# DEVELOPMENT AND ANALYSIS OF LIPIDOMICS PROCEDURES FOR THE CAUSAL INVESTIGATION OF ALZHEIMER'S DISEASE

by

Max Richard Koch

A thesis submitted in partial fulfillment
of the requirements for the degree

of

Master of Science

in

Biochemistry

MONTANA STATE UNIVERSITY
Bozeman, Montana

December 2022

# ACKNOWLEDGEMENTS

I would like to thank my mentor and advisor Dr. Edward Dratz for providing guidance and support during my graduate career. Additionally, I would like to thank the various members of the Dratz lab who taught me much about navigating the academic environment as well as the undergrads I have worked who all worked hard to progress the projects in this work. Lastly, I would like to thank my friends and family without whom I would never have made it this far.

TABLE OF CONTENTS

TABLE OF CONTENTS CONTINUED

# LIST OF TABLES

Table Page

# LIST OF FIGURES

Figure Page

# ABSTRACT

Uncovering sets of molecular features which cause a healthy metabolic state to transition to one of disease, requires extensive experimentation and often presents a difficult analysis. In the case of neurodegenerative diseases, such as Alzheimer's Disease, simply obtaining suitable samples can be a challenging endeavor. Many current "Omics" techniques excel at profiling a vast array of molecules, such as water-soluble metabolites, lipids, and proteins, in order to compare groups of samples from healthy and diseased organisms. Such approaches primarily use various associations between molecules and disease to identify biomarkers. However, these "omics" experiments frequently result in intriguing biological hypotheses, but to date have rarely provided mechanistic explanations. How then, can mechanistic explanations be recovered from metabolite or lipid profile data? In our work, we applied these methods to 6 Alzheimer's diseased brain samples and 6 age matched controls. When analyzed via mass spectrometry, lipids which differed significantly between control and disease were identified, but this information was not able to provide mechanistic insight. The beginning of any "omics" based experiment starts with the extraction of the desired molecules. In order to assess the efficiency of three different lipid extraction methods, a lipid standard was extracted from a matrix composed of rat liver tissue and analyzed by mass spectrometry. The classic Folch extraction was found to be best at reproducibly extracting a wide range of lipids. Several of the lipids identified from human brains showed oxidative damage. Lastly, 5 statistical measures of dependence and 3 network algorithms were investigated for their ability to reconstruct mechanistic relationships in a dynamic model of arachidonic acid metabolism. Many of the metabolites of arachidonic acid are oxidation products. Under conditions of high noise and relatively few samples, standard measures of correlation, such as Pearson's correlation, Spearman's correlation and Kendall's Tau were found to perform the best. Metrics which incorporate nonlinear metabolic relations and network algorithms were found to be applicable, when sample size is large and the signal to noise ratio is close to l.

# CHAPTER ONE

## INTRODUCTION

In 1901, the German psychiatrist Alois Alzheimer encountered an unusual patient, a 51-year-old woman who was plagued by paranoid delusions, such as a cheating spouse or someone plotting her death. She was frequently bewildered in her own home. She shuffled objects around and often concealed them. When shown items, she would accurately name them, yet quickly forget what she had seen. Her conduct became increasingly unpredictable as the sickness advanced, and her memory deteriorated to the point that she was unable to grasp the attempts of the attending physician to check her. Dr. Alzheimer observed her case until her death in 1906. In 1907, he published the first documented description of Alzheimer's Disease[1].

Over the years Alzheimer's Disease has been increasingly diagnosed and in 2021, the occurrence of Alzheimer's disease has increased to impact 6.2 million people in the United States alone. By 2060, it is projected to increase to more than 13 million individuals in the US[2]. According to the most recent year for which statistics are available, 121,499 persons died from Alzheimer's disease in 2019[3]. Between 2000 and 2019, the number of deaths from Alzheimer's disease, as recorded on death certificates increased by 145.2%[3]. For comparison deaths from heart disease, which is the number one cause of death decreased 7.3%. The increase of death listing Alzheimer's as the underlying cause of death is likely a genuine increase, as the population ages. Furthermore, the care required by people with Alzheimer's disease and Alzheimer's-like

dementia imposes a very significant financial burden on both caregivers and society. In 2020 dollars, the total lifetime cost of dementia care was calculated at $373,527[4] per person.

## Characterization of Alzheimer's Disease

Alzheimer's disease (AD) causes progressive memory loss and structural atrophy of the brain. Oxidative stress[5], neuroinflammation[6], and aberrant lipid metabolism have all been linked to the development of Alzheimer's disease[7,8]. The only way to definitively diagnose Alzheimer's disease is by a postmortem brain biopsy, however imaging of A-β deposits and tau neurofibrillary tangles in the brain are highly correlated with AD status.

90-95% of Alzheimer's disease patients exhibit symptoms after the age of 65. The remaining 5-10% develop symptoms before the age of 65 and the disease can appear as early as the mid-forties[8,9]. The cases before 65 are classified as early-onset Alzheimer's disease. The pathophysiology of early-onset Alzheimer's disease is very similar to that of late onset Alzheimer's disease, although the causal agents are likely to be at least somewhat different.

In addition to the presence of amyloid-beta plaques in the brain, the presence of amyloid-beta (A-β) peptides in the CNS is another feature of Alzheimer's disease that distinguishes it from other forms of dementia. Different forms of A-β: amyloid 40 and 42 (A-40/42) peptides are cleavage products of the amyloid precursor protein (APP)[10-12]. High levels of these peptides are thought to be the major cause of A-β plaque formation in the brain, which arise as a result of both overproduction and accumulation of amyloid precursor protein (APP) cleavage products[10–12]. These primary A-β isoforms are

generated by the γ-secretase[13] enzyme's imperfect cleavage of A-β at the C-terminus, resulting in the synthesis of A-β peptides. Early onset Alzheimer's disease has been associated with mutations in the catalytic components of γ-secretase, presenilin 1, and presenilin 2, as well as point mutations in APP[14,15]. Among the numerous genetic risk factors implicated in the pathogenesis of late-onset Alzheimer's disease, decreased A-β clearance[16] and the ε4 allele of the apolipoprotein E (APOE)[8] gene are the most potent risk factors. The prevalence of the ε4 allele of the APOE gene is especially high in patients with late-onset Alzheimer's disease. Despite this, most animal models are based on γ-secretase changes and point mutations in APP, that are seen in early-onset Alzheimer's disease[17]. A-β peptides complete with insulin for decomposition by the insulin degrading enzyme (IDE) which may be and maybe a pathway for insulin elevation in type 2 diabetes being a risk factor for Alzheimer's Disease[18].

A substantial portion of Alzheimer's disease research has concentrated on early detection rather than late-stage therapy. Treatment in the later stages of Alzheimer's disease has had minimal success, even though a large number of treatments have been tested[19,20]. There is widespread agreement by investigators that treating the disease in the early stages, before symptoms appear, has the greatest potential. However, this has not yet been shown[20,21]. Because the pathogenesis mechanism is not yet clear, Alzheimer's disease is difficult to diagnose in its early stages. However, several biomarkers have been found to clearly demonstrate that the disease begins well before the onset of symptoms[22]. Neuroimaging biomarkers are capable of detecting significant increases in brain atrophy 13 years before expected symptoms and significant deposition of A-β 22 years before

expected onset[22]. Tau phosphorylation state changes, specifically p-tau217 and p-tau181, can be seen with initial increases in aggregate A-β as early as two decades before the development of aggregated tau pathology[23]. However, the imaging techniques needed to recognize early changes in the brain are extremely expensive, not widely available, and are thus not suitable for wide application as screening for early diagnosis

## The Multi-Omics Approach

It is hard to over emphasize the metabolic intricacy of Alzheimer's disease[25]. In recent years there has been particular interest in the relation of lipid metabolism to Alzheimer's disease as many lipids have been linked to the disease in some fashion, as explained below. Patients with Alzheimer's disease have been found to have lower levels of phospholipids and sphingolipids, which has been correlated to an increase in ceramide[24]. A study involving 1912 participants demonstrated a positive connection between AD and species of n-acylated ceramides containing 18:0, 20:0 and 24:1 fatty acids[25]. Negative or neutral interactions were identified between species containing 22:0, 24:0, and 26:0 fatty acids[25]. The same study revealed links between diacylglycerol and triacylglycerol phosphatidylethanolamines (PE) and Alzheimer's disease[25]. PE species have been found to correlate closely with elevated levels of TGs and dyslipidemia[25]. These correlations closely parallel those observed in type 2 diabetes[26] while both dyslipidemia and insulin resistance have been identified as significant risk factors for AD[27,28].

An analysis of the serum metabolome revealed that AD patients have lower levels of ether sterols, PCs, SMs, and phospholipids than healthy controls[29]. Kim et al.,

identified 14 significantly elevated lipids in the plasma of AD including PE, diacylglyceride (DAG), triglyceride (TG) and ceramide[30]. TG 50:1, DAG 18:1_18:1, and PE 36:2 lipid species were found to be highly correlated with the degree of brain atrophy[30]. Trushina et al. discovered alterations in sphingolipids as well as cholesterol transport when comparing both plasma and cerebral spinal fluid of AD positive individuals to controls via untargeted LCMS based metabolomics[31]. Sphingolipids in the blood have been shown to influence the activity of γ-secretase and ceramides have been discovered that are capable of stabilizing the β-site of BACE-1[32], which is responsible for the early cleavage of A-21 peptide to generate biomarker peptides.

It is becoming increasingly clear that Alzheimer's disease is the result of a complex interplay between genes, proteins, metabolites, and lipids. A snapshot of this complexity is revealed in further detail by a comprehensive map of signaling pathways that have been implicated in the illness, which was generated by Mizuo et al[33]. This map currently contains 1347 molecules and 1070 reactions! It follows that any description of Alzheimer's disease sufficient to design effective interventions most likely must incorporate a complex interplay between a variety of pathways, across the span of human metabolism.

Approaches exist to examine many molecules in a particular class simultaneously over the course of a single tissue extraction and analysis. These approaches are typically referred to using the "omics" suffix. Of these methods, genomics, transcriptomics and proteomics have been applied with the greatest frequency. However, in recent years metabolomics and lipidomics have gained much attention. These "omics" methodologies

hold considerable promise for uncovering new systems-based metabolic and mechanistic information about Alzheimer's disease, as well as in a wide range of other health and disease systems[34–36].

As the interest in systems-based analysis grows, so to do the number and specificity of the methods being applied. Many such -omics methods now exist to analyze multiple classes of molecules from samples. When combined with LCMS analysis it is very feasible to profile a wide range of molecules across multiple molecular classes. This has led to so-called “multi-omics” approaches as a natural extension, which examine not only multiple classes of molecules, such as genes, proteins and metabolites, but also seeks to profile the interactions between these classes. Multi-omics approaches are particularly advantageous for the characterization of complex biological outcomes and have proven promising in a variety of systems[37–39]. However, the majority of the combined omics literature focuses primarily on combining genomics and transcriptomics with either proteomics or metabolomics.

One of the main considerations when attempting to profile analytes with widely different molecular properties is how to obtain these different types of analytes from the sample of interest. Biological samples, particularly those derived from human disease states, can be difficult to obtain in large quantities and/or prohibitively expensive. Consequently, sample amounts are frequently very small, ranging from a few milligrams to as little as a few micrograms in weight, depending on the circumstances. As a result of the high level of complexity and often limited resources available, even a modest increase in analysis efficiency and/or throughput can have a significant impact on both improving

the overall quality of the results obtained and reducing the amount of time spent on the analysis process.

When only one class of molecules are to be examined: metabolites, nucleic acids, lipids or protein, the desired, extraction methods are readily available and simple[40]. When more than two of these classes of molecules are to be targeted, it is generally necessary to design a biphasic extraction and/or several sets of complicated filtration procedures. Obtaining lipids, which are the least polar of molecules generally profiled, makes a biphasic extraction a necessity.

