3-4. Pie chart of lipid classes and the number of each identified by LipidMatch in negative polarity

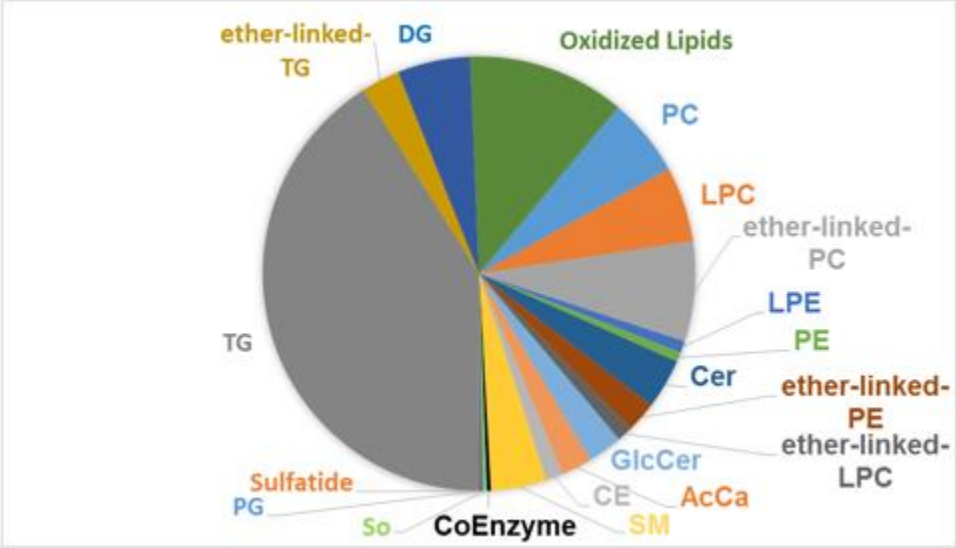


Figure S3-5. Pie chart of lipid classes and the number of each identified by LipidMatch in positive polarity

Chapter 4

Table S4-1. Liquid chromatography gradient

Time (min)	0	1	3	4	6	8	10	15	17	18	19	23
C (%)	80	80	70	55	40	35	35	10	2	2	80	80
D (%)	20	20	30	45	60	65	65	90	98	98	20	20

Table S4-2. Mass spectrometry scan parameters

Parameter	Full Scan	AIF scan	Targeted MS^2
Resolution	70,000	70,000	17,500
AGC	3×10^6	3×10^6	3×10^6
Injection time (ms)	256	256	256
Range (*m/z*)	120 - 1,200	80 - 1,200	automatic
NCE	NA	25	25
Isolation window (*m/z*)	NA	1120	1

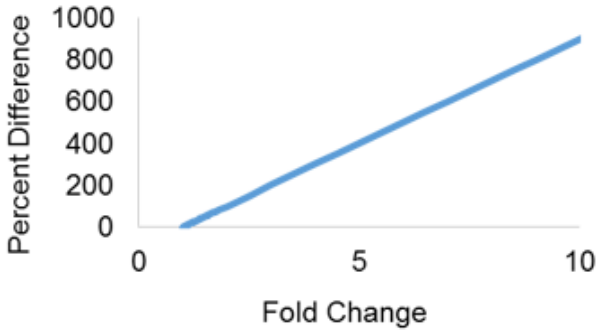
Table S4-3. Comparisons of LMQ derived concentrations using various data processing techniques and ions for quantification

Concentrations calculated after smoothing versus not smoothing peak heights				
a)	Species Both*	Avg % Diff**	Med % Diff***	log_2 fold change ****
Cer	11	5 ± 6	2.6	0 ± 0.1
LPC	15	6 ± 5	5.4	-0.1 ± 0.1
OxTG	4	2 ± 1	2.6	0 ± 0
PC	37	8 ± 6	7.3	-0.1 ± 0.1
PE	14	9 ± 7	8.1	-0.1 ± 0.1
SM	3	9 ± 3	7.7	0 ± 0.2
TG	59	4 ± 3	3.5	0 ± 0.1
ether-LPC	5	11 ± 9	14	-0.2 ± 0.1
ether-PC	15	4 ± 3	3.9	-0.1 ± 0
ether-TG	15	4 ± 3	4.4	-0.1 ± 0.1
ether-PE	6	6 ± 2	6.9	-0.1 ± 0
Concentrations calculated after smoothing versus not smoothing peak areas				
b)				
Cer	11	2 ± 2	1.2	0 ± 0.1
LPC	15	1 ± 1	0.3	-0.1 ± 0.1
OxTG	4	3 ± 2	2.6	0 ± 0
PC	37	1 ± 1	0.2	-0.1 ± 0.1
PE	14	2 ± 2	1	-0.1 ± 0.1
SM	3	1 ± 1	0.3	0 ± 0.2
TG	59	2 ± 2	1	0 ± 0.1
ether-LPC	5	0 ± 0	0.6	-0.2 ± 0.1
ether-PC	15	1 ± 1	0.3	-0.1 ± 0
ether-TG	15	3 ± 2	2.8	-0.1 ± 0.1
ether-PE	6	1 ± 1	0.4	-0.1 ± 0

Table S4-3. Continued

Concentrations calculated using peak height versus peak area c)				
Cer	11	25 ± 32	26	-0.2 ± 0.3
LPC	15	24 ± 18	13	0.1 ± 0.3
OxTG	4	78 ± 25	29	0.3 ± 0.2
PC	37	28 ± 19	18	0.1 ± 0.3
PE	14	28 ± 24	32	-0.1 ± 0.5
SM	3	45 ± 31	54	-0.5 ± 0.3
TG	59	53 ± 44	42	-0.5 ± 0.4
ether-LPC	5	31 ± 18	40	0.3 ± 0.4
ether-PC	15	17 ± 13	28	0.3 ± 0.4
ether-TG	15	16 ± 9	18	0.3 ± 0.3
ether-PE	6	16 ± 15	15	0.2 ± 0.2
Concentrations calculated using positive versus negative polarity data d)				
Cer	7	53 ± 23	31	-0.6 ± 0.3
ether-PC	4	41 ± 7.4	40	0.5 ± 0.1
ether-PE	4	33 ± 30	22	0.3 ± 0.4
LPC	14	43 ± 18	41	-0.5 ± 0.2
PC	22	27 ± 24	20	0.2 ± 0.4
Concentrations calculated using major versus $[M+Na]^+$ adducts e)				
Cer	7	570 ± 210	180	-3 ± 0
LPC	5	220 ± 140	220	1.6 ± 0.6
PC	26	180 ± 320	120	-1 ± 1
TG	38	4300 ± 11300	260	-1 ± 4

A



B

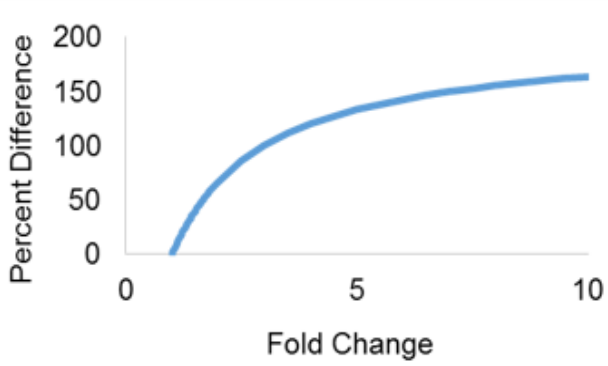


Figure S4-1. A comparison of fold change (greater than 1) versus percent difference for the traditional equation for percent difference and the one used here.
A) Fold change (greater than 1) versus percent difference calculated using the average in the formula: $Relative\ percent\ difference = \frac{x-y}{\mathrm{Avg}(x,y)} \times 100$
B) Fold change (greater than 1) versus percent difference calculated using the minimum in the formula: $Relative\ percent\ difference = \frac{x-y}{\min(x,y)} \times 100$

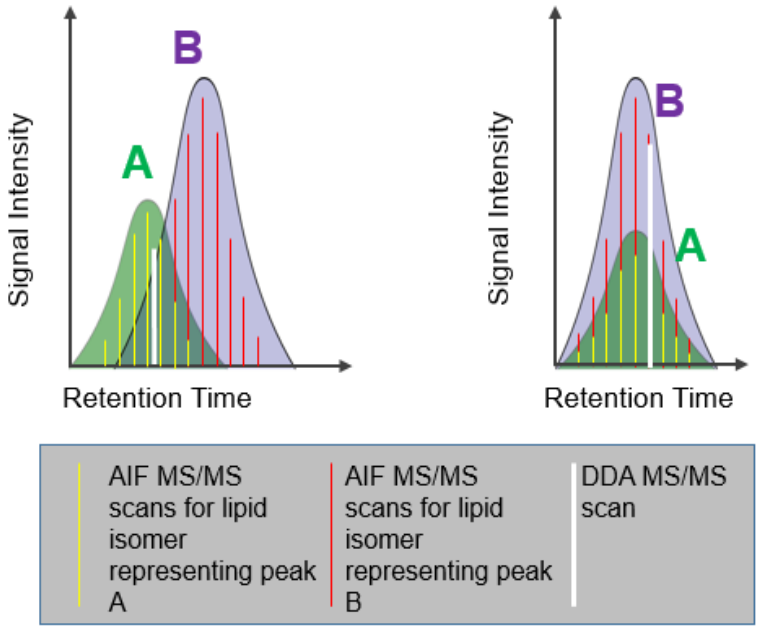


Figure S4-2. Depiction of MS/MS scans using AIF (red and yellow) and using DDA (white) for two lipid isomers (A and B). An advantage of AIF over data-dependent MS/MS for annotation is that via deconvolution, the correct peak can be assigned to the correct lipid (Figure S4-1A). In DDA in in Figure S4-1A, if only one MS/MS scan is obtained, there is not enough information to correctly assign which lipid isomer belongs to which peak. If both isomers completely overlap, AIF cannot be used to distinguish the isomers.