When attempting to profile lipids, two main extraction methods are commonly used; Folch[41] and Bligh-Dyer[42]. Both of these extractions employ a mixture of chloroform and methanol, followed by the addition of water to induce phase separation. Both of these extraction methods result in an upper layer of methanol:water and a lower, denser layer of chloroform:methanol at the bottom of the extraction vessel with a layer of protein between the phases. Most biphasic systems applied to lipids contain methanol, which is useful because it serves to disrupt hydrogen bonding and hydrophobic interactions between lipids, as well as between lipids and other biomolecules[43]. Additionally, methanol is often preferred for the extraction of polar metabolites.

Most extractions applied in multi-omics studies use a similar combination of chloroform and methanol to obtain polar metabolites from the upper methanol water layer, lipids from the lower chloroform layer and the protein material from the interphase between the layers. The polar and protein layers can also be further processed to obtain nucleic acids through various methods of filtration and precipitation. While these

methods are effective, they are far from high throughput[44,45]. Obtaining the statistical power required to examine interactions between many different variables generally requires a relatively high number of replicates, as will be discussed. This makes high throughput a near necessity when these analyses are conducted.

Most obstacles to high throughput in these extractions are a result of the chloroform. Chloroform makes the use of plastic materials problematic as plasticizers and lubricants used in their manufacture can contaminate samples. Additionally, the protein and lipid material are somewhat time consuming to acquire since the lipids are in the lower, dense chloroform phase and the partitioning of the protein to the interface between the polar and non-polar fraction prevents the formation of a pellet. In recent years, the BUME[46] and methyl-tert-butyl ether[47] (MTBE) extractions have become popular, as more high throughput alternatives to existing extractions. These latter methods employ solvent systems, butanol:MeOH / heptane:ethyl acetate and MTBE:MeOH:H2O respectively, that move the non-polar lipid-containing phase to the top of the extraction where it is more easily removed and also allows the protein to pellet when centrifuged. Both solvent systems also produce less contaminants if plastic materials are employed. However, these extractions are primarily targeted towards lipids and have yet to be properly evaluated in terms of metabolite or protein recovery. Our work has compared the effectiveness of alternative lipid extraction methods applied to tissues, as will be explained.

## Making Sense of Metabolic Complexity

The greatest challenge when profiling a wide range of molecules of any class lies in the processing and interpretation of the data itself. Large data sets generated from the -omics disciplines require the application of different types of multivariate statistics. The methods applied vary widely, depending on the class of molecules examined. The analysis of genomics data is very different from metabolomics data, for example. This is a result of the flow of information within biological systems from genes to proteins to small molecule metabolites. Since levels of these small molecule metabolites are nearly always influenced by gene and protein function, they are simultaneously the most susceptible to changes in the system, as well as the most difficult to properly analyze. Metabolomics and lipidomics data sets consist of a matrix composed of many variables measured across several biological replicates and usually utilize samples from several groups of subjects. When mass spectroscopy is used to acquire data, the output variables correspond to mass features, which appear consistently across samples and may or may not be able to be matched to molecular identifications. These molecular features are generally filtered according to fold change and significance between groups, before being passed on to multivariate analysis.

The first step in most data analyses is to apply a dimensionality reduction technique to present the multivariate data in a human readable format and to examine which features contribute most to the variance of the separation between groups. Although there are many dimensionality reduction techniques, the most commonly used in -omics is principal component analysis (PCA). In brief, PCA works by constructing

linear combinations of variables, weighted according to variance (principal components) and transforming the variables to a space defined by these principal components. These principal components can be examined to determine which variables are responsible for the greatest effect on the variance between groups. This is helpful in the identification of biomarkers, as it allows identification of features that can be leveraged to classify samples between groups, such as disease and control. Unfortunately, a feature that is suitable as a biomarker is not guaranteed to be directly related to the causal factors that produce the observed differences between the groups of interest. If the biomarker is a downstream reporter that shows a large response after a very long chain of effects, it may not be possible to identify the causal pathway from the biomarker alone.

The most ideal way to draw useful conclusions from omics data would be to place the data in the context of molecular pathways. Considering that many of the features in metabolomics and/or lipidomic data will often not be matched to molecular IDs, such a task can be more difficult than it first appears. The simplest approach to this problem is to rely only on the features with molecular IDs and what is known about the groups being compared. In this approach, the main features with a significant difference between groups are compared to a reference set taken from a pathway of interest. If many of these features match to metabolites in the pathway, then there is likely to be an association between the pathway and the observed differences. In genomics this is known as enrichment and often conducted using Fisher's exact test[48].

The primary downside of this approach is the need for features with matched molecular IDs. If there are too few IDs within a dataset, enrichment may not be useful. A

commonly used variation on this approach, mummichog[49], can also incorporate significant features with missing IDs by considering all possible matches and using the list of features that were not significantly different between groups to estimate a null distribution. However, this still requires a reference pathway network to be identified.

If MS2 data is acquired, significant features which lack IDs can be matched to features with IDs, based on the similarity between the MS2 spectra in databases. Similarity between these spectra can be used to infer structural similarity, when molecular fragments are shared between the features in the MS2 data[50]. If the features with IDs are matched to pathways, this information can be used to make inferences about how the features that lack full IDs may be related, as well as providing information about the structure of the molecule implicated.

While the methods discussed thus far rely on prior knowledge about both the system of study and the features observed in the experiment, this problem can also be approached from a completely unsupervised perspective[51]. In the absence of prior information, a measure of dependency can be calculated between the individual features to assess the degree to which features share information. There are many metrics which can serve as a measure of dependency. A few of the most notable are distance correlation[52] (DC), Heller-Heller-Gorfine[53] (HHG), mutual information[54] (MI), and Hilbert–Schmidt independence criterion[55] (HISC). Any of these metrics can be used to construct a network of dependency relations between variables. However, the simplest metric, as well as the most often applied, is partial correlation.

Partial correlation is the measure of the degree of association between two variables, with the effect of a set of confounding variables removed and can be used to assess conditional independence. Conditional independence describes situations where a set of observations contains redundant information. For example, given observations A, B, and C if one can predict A with only information from C then B contributes no useful information. In this scenario it may still be possible to predict A from B even if B is unrelated to A and so there would be a dependence edge between A and B, if conditional independence is not assessed. If the goal of the analysis is to assess causal relations, conditional independence is a necessary property, as having redundant relations would only complicate the analysis without providing any useful information. Theoretically, a measure of conditional independence can be derived for any of the relations mentioned earlier. Of the measures mentioned, only HISC and MI are commonly used in this capacity.

Once a network has been constructed using a dependency relation such as partial correlation, confounding variables are removed via a statistical test for conditional independence[56]. The remaining relations form the skeleton of what can be considered a causal graph, as the relations only constitute direct effects between variables. The final step of this analysis is to determine a directionality of the relations in the graph. The resulting directed acyclic graph provides a clear chain of causality between the features in the graph[56].

These directed acyclic graphs provide extremely effective use of information from molecular identifications as well as any prior information about group differences, as the

causal relations can be compared directly with pathway networks[57]. This form of analysis is not mutually exclusive with enrichment analysis or interaction networks composed of structural relations from MS2[58]. Information from these methods can either be compared directly, incorporated into the graph, or used to inform the construction of the network by removing edges[59]. Directed acyclic graphs can readily be constructed as Bayesian networks, lending the potential to add predictive capacity to the resulting graph, making them useful for pathway modeling and design or assessment of interventions[60,61].

# CHAPTER TWO

## UNTARGETED LIPIDOMIC ANALYSIS OF HUMAN BRAIN CORTEX IN ALZHEIMER'S DISEASE REVEALS POTENTIAL BIOMARKERS

In order to investigate the relationship between lipids and Alzheimer's disease, a lipidomics pilot study was designed. Twelve Human cerebral cortex brain tissue samples were obtained from the Victorian Brain Bank, Melbourne AU. Six of the samples were age-matched controls that did not meet the criteria for inclusion in AD diagnosis (NIA Reagan criteria)[62]. The other six samples met the standard criteria for AD diagnosis. The goal of this work was to extract the lipid material from these samples and identify the lipid species that change significantly between the AD and aged matched controls, using mass spectrometry.

Table 1. Sample information for the 6 AD and 6 age matched control samples obtained from Victorian Brain Bank, Melbourne AU. Disease status indicates whether the individual tested positive or negative for AD after death, PMI is postmortem interval which indicates the number of hours after death that the sample was taken, and Genotype indicates the variant of the ApoE alleles.

| ID | Disease Status | Gender | Age | PMI | Genotype |
|---|---|---|---|---|---|
| 05-317 | AD - | Male | 73.6 | 49 | E2/E3 |
| 06-144 | AD - | Male | 69.4 | 24 | E3/E4 |
| 06-993 | AD - | Female | 80.7 | 59 | E3/E3 |
| 07-295 | AD - | Male | 82 | 29 | E3/E3 |
| 07-635 | AD - | Male | 72.6 | 42.5 | E3/E3 |
| 08-026 | AD - | Female | 67.3 | 24 | E3/E3 |
| 07-757 | AD + | Male | 70 | 9 | E2/E4 |
| 07-835 | AD + | Female | 85.3 | 60 | E3/E4 |
| 08-007 | AD + | Male | 89.4 | 13 | E4/E4 |

Table 1 Continued

| | | | | | |
|---|---|---|---|---|---|
| 08-312 | AD + | Male | 72.1 | 13 | E3/E4 |
| 08-334 | AD + | Female | 102.7 | 69.5 | E3/E3 |
| 09-006 | AD + | Male | 74.6 | 30 | E3/E4 |

## Lipid Extraction

Extraction of lipid material was accomplished using a methyl-tert-butyl-ether (MTBE) method[47]. While a somewhat recent addition to the literature in comparison to traditional Folch[41] and Bligh-Dyer[42] based extractions, the MTBE method has been shown to be effective at extracting a wide range of lipids[43,47]. The MTBE method has two main advantages over more traditional methods. The first is that plastic tubes and pipet tips can be used during the extraction without risk of introducing excessive amounts of polymers, that can make the data produced by mass spectrometry difficult to analyze. The second, and even more important advantage, is that the solvent system moves the protein layer in the extraction to the bottom pellet as opposed to an interface between the polar and nonpolar layers in the Folch and Bligh-Dyer extraction methods. The avoidance of the protein interface facilitates a cleaner removal of the lipid material.

The MTBE procedure is as follows. Brain samples were ground under liquid nitrogen, using a mortar and pestle. For each sample, 15 mg of ground tissue was transferred to a 2 mL microcentrifuge tube and mixed with 240 μL of MeOH and 800 μL of MTBE. The vials were vortexed for 30s, followed by homogenization, using a Polytron Homogenizer with a 7 mm tip diameter. 200 μL of 0.15 mM ammonium acetate pH 8 was then added to induce phase separation. The vials were again vortexed and then

centrifuged at 6000 g for 15 min. The upper MTBE phase was then transferred to an autosampler vial. A second wash step of the lower phase was conducted by adding 800 µL MTBE, followed by 30 s of vortexing and 15 min centrifugation. The resulting MTBE layer was combined with the first fraction in the autosampler vial. The combined extract was then dried under nitrogen, and resuspended in 50 uL of methanol.

## LCMS

Lipids were separated using an Agilent 1290 UHPLC system, coupled to an Agilent 6538 Q-ToF MS. The UHPLC was equipped with a Zorbax Eclipse XDB-C18 (2.1x150mm 1.8µm) RRHD column, supplied by Agilent. The mobile phase solvent A contained 0.1% formic acid in water and B was chloroform:methanol:isopropanol 1:2:4 (v/v/v). A linear solvent gradient of 65%B to 95%B at 300 µL/min was used to elute lipids over 12 minutes, followed by a 95%B hold time of 2 minutes. The Agilent 6538 was run in alternating positive and negative mode over a mass range of 150 m/z to 1700 m/z. The MS capillary was set to 3500 volts, drying gas to 4L/min, nebulizer 45 psi, fragmentor 45 volts, and skimmer set to 500 volts. Data was collected at a rate of 1Hz from 1 minute to 15 minutes. MS2 data was collected using iterative MS2[63] in positive mode with a mass tolerance of 20 ppm and collision energy of 30 eV.

## Analysis of MS data

Preprocessing of the raw data was conducted using MS-DIAL[64]. In the data collection tab MS2 tolerance was set to 0.05 Da. The MS1 mass range was set to begin at 300 Da and MS2 mass range was set to begin at 50. Under the MS2 deconvolution tab,

the MS2 abundance cut off amplitude was set to 10. For identification, accurate mass tolerance was set to 0.01 Da for MS1 and 0.05 Da for MS2, with a score cut off of 80%. Features were removed based on blank information with the fold change between sample and blank set to 5 under alignment. All other options were left at the default settings.

Lipid mass feature intensities were normalized across samples, using the total ion chromatogram (TIC) intensity in MS-DIAL, followed by normalization by sample mass in Python. Analysis of variance (ANOVA) and principal component analysis (PCA) were conducted in Python, using the statsmodels[65] and Scikit-learn[66] packages. Factor analysis, followed by varimax rotation, was also conducted in Python using the factor_analyzer package. Plots were created using the matplotlib and seaborn Python packages.

When the full dataset is taken into account, 7346 features were found in positive mode and 4197 in negative mode, as a result of the MS-DIAL analysis. Of these features 649 of the positive mode features and 163 of the negative mode features were identified via MS2. When mass features were compared with principal component analysis, followed by factor analysis, 59 lipid species in positive mode and 77 in negative mode were identified as having expression levels capable of distinguishing AD from control samples. Both of the principal component plots in Figure 1 show good separation between the AD and age matched control samples, indicating that the disease samples could be reliably distinguished with the significant features.

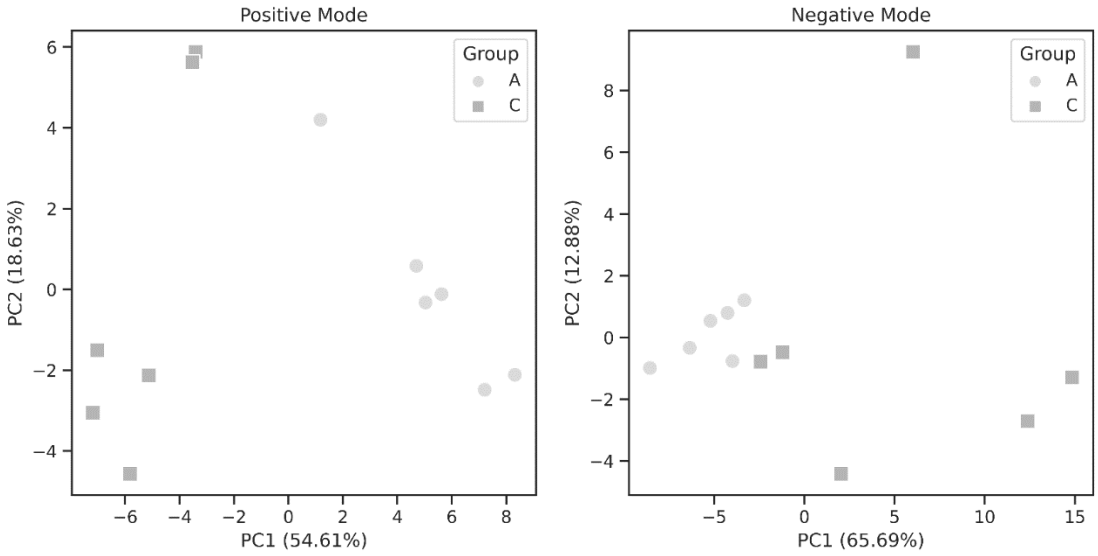


Figure 1. Principal Component Analysis (PCA) created using the intensities of the most significant mass features as determined by ANOVA and effect size in positive and negative mode. The circles labeled A, correspond to the 6 AD samples and the squares labeled C correspond to the 6 age matched control samples.

Unfortunately, many of these lipid feastures that differed most significantly between AD and control could not be identified with the MS data available and very few matched MS2 spectra in databases. Of the significant lipids that have an identification in Table 2 only SM 43:5;3O|SM 25:3;2O(FA 18:1) in positive mode and PE 40:5|PE 18:1_22:4 in negative mode were matched to MS2 spectra, all others rely exclusively on the MS1 precursor mass.

Table 2: Significant lipids with P-values < 0.05 and η2 > 0.5. The P-values were obtained from ANOVA. F1 and F2 represent the two factors obtained form factor analysis. Fold change was calculated as the ratio of the difference between intensity in the disease sample and the control sample value to the intensity in the control sample.

| Metabolite name | Mode | Ave. Mz | Fold Change | η2 | P-Value | F1 | F2 |
|---|---|---|---|---|---|---|---|
| AHexCer (O-17:0)12:0;2O/12:0;O | Pos | 812.66 | -0.67 | 0.61 | 0.03 | -0.48 | 0.48 |
| AHexCer (O-18:1)12:1;2O/14:0;O | Pos | 868.69 | 9.31 | 0.76 | 0.01 | 0.88 | -0.40 |
| AHexCer (O-20:4)12:1;2O/12:0;O | Pos | 862.64 | -0.89 | 0.78 | 0.01 | -0.31 | 0.72 |
| Cer 12:0;2O/29:1 | Pos | 618.62 | -0.39 | 0.62 | 0.01 | -0.02 | 0.72 |
| Cer 12:0;2O/32:1 | Pos | 660.67 | -0.78 | 0.64 | 0.03 | -0.59 | 0.52 |
| Cer 12:0;2O/38:2 | Pos | 782.73 | 2.31 | 0.72 | 0.00 | 0.97 | -0.17 |
| Cer 12:2;2O/43:12;2O | Pos | 820.62 | -0.55 | 0.62 | 0.03 | -0.09 | 0.72 |
| Cer 13:1;2O/44:13;2O | Pos | 848.66 | -0.40 | 0.60 | 0.01 | -0.08 | 0.66 |
| Cer 13:1;3O/30:5;(2OH) | Pos | 668.57 | -0.91 | 0.69 | 0.02 | -0.72 | 0.59 |
| Cer 17:3;2O/38:6 | Pos | 798.71 | 2.99 | 0.73 | 0.00 | 0.96 | -0.20 |
| Cer 18:0;3O/38:0;(2OH) | Pos | 862.86 | -0.91 | 0.61 | 0.04 | -0.15 | 0.86 |
| DG 43:10 | Pos | 725.51 | -0.83 | 0.65 | 0.03 | -0.11 | 0.83 |
| DG 45:1 | Pos | 771.68 | -0.42 | 0.65 | 0.03 | -0.03 | 0.71 |
| DGCC 9:0_22:2 | Pos | 710.55 | -0.71 | 0.74 | 0.01 | -0.09 | 0.84 |
| DGDG O-9:0_26:2 | Pos | 934.68 | 10.05 | 0.75 | 0.01 | 0.88 | -0.40 |
| Hex2Cer 17:1;2O/24:5 | Pos | 950.65 | 2.06 | 0.77 | 0.01 | 0.82 | -0.22 |
| MGDG O-9:0_28:2 | Pos | 800.66 | -0.49 | 0.62 | 0.03 | -0.41 | 0.59 |
| NAGlySer 13:0;O | Pos | 375.25 | -0.48 | 0.66 | 0.03 | -0.13 | 0.64 |
| NAGlySer 17:0;O(FA 28:2) | Pos | 850.72 | 12.18 | 0.75 | 0.01 | 0.87 | -0.39 |
| PC 40:0 | Pos | 868.67 | -0.91 | 0.63 | 0.03 | -0.42 | 0.62 |
| PC O-41:6 | Pos | 834.64 | -0.84 | 0.63 | 0.03 | -0.17 | 0.85 |
| PC O-43:4 | Pos | 866.69 | 1.69 | 0.65 | 0.00 | 0.89 | -0.11 |
| SM 37:7;3O | Pos | 749.52 | -0.52 | 0.65 | 0.03 | -0.21 | 0.50 |
| SM 43:5;3O\|SM 25:3;2O(FA 18:1) | Pos | 837.62 | -0.63 | 0.60 | 0.02 | -0.16 | 0.47 |
| AHexCer (O-14:1)12:2;2O/13:1;O | Neg | 792.56 | -0.56 | 0.76 | 0.00 | 0.73 | 0.58 |
| Cer 22:0;3O/38:0;(2OH) | Neg | 994.94 | -0.50 | 0.67 | 0.02 | 0.61 | 0.72 |
| CerP 12:1;2O/15:1 | Neg | 516.35 | -0.39 | 0.61 | 0.02 | 0.88 | 0.19 |
| DGGA 8:0_21:4 | Neg | 693.42 | -0.74 | 0.72 | 0.01 | 0.64 | 0.65 |

Table 2 Continued

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| HexCer 17:0;2O/18:3;O | Neg | 784.56 | -0.73 | 0.62 | 0.03 | 0.89 | -0.11 |
| HexCer 17:0;2O/18:5;O | Neg | 780.53 | -0.64 | 0.60 | 0.03 | 0.23 | 0.86 |
| LPC 16:2 | Neg | 550.31 | -0.63 | 0.69 | 0.02 | 0.53 | 0.75 |
| LPE 20:3 | Neg | 502.29 | -0.37 | 0.67 | 0.02 | 0.93 | 0.22 |
| LPI 17:2 | Neg | 581.27 | -0.47 | 0.67 | 0.00 | 0.68 | 0.56 |
| LPS-N (FA 4:0)20:5 | Neg | 612.29 | -0.66 | 0.60 | 0.04 | 0.96 | -0.17 |
| MGDG 13:0_22:5 | Neg | 821.54 | -0.50 | 0.64 | 0.02 | 0.14 | 0.92 |
| PC O-20:3_22:6;2O | Neg | 932.60 | -0.70 | 0.65 | 0.02 | -0.14 | 0.79 |
| PC O-8:0_7:0;1O | Neg | 556.32 | -0.60 | 0.70 | 0.01 | 0.38 | 0.87 |
| PE 15:3_36:10 | Neg | 930.60 | -0.67 | 0.61 | 0.03 | -0.18 | 0.76 |
| PE 20:4_20:3;1O | Neg | 804.52 | -0.61 | 0.66 | 0.02 | 0.11 | 0.96 |
| PE 22:5_18:1;1O | Neg | 806.54 | -0.64 | 0.61 | 0.03 | -0.14 | 0.84 |
| PE 40:5\|PE 18:1_22:4 | Neg | 792.56 | -0.55 | 0.62 | 0.02 | 0.07 | 0.95 |
| PE O-22:4_28:7 | Neg | 906.64 | -0.62 | 0.75 | 0.00 | 0.68 | 0.61 |
| PE-Cer 13:1;2O/18:3;O | Neg | 627.41 | -0.56 | 0.65 | 0.00 | 0.80 | 0.40 |
| PG 14:0_18:1;4O | Neg | 783.46 | 13.02 | 0.70 | 0.01 | -0.44 | -0.34 |
| PG 6:0_16:1 | Neg | 579.33 | -0.75 | 0.67 | 0.02 | 0.92 | 0.01 |
| PG 6:0_32:4 | Neg | 797.54 | -0.53 | 0.63 | 0.02 | 0.23 | 0.86 |
| PG 9:0_20:3;1O | Neg | 689.40 | -0.55 | 0.62 | 0.01 | 0.79 | 0.40 |
| PG O-8:0_8:0 | Neg | 483.28 | -0.77 | 0.64 | 0.02 | 0.55 | 0.67 |
| PI 8:0_12:0;1O | Neg | 657.33 | -0.28 | 0.65 | 0.02 | 0.25 | 0.79 |
| PI 8:0_17:1;2O | Neg | 741.38 | -0.71 | 0.60 | 0.03 | 0.81 | 0.00 |
| PI O-8:0_10:0 | Neg | 599.33 | -0.72 | 0.69 | 0.02 | 0.61 | 0.68 |
| PS 8:0_14:1;2O | Neg | 624.31 | -0.46 | 0.61 | 0.03 | 0.69 | 0.57 |
| PS O-8:0_16:4 | Neg | 600.33 | -0.67 | 0.68 | 0.02 | 0.59 | 0.72 |
| SL 12:2;O/12:0;O | Neg | 474.29 | -0.51 | 0.73 | 0.00 | 0.73 | 0.57 |
| SL 12:2;O/34:6 | Neg | 754.54 | -0.86 | 0.62 | 0.03 | 0.07 | 0.71 |
| SL 13:1;O/32:6;O | Neg | 758.54 | -0.71 | 0.70 | 0.02 | 0.16 | 0.92 |
| SM 12:1;2O/21:0 | Neg | 747.57 | -0.50 | 0.69 | 0.00 | 0.51 | 0.68 |
| SMGDG O-15:2_5:0 | Neg | 623.31 | -0.50 | 0.65 | 0.01 | 0.86 | 0.00 |

Several of the species identified were highly polyunsaturated, oxidized, or highly unusual lipids. There are three probable explanations for these strange lipids. First, there was significant variance in the post mortem interval among the samples which is shown

in Table 1. This could possibly explain the significantly different lipids as degraded products, resulting from changes that occurred after death. Second, these identifications could be poor or incorrect matches as a result of only being able to match the MS1 data. Lastly, the MTBE extraction that was used resulted in a surprisingly large amount of lipid material, some of which precipitated when the samples were stored after extraction. The samples were dried under nitrogen and then resuspended a second time in 100 uL of chloroform : methanol 1:1. It is possible that some degradation occurred as a result of having to dry the samples twice or that not everything was properly resuspended resulting in an altered distribution of lipids in solution. As a result of this final point and previous experimentation with different lipid extractions, several different extraction methods are tested and compared in Chapter 2 to seek a more optimum method.