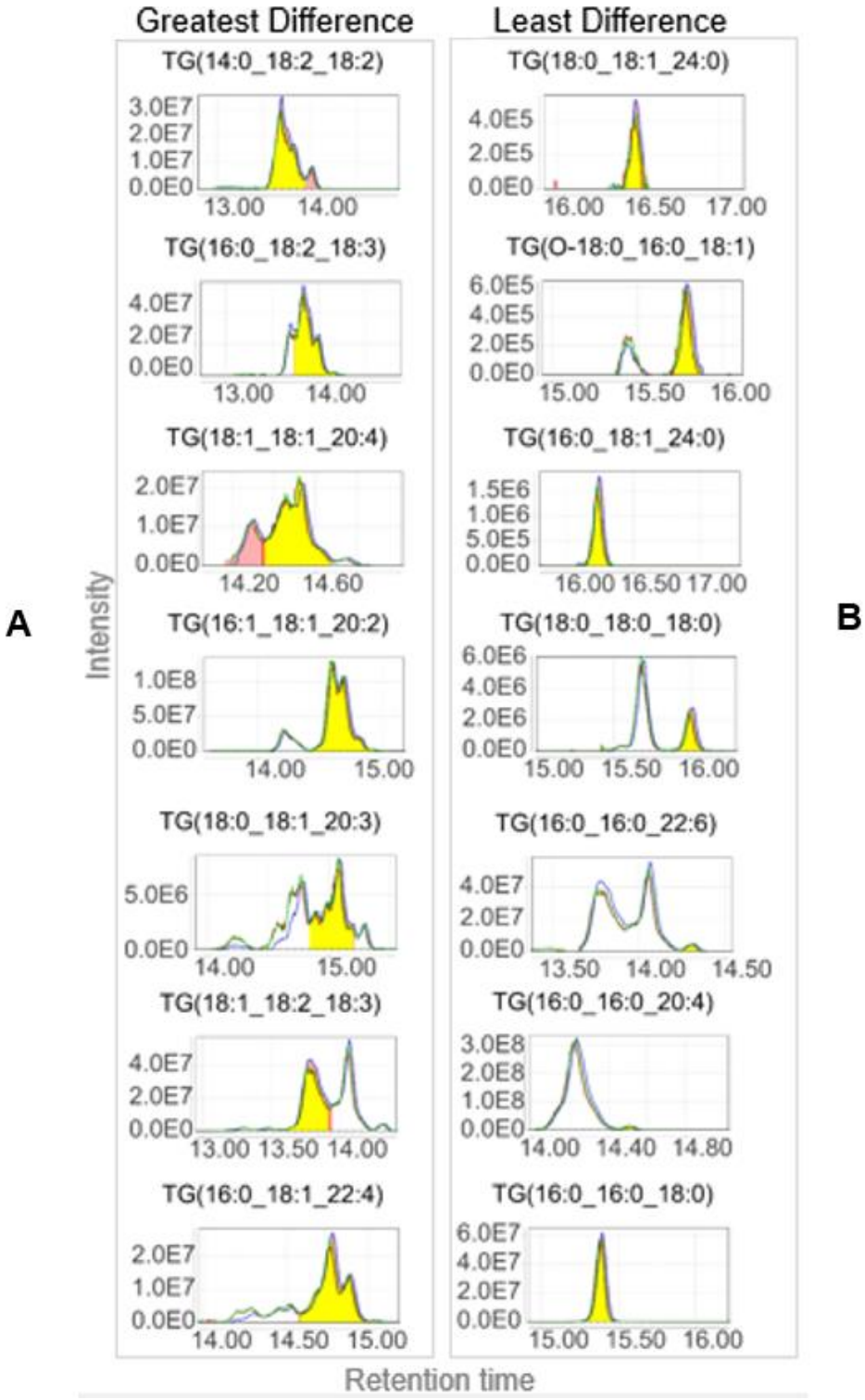
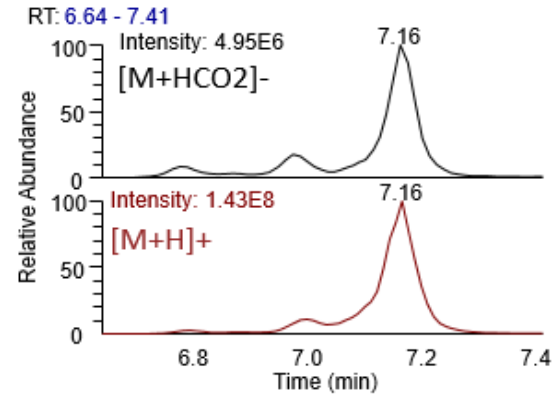


Figure S4-3. Extracted ion chromatograms (EICs) and peak integration by MZmine of the triglycerides (TGs) with: A) the most and B) least percent difference when comparing quantification using peak height versus peak area.



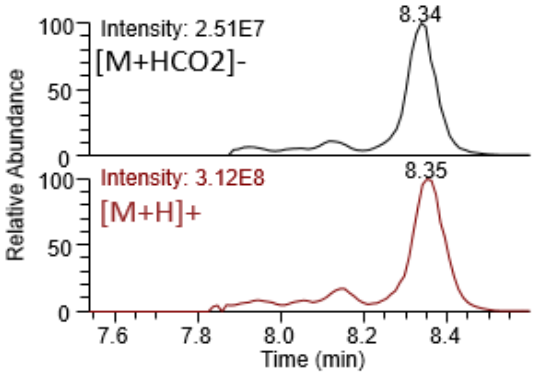


Figure S4-4. Extracted ion chromatograms (EICs) of PC(16:0_20:5) and PC(18:0_20:4) (the predominant peak at 7.16 and 8.34, respectively). Peaks had similar looking EICs in negative and positive mode, but very different percent differences in concentration between the two polarities, although the same molecular lipid was used for quantification.

Chapter 5

Object 5-1. Additional supplemental information for Chapter 5 (Table S5-3 through Table S5-7 as .xlsx, and all supplemental Figures and Tables in power point, and LipidMatch software, 11.2 MB)

Table S5-1. Ultra-high-performance liquid chromatography (UHPLC) gradient

Retention Time (min)	0	1	3	4	6	8	10	15	17	18	19	23
% A (60:40 ACN:H_2O)	80	80	70	55	40	35	35	10	2	2	80	80
% B (90:8:2 IPA:ACN:H_2O)	20	20	30	45	60	65	65	90	98	98	20	20

Table S5-2. Q-Exactive scanning parameters

	Res	AGC	IT (ms)	Range *(m/z)*	NCE
Full Scan	70k	$5x10^6$	256	200-1200	NA
ddMS2	35k	$5x10^6$	175	80-1200	25 ± 5
	Iso *(m/z)*	Underfill (%)	Appex (s)	Dyn Excl (s)	top N
ddMS2	1	1	10-20	4	10

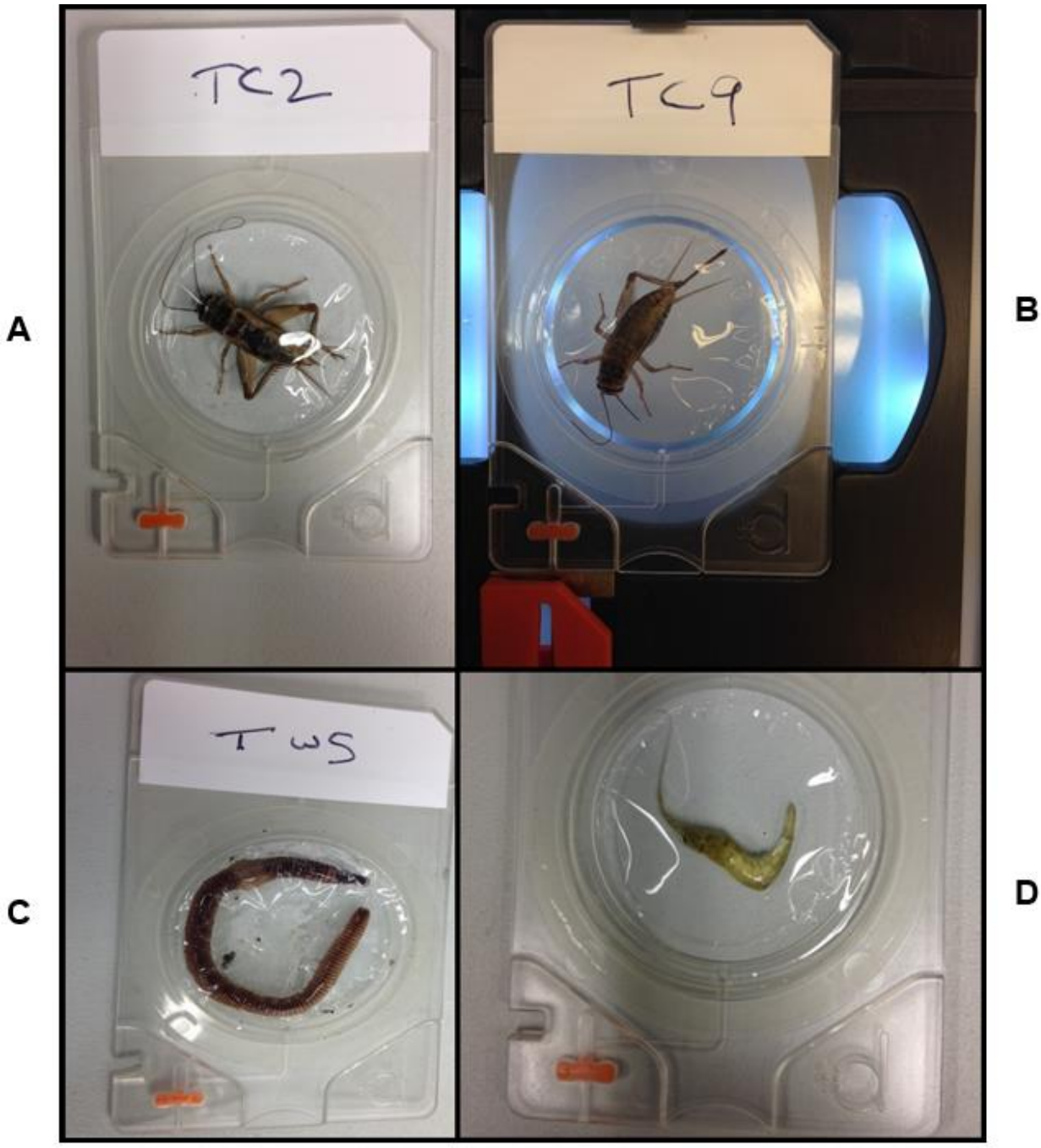


Figure S5-1. Denator Maintainor cartridges containing: A) common house crickets before and B) after heat treatment, C) earthworm after heat treatment and D) ghost shrimp after heat treatment.

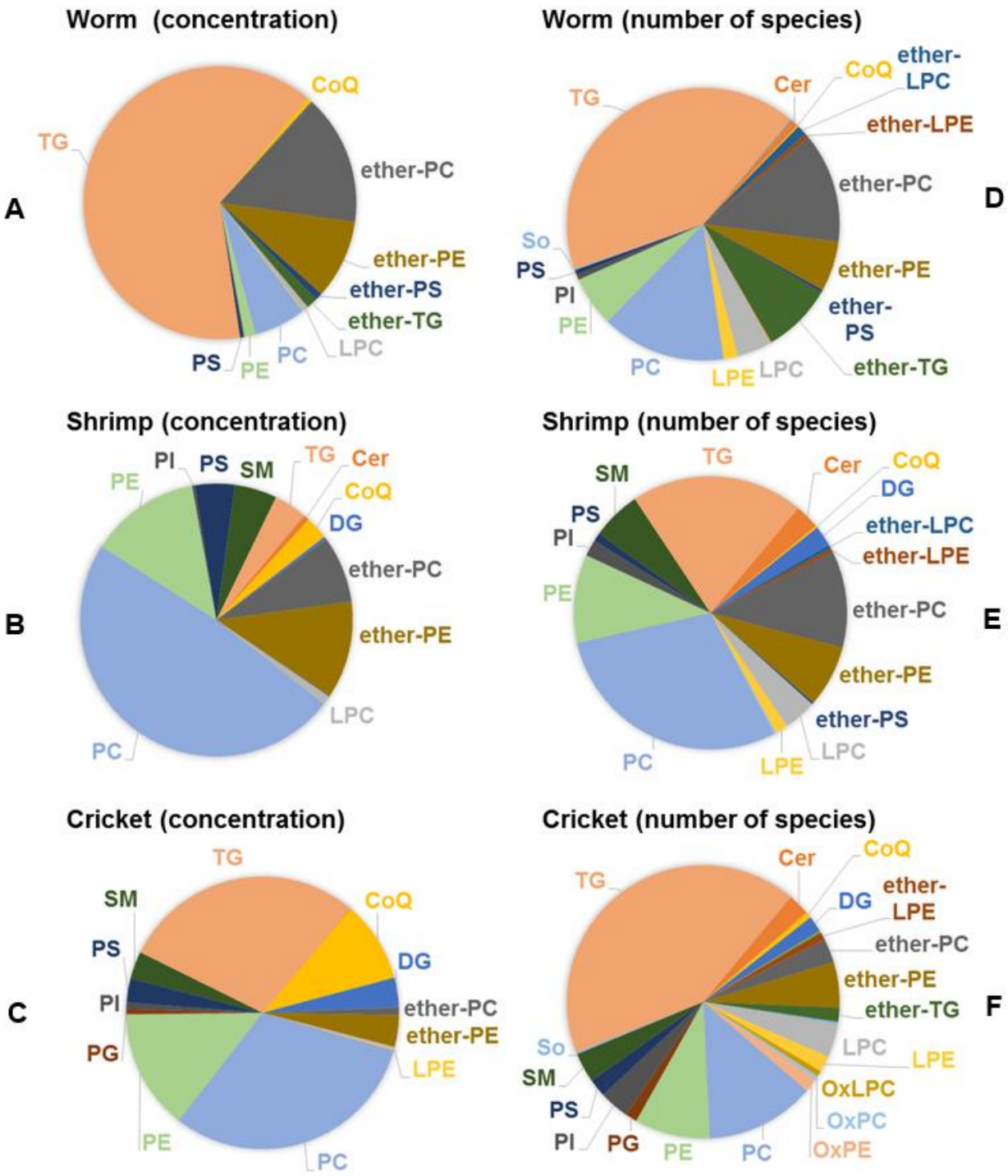


Figure S5-2. Relative concentrations and numbers of species of each lipid type in earthworm, ghost shrimp, and common house cricket. Lipid class/type acronyms and definitions are provided in Supplementary Table S-6. Ether stands for either plasmanyl or plasmenyl species, or both. The panels are: A) Concentration of each lipid class for earthworm, B) ghost shrimp, and C) common house cricket, and D) number of lipid species per class for earthworm, E) ghost shrimp, and F) common house cricket.

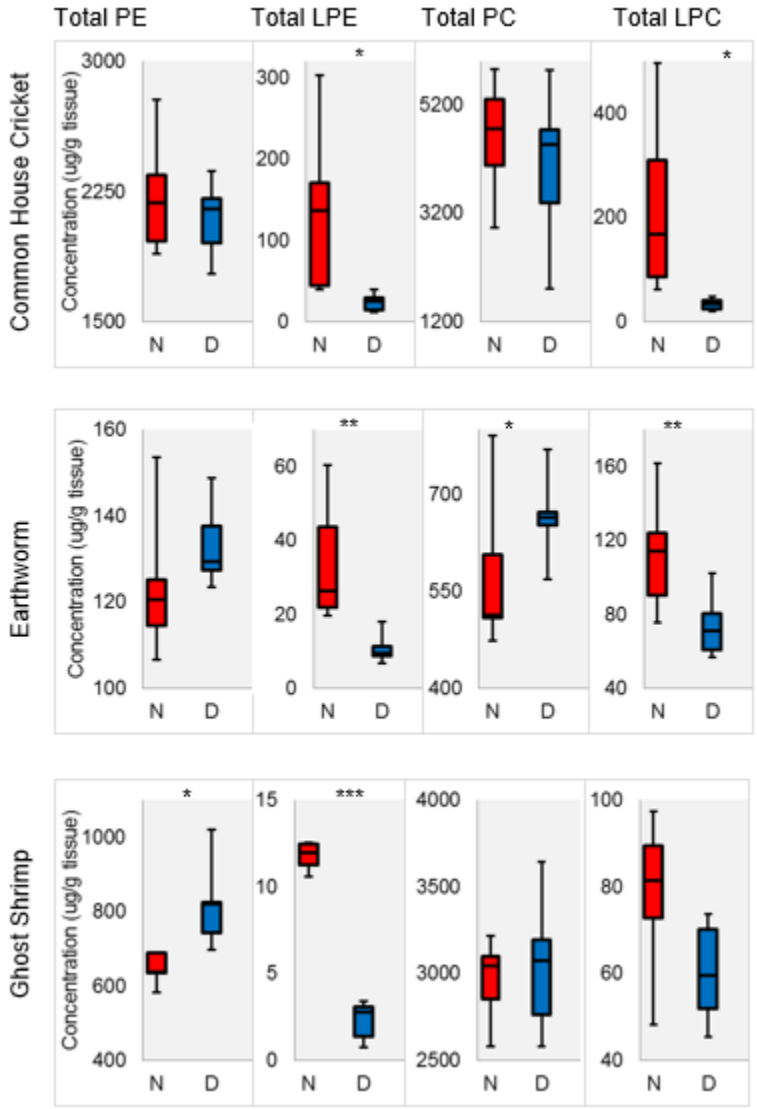


Figure S5-3. Total PE and LPE $[M-H]^-$ and PC and LPC $[M+HCO_2]^-$ concentrations in cricket, earthworm, and shrimp samples that are flash-frozen (N) and with heat treatment (D), including samples that were incubated for one hour on ice. Three significance levels are: $p < 0.05$ (*), $p < 0.005$ (**), and $p < 0.0005$ (***).

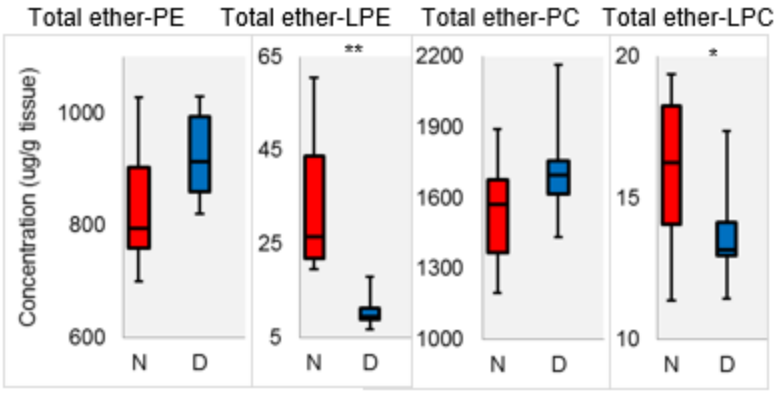


Figure S5-4. Total ether-linked (plasmanyl and plasmenyl species) PE and LPE $[M-H]^-$ and PC and LPC $[M+HCO_2]^-$ concentrations in frozen (N) and heat-treated (D). The three significance levels are: $p < 0.05$ (*), $p < 0.005$ (**), and $p < 0.0005$ (***).

LIST OF REFERENCES

1. Joyce, A.R., Palsson, B.Ø.: The model organism as a system: integrating “omics” data sets. Nat. Rev. Mol. Cell Biol. 7, 198–210 (2006). doi:10.1038/nrm1857

2. Kaddurah-Daouk, R., Kristal, B.S., Weinshilboum, R.M.: Metabolomics: A Global Biochemical Approach to Drug Response and Disease. Annu. Rev. Pharmacol. Toxicol. 48, 653–683 (2008). doi:10.1146/annurev.pharmtox.48.113006.094715

3. Van Assche, R., Broeckx, V., Boonen, K., Maes, E., De Haes, W., Schoofs, L., Temmerman, L.: Integrating -Omics: Systems Biology as Explored Through C. elegans Research. J. Mol. Biol. doi:10.1016/j.jmb.2015.03.015

4. van Meer, G., Voelker, D.R., Feigenson, G.W.: Membrane lipids: where they are and how they behave. Nat. Rev. Mol. Cell Biol. 9, 112–124 (2008). doi:10.1038/nrm2330

5. Pablo, M.A. de, Puertollano, M.A., Cienfuegos, G.Á. de: Biological and Clinical Significance of Lipids as Modulators of Immune System Functions. Clin. Diagn. Lab. Immunol. 9, 945–950 (2002). doi:10.1128/CDLI.9.5.945-950.2002

6. Cooper, G.M.: The Cell: A Molecular Approach. Sinauer Associates, Sunderland, MA, Sunderland, MA (2000)

7. Edwards, P.A., Ericsson, J.: Sterols and Isoprenoids: Signaling Molecules Derived from the Cholesterol Biosynthetic Pathway. Annu. Rev. Biochem. 68, 157–185 (1999). doi:10.1146/annurev.biochem.68.1.157

8. Martin, T.F.J.: Phosphoinositide lipids as signaling molecules: Common Themes for Signal Transduction, Cytoskeletal Regulation, and Membrane Trafficking. Annu. Rev. Cell Dev. Biol. 14, 231–264 (1998). doi:10.1146/annurev.cellbio.14.1.231

9. Simons, K., Toomre, D.: Lipid rafts and signal transduction. Nat. Rev. Mol. Cell Biol. 1, 31–39 (2000). doi:10.1038/35036052

10. Sies, H.: Oxidative stress: oxidants and antioxidants. Exp. Physiol. 82, 291–295 (1997)

11. Veldhuizen, R., Nag, K., Orgeig, S., Possmayer, F.: The role of lipids in pulmonary surfactant. Biochim. Biophys. Acta BBA - Mol. Basis Dis. 1408, 90–108 (1998). doi:10.1016/S0925-4439(98)00061-1

12. Imokawa, G., Kuno, H., Kawai, M.: Stratum Corneum Lipids Serve as a Bound-Water Modulator. J. Invest. Dermatol. 96, 845–851 (1990). doi:10.1111/1523-1747.ep12474562

13. Bron, A.J., Tiffany, J.M., Gouveia, S.M., Yokoi, N., Voon, L.W.: Functional aspects of the tear film lipid layer. Exp. Eye Res. 78, 347–360 (2004). doi:10.1016/j.exer.2003.09.019

14. Ridker, P., Rifai, N., Cook, N.R., Bradwin, G., Buring, J.E.: NOn–hdl cholesterol, apolipoproteins a-i and b100, standard lipid measures, lipid ratios, and crp as risk factors for cardiovascular disease in women. JAMA. 294, 326–333 (2005). doi:10.1001/jama.294.3.326

15. Oberg, B.P., Mcmenamin, E., Lucas, F.L., Mcmonagle, E., Morrow, J., Ikizler, T.A., Himmelfarb, J.: Increased prevalence of oxidant stress and inflammation in patients with moderate to severe chronic kidney disease. Kidney Int. 65, 1009–1016 (2004). doi:10.1111/j.1523-1755.2004.00465.x

16. Fernandis, A.Z., Wenk, M.R.: Lipid-based biomarkers for cancer. J. Chromatogr. B. 877, 2830–2835 (2009). doi:10.1016/j.jchromb.2009.06.015

17. Dalle-Donne, I., Rossi, R., Colombo, R., Giustarini, D., Milzani, A.: Biomarkers of Oxidative Damage in Human Disease. Clin. Chem. 52, 601–623 (2006). doi:10.1373/clinchem.2005.061408