# CHAPTER THREE

# COMPARISON OF THREE LIPID EXTRACTION WITH DIVERSE INTERNAL STANDARDS IN RAT LIVER TEST MATERIAL

## Contribution of Authors and Co-Authors

Manuscript in Chapter 2

Author: Max R. Koch

Contributions: Wrote and edited the manuscript. Conducted lipid extractions and MS experiments. Processed data and conducted statistical analysis.

Co-Author: Daniel Willems

Contributions: Contributed to the selection of lipid extraction methods as well as the design of the MS and chromatography methods.

Co-Author: Jason L. Burkhead

Contributions: Contributed the rat liver samples used for the analysis.

Co-Author: Edward A. Dratz

Contributions: Reviewed and edited the manuscript. Supervised the lipid extractions and MS experiments.

Manuscript Information

Max R. Koch, Daniel WIllems, Jason L. Burkhead, Edward A. Dratz

Biomedical Chromatography

Status of Manuscript:
____ Prepared for submission to a peer-reviewed journal
_X__ Officially submitted to a peer-reviewed journal
____ Accepted by a peer-reviewed journal
____ Published in a peer-reviewed journal

Biomedical Chromatography
June 2nd 2022

## Abstract

Increasing interest in lipidomics has led to several methods being used to obtain lipid material. We compared three popular lipid extraction methods, Folch, Bligh-Dyer, and BUME to assess possible differences between methods. Peak areas of 12 deuterium labeled internal standards, representing 12 distinct lipid classes, were added to the rat liver test material and the results averaged across 25 replicates for each method. Centrifuge adapters were made by 3D printing for small glass vials with teflon lined caps, that were suitable for the small sample sizes employed. Results were compared statistically, using Welch's ANOVA, followed by the Games-Howell post-hoc test. Peak areas were converted to concentrations by means of regression on dilutions of the standards and the repeatability coefficients produced by Bland-Altman analysis were compared. The Folch and Bligh-Dyer methods were found to agree quite well across the PC, TG, MG, and Cer standards with mean bias below 0.1 ng/μL ($p > 0.05$) and repeatability coefficients (RC) around 0.4 ng/μL. The primary differences between the methods were seen in LPC, PS, and PE, with reasonable agreeability across the other classes. Endogenous lipids extracted from rat liver were analyzed by all three methods with 25 replicates for each method. Folch produced the most endogenous lipid features identified (486) and the highest peak areas.


## Introduction

The '-omics' fields (e.g. genomics, epigenomics, transcriptomics, proteomics and metabolomics) have been increasingly utilized in recent years and a great deal of

biological information has emerged from this work [67,68]. While much of the activity in metabolomics has focused on water soluble metabolites, lipid metabolism has been linked to an increasing number of biological mechanisms and diseases in recent years [69–73]. In addition to membrane structure and energy storage, lipids are essential for many other cellular functions, including: signal transduction, second messenger activity, control of lipid rafts, and hormone synthesis [74–76]. Lipids are an extremely diverse class of compounds and as the list of diseases linked to aberrant lipid metabolism and lipid levels grows, so too does the number of analytical techniques used to investigate these molecules.

Lipidomics, as with all -omics methodologies, involves the simultaneous analysis of hundreds to thousands of analytes acquired from biological samples of interest. The biological samples may be costly and/or difficult to obtain, particularly from tissues isolated from human disease states. Therefore, the sample amounts are often quite small, sometimes just a few milligrams or even micrograms. With such a high degree of complexity and often very limited amounts, even a modest improvement in analysis efficiency and/or throughput can greatly enhance the results obtained and/or decrease the time required for analysis.

Lipids are isolated from biological materials through exposure to a mixture of organic and polar solvents, that separate into two phases and partition the lipids into the organic rich layer [77,78]. A large fraction of the lipid extraction techniques in the literature rely on foundational work of Folch, et al [41] or Bligh and Dyer [42], typically with numerous modifications. Both the original Folch and Bligh and Dyer methods rely on a solvent

system composed of chloroform, methanol, and water. Chloroform:methanol systems separate into two phases after the addition of water, resulting in an upper layer of methanol:water and a lower, more dense layer of chloroform:methanol containing the lipids. Chloroform and methanol are frequently used to extract a wide range of lipid classes from a variety of sample matrices. In particular, methanol is used in the majority of biphasic systems applied to lipids, as it serves to disrupt hydrogen bonding and electrostatic interactions between lipids or lipids and other biomolecules[43,79].

While there is little disadvantage in using methanol in these extractions, the use of chloroform introduces some complications. It can be difficult to cleanly draw the lipid extract from the bottom chloroform-rich layer through the upper layer and a protein rich interface. Additionally, plastic microcentrifuge tubes and other polymer-containing products, like syringes can contaminate the samples when exposed to chloroform. Chloroform tends to dissolve plasticizers and mold release lubricants from plastics, which can complicate mass spectral detection of some lipids. The use of glass containers and glass syringes can eliminate plastic contamination issues, however they are often more cumbersome to employ. We have found that 1 dram glass vials, with teflon lined caps, are excellent for lipid extraction of small samples. No commercial centrifuge adapters are available for the small glass vials, and we made suitable centrifuge adapters using 3D printing, as will be described.

Alternative lipid extraction procedures, such as the Matyash [47] or BUME [46] methods, have recently gained favor. The Matayash extraction employs methyl tert-butyl ether (MTBE) in place of chloroform. The BUME method, named for the use of

butanol:methanol 3:1, uses a mixture of heptane:ethyl acetate 3:1 as the organic solvent. Both the MTBE and BUME methods have shown similar extraction profiles of primary lipid classes [79,80]. Both the MTBE and the BUME methods eliminate the difficulty of removing the lipid extract from the bottom phase as the organic solvent since the lipids partition to the top layer in both methods.

Chloroform:methanol based extractions are widely adopted and are frequently modified to increase yield, lipid class specificity, or to optimize the extractions for a particular system [81–83]. Likewise, many variations of the MTBE and BUME methods exist for similar purposes. Publications focusing primarily on a particular biological system of study, often do not clearly describe what modifications have been made or show evidence of improved lipid extraction performance due to the modifications. Some investigators have evaluated the effectiveness of specific lipid class extraction techniques [84,85] and others have evaluated more general lipid extraction techniques [86–88]. With several different solvent systems becoming widely used we found it important to examine the extraction efficiency, and how many different types of endogenous lipids can be obtained from a liver test sample with the three different methods we tested.

## Materials & Methods

### Materials

Organic solvents (chloroform, methanol, 1-butanol, and ethyl acetate) were Fisher Optima, LC/MS grade. The formic acid, acetic acid, and water were Fisher Optima LC/MS grade. We found that 4ml/1 dram glass vials with teflon lined caps were highly suitable for lipid extraction at an appropriate scale, but no centrifuge adapters were

available for similar glass vials. Therefore, we used 3D printing to create custom centrifuge adaptors for these 4 mL glass vials (Kimble, P/N 60811B1) that were found to be suitable in size, solvent resistance, and strength for the lipid extraction protocols tested. The glass vials selected reliably survive the centrifugation steps in the custom 3D printed centrifuge adapters. The autosampler vials had press fit PTFE faced rubber liners (P/N 14-812-49) with no adhesive and were purchased from Fisher Scientific. Teflon lined caps were important to minimize contamination from exposure to the lipid solvents. In situations where the liner would be pierced, crimp top Viton liners were used (P/N 03-378-343), with compatible glass vials (P/N 03-391-6). When necessary, micro volume inserts were used in the autosampler vials (P/N 14-823-374).

The LIPIDOMIX EquiSPLASH lipid standard was purchased from Avanti Polar lipids. The EquiSPLASH standard contains 100 μg/mL) of each of 13 deuterated lipids; 15:0-18:1(d7) PC, 18:1(d7) Lyso PC, 15:0-18:1(d7) PE, 18:1(d7) Lyso PE, 15:0-18:1(d7) PG, 15:0-18:1(d7) PI, 15:0-18:1(d7) PS, 15:0-18:1(d7)-15:0 TG, 15:0-18:1(d7) DG, 18:1(d7) MG, 18:1(d7) Chol Ester, d18:1-18:1(d9) SM, and C15 Ceramide-d7.

Sample Collection & Preparation

Rat livers were harvested after cervical dislocation, followed by dissection into liver lobes and flash frozen in liquid nitrogen. All animal husbandry and experimental procedures were reviewed and approved by the University of Alaska Anchorage Institutional Animal Care and Use Committee under OLAW Assurance A3710–1. Rats were raised to 24 weeks with ad libitum access to Mazuri Rodent diet (PMI Nutrition, St. Louis, MO, USA) and were part of an experiment where animals were fed experimental

diets for the following 12 weeks, based on the Purified AIN76A formulation, modified for variations in target sucrose and Cu content (Custom Animal Diets, Bangor, NJ, USA), as described in Tallino et al. [89]. Liver from one dietary group, CuA (125 mg/kg Cu and 30% sucrose enriched diet) were used. At the end of the feeding trial, animals were euthanized, according to American Veterinary Medical Association guidelines and livers were harvested, dissected and frozen in liquid nitrogen. Whole liver lobes were used to average over variation of lipid content in different regions of the liver [90,91] and were ground with a pestle in a 55mm mortar in liquid nitrogen to a fine, sand like consistency. The mortar and pestle were prepared by rinsing with high purity water, followed by methanol and finally with chloroform and dried before immersing in liquid nitrogen. Six livers were pooled, and 15 mg replicates of the frozen ground liver powder were measured for analysis.

Lipid Extraction Methods

Two lipid extraction techniques were taken from the recent literature, Folch and BUME. These methods were compared to each other and to an in-house extraction procedure modified from the Bligh & Dyer extraction. A 6 µL aliquot of the lipid standard solution (Avanti EquiSPLASH Lipidomix) was added to the glass vials or microcentrifuge tubes, using a Hamilton Syringe (P/N 1750), and dried under nitrogen prior to addition of 15mg liver powder. Each extraction method was carried out on 25 independently prepared samples, taken from the same pooled liver sample that had been ground under liquid nitrogen.

*Folch* [41]*:* 15 mg of ground tissue was transferred to a 1 dram glass vial containing dried standards and mixed with 960 µL chloroform:MeOH (2:1, v/v). The vials were vortexed for 30 s and 240 µL H2O was added to produce a solvent ratio of 8:4:3 (v/v/v) CHCl3:MeOH:H2O. The sample was homogenized using a polytron homogenizer with a 7 mm tip diameter and centrifuged at 6000 g at 4 °C for 15 min. The lower chloroform phase was transferred to an autosampler vial using a glass Hamilton syringe. A second wash step of the upper phase was conducted by adding 640 µL chloroform followed by 30 s of vortexing and centrifugation. The resulting chloroform layer was combined with the first fraction in the autosampler vial.

*Modified Bligh & Dyer* [42]*:* 15 mg of ground tissue was transferred to a 1 dram glass vial containing dried standards and mixed with 900 µL chloroform:MeOH (1:2, v/v). The vials were vortexed for 30 s and the sample was homogenized using a polytron homogenizer with a 7 mm tip diameter. 300 µL chloroform was added, followed by 540 µL of 20 mM KPHO4 pH 8 creating a solvent ratio of 2:2:1.8 (v/v/v) CHCl3:MeOH:H2O. The vials were vortexed for 30 s, then centrifuged at 6000 g at 4 °C for 15 min, and the lower chloroform phase was transferred to an autosampler vial. A second wash step of the upper phase was conducted by adding 600 µL chloroform followed by 30 s of vortexing and centrifugation. The resulting chloroform layer was combined with the first fraction in the autosampler vial.

*BUME* [46]*:* 15 mg of ground tissue was transferred to a 2 mL plastic microcentrifuge tube containing dried standards and mixed with 500 µL of 3:1 (v/v) Butanol : MeOH. 500 µL 3:1 (v/v) heptane : ethyl acetate was added and the vials were vortexed for 30 s. The

sample was homogenized using a polytron homogenizer with a 7 mm tip diameter and 500 µL 1% acetic acid was added to induce phase separation. The vials were vortexed, centrifuged at 6000 g at 4 °C for 15 min, and the upper phase was transferred to an autosampler vial. A second wash step of the lower layer was conducted by adding 500 µL of 3:1 (v/v) heptane : ethyl acetate followed by 30 s of vortexing and centrifugation. The resulting upper phase was combined with the first fraction in the autosampler vial.

After extraction, the organic fractions were dried under a gentle stream of N2 and resuspended in 60 µL MeOH prior to analysis. 10 µL MeOH was removed from each sample and pooled to create a quality control (QC) sample.

LC-MS

Lipids were separated and detected using an Agilent 1290 UPLC system, coupled to an Agilent 6538 Q-ToF MS. The UPLC was equipped with a Zorbax Eclipse XDB-C18 (2.1x150mm 1.8µm) RRHD column, supplied by Agilent. The mobile phase solvent A, contained 0.1% formic acid in water and B was chloroform:methanol:isopropanol 1:2:4 (v/v/v). A linear solvent gradient of 65%B to 95%B at 300 µL/min was used to elute lipids over 12 minutes, followed by a 95%B hold time of 2 minutes. The Agilent 6538 was run in positive and negative mode over a mass range of 150 m/z to 1700 m/z. The MS capillary was set to 3500 volts, drying gas to 4L/min, nebulizer 45 psi, fragmentor 45 volts, and skimmer set to 500 volts. Data was collected at a rate of 1Hz from 1 minute to 15 minutes. MS2 data was collected using iterative MS2 in positive mode with a mass tolerance of 20 ppm and collision energy of 20 eV.

MS Data Analysis

MS data were initially processed by Agilent MassHunter Qualitative Analysis to extract peak areas for the internal standards. Raw data files were centroided and converted to a mzML format, using MSConvert. The mzML files were processed and MS2 spectra were matched to known lipids, using the internal libraries (LipidBlast [92], GNPS [93], and MassBank [94]) of MS-DIAL [95] with a mass tolerance of 0.01 Da and an identification score cutoff of 80%. The resulting peak areas obtained for the internal standards and the output from MS-DIAL were normalized to slight variations in sample weights.

Statistical analysis was conducted in python using the scipy, sklearn, statsmodels, and pingouin libraries. Levene's test was used to assess heteroscedasticity (unequal variance between groups). Since most lipid classes tested positive for heteroscedasticity between groups ($p < 0.0001$), Welch's ANOVA was applied, followed by the Games-Howell post-hoc test. For ease of interpretation, concentrations of the lipids detected were estimated from a dilution series of the standard. A logarithmic function, $y = m\log(x+a)$, was fit to data produced from a dilution series of the standard. The comparisons between groups were assessed by applying Bland-Altman analysis to the resulting concentrations [96]. The MS-DIAL output was filtered to contain only lipid IDs which appeared in all extraction methods for comparison of the methods. Individual lipids identified were grouped into major lipid classes and the peak areas were combined and averaged. All figures were created in python using the matplotlib and seabron libraries.

## Results and Discussion

### Intensity of Internal Standards in the BUME, Bligh-Dyer, and Folch procedures

A number of statistical tests were applied to evaluate differences between the methods. Initially, one-way ANOVA, followed by Tukey's post-hoc test was applied to examine differences between the methods. However, examination of the Bland-Altman plots produced in subsequent analyses revealed a clear trend when comparing individual methods. This is generally caused by unequal variance between groups (heteroscedasticity). Since homosecedasticity is a key assumption of ANOVA, Levene's test was applied. Several lipids were found to violate the homosecedasticity assumption, primarily LPE, PS, MG, DG, and TG ($p < 0.05$). Welch's ANOVA and the Games-Howell post-hoc test were applied in place of one-way ANOVA and Tukey's post-hoc test. The p values resulting from the Games-Howell post-hoc test can be found in Table 1, along with bias (mean difference) and repeatability coefficients from Bland-Altman analysis in units of concentration.

Table 3: Mean difference of peak areas (bias), within subject standard deviations, repeatability coefficients for lipids in the internal standard in units of ng / μL. SM and Cholesterol Ester are not shown, as an $R^2$ above 0.7 could not be achieved when conducting regression. Games-Howell p-values are also reported.

| Lipid | Method 1 | Method 2 | Mean Diff (Bias) | Within-Subject SD | Repeatability Coefficient (RC) | Games-Howell pval |
|---|---|---|---|---|---|---|
| C15 Ceramide-d7 | BUME | Folch | 0.00 | 0.09 | 0.2 | 0.900 |
| C15 Ceramide-d7 | Bligh-Dyer | Folch | -0.07 | 0.13 | 0.4 | 0.104 |
| C15 Ceramide-d7 | BUME | Bligh-Dyer | 0.06 | 0.09 | 0.3 | 0.019 |
| DG 15:0-18:1(d7) | BUME | Folch | 0.02 | 0.03 | 0.1 | 0.248 |
| DG 15:0-18:1(d7) | Bligh-Dyer | Folch | 0.00 | 0.04 | 0.1 | 0.900 |
| DG 15:0-18:1(d7) | BUME | Bligh-Dyer | 0.02 | 0.05 | 0.1 | 0.371 |

Table 3 Continued

| | | | | | | |
|---|---|---|---|---|---|---|
| LPC 18:1(d7) | BUME | Folch | -1.53 | 0.51 | 3.4 | 0.001 |
| LPC 18:1(d7) | Bligh-Dyer | Folch | -0.45 | 1.22 | 1.4 | 0.001 |
| LPC 18:1(d7) | BUME | Bligh-Dyer | -1.08 | 1.10 | 3.1 | 0.001 |
| LPE 18:1(d7) | BUME | Folch | -0.30 | 0.21 | 0.8 | 0.001 |
| LPE 18:1(d7) | Bligh-Dyer | Folch | -0.18 | 0.27 | 0.6 | 0.001 |
| LPE 18:1(d7) | BUME | Bligh-Dyer | -0.12 | 0.29 | 0.8 | 0.186 |
| MG 18:1(d7) | BUME | Folch | -0.19 | 0.14 | 0.5 | 0.001 |
| MG 18:1(d7) | Bligh-Dyer | Folch | -0.05 | 0.18 | 0.4 | 0.301 |
| MG 18:1(d7) | BUME | Bligh-Dyer | -0.14 | 0.17 | 0.5 | 0.001 |
| PC 15:0-18:1(d7) | BUME | Folch | 0.08 | 0.07 | 0.2 | 0.001 |
| PC 15:0-18:1(d7) | Bligh-Dyer | Folch | -0.04 | 0.07 | 0.2 | 0.089 |
| PC 15:0-18:1(d7) | BUME | Bligh-Dyer | 0.11 | 0.09 | 0.3 | 0.001 |
| PE 15:0-18:1(d7) | BUME | Folch | 1.55 | 1.24 | 3.4 | 0.001 |
| PE 15:0-18:1(d7) | Bligh-Dyer | Folch | 1.65 | 1.32 | 3.7 | 0.001 |
| PE 15:0-18:1(d7) | BUME | Bligh-Dyer | -0.10 | 0.13 | 0.3 | 0.008 |
| PI 15:0-18:1(d7) | BUME | Folch | -2.70 | 2.77 | 7.7 | 0.001 |
| PI 15:0-18:1(d7) | Bligh-Dyer | Folch | -2.09 | 3.32 | 9.2 | 0.031 |
| PI 15:0-18:1(d7) | BUME | Bligh-Dyer | -0.60 | 2.77 | 7.7 | 0.731 |
| PS 15:0-18:1(d7) | BUME | Folch | -0.10 | 0.12 | 0.3 | 0.001 |
| PS 15:0-18:1(d7) | Bligh-Dyer | Folch | -5.49 | 7.11 | 19.7 | 0.008 |
| PS 15:0-18:1(d7) | BUME | Bligh-Dyer | 5.39 | 7.08 | 19.6 | 0.010 |
| TG 15:0-18:1(d7)-15:0 | BUME | Folch | 0.14 | 0.11 | 0.4 | 0.001 |
| TG 15:0-18:1(d7)-15:0 | Bligh-Dyer | Folch | -0.04 | 0.13 | 0.3 | 0.424 |
| TG 15:0-18:1(d7)-15:0 | BUME | Bligh-Dyer | 0.18 | 0.15 | 0.4 | 0.001 |

Average peak areas for the deuterated lipid standards recovered by the different extraction methods are compared in Figure 2 with the exception of SM. SM was excluded from this comparison since the average peak area of SM was much greater than other lipids in all three extractions and inclusion of SM made it difficult to view the results of the other lipids. SM had the highest average peak area across all extractions and significantly different in the BUME and Folch extractions as well as in Bligh-Dyer. The

lowest peak areas were observed in TG, and DG which appear to be fairly comparable between all three extractions. Of the lipids tested, only the peak area of DG was insignificant ($p > 0.05$) across all three methods, presumably because the DG did not ionize efficiently under the MS conditions used. Comparing Folch and modified Bligh-Dyer, average peak areas of PC, PI, TG, MG, Cer, and Chol-ester were also not significantly different between methods ($p > 0.05$, Games-Howell post-hoc). This can be clearly seen in Figure 2, as the average peak areas appear to be very comparable between Folch and Bligh-Dyer across most lipids. When comparing BUME and modified Bligh-Dyer, LPE, PI, and SM did not differ significantly ($p > 0.05$, Games-Howell post-hoc). In BUME and Folch, only Cer was below the cutoff for significant difference between groups ($p > 0.05$, Games-Howell post-hoc).