18. Dowhan, W.: Molecular Basis for Membrane Phospholipid Diversity: Why Are There So Many Lipids? Annu. Rev. Biochem. 66, 199–232 (1997). doi:10.1146/annurev.biochem.66.1.199

19. Watanabe, K., Yasugi, E., Oshima, M.: How to Search the Glycolipid data in “LIPID BANK for the Web”, the Newly Developed Lipid Database in Japan. Trends Glycosci. Glycotechnol. 12, 175–184 (2000). doi:10.4052/tigg.12.175

20. Dove, A.: Greasing the Wheels of Lipidomics, http://www.sciencemag.org/site/products/lst_20150213.xhtml, (2015)

21. Schmelzer, K., Fahy, E., Subramaniam, S., Dennis, E.A.: The Lipid Maps Initiative in Lipidomics. Elsevier, Amsterdam, Netherlands (2007)

22. Spener, F.: European Commission funds lipidomics project. Eur. J. Lipid Sci. Technol. 107, 1–2 (2005). doi:10.1002/ejlt.200590000

23. 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.: LMSD: LIPID MAPS structure database. Nucleic Acids Res. 35, D527-532 (2007). doi:10.1093/nar/gkl838

24. Sud, M., Fahy, E., Subramaniam, S.: Template-based combinatorial enumeration of virtual compound libraries for lipids. J. Cheminformatics. 4, 23 (2012). doi:10.1186/1758-2946-4-23

25. Yetukuri, L., Ekroos, K., Vidal-Puig, A., Orešič, M.: Informatics and computational strategies for the study of lipids. Mol. Biosyst. 4, 121–127 (2008). doi:10.1039/B715468B

26. van Meer, G.: Cellular lipidomics. EMBO J. 24, 3159–3165 (2005). doi:10.1038/sj.emboj.7600798

27. Wenk, M.R.: The emerging field of lipidomics. Nat. Rev. Drug Discov. 4, 594–610 (2005). doi:10.1038/nrd1776

28. Shevchenko, A., Simons, K.: Lipidomics: coming to grips with lipid diversity. Nat. Rev. Mol. Cell Biol. 11, 593–598 (2010). doi:10.1038/nrm2934

29. Patterson, N.H., Thomas, A., Chaurand, P.: Monitoring time-dependent degradation of phospholipids in sectioned tissues by MALDI imaging mass spectrometry. J. Mass Spectrom. 49, 622–627 (2014). doi:10.1002/jms.3382

30. Wang, X., Gu, X., Song, H., Song, Q., Gao, X., Lu, Y., Chen, H.: Phenylmethanesulfonyl fluoride pretreatment stabilizes plasma lipidome in lipidomic and metabolomic analysis. Anal. Chim. Acta. 893, 77–83 (2015). doi:10.1016/j.aca.2015.08.049

31. Jentzsch, A.M., Bachmann, H., Fürst, P., Biesalski, H.K.: Improved analysis of malondialdehyde in human body fluids. Free Radic. Biol. Med. 20, 251–256 (1996). doi:10.1016/0891-5849(95)02043-8

32. Stumpf, P.K.: Lipids: Structure and Function: The Biochemistry of Plants. Elsevier, Amsterdam, Netherlands (2014)

33. Zhao, Z., Xu, Y.: An extremely simple method for extraction of lysophospholipids and phospholipids from blood samples. J. Lipid Res. 51, 652–659 (2010). doi:10.1194/jlr.D001503

34. Roughan, P.G., Slack, C.R., Holland, R.: Generation of phospholipid artefacts during extraction of developing soybean seeds with methanolic solvents. Lipids. 13, 497–503. doi:10.1007/BF02533620

35. Manirakiza, P., Covaci, A., Schepens, P.: Comparative Study on Total Lipid Determination using Soxhlet, Roese-Gottlieb, Bligh & Dyer, and Modified Bligh & Dyer Extraction Methods. J. Food Compos. Anal. 14, 93–100 (2001). doi:10.1006/jfca.2000.0972

36. Lee, J.-Y., Yoo, C., Jun, S.-Y., Ahn, C.-Y., Oh, H.-M.: Comparison of several methods for effective lipid extraction from microalgae. Bioresour. Technol. 101, S75–S77 (2010). doi:10.1016/j.biortech.2009.03.058

37. Patterson, R.E., Ducrocq, A.J., McDougall, D.J., Garrett, T.J., Yost, R.A.: Comparison of blood plasma sample preparation methods for combined LC–MS lipidomics and metabolomics. J. Chromatogr. B. 1002, 260–266 (2015). doi:10.1016/j.jchromb.2015.08.018

38. Bligh, E.G., Dyer, W.J.: A Rapid Method of Total Lipid Extraction and Purification. Can. J. Biochem. Physiol. 37, 911–917 (1959). doi:10.1139/o59-099

39. Folch, J., Lees, M., Sloane Stanley, G.H.: A simple method for the isolation and purification of total lipides from animal tissues. J. Biol. Chem. 226, 497–509 (1957)

40. Holland, W.L., Stauter, E.C., Stith, B.J.: Quantification of phosphatidic acid and lysophosphatidic acid by HPLC with evaporative light-scattering detection. J. Lipid Res. 44, 854–858 (2003). doi:10.1194/jlr.D200040-JLR200

41. Bromke, M.A., Hochmuth, A., Tohge, T., Fernie, A.R., Giavalisco, P., Burgos, A., Willmitzer, L., Brotman, Y.: Liquid chromatography high-resolution mass spectrometry for fatty acid profiling. Plant J. Cell Mol. Biol. 81, 529–536 (2015). doi:10.1111/tpj.12739

42. Han, X., Gross, R.W.: Global analyses of cellular lipidomes directly from crude extracts of biological samples by ESI mass spectrometry a bridge to lipidomics. J. Lipid Res. 44, 1071–1079 (2003). doi:10.1194/jlr.R300004-JLR200

43. Houjou, T., Yamatani, K., Imagawa, M., Shimizu, T., Taguchi, R.: A shotgun tandem mass spectrometric analysis of phospholipids with normal-phase and/or reverse-phase liquid chromatography/electrospray ionization mass spectrometry. Rapid Commun. Mass Spectrom. RCM. 19, 654–666 (2005). doi:10.1002/rcm.1836

44. Okazaki, Y., Kamide, Y., Hirai, M.Y., Saito, K.: Plant lipidomics based on hydrophilic interaction chromatography coupled to ion trap time-of-flight mass spectrometry. Metabolomics. 9, 121–131 (2013). doi:10.1007/s11306-011-0318-z

45. Hines, K.M., Herron, J., Xu, L.: Assessment of Altered Lipid Homeostasis by HILIC-Ion Mobility-Mass Spectrometry-Based Lipidomics. J. Lipid Res. jlr.D074724 (2017). doi:10.1194/jlr.D074724

46. Han, X., Gross, R.W.: Shotgun lipidomics: Electrospray ionization mass spectrometric analysis and quantitation of cellular lipidomes directly from crude extracts of biological samples. Mass Spectrom. Rev. 24, 367–412 (2005). doi:10.1002/mas.20023

47. Murphy, R.C., Fiedler, J., Hevko, J.: Analysis of nonvolatile lipids by mass spectrometry. Chem. Rev. 101, 479–526 (2001). doi:10.1021/cr9900883

48. Cai, S.-S., Syage, J.A.: Comparison of Atmospheric Pressure Photoionization, Atmospheric Pressure Chemical Ionization, and Electrospray Ionization Mass Spectrometry for Analysis of Lipids. Anal. Chem. 78, 1191–1199 (2006). doi:10.1021/ac0515834

49. Leinonen, A., Kuuranne, T., Kostiainen, R.: Liquid chromatography/mass spectrometry in anabolic steroid analysis—optimization and comparison of three ionization techniques: electrospray ionization, atmospheric pressure chemical ionization and atmospheric pressure photoionization. J. Mass Spectrom. 37, 693–698 (2002). doi:10.1002/jms.328

50. Cech, N.B., Enke, C.G.: Practical implications of some recent studies in electrospray ionization fundamentals. Mass Spectrom. Rev. 20, 362–387 (2001). doi:10.1002/mas.10008

51. Fernandez de la Mora, J.: Electrospray ionization of large multiply charged species proceeds via Dole's charged residue mechanism. Anal. Chim. Acta. 406, 93–104 (2000). doi:10.1016/S0003-2670(99)00601-7

52. Konermann, L., Ahadi, E., Rodriguez, A.D., Vahidi, S.: Unraveling the Mechanism of Electrospray Ionization. Anal. Chem. 85, 2–9 (2013). doi:10.1021/ac302789c

53. Fernandez de la Mora, J.: Electrospray ionization of large multiply charged species proceeds via Dole's charged residue mechanism. Anal. Chim. Acta. 406, 93–104 (2000). doi:10.1016/S0003-2670(99)00601-7

54. Beveridge, R., Phillips, A.S., Denbigh, L., Saleem, H.M., MacPhee, C.E., Barran, P.E.: Relating gas phase to solution conformations: Lessons from disordered proteins. Proteomics. 15, 2872–2883 (2015). doi:10.1002/pmic.201400605

55. Cech, N.B., Enke, C.G.: Practical implications of some recent studies in electrospray ionization fundamentals. Mass Spectrom. Rev. 20, 362–387 (2001). doi:10.1002/mas.10008

56. Yang, K., Han, X.: Accurate Quantification of Lipid Species by Electrospray Ionization Mass Spectrometry — Meets a Key Challenge in Lipidomics. Metabolites. 1, 21–40 (2011). doi:10.3390/metabo1010021

57. Han, X., Yang, K., Gross, R.W.: Multi-dimensional mass spectrometry-based shotgun lipidomics and novel strategies for lipidomic analyses. Mass Spectrom. Rev. 31, 134–178 (2012). doi:10.1002/mas.20342

58. O'Connor, P.B., Pittman, J.L., Thomson, B.A., Budnik, B.A., Cournoyer, J.C., Jebanathirajah, J., Lin, C., Moyer, S., Zhao, C.: A new hybrid electrospray Fourier transform mass spectrometer: design and performance characteristics. Rapid Commun. Mass Spectrom. 20, 259–266 (2006). doi:10.1002/rcm.2307

59. Makarov, A., Denisov, E., Kholomeev, A., Balschun, W., Lange, O., Strupat, K., Horning, S.: Performance Evaluation of a Hybrid Linear Ion Trap/Orbitrap Mass Spectrometer. Anal. Chem. 78, 2113–2120 (2006). doi:10.1021/ac0518811

60. Schwudke, D., Hannich, J.T., Surendranath, V., Grimard, V., Moehring, T., Burton, L., Kurzchalia, T., Shevchenko, A.: Top-Down Lipidomic Screens by Multivariate Analysis of High-Resolution Survey Mass Spectra. Anal. Chem. 79, 4083–4093 (2007). doi:10.1021/ac062455y

61. Olsen, J.V., Godoy, L.M.F. de, Li, G., Macek, B., Mortensen, P., Pesch, R., Makarov, A., Lange, O., Horning, S., Mann, M.: Parts per Million Mass Accuracy on an Orbitrap Mass Spectrometer via Lock Mass Injection into a C-trap. Mol. Cell. Proteomics. 4, 2010–2021 (2005). doi:10.1074/mcp.T500030-MCP200