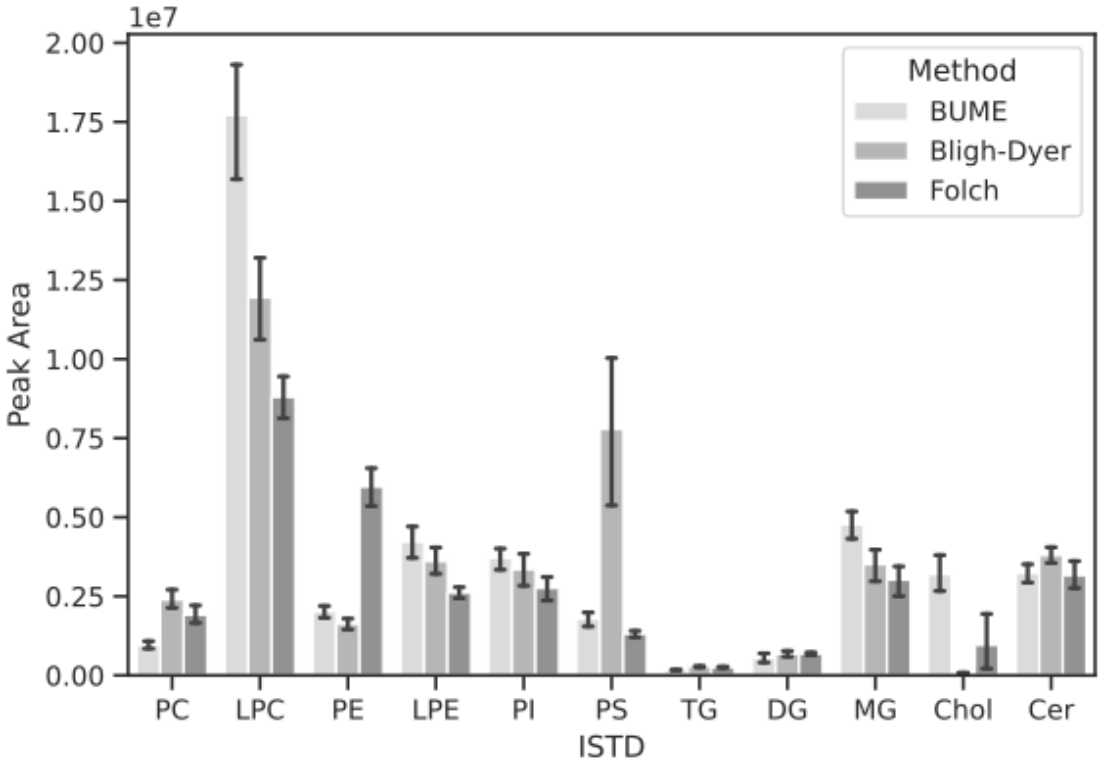


Figure 2: Comparison of average peak areas representing the internal standards across 25 replicates observed in the BUME, Bligh-Dyer, and MTBE extractions. Error bars reflect standard deviations

The average peak areas of PE and LPC is significantly higher in Folch compared to both BUME and Bligh-Dyer. PI was also significantly higher in Folch when compared to BUME. In Bligh-Dyer the peak area of PS is noticeably higher than either Folch or BUME. Additionally, SM was significantly higher in Bligh-Dyer when compared to Folch.

Bland-Altman Analysis

In order to examine the interchangeability of these methods it is useful to examine both the mean bias and repeatability coefficients (RC) produced from Bland-Altman Analysis [96]. When examined as a function of peak area it can be quite difficult to reason

about the recovery of standards between the extractions. To facilitate analysis, relative concentrations were estimated from the peak areas of the internal standard dilution series, using regression. When examined in this manner it can be seen that the Folch and Bligh-Dyer methods agree quite well across the PC, TG, MG, and Cer standards with RCs of 0.2, 0.3, 0.4, and 0.4 ng/μL respectively. This indicates that one should not expect these lipids to differ more than 0.4 ng/μL between replicates. This is also supported with a maximum bias of 0.07 ng/μL, which is not a significant difference, across these four lipids. These results also line up quite well with a visual assessment of Figure 2 and align well with reports of high similarity between results from these two methods [43,97].

When comparing LPC, LPE, and SM, Folch and Bligh-Dyer compare more favorably to each other than to BUME, in terms of mean bias and RC. This can likely be attributed to both Folch and Bligh-Dyer being chloroform:methanol based extractions. The BUME extraction also compared well to the Folch and Bligh-Dyer extractions of PC, TG, MG, and Cer with a maximum RC of 0.5 ng/μL in MG.

Although most lipid classes were similarly extracted across all three methods, a few were different enough that they should be considered when comparing analyses. The Folch method seems to result in a higher extraction efficiency of PE, while Bligh-Dyer and BUME result in improved extraction of PS and LPC, respectively. The peak area of LPC and to a lesser extent LPE is of particular note as several studies have reported lower recoveries of polar lipids in Folch extractions with higher recoveries in BUME and Bligh-Dyer[46,80,97].

Identification and Analysis of MS2 features

The endogenous lipids from rat liver are compared in Figure 3A and the number of features detected using MS-DIAL can be seen in Figure 3B. BUME and Bligh-Dryer produced a comparable number of features, 353 and 314 features respectively, while Folch produced the most features with 486. Of the 486 features found in the Folch, 151 of these were unique to Folch compared to 29 unique in BUME and 1 unique in Bligh-Dyer respectively. Out of the 286 features present in all three methods, 162 were matched to lipids in databases using the MS2 data and the internal spectral libraries in MS-DIAL (LipidBlast, GNPS, MassBank). The individual 162 lipids were grouped by class and averaged. Log normalized peak areas of the identified lipids are shown in Figure 3A. The greatest peak areas were seen in Folch across all classes except PC. BUME and Bligh-Dyer were comparable to Folch in total PC intensity but were significantly lower in all other classes.

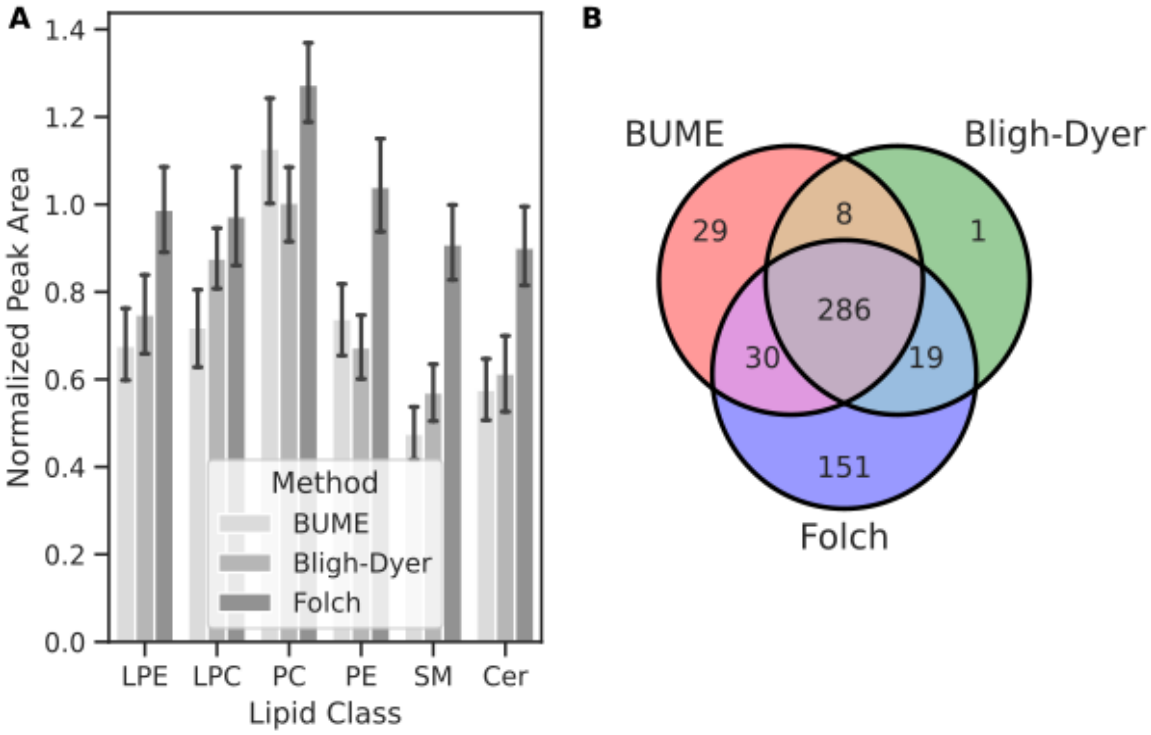


Figure 3: Comparison of endogenous lipid species extracted from rat liver. A: Average log2 normalized peak areas for lipids identified in the MS2 data, grouped by class. Error bars represent standard deviations. B: Comparison of the number of features detected by each extraction method

In addition to the BUME, Bligh-Dyer, and Folch extractions, the MTBE extraction was also tested. However, when examining the extract produced before MS analysis, there were clear signs of lipid material settling out of solution. Since this was likely to affect any comparison attempted with MTBE, we excluded it from detailed analysis.

## Conclusion

In conclusion, the traditional Folch method remains an excellent choice for wide coverage of the lipidome. The Folch extraction was the most efficient when examining the endogenous lipids extracted from the rat liver samples with a greater number of stable features and higher average peak areas. The Bligh-Dyer method serves as a suitable replacement, when high coverage is desired and may perform when aiming to quantify specific analytes. When high throughput is required it is difficult to beat the BUME method which allows for easy recovery of the lipid fraction, the use of plastic materials, high agreeability with the Folch and Bligh-Dyer methods as well as the potential for automation[46].

## Acknowledgements

This work was supported in part by National Institutes of Health Grants P20RR024237 (EAD); P20 GM 104935-06A1 (EAD) and an MSU Kopriva Graduate Fellowship (DLW). The Mass Spectrometry Core Facility at Montana State University was supported in part the by MJ Murdock Charitable Trust, 2009084 and the National Institutes of Health Grants P20RR024237; and P20GM104935-06A1. We thank Dr. Don Smith and Jesse Thomas for assistance with the mass spectrometry analysis.

## Conflict of Interest

The authors declare that there are no conflicts of interest.

# CHAPTER FOUR

# ASSESSMENT OF ALGORITHMS FOR RECOVERY OF METABOLIC PATHWAY STRUCTURE

As the end point of biological activities, the metabolome carries information regarding genetic, epigenetic, and environmental factors. This information can thus serve as a bridge between genotype and phenotype[98,99]. As such, the ability to reconstruct metabolic pathways from concentration profiles is the subject of much investigation[51,100,101].

A natural starting point in such investigations is to examine the relations between metabolite concentrations, using simple measures of dependency such as correlations to construct a network of associations[102,103]. To construct such a network, metabolites are represented as nodes and their relationships between the metabolites as edges. Such networks are conveniently described using an adjacency matrix. Both the rows and columns in an adjacency matrix contain all nodes in the network and thus is a square matrix. In the simplest form, edges connecting two nodes are represented as a 1 in the corresponding column and row of the connected row.

When correlation or another measure of association is computed for every possible pair of variables, the resulting matrix can be used as a weighted adjacency matrix[104]. In a weighted adjacency matrix, the values can vary with edges taking on the value of the statistic being used. However, this results in a fully connected network with weighted connections between all nodes in the network. In order to use such a network to form useful conclusions it is necessary to apply a threshold, by setting all values below a

threshold to zero and all values above the threshold to 1, resulting in a binary adjacency matrix[104].

While creating a network from metabolomics data is simple, aligning connections in the network with known reactions can be quite complicated. Furthermore, there has not been much investigation into the sample size required for recovering such connections or the susceptibility of the algorithms used to noise.

## Dependency Metrics and Network Algorithms

In total 8 algorithms were chosen for investigation of network connections. Of these algorithms, Pearson's correlation and Spearman's correlation were chosen for their wide use in the literature. Spearman's correlation differs from Pearson's correlation by using the ranks of the values in place of their magnitude. Both Pearson's and Spearman's correlations take values between 1 and -1.

Kendall's Tau[105] evaluates the strength of a relationship based on the pattern of concordance and discordance between pairs of observations. A pair of variables are considered concordant if the order of their ranks agree, otherwise they are considered discordant[105]. Kendall's Tau has the advantage of being a non-parametric test and can be applied in situations where the variables may not follow a Gaussian distribution[105].

Adjusted Mutual Information (AMI)[106,107] is a dependency criterion made using binning and mutual information. Mutual information is a commonly used dependency metric and has shown promise as a metric for reconstructing gene regulatory networks[108,109]. AMI uses an adjustment of the mutual information score to account for chance. It compensates for the fact that the MI is typically larger for two clusters when

there are more total clusters, regardless of whether more information is actually shared[106]. The normalized version of AMI takes values between 0 and 1.