62. Schwudke, D., Schuhmann, K., Herzog, R., Bornstein, S.R., Shevchenko, A.: Shotgun Lipidomics on High Resolution Mass Spectrometers. Cold Spring Harb. Perspect. Biol. 3, (2011). doi:10.1101/cshperspect.a004614

63. Biocompare: Talking About a Revolution: FT-ICR Mass Spectrometry Offers High Resolution and Mass Accuracy for Pr, http://www.biocompare.com/Editorial-Articles/41589-Talking-About-a-Revolution-FT-ICR-Mass-Spectrometry-Offers-High-Resolution-and-Mass-Accuracy-for-Pr/

64. Ahlf, D.R., Compton, P.D., Tran, J.C., Early, B.P., Thomas, P.M., Kelleher, N.L.: Evaluation of the Compact High-Field Orbitrap for Top-Down Proteomics of Human Cells. J. Proteome Res. 11, 4308–4314 (2012). doi:10.1021/pr3004216

65. Hardman, M., Makarov, A.A.: Interfacing the Orbitrap Mass Analyzer to an Electrospray Ion Source. Anal. Chem. 75, 1699–1705 (2003). doi:10.1021/ac0258047

66. Gorshkov, M.V., Fornelli, L., Tsybin, Y.O.: Observation of ion coalescence in Orbitrap Fourier transform mass spectrometry. Rapid Commun. Mass Spectrom. 26, 1711–1717 (2012). doi:10.1002/rcm.6289

67. Lange, O., Damoc, E., Wieghaus, A., Makarov, A.: Enhanced Fourier transform for Orbitrap mass spectrometry. Int. J. Mass Spectrom. 369, 16–22 (2014). doi:10.1016/j.ijms.2014.05.019

68. Hu, Q., Noll, R.J., Li, H., Makarov, A., Hardman, M., Cooks, G.R.: The Orbitrap: a new mass spectrometer. J. Mass Spectrom. JMS. 40, 430–443 (2005). doi:10.1002/jms.856

69. Makarov, A.: Electrostatic Axially Harmonic Orbital Trapping: A High-Performance Technique of Mass Analysis. Anal. Chem. 72, 1156–1162 (2000). doi:10.1021/ac991131p

70. Denisov, E., Damoc, E., Lange, O., Makarov, A.: Orbitrap mass spectrometry with resolving powers above 1,000,000. Int. J. Mass Spectrom. 325–327, 80–85 (2012). doi:10.1016/j.ijms.2012.06.009

71. Michalski, A., Damoc, E., Hauschild, J.-P., Lange, O., Wieghaus, A., Makarov, A., Nagaraj, N., Cox, J., Mann, M., Horning, S.: Mass Spectrometry-based Proteomics Using Q Exactive, a High-performance Benchtop Quadrupole Orbitrap Mass Spectrometer. Mol. Cell. Proteomics. 10, M111.011015 (2011). doi:10.1074/mcp.M111.011015

72. Olsen, J.V., Macek, B., Lange, O., Makarov, A., Horning, S., Mann, M.: Higher-energy C-trap dissociation for peptide modification analysis. Nat. Methods. 4, 709–712 (2007). doi:10.1038/nmeth1060

73. Olsen, J.V., Macek, B., Lange, O., Makarov, A., Horning, S., Mann, M.: Higher-energy C-trap dissociation for peptide modification analysis. Nat. Methods. 4, 709–712 (2007). doi:10.1038/nmeth1060

74. Kim, D.-H., Achcar, F., Breitling, R., Burgess, K.E., Barrett, M.P.: LC–MS-based absolute metabolite quantification: application to metabolic flux measurement in trypanosomes. Metabolomics. 11, 1721–1732 (2015). doi:10.1007/s11306-015-0827-2

75. Ivanova, P.T., Milne, S.B., Myers, D.S., Brown, H.A.: Lipidomics: a mass spectrometry based systems level analysis of cellular lipids. Curr. Opin. Chem. Biol. 13, 526–531 (2009). doi:10.1016/j.cbpa.2009.08.011

76. Kelley, N.S., Hubbard, N.E., Erickson, K.L.: Conjugated Linoleic Acid Isomers and Cancer. J. Nutr. 137, 2599–2607 (2007)

77. Tricon, S., Burdge, G.C., Kew, S., Banerjee, T., Russell, J.J., Jones, E.L., Grimble, R.F., Williams, C.M., Yaqoob, P., Calder, P.C.: Opposing effects of cis-9,trans-11 and trans-10,cis-12 conjugated linoleic acid on blood lipids in healthy humans. Am. J. Clin. Nutr. 80, 614–620 (2004)

78. Churruca, I., Fernández-Quintela, A., Portillo, M.P.: Conjugated linoleic acid isomers: Differences in metabolism and biological effects. BioFactors. 35, 105–111 (2009). doi:10.1002/biof.13

79. Shinzawa-Itoh, K., Aoyama, H., Muramoto, K., Terada, H., Kurauchi, T., Tadehara, Y., Yamasaki, A., Sugimura, T., Kurono, S., Tsujimoto, K., Mizushima, T., Yamashita, E., Tsukihara, T., Yoshikawa, S.: Structures and physiological roles of 13 integral lipids of bovine heart cytochrome c oxidase. EMBO J. 26, 1713–1725 (2007). doi:10.1038/sj.emboj.7601618

80. Kanehisa, M., Goto, S., Hattori, M., Aoki-Kinoshita, K.F., Itoh, M., Kawashima, S., Katayama, T., Araki, M., Hirakawa, M.: From genomics to chemical genomics: new developments in KEGG. Nucleic Acids Res. 34, D354–D357 (2006). doi:10.1093/nar/gkj102

81. Fahy, E., Subramaniam, S., Murphy, R.C., Nishijima, M., Raetz, C.R.H., Shimizu, T., Spener, F., van Meer, G., Wakelam, M.J.O., Dennis, E.A.: Update of the LIPID MAPS comprehensive classification system for lipids. J. Lipid Res. 50, S9–S14 (2009). doi:10.1194/jlr.R800095-JLR200

82. Knittelfelder, O.L., Weberhofer, B.P., Eichmann, T.O., Kohlwein, S.D., Rechberger, G.N.: A versatile ultra-high performance LC-MS method for lipid profiling. J. Chromatogr. B. 951–952, 119–128 (2014). doi:10.1016/j.jchromb.2014.01.011

83. Andrews, G.L., Dean, R.A., Hawkridge, A.M., Muddiman, D.C.: Improving Proteome Coverage on a LTQ-Orbitrap Using Design of Experiments. J. Am. Soc. Mass Spectrom. 22, 773–783 (2011). doi:10.1007/s13361-011-0075-2

84. K Hirasawa, Nishizuka, and Y.: Phosphatidylinositol Turnover in Receptor Mechanism and Signal Transduction. Annu. Rev. Pharmacol. Toxicol. 25, 147–170 (1985). doi:10.1146/annurev.pa.25.040185.001051

85. Tang, X., Edwards, E.M., Holmes, B.B., Falck, J.R., Campbell, W.B.: Role of phospholipase C and diacylglyceride lipase pathway in arachidonic acid release and acetylcholine-induced vascular relaxation in rabbit aorta. Am. J. Physiol. - Heart Circ. Physiol. 290, H37–H45 (2006). doi:10.1152/ajpheart.00491.2005

86. Bendall, S.C., Hughes, C., Campbell, J.L., Stewart, M.H., Pittock, P., Liu, S., Bonneil, E., Thibault, P., Bhatia, M., Lajoie, G.A.: An Enhanced Mass Spectrometry Approach Reveals Human Embryonic Stem Cell Growth Factors in Culture. Mol. Cell. Proteomics. 8, 421–432 (2009). doi:10.1074/mcp.M800190-MCP200

87. Rudomin, E.L., Carr, S.A., Jaffe, J.D.: Directed Sample Interrogation Utilizing an Accurate Mass Exclusion-Based Data-Dependent Acquisition Strategy (AMEx). J. Proteome Res. 8, 3154–3160 (2009). doi:10.1021/pr801017a

88. Rolland, A.D., Lavigne, R., Dauly, C., Calvel, P., Kervarrec, C., Freour, T., Evrard, B., Rioux-Leclercq, N., Auger, J., Pineau, C.: Identification of genital tract markers in the human seminal plasma using an integrative genomics approach. Hum. Reprod. 28, 199–209 (2013). doi:10.1093/humrep/des360

89. Seo, J., Jeong, J., Kim, Y.M., Hwang, N., Paek, E., Lee, K.-J.: Strategy for Comprehensive Identification of Post-translational Modifications in Cellular Proteins, Including Low Abundant Modifications: Application to Glyceraldehyde-3-phosphate Dehydrogenase. J. Proteome Res. 7, 587–602 (2008). doi:10.1021/pr700657y

90. Hughes, C.S., Postovit, L.M., Lajoie, G.A.: Matrigel: A complex protein mixture required for optimal growth of cell culture. Proteomics. 10, 1886–1890 (2010). doi:10.1002/pmic.200900758

91. Vahidi, S., Stocks, B.B., Liaghati-Mobarhan, Y., Konermann, L.: Mapping pH-Induced Protein Structural Changes Under Equilibrium Conditions by Pulsed Oxidative Labeling and Mass Spectrometry. Anal. Chem. 84, 9124–9130 (2012). doi:10.1021/ac302393g

92. Lämmerhofer, M., Weckwerth, W.: Metabolomics in Practice: Successful Strategies to Generate and Analyze Metabolic Data. John Wiley & Sons (2013)

93. Edmands, W.M., Ferrari, P., Rothwell, J.A., Rinaldi, S., Slimani, N., Barupal, D.K., Biessy, C., Jenab, M., Clavel-Chapelon, F., Fagherazzi, G., Boutron-Ruault, M.-C., Katzke, V.A., Kühn, T., Boeing, H., Trichopoulou, A., Lagiou, P., Trichopoulos, D., Palli, D., Grioni, S., Tumino, R., Vineis, P., Mattiello, A., Romieu, I., Scalbert, A.: Polyphenol metabolome in human urine and its association with intake of polyphenol-rich foods across European countries. Am. J. Clin. Nutr. 102, 905–913 (2015). doi:10.3945/ajcn.114.101881

94. Nazari, M., Muddiman, D.C.: Enhanced Lipidome Coverage in Shotgun Analyses by using Gas-Phase Fractionation. J. Am. Soc. Mass Spectrom. 27, 1735–1744 (2016). doi:10.1007/s13361-016-1446-5

95. Schwudke, D., Oegema, J., Burton, L., Entchev, E., Hannich, J.T., Ejsing, C.S., Kurzchalia, T., Shevchenko, A.: Lipid Profiling by Multiple Precursor and Neutral Loss Scanning Driven by the Data-Dependent Acquisition. Anal. Chem. 78, 585–595 (2006). doi:10.1021/ac051605m

96. Lin, L., Yu, Q., Yan, X., Hang, W., Zheng, J., Xing, J., Huang, B.: Direct infusion mass spectrometry or liquid chromatography mass spectrometry for human metabonomics? A serum metabonomic study of kidney cancer. The Analyst. 135, 2970–2978 (2010). doi:10.1039/c0an00265h

97. R Development Core Team: R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria (2016)

98. Kessner, D., Chambers, M., Burke, R., Agus, D., Mallick, P.: ProteoWizard: open source software for rapid proteomics tools development. Bioinforma. Oxf. Engl. 24, 2534–2536 (2008). doi:10.1093/bioinformatics/btn323

99. Taguchi, R., Ishikawa, M.: Precise and global identification of phospholipid molecular species by an Orbitrap mass spectrometer and automated search engine Lipid Search. J. Chromatogr. A. 1217, 4229–4239 (2010). doi:10.1016/j.chroma.2010.04.034

100. Pluskal, T., Castillo, S., Villar-Briones, A., Orešič, M.: MZmine 2: Modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. BMC Bioinformatics. 11, 395 (2010). doi:10.1186/1471-2105-11-395

101. Turner, J.D., Rouser, G.: Precise quantitative determination of human blood lipids by thin-layer and triethylaminoethylcellulose column chromatography. Anal. Biochem. 38, 437–445 (1970). doi:10.1016/0003-2697(70)90468-9

102. 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., McDonald, J.G., Subramaniam, S., Fahy, E., Dennis, E.A.: Lipidomics reveals a remarkable diversity of lipids in human plasma,. J. Lipid Res. 51, 3299–3305 (2010). doi:10.1194/jlr.M009449

103. Wang, M., Hayakawa, J., Yang, K., Han, X.: Characterization and quantification of diacylglycerol species in biological extracts after one-step derivatization: a shotgun lipidomics approach. Anal. Chem. 86, 2146–2155 (2014). doi:10.1021/ac403798q

104. Jenkins, B., West, J.A., Koulman, A.: A Review of Odd-Chain Fatty Acid Metabolism and the Role of Pentadecanoic Acid (C15:0) and Heptadecanoic Acid (C17:0) in Health and Disease. Molecules. 20, 2425–2444 (2015). doi:10.3390/molecules20022425

105. Gilbertson, J.R., Karnovsky, M.L.: Nonphosphatide fatty acyl esters of alkenyl and alkyl ethers of glycerol. J. Biol. Chem. 238, 893–897 (1963)

106. Kind, T., Liu, K.-H., Lee, D.Y., DeFelice, B., Meissen, J.K., Fiehn, O.: LipidBlast in silico tandem mass spectrometry database for lipid identification. Nat. Methods. 10, 755–758 (2013). doi:10.1038/nmeth.2551

107. Liebisch, G., Vizcaíno, J.A., Köfeler, H., Trötzmüller, M., Griffiths, W.J., Schmitz, G., Spener, F., Wakelam, M.J.O.: Shorthand notation for lipid structures derived from mass spectrometry. J. Lipid Res. 54, 1523–1530 (2013). doi:10.1194/jlr.M033506

108. Koelmel, J.P., Ulmer, C.Z., Jones, C.M., Yost, R.A., Bowden, J.A.: Common cases of improper lipid annotation using high-resolution tandem mass spectrometry data and corresponding limitations in biological interpretation. Biochim. Biophys. Acta BBA - Mol. Cell Biol. Lipids. 1862, 766–770 (2017). doi:10.1016/j.bbalip.2017.02.016

109. Ekroos, K., Ejsing, C.S., Bahr, U., Karas, M., Simons, K., Shevchenko, A.: Charting molecular composition of phosphatidylcholines by fatty acid scanning and ion trap MS3 fragmentation. J. Lipid Res. 44, 2181–2192 (2003). doi:10.1194/jlr.D300020-JLR200

110. Kyle, J.E., Zhang, X., Weitz, K.K., Monroe, M.E., Ibrahim, Y.M., Moore, R.J., Cha, J., Sun, X., Lovelace, E.S., Wagoner, J., Polyak, S.J., Metz, T.O., Dey, S.K., Smith, R.D., Burnum-Johnson, K.E., Baker, E.S.: Uncovering biologically significant lipid isomers with liquid chromatography, ion mobility spectrometry and mass spectrometry. Analyst. 141, 1649–1659 (2016). doi:10.1039/C5AN02062J

111. IUPAC-IUB Joint Commission on Biochemical Nomenclature (JCBN) Nomenclature of glycolipids. Eur. J. Biochem. 257, 293–298 (1998). doi:10.1046/j.1432-1327.1998.2570293.x

112. The nomenclature of lipids (recommendations 1976). IUPAC-IUB Commission on Biochemical Nomenclature. J. Lipid Res. 19, 114–128 (1978)

113. IUPAC-IUB Commission on Biochemical Nomenclature (CBN): The Nomenclature of Lipids. Eur. J. Biochem. 2, 127–131 (1967). doi:10.1111/j.1432-1033.1967.tb00116.x

114. Fahy, E., Subramaniam, S., Brown, H.A., Glass, C.K., Merrill, A.H., Murphy, R.C., Raetz, C.R.H., Russell, D.W., Seyama, Y., Shaw, W., Shimizu, T., Spener, F., van Meer, G., VanNieuwenhze, M.S., White, S.H., Witztum, J.L., Dennis, E.A.: A comprehensive classification system for lipids. J. Lipid Res. 46, 839–861 (2005). doi:10.1194/jlr.E400004-JLR200

115. Gerspach, C., Imhasly, S., Gubler, M., Naegeli, H., Ruetten, M., Laczko, E.: Altered plasma lipidome profile of dairy cows with fatty liver disease. Res. Vet. Sci. 110, 47–59 (2017). doi:10.1016/j.rvsc.2016.10.001

116. Smith, R., Mathis, A.D., Ventura, D., Prince, J.T.: Proteomics, lipidomics, metabolomics: a mass spectrometry tutorial from a computer scientist's point of view. BMC Bioinformatics. 15, S9 (2014). doi:10.1186/1471-2105-15-S7-S9

117. Ding, J., Sorensen, C.M., Jaitly, N., Jiang, H., Orton, D.J., Monroe, M.E., Moore, R.J., Smith, R.D., Metz, T.O.: Application of the accurate mass and time tag approach in studies of the human blood lipidome. J. Chromatogr. B Analyt. Technol. Biomed. Life. Sci. 871, 243–252 (2008). doi:10.1016/j.jchromb.2008.04.040

118. Liu, Y., Chen, Y., Momin, A., Shaner, R., Wang, E., Bowen, N.J., Matyunina, L.V., Walker, L.D., McDonald, J.F., Sullards, M.C., Merrill, A.H.: Elevation of sulfatides in ovarian cancer: An integrated transcriptomic and lipidomic analysis including tissue-imaging mass spectrometry. Mol. Cancer. 9, 186 (2010). doi:10.1186/1476-4598-9-186

119. Jin, R., Li, L., Feng, J., Dai, Z., Huang, Y.-W., Shen, Q.: Zwitterionic hydrophilic interaction solid-phase extraction and multi-dimensional mass spectrometry for shotgun lipidomic study of Hypophthalmichthys nobilis. Food Chem. 216, 347–354 (2017). doi:10.1016/j.foodchem.2016.08.074

120. Axelsen, P.H., Murphy, R.C.: Quantitative analysis of phospholipids containing arachidonate and docosahexaenoate chains in microdissected regions of mouse brain. J. Lipid Res. 51, 660–671 (2010). doi:10.1194/jlr.D001750

121. Liebisch, G., Ejsing, C.S., Ekroos, K.: Identification and Annotation of Lipid Species in Metabolomics Studies Need Improvement. Clin. Chem. 61, 1542–1544 (2015). doi:10.1373/clinchem.2015.244830

122. Narváez-Rivas, M., Vu, N., Chen, G.-Y., Zhang, Q.: Off-line mixed-mode liquid chromatography coupled with reversed phase high performance liquid chromatography-high resolution mass spectrometry to improve coverage in lipidomics analysis. Anal. Chim. Acta. 954, 140–150 (2017). doi:10.1016/j.aca.2016.12.003

123. Li, Q., Zhao, Y., Zhu, D., Pang, X., Liu, Y., Frew, R., Chen, G.: Lipidomics profiling of goat milk, soymilk and bovine milk by UPLC-Q-Exactive Orbitrap Mass Spectrometry. Food Chem. 224, 302–309 (2017). doi:10.1016/j.foodchem.2016.12.083

124. Skotland, T., Ekroos, K., Kauhanen, D., Simolin, H., Seierstad, T., Berge, V., Sandvig, K., Llorente, A.: Molecular lipid species in urinary exosomes as potential prostate cancer biomarkers. Eur. J. Cancer. 70, 122–132 (2017). doi:10.1016/j.ejca.2016.10.011

125. Gallego, S.F., Sprenger, R.R., Neess, D., Pauling, J.K., Færgeman, N.J., Ejsing, C.S.: Quantitative lipidomics reveals age-dependent perturbations of whole-body lipid metabolism in ACBP deficient mice. Biochim. Biophys. Acta BBA - Mol. Cell Biol. Lipids. 1862, 145–155 (2017). doi:10.1016/j.bbalip.2016.10.012

126. Zhong, H., Xiao, M., Zarkovic, K., Zhu, M., Sa, R., Lu, J., Tao, Y., Chen, Q., Xia, L., Cheng, S., Waeg, G., Zarkovic, N., Yin, H.: Mitochondrial control of apoptosis through modulation of cardiolipin oxidation in hepatocellular carcinoma: A novel link between oxidative stress and cancer. Free Radic. Biol. Med. 102, 67–76 (2017). doi:10.1016/j.freeradbiomed.2016.10.494

127. Lerner, R., Post, J., Loch, S., Lutz, B., Bindila, L.: Targeting brain and peripheral plasticity of the lipidome in acute kainic acid-induced epileptic seizures in mice via quantitative mass spectrometry. Biochim. Biophys. Acta BBA - Mol. Cell Biol. Lipids. 1862, 255–267 (2017). doi:10.1016/j.bbalip.2016.11.008

128. Tsugawa, H., Cajka, T., Kind, T., Ma, Y., Higgins, B., Ikeda, K., Kanazawa, M., VanderGheynst, J., Fiehn, O., Arita, M.: MS-DIAL: data-independent MS/MS deconvolution for comprehensive metabolome analysis. Nat. Methods. 12, 523–526 (2015). doi:10.1038/nmeth.3393

129. Kochen, M.A., Chambers, M.C., Holman, J.D., Nesvizhskii, A.I., Weintraub, S.T., Belisle, J.T., Islam, M.N., Griss, J., Tabb, D.L.: Greazy: Open-Source Software for Automated Phospholipid Tandem Mass Spectrometry Identification. Anal. Chem. 88, 5733–5741 (2016). doi:10.1021/acs.analchem.6b00021

130. Zemski Berry, K.A., Murphy, R.C.: Electrospray ionization tandem mass spectrometry of glycerophosphoethanolamine plasmalogen phospholipids. J. Am. Soc. Mass Spectrom. 15, 1499–1508 (2004). doi:10.1016/j.jasms.2004.07.009

131. Koelmel, J.P., Kroeger, N.M., Gill, E.L., Ulmer, C.Z., Bowden, J.A., Patterson, R.E., Yost, R.A., Garrett, T.J.: Expanding lipidome coverage using LC-MS/MS data-dependent acquisition with automated exclusion list generation. J. Am. Soc. Mass Spectrom. Manuscript Submitted, (2017)