The Hilbert Schmidt Independence Criterion (HSIC)[110] is a kernel-based independence test, that measures multivariate nonlinear relationships based on a specific kernel[110]. The default kernel is the Gaussian kernel, which utilizes the median distance as the bandwidth. This kernel is a defining kernel that ensures HSIC is a consistent test[111,112]. HSIC is related to distance correlation, a generalization of correlation that takes into account nonlinear interactions[112]. HSIC and distance correlation are identical, in that every distance correlation test is also an HSIC test, and vice versa[112,113]. The normalized version of this metric takes values between 0 and 1.

Three algorithms further remove indirect effects given a matrix of dependency metrics, the network deconvolution (ND) algorithm introduced by Feizi et al.[114], the Aracne[115] algorithm, and the context likelihood or relatedness[116] (CLR) algorithm. Each of these methods attempt to remove the effects of all indirect relationships in a graph. These algorithms have proven effective at recovering gene and protein interactions[109,114–116].

## Simulation of Metabolic Profiles with a Dynamic Model of Arachidonic Acid Metabolism

In order to create a suitable set of data for testing the ability of dependency metrics to recover metabolic relations, a dynamic model of arachidonic acid (AA) metabolism was taken from the literature[101] in order to simulate profiles of metabolite concentrations. This model was chosen due to its use in a similar capacity in the original

publication as well as the model being of adequate size. The model is composed of ordinary differential equations describing the synthesis, conversion and degradation of the 83 metabolites in the model. Enzymatic reactions are described using Michaelis-Menten kinetics, while non-enzymatic reactions are described by simple mass action laws.

In order to simulate different metabolite profiles that mimic the differences between individuals, equilibrium constants for each reaction were sampled from a uniform distribution. The maximum of this distribution was defined as the 110% of the reference constant and the minimum was defined as 90% of the reference constant. Using the Tellurium Python package, a dynamic simulation over the course of 100 hours was conducted for each of the 1000 sets of constants[117,118]. When using the optimal parameters, provided with the model, we found that the system achieved a steady state after around 70 hours. Since differences between metabolite concentrations are likely to be much greater at non-steady state than at steady state, the strength of dependency metrics will be affected by the proximity of the system to steady state. To account for this, metabolite concentrations at 10 and 90 hours were saved for downstream analysis. These time points were chosen to remain consistent with the study which provided the model[101] and to provide a comparison between steady state (90 hrs) and non-steady state (10 hrs).

Experimental metabolomic data, often generated via mass spectrometry or nuclear magnetic resonance, is subject to some level of noise as a result of the measurement process. In order to account for noise, Gaussian uncorrelated noise with zero mean was

added to the data sets. To investigate the effect of such noise on the ability of the dependence metrics to recover genuine relationships between metabolites, the data sets were treated with 11 levels of noise, ranging from a signal to noise ratio (SNR) of 1% to 99% in 10% steps The signal-to-noise ratio is the ratio of signal power to the background noise, defined here as the mean square of metabolite concentration or peak height divided by the standard deviation of peak height.

## Performance Assessment

In order to assess the performance of different methods for reconstructing the metabolic network defined by the AA model, the receiver operating characteristic (ROC) was used. The ROC is a plot of true positive rate vs the false positive rate. The true positive rate is the ratio of true positives to the combination of true positives and false negatives. Here, true positives are the edges in the AA model that the algorithm correctly predicted using either the steady state or non-steady state concentrations, and false negatives are the edges in the AA model the algorithm failed to predict. The false positive rate is the ratio of false positives to the combination of false positives and true negatives. In this case, false positives are the edges predicted by the algorithm not found in the AA model and true negatives are the edges that the algorithm failed to recover, that are also not present in the AA model.

The results of the ROC analysis can be summarized as a single metric by calculating the area under the curve (AUC). An AUC close to 1 indicates a good classifier or in this case that the majority of the edges were correctly predicted by the algorithm. An AUC close to 0.5 indicates that correct edges are only recovered randomly.

Constructing an ROC curve requires varying a threshold for an algorithm and plotting the TPR and FPR to obtain the curve[119]. Since the algorithms tested produce values over different ranges, 300 threshold values were chosen the between minimum and maximum values, resulting from each algorithm individually. The network deconvolution, Aracne and CLR algorithms, which require an adjacency matrix as an input were used. For these algorithms an adjacency matrix was first computed for one of the dependency metrics and the threshold was applied before running the algorithm. In Figure 4 some example ROC curves are given for Pearson's correlation, compared with the application of the downstream network algorithms discussed above.

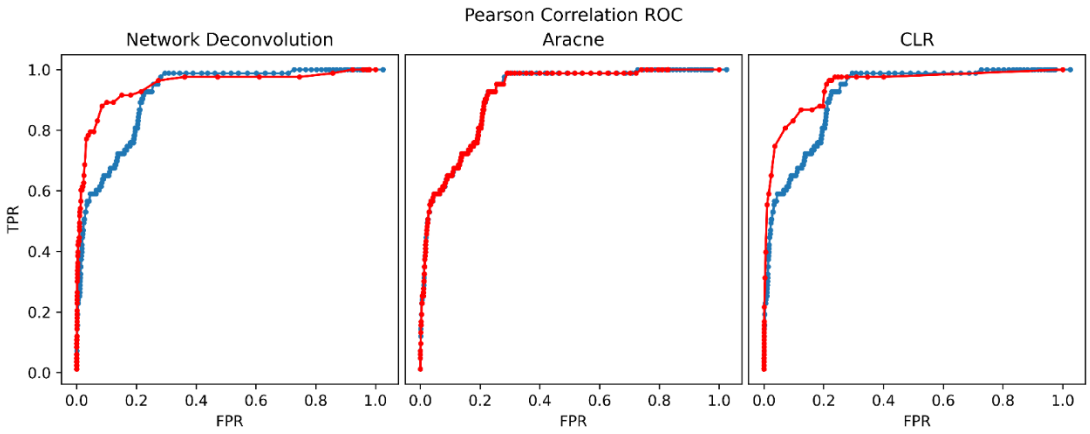


Figure 4: Plots of Receiver Operating Characteristic (ROC) comparing adjacency matrices composed of Pearson's correlation alone (blue) and matrices after application of network algorithms (red) at a sample size of 1000 and SNR of 99% at the 90 hr timepoint.

The AUCs for the plots in Figure 4 are shown as the first four values in the bottom row of the first heatmap of Figure 5. As can be seen, when examining Figures 4 and 5, both network deconvolution and CLR generally improve the performance of Pearson's correlation, under the conditions of high sample size and low noise.

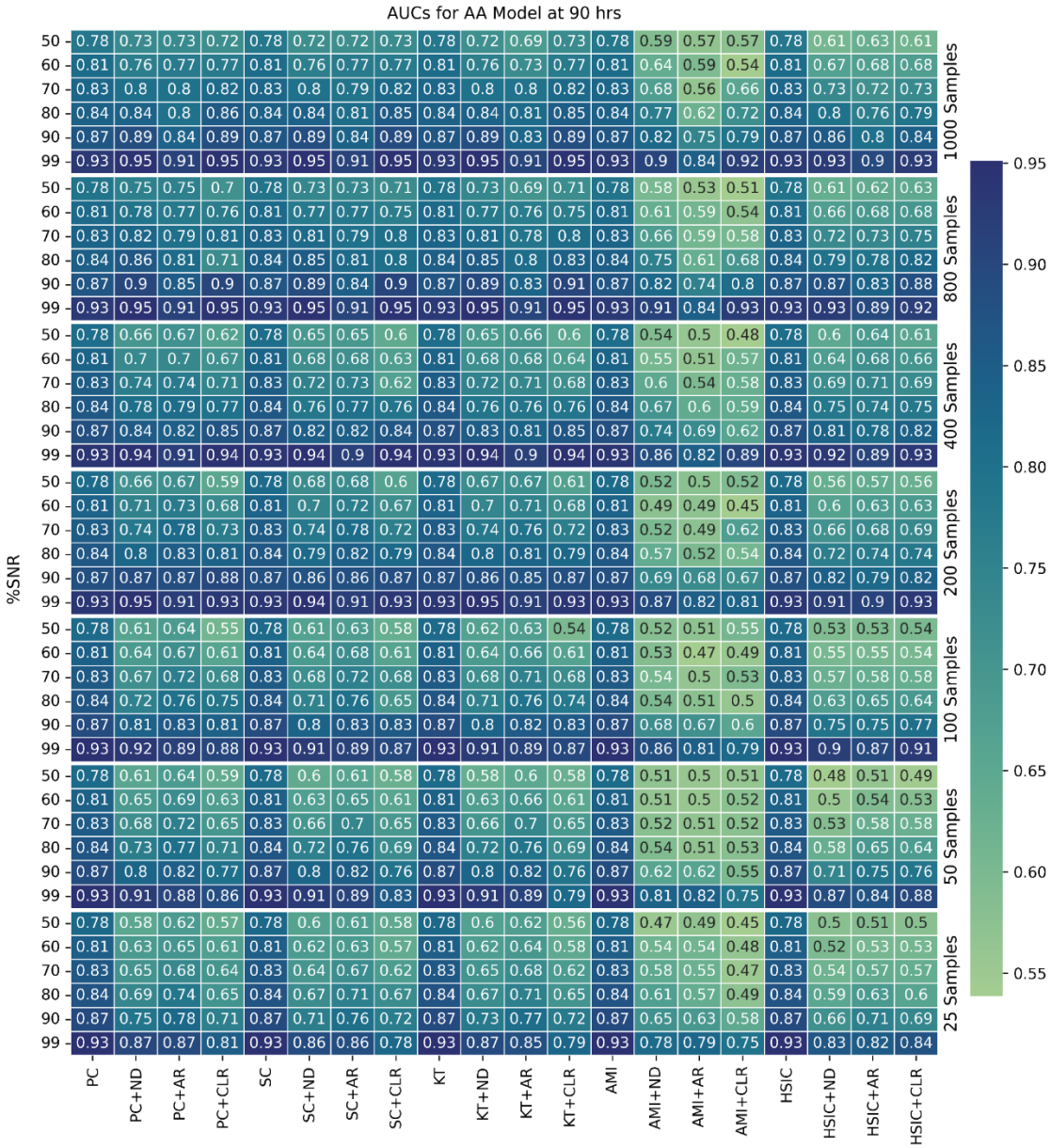


Figure 5: Individual heat maps for the steady state (90 hr) time point for each sample size tested. The Y axis corresponds to the SNR, given as a percent. The X axis lists the dependency metric used and network algorithm applied (if any) from Pearson's Correlation (PC), Spearman's Correlation (SC), Kendall's Tau (KT), Adjusted Mutual Information (AMI), Hilbert Schmidt Independence Criterion (HISC), Network Deconvolution (ND), Aracne (AR), and Context Likelihood or Relatedness (CLR).

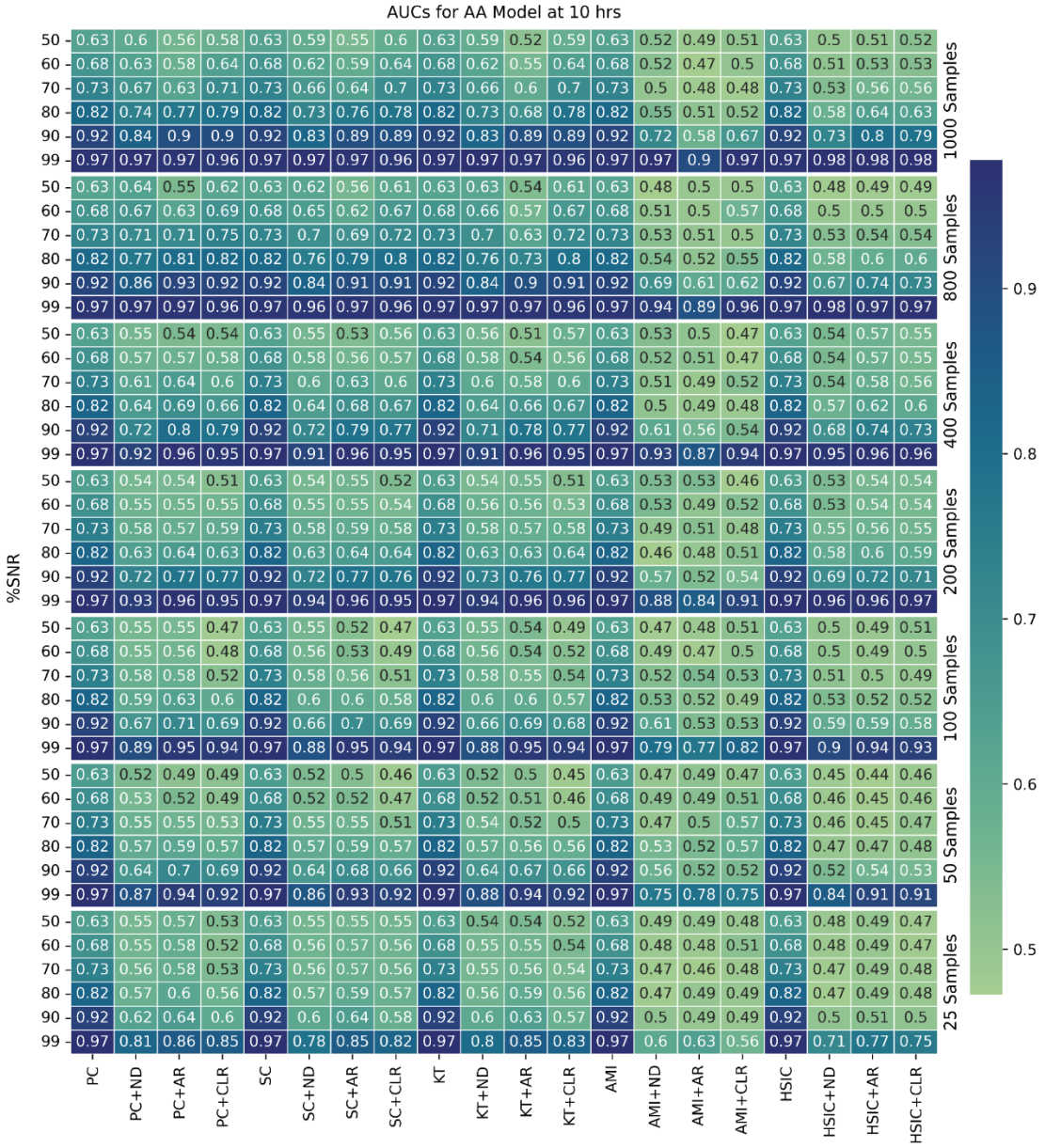


Figure 6: Individual heat maps for the non-steady-state (10 hr) time point for each sample size tested. The Y axis corresponds to the SNR given as a percent. The X axis lists the dependency metric used and network algorithm applied (if any) from Pearson's Correlation (PC), Spearman's Correlation (SC), Kendall's Tau (KT), Adjusted Mutual Information (AMI), Hilbert Schmidt Independence Criterion (HISC), Network Deconvolution (ND), Aracne (AR), and Context Likelihood or Relatedness (CLR).

Network algorithms were observed to improve performance at 200 samples or above and a SNR above 80% for the metrics that only take linear relations into account, when examining the steady state time point. In the non-steady-state case, none of the network algorithms were able to significantly improve performance when applied to PC, SC, or KT, in any of the conditions tested.

For metrics that can detect nonlinear relations an improvement in performance, due to the network algorithms, was not seen even at a sample size of 1000 in the steady-state-time point. However, a slight improvement was seen when network algorithms were applied to the HSIC threshold matrix at 99% SNR and a sample size over 800.

Interestingly, all the dependence metrics tested are nearly identical in terms of AUC when no network algorithm is applied. This is likely primarily due to the AA model used for simulation. Since the model did not explicitly include any complicated indirect relations such as allosteric regulation, it makes sense that the metrics perform similarly as both HSIC and AMI should report the same information as the correlation measures, if there are no nonlinear relations. Additionally, the data is simulated, and Gaussian noise is directly added to the data, thus the data produced will always take the form of a Gaussian distribution. If this is the case, one would expect KT to also perform similarly.

However, in both-steady state and non-steady-state these metrics greatly affect the performance of the network algorithms. For example, in Figure 5 it can be seen that the application of network deconvolution and CLR improve the performance of Pearson's correlation at an SNR of 99% and 400 samples (an AUC of 0.93 compared to 0.94) while the performance of AMI in combination with these algorithms drops from 0.93 to 0.86

and 0.79 respectively. Since the AA model does not include allosteric or other nonlinear metabolic relations, the nonlinear information captured by these metrics is likely a result of noise. This then results in more connections being recognized as indirect by the network algorithms causing a drop in overall performance. This is especially evident in Figure 6, when the non-steady-state time point is examined. Since the system is less homogenous both linear and nonlinear relations are easier to recognize, which likely results in more false positives when the network algorithms are applied to AMI and HSIC. At the highest sample size and lowest noise level, network algorithms applied to HSIC outperform the base metric in the non-steady-state data, 0.97 compared to 0.98 in the first heat map in Figure 6. This suggests that nonlinear relations have the potential to improve the performance of the network algorithms, if enough information is present for them to be reliably detected. Given that all three network algorithms tested improved performance when applied to HSIC at high sample size and low noise in the non-steady-state case, it is possible that HISC and other measures that incorporate nonlinear relations may improve performance, if these relations are known to exist in the data set.

As can be seen in Figures 5 and 6 all three network algorithms result in a drop in performance when sample size is below 200 or when SNR is below 80%. Thus, when sample size is low (below 200 to 400) or SNR is below 80%, network algorithms are unlikely to improve the recovery of mechanistic information. This is consistent with the literature as higher sample sizes are recommended in the original publications for these algorithms[114–116]. At least 1000 samples are recommended in the case of Aracne[115]. If data is not expected to be homogenous or similar to steady-state, network deconvolution,

Aracne, and CLR should not be applied as around 1000 samples and data with very low noise are required to obtain even a modest improvement (Figure 6).

# CHAPTER FIVE

## SUMMARY AND CONCLUSION

In Chapter 2, the Alzheimer's disease brain tissue showed a high prevalence of highly polyunsaturated, oxidized, and unusual lipids. Highly unsaturated lipids are expected in brain tissues. Oxidized and perhaps unusual lipids might be explained by the wide range in postmortem intervals, as seen in Table 1, or relying solely on the MS1 data for identification. However, a large amount of lipid material precipitated out of the samples after they were stored due to the MTBE extraction. The lipids in the solution analyzed may have changed due to the precipitation and resuspension process.

Due to the recovery of unexpected lipids in the human brain samples, other lipid extraction methods were investigated in Chapter 3. Folch extraction appears to be the most successful method for separating endogenous lipids from rat liver samples, with greater coverage and larger average peak areas than the other two methods. Both the Bligh-Dyer and BUME methods serve as suitable alternatives with comparable recovery of lipid standards. The BUME method has the added benefit of allowing the use of plastic materials and easy recovery of the lipid fraction. Unfortunately, problems similar to those encountered with the human brain samples prevented the detailed analysis of the MTBE. While the MTBE procedure resulted in the precipitation of lipid material upon storage, analyzing the MTBE lipid extracts produced from the rat liver samples did not result in strange lipids similar in any way to those encountered in the human brain samples. Since the precipitation did result in unreliable detection of the lipid standards, it is extremely probable that the distribution of lipids in the solution extracted from the human brain

samples by MTBE was also affected. Ideally, the Folch method should be applied to a much larger number of fresh human brain samples, as summarized below.

Lastly, Chapter 4 investigated the ability of network methods to recover mechanistic information from metabolomic data. Figures 5 and 6 show that the performance of all three network techniques is mediocre below 200 samples or 80% SNR. Thus, in the case of minimal samples it is best to rely on simple measures of dependence such as Pearson's or Spearman's correlation. With AUCs consistently above 0.8 in the case of low noise correlation bases measures show great promise for generating mechanistic hypothesis. When samples are abundant, more mechanistic information may be recovered through the application of network algorithms such as CLR.

Taken together, these results indicate that many samples are required to generate mechanistic hypothesis from lipidomic or metabolomic data. If mechanistic hypotheses are to be sought, a minimum of 200 samples with a signal to noise ratio greater than 80% is required. Since studies involving a sufficient number of human brain samples may be unrealistic, great care must be taken to ensure extremely low levels of noise, in order to obtain meaningful conclusions. The extraction procedure for the molecules of interest and the state of the samples during data acquisition greatly affect the noise in the detection system. It would also be desirable to employ mass spectrometers with much higher mass accuracy.

# REFERENCES CITED


(1) Alzheimer, A.; Stelzmann, R. A.; Schnitzlein, H. N.; Murtagh, F. R. An English Translation of Alzheimer's 1907 Paper, "Uber Eine Eigenartige Erkankung Der Hirnrinde." *Clin. Anat. N. Y. N* **1995**, *8* (6), 429–431. https://doi.org/10.1002/ca.980080612.

(2) Rajan, K. B.; Weuve, J.; Barnes, L. L.; McAninch, E. A.; Wilson, R. S.; Evans, D. A. Population Estimate of People with Clinical Alzheimer's Disease and Mild Cognitive Impairment in the United States (2020–2060). *Alzheimers Dement. n/a* (n/a). https://doi.org/10.1002/alz.12362.

(3) *Facts and Figures*. Alzheimer's Disease and Dementia. https://www.alz.org/alzheimers-dementia/facts-figures (accessed 2022-01-16).

(4) Jutkowitz, E.; Kane, R. L.; Gaugler, J. E.; MacLehose, R. F.; Dowd, B.; Kuntz, K. M. SOCIETAL AND FAMILY LIFETIME COST OF DEMENTIA: IMPLICATIONS FOR POLICY. *J. Am. Geriatr. Soc.* **2017**, *65* (10), 2169–2175. https://doi.org/10.1111/jgs.15043.

(5) Christen, Y. Oxidative Stress and Alzheimer Disease. *Am. J. Clin. Nutr.* **2000**, *71* (2), 621S-629S. https://doi.org/10.1093/ajcn/71.2.621s.

(6) Heneka, M. T.; Carson, M. J.; El Khoury, J.; Landreth, G. E.; Brosseron, F.; Feinstein, D. L.; Jacobs, A. H.; Wyss-Coray, T.; Vitorica, J.; Ransohoff, R. M.; Herrup, K.; Frautschy, S. A.; Finsen, B.; Brown, G. C.; Verkhratsky, A.; Yamanaka, K.; Koistinaho, J.; Latz, E.; Halle, A.; Petzold, G. C.; Town, T.; Morgan, D.; Shinohara, M. L.; Perry, V. H.; Holmes, C.; Bazan, N. G.; Brooks, D. J.; Hunot, S.; Joseph, B.; Deigendesch, N.; Garaschuk, O.; Boddeke, E.; Dinarello, C. A.; Breitner, J. C.; Cole, G. M.; Golenbock, D. T.; Kummer, M. P. Neuroinflammation in Alzheimer's Disease. *Lancet Neurol.* **2015**, *14* (4), 388–405. https://doi.org/10.1016/S1474-4422(15)70016-5.

(7) Bremer, A. A.; Mietus-Snyder, M.; Lustig, R. H. Toward a Unifying Hypothesis of Metabolic Syndrome. *Pediatrics* **2012**, *129* (3), 557–570. https://doi.org/10.1542/peds.2011-2912.

(8) Liu, C.-C.; Kanekiyo, T.; Xu, H.; Bu, G. Apolipoprotein E and Alzheimer Disease: Risk, Mechanisms, and Therapy. *Nat. Rev. Neurol.* **2013**, *9* (2), 106–118. https://doi.org/10.1038/nrneurol.2012.263.

(9) Brookmeyer