132. Gil, A., Siegel, D., Permentier, H., Reijngoud, D.-J., Dekker, F., Bischoff, R.: Stability of energy metabolites? An often overlooked issue in metabolomics studies: A review. Electrophoresis. 36, 2156–2169 (2015). doi:10.1002/elps.201500031

133. Lorenz, M.A., Burant, C.F., Kennedy, R.T.: Reducing Time and Increasing Sensitivity in Sample Preparation for Adherent Mammalian Cell Metabolomics. Anal. Chem. 83, 3406–3414 (2011). doi:10.1021/ac103313x;

134. Plueckthun, A., Dennis, E.A.: Acyl and phosphoryl migration in lysophospholipids: importance in phospholipid synthesis and phospholipase specificity. Biochemistry (Mosc.). 21, 1743–1750 (1982). doi:10.1021/bi00537a007

135. R Development Core Team: R: a language and environment for statistical computing. Vienna, Austria : the R Foundation for Statistical Computing. ISBN: 3-900051-07-0. Available online at http://www.R-project.org/. (2011)

136. Fahy, E., Sud, M., Cotter, D., Subramaniam, S.: LIPID MAPS online tools for lipid research. Nucleic Acids Res. 35, W606-612 (2007). doi:10.1093/nar/gkm324

137. Kind, T., Okazaki, Y., Saito, K., Fiehn, O.: LipidBlast Templates As Flexible Tools for Creating New in-Silico Tandem Mass Spectral Libraries. Anal. Chem. 86, 11024–11027 (2014). doi:10.1021/ac502511a

138. Tautenhahn, R., Patti, G.J., Rinehart, D., Siuzdak, G.: XCMS Online: a web-based platform to process untargeted metabolomic data. Anal. Chem. 84, 5035–5039 (2012). doi:10.1021/ac300698c

139. Husen, P., Tarasov, K., Katafiasz, M., Sokol, E., Vogt, J., Baumgart, J., Nitsch, R., Ekroos, K., Ejsing, C.S.: Analysis of Lipid Experiments (ALEX): A Software Framework for Analysis of High-Resolution Shotgun Lipidomics Data. PLOS ONE. 8, e79736 (2013). doi:10.1371/journal.pone.0079736

140. Herzog, R., Schuhmann, K., Schwudke, D., Sampaio, J.L., Bornstein, S.R., Schroeder, M., Shevchenko, A.: LipidXplorer: A Software for Consensual Cross-Platform Lipidomics. PLoS ONE. 7, e29851 (2012). doi:10.1371/journal.pone.0029851

141. Sabareesh, V., Singh, G.: Mass spectrometry based lipid(ome) analyzer and molecular platform: a new software to interpret and analyze electrospray and/or matrix-assisted laser desorption/ionization mass spectrometric data of lipids: a case study from Mycobacterium tuberculosis. J. Mass Spectrom. JMS. 48, 465–477 (2013). doi:10.1002/jms.3163

142. Haimi, P., Chaithanya, K., Kainu, V., Hermansson, M., Somerharju, P.: Instrument-independent software tools for the analysis of MS-MS and LC-MS lipidomics data. Methods Mol. Biol. Clifton NJ. 580, 285–294 (2009). doi:10.1007/978-1-60761-325-1_16

143. Collins, J.R., Edwards, B.R., Fredricks, H.F., Van Mooy, B.A.S.: LOBSTAHS: An Adduct-Based Lipidomics Strategy for Discovery and Identification of Oxidative Stress Biomarkers. Anal. Chem. 88, 7154–7162 (2016). doi:10.1021/acs.analchem.6b01260

144. Hartler, J., Trötzmüller, M., Chitraju, C., Spener, F., Köfeler, H.C., Thallinger, G.G.: Lipid Data Analyzer: unattended identification and quantitation of lipids in LC-MS data. Bioinformatics. 27, 572–577 (2011). doi:10.1093/bioinformatics/btq699

145. Song, H., Hsu, F.-F., Ladenson, J., Turk, J.: 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 (2007). doi:10.1016/j.jasms.2007.07.023

146. Ahmed, Z., Mayr, M., Zeeshan, S., Dandekar, T., Mueller, M.J., Fekete, A.: Lipid-Pro: a computational lipid identification solution for untargeted lipidomics on data-independent acquisition tandem mass spectrometry platforms. Bioinformatics. 31, 1150–1153 (2015). doi:10.1093/bioinformatics/btu796

147. Lintonen, T.P.I., Baker, P.R.S., Suoniemi, M., Ubhi, B.K., Koistinen, K.M., Duchoslav, E., Campbell, J.L., Ekroos, K.: Differential Mobility Spectrometry-Driven Shotgun Lipidomics. Anal. Chem. 86, 9662–9669 (2014). doi:10.1021/ac5021744

148. Koivusalo, M., Haimi, P., Heikinheimo, L., Kostiainen, R., Somerharju, P.: Quantitative determination of phospholipid compositions by ESI-MS: effects of acyl chain length, unsaturation, and lipid concentration on instrument response. J. Lipid Res. 42, 663–672 (2001)

149. Lam, S.M., Tian, H., Shui, G.: Lipidomics, en route to accurate quantitation. Biochim. Biophys. Acta BBA - Mol. Cell Biol. Lipids. 1862, 752–761 (2017). doi:10.1016/j.bbalip.2017.02.008

150. Saito, K., Ohno, Y., Saito, Y.: Enrichment of resolving power improves ion-peak quantification on a lipidomics platform. J. Chromatogr. B. 1055, 20–28 (2017). doi:10.1016/j.jchromb.2017.04.019

151. Haimi, P., Uphoff, A., Hermansson, M., Somerharju, P.: Software tools for analysis of mass spectrometric lipidome data. Anal. Chem. 78, 8324–8331 (2006). doi:10.1021/ac061390w

152. Fauland, A., Köfeler, H., Trötzmüller, M., Knopf, A., Hartler, J., Eberl, A., Chitraju, C., Lankmayr, E., Spener, F.: A comprehensive method for lipid profiling by liquid chromatography-ion cyclotron resonance mass spectrometry. J. Lipid Res. 52, 2314–2322 (2011). doi:10.1194/jlr.D016550

153. Koelmel, J.P., Kroeger, N.M., Ulmer, C.Z., Bowden, J.A., Patterson, R.E., Cochran, J.A., Beecher, C.W.W., Garrett, T.J., Yost, R.A.: LipidMatch: an automated workflow for rule-based lipid identification using untargeted high-resolution tandem mass spectrometry data. BMC Bioinformatics. 18, 331 (2017). doi:10.1186/s12859-017-1744-3

154. Zacarias, A., Bolanowski, D., Bhatnagar, A.: Comparative measurements of multicomponent phospholipid mixtures by electrospray mass spectroscopy: relating ion intensity to concentration. Anal. Biochem. 308, 152–159 (2002). doi:10.1016/S0003-2697(02)00209-9

155. Phinney, K.W., Ballihaut, G., Bedner, M., Benford, B.S., Camara, J.E., Christopher, S.J., Davis, W.C., Dodder, N.G., Eppe, G., Lang, B.E., Long, S.E., Lowenthal, M.S., McGaw, E.A., Murphy, K.E., Nelson, B.C., Prendergast, J.L., Reiner, J.L., Rimmer, C.A., Sander, L.C., Schantz, M.M., Sharpless, K.E., Sniegoski, L.T., Tai, S.S.-C., Thomas, J.B., Vetter, T.W., Welch, M.J., Wise, S.A., Wood, L.J., Guthrie, W.F., Hagwood, C.R., Leigh, S.D., Yen, J.H., Zhang, N.-F., Chaudhary-Webb, M., Chen, H., Fazili, Z., LaVoie, D.J., McCoy, L.F., Momin, S.S., Paladugula, N., Pendergrast, E.C., Pfeiffer, C.M., Powers, C.D., Rabinowitz, D., Rybak, M.E., Schleicher, R.L., Toombs, B.M.H., Xu, M., Zhang, M., Castle, A.L.: Development of a Standard Reference Material for Metabolomics Research. Anal. Chem. 85, 11732–11738 (2013). doi:10.1021/ac402689t

156. Matyash, V., Liebisch, G., Kurzchalia, T.V., Shevchenko, A., Schwudke, D.: Lipid extraction by methyl-tert-butyl ether for high-throughput lipidomics. J. Lipid Res. 49, 1137–1146 (2008). doi:10.1194/jlr.D700041-JLR200

157. Ulmer, C.Z., Koelmel, J.P., Ragland, J.M., Garrett, T.J., Bowden, J.A.: LipidPioneer : A Comprehensive User-Generated Exact Mass Template for Lipidomics. J. Am. Soc. Mass Spectrom. 28, 1–4 (2017). doi:10.1007/s13361-016-1579-6

158. Altman, D.G., Bland, J.M.: Measurement in Medicine: The Analysis of Method Comparison Studies. J. R. Stat. Soc. Ser. Stat. 32, 307–317 (1983). doi:10.2307/2987937

159. Giavarina, D.: Understanding Bland Altman analysis. Biochem. Medica. 25, 141–151 (2015). doi:10.11613/BM.2015.015

160. Xia, J., Sinelnikov, I.V., Han, B., Wishart, D.S.: MetaboAnalyst 3.0—making metabolomics more meaningful. Nucleic Acids Res. gkv380 (2015). doi:10.1093/nar/gkv380

161. Tang, S., Zhang, H., Lee, H.K.: Advances in Sample Extraction. Anal. Chem. 88, 228–249 (2016). doi:10.1021/acs.analchem.5b04040

162. Checa, A., Bedia, C., Jaumot, J.: Lipidomic data analysis: Tutorial, practical guidelines and applications. Anal. Chim. Acta. 885, 1–16 (2015). doi:10.1016/j.aca.2015.02.068

163. Astarita, G., Ahmed, F., Piomelli, D.: Lipidomic Analysis of Biological Samples by Liquid Chromatography Coupled to Mass Spectrometry. In: Armstrong, D. (ed.) Lipidomics. pp. 201–219. Humana Press (2009)

164. Han, X.: Lipidomics: Comprehensive Mass Spectrometry of Lipids. John Wiley & Sons (2016)

165. Teo, C.C., Chong, W.P.K., Tan, E., Basri, N.B., Low, Z.J., Ho, Y.S.: Advances in sample preparation and analytical techniques for lipidomics study of clinical samples. TrAC Trends Anal. Chem. 66, 1–18 (2015). doi:10.1016/j.trac.2014.10.010

166. Yin, P., Lehmann, R., Xu, G.: Effects of pre-analytical processes on blood samples used in metabolomics studies. Anal. Bioanal. Chem. 407, 4879–4892 (2015). doi:10.1007/s00216-015-8565-x

167. Jernerén, F., Söderquist, M., Karlsson, O.: Post-sampling release of free fatty acids — effects of heat stabilization and methods of euthanasia. J. Pharmacol. Toxicol. Methods. 71, 13–20 (2015). doi:10.1016/j.vascn.2014.11.001

168. Saigusa, D., Okudaira, M., Wang, J., Kano, K., Kurano, M., Uranbileg, B., Ikeda, H., Yatomi, Y., Motohashi, H., Aoki, J.: Simultaneous Quantification of Sphingolipids in Small Quantities of Liver by LC-MS/MS. Mass Spectrom. 3, S0046–S0046 (2014). doi:10.5702/massspectrometry.S0046

169. Kamthe, P.N., University, F.A.: Acute toxicity of the agricultural chemicals endosulfan and copper sulfate to a freshwater shrimp, Palaemonetes paludosus. MyScienceWork.

170. Li, M.-H., McKee, M.J.: Toxicokinetics of 2,2′,4,4′- and 3,3′,4,4′-tetrachlorobiphenyl congeners in the house cricket. Ecotoxicol. Environ. Saf. 24, 309–318 (1992). doi:10.1016/0147-6513(92)90007-P

171. Beeby, A.: What do sentinels stand for? Environ. Pollut. 112, 285–298 (2001). doi:10.1016/S0269-7491(00)00038-5

172. Liebeke, M., Bundy, J.G.: Tissue disruption and extraction methods for metabolic profiling of an invertebrate sentinel species. Metabolomics. 8, 819–830 (2011). doi:10.1007/s11306-011-0377-1

173. Schock, T.B., Strickland, S., Steele, E.J., Bearden, D.W.: Effects of heat-treatment on the stability and composition of metabolomic extracts from the earthworm Eisenia fetida. Metabolomics. 12, (2016). doi:10.1007/s11306-016-0967-z

174. Koelmel, J., Kroeger, N.M., Ulmer, C.Z., Bowden, J.A., Patterson, R.E., Cochran, J., Beecher, C., Garrett, T.J., Yost, R.A.: LipidMatch: an automated workflow for rule-based lipid identification using untargeted high-resolution tandem mass spectrometry data. BMC Bioinformatics. (2017)

175. Benjamini, Y., Hochberg, Y.: Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J. R. Stat. Soc. Ser. B Methodol. 57, 289–300 (1995)

176. Stevens, A., Lowe, J.S., Scott, I.: Core Pathology. Elsevier Health Sciences, Amsterdam, Netherlands (2008)

177. Six, D.A., Dennis, E.A.: The expanding superfamily of phospholipase A2 enzymes: classification and characterization. Biochim. Biophys. Acta BBA - Mol. Cell Biol. Lipids. 1488, 1–19 (2000). doi:10.1016/S1388-1981(00)00105-0

178. Broniec, A., Klosinski, R., Pawlak, A., Wrona-Krol, M., Thompson, D., Sarna, T.: Interactions of plasmalogens and their diacyl analogs with singlet oxygen in selected model systems. Free Radic. Biol. Med. 50, 892–898 (2011). doi:10.1016/j.freeradbiomed.2011.01.002

179. Sassa, T., Kihara, A.: Metabolism of Very Long-Chain Fatty Acids: Genes and Pathophysiology. Biomol. Ther. 22, 83–92 (2014). doi:10.4062/biomolther.2014.017

180. Abdelkafi, S., Abousalham, A.: The substrate specificities of sunflower and soybean phospholipases D using transphosphatidylation reaction. Lipids Health Dis. 10, 196 (2011). doi:10.1186/1476-511X-10-196

181. Armand, M., Borel, P., Ythier, P., Dutot, G., Melin, C., Senft, M., Lafont, H., Lairon, D.: Effects of droplet size, triacylglycerol composition, and calcium on the hydrolysis of complex emulsions by pancreatic lipase: an in vitro study. J. Nutr. Biochem. 3, 333–341 (1992). doi:10.1016/0955-2863(92)90024-D

182. Jenkins-Kruchten, A.E., Bennaars-Eiden, A., Ross, J.R., Shen, W.-J., Kraemer, F.B., Bernlohr, D.A.: Fatty Acid-binding Protein-Hormone-sensitive Lipase Interaction. J. Biol. Chem. 278, 47636–47643 (2003). doi:10.1074/jbc.M307680200

183. Reis, A., Spickett, C.M.: Chemistry of phospholipid oxidation. Biochim. Biophys. Acta BBA - Biomembr. 1818, 2374–2387 (2012). doi:10.1016/j.bbamem.2012.02.002

184. Brown, S.H.J., Mitchell, T.W., Blanksby, S.J.: Analysis of unsaturated lipids by ozone-induced dissociation. Biochim. Biophys. Acta BBA - Mol. Cell Biol. Lipids. 1811, 807–817 (2011). doi:10.1016/j.bbalip.2011.04.015

185. Cajka, T., Fiehn, O.: Comprehensive analysis of lipids in biological systems by liquid chromatography-mass spectrometry. Trends Anal. Chem. TRAC. 61, 192–206 (2014). doi:10.1016/j.trac.2014.04.017

186. Zhang, X., Garimella, S.V.B., Prost, S.A., Webb, I.K., Chen, T.-C., Tang, K., Tolmachev, A.V., Norheim, R.V., Baker, E.S., Anderson, G.A., Ibrahim, Y.M., Smith, R.D.: Ion Trapping, Storage, and Ejection in Structures for Lossless Ion Manipulations. Anal. Chem. 87, 6010–6016 (2015). doi:10.1021/acs.analchem.5b00214

187. Heller, S.R., McNaught, A., Pletnev, I., Stein, S., Tchekhovskoi, D.: InChI, the IUPAC International Chemical Identifier. J. Cheminformatics. 7, 23 (2015). doi:10.1186/s13321-015-0068-4

188. Homer, R.W., Swanson, J., Jilek, R.J., Hurst, T., Clark, R.D.: SYBYL Line Notation (SLN): A Single Notation To Represent Chemical Structures, Queries, Reactions, and Virtual Libraries. J. Chem. Inf. Model. 48, 2294–2307 (2008). doi:10.1021/ci7004687

BIOGRAPHICAL SKETCH

Jeremy P. Koelmel grew up in the borough of Queens in New York City. He was always passionate about the natural world, and it was his greatest pleasure to be in the woods of upstate New York or spend time in his backyard as a child. He would spend countless hours getting lost in the worlds of the small creatures and plants. Jeremy flourished academically in high school due to the project-based learning approach at the Baccalaureate School for Global Education (BSGE) and his passionate and invested teachers. Jeremy Koelmel's first exposure to rigorous scientific research was as a high school senior when he determined novel lichen species as indicators of vehicle pollution in his home county of Queens, NYC, receiving the Young Naturalist Award from the American Museum of Natural History (2007).

Jeremy focused on environmental chemistry during his years at Hampshire College (2007-2011). In the summer of 2010, Jeremy was awarded an NSF REU grant culminating in a paper (unpublished) on diurnal isoprene nitrate chemistry in polluted and unpolluted air parcels as part of his work at the University of Michigan Biological Station. The following year he completed a 122 page senior thesis modeling metal uptake and imaging metal distributions in ferns growing on shooting range soils in Chesterfield, MA, under chemistry professor Dulasiri Amarasiriwardena and botany professor Lawrence Winship. Results were published in the journal *Environment and Pollution*. In the summer of 2011, he continued research with Dulasiri, researching gold nanoparticle uptake by rice plants (manuscript also published in *Environment and Pollution*) and synthesizing iron nanoparticles for adsorption of trace metals.

During Jeremy's time at Hampshire College, he began to practice Vipassana meditation as taught by S. N. Goenka and became very interested in Buddhist

philosophy from various traditions. Therefore, he took part in the 5 college exchange program to the Central University of Tibetan Studies, in Sarnath, India, where he was shocked by the level of environmental pollution in India. After he graduated from Hampshire College, he applied and was accepted as a Fulbright-Nehru scholar to study phytoremediation (low-cost remediation of contamination using plants and bacteria) with an expert in the field, Dr. M. N. V. Prasad. Jeremy had reached out to Dr. Prasad after being inspired after reading one of Dr. Prasad's many books on the subject. As a Fulbright-Nehru scholar (2012), Jeremy authored two book chapters and one journal article on the topic of phytoremediation, with emphasis on prospects for phytoremediation in India. As part of Jeremy's Fulbright scholarship, Jeremy used Google maps to map out possible pollution sources to the local water systems of Ranipet, Tamil Nadu, India. Jeremy hired a translator and retired Indian military personnel as a driver, interviewing numerous villagers, taking industrial sludge and water samples for trace metal analysis, and investigating the flora and microbial organism growing in the most polluted locations. Ranipet is designated as one of the top most polluted locations on earth by Blacksmith Institute in 2006.

Jeremy's experience in Ranipet and India made him realize that trace metal analysis was limited in scope, and if he was to better characterize and investigate polluted areas he would benefit from learning organic mass spectrometry. After meeting Dr. Richard Yost at the University of Florida, he was attracted by the freedom to follow his interests under Dr. Yost's mentorship. Dr. Yost's similar interests in environmental chemistry also attracted Jeremy. In addition, Dr. Yost's expertise in mass spectrometry, and financial and networking connections, were a valuable asset for Jeremy to realize

his career goals. Soon after he joined the Yost group in 2013, Jeremy began working on investigating wildlife disease via lipidomics with a collaboration with Dr. John Bowden, which culminated in a trip to South Africa to investigate the devastating spread of a disease called pansteatitis. After taking interest in lipids, Jeremy quickly realized there were limited software tools to do the required data processing and analysis, which would be very tedious. Jeremy began to rekindle an interest in coding in R, and began writing scripts for collaborators and lab-mates to help them with their lipidomics workflows.

Inspired by the utility of the research to the broader lipidomics community, Jeremy developed an entire lipidomics open source workflow, from data-acquisition to relative quantification, working with a team of undergraduate computer programmers (most importantly Jason Cochran and Nicholas Kroeger), UF faculty Richard Yost and Timothy Garret, and researchers at the National Institute for Standards and Technology (NIST) at Hollings Marine Lab (John Bowden, Candice Ulmer, and Jared Ragland). Lipidomics tools developed during his PhD, in which he is first author or coauthor on respective manuscripts, include LipidMatch, LipidMatch Quant, LipidPioneer, LipidQC, and IE-Omics. Jeremy received his PhD from the University of Florida in the fall of 2017.

Beyond scientific research, Jeremy is very active in the Gainesville community. He is trained in life coaching by the Satvatove Institute, and has facilitated numerous 3-4 hour communication courses, and one-on-one coaching sessions, mainly for undergraduates of University of Florida. He loves to join and teach dance (especially in the tradition of Barbara Mettler), meditation, transformative communication, and loves cooking vegan dishes. He is dating Harmony Miller, midwife, founder, and business

owner of Rosemary Birthing Home, and enjoys spending time with her two boys, Rio and Cairo Ortiz.

As of 2018, Jeremy serves as a consultant for designing both proprietary and open source lipidomics software. After completing the lipidomics software tools, Jeremy plans on returning to his passion of environmental analytical chemistry, applying his knowledge acquired during his PhD on state of the art organic mass spectrometric techniques and data-handling tools to develop techniques in the field of exposomics. Specifically, Jeremy is interested in designing software and developing data-acquisition strategies for untargeted detection of anthropogenic chemicals released into the environment. These strategies would be important to detect releases in real time, and quickly assess environmental risk before the levels of persistent chemicals in the environment are raised to dangerous levels. This is important, as once persistent chemicals are released into the environment, the successful remediation of the environment to pre-release conditions is often cost-prohibitive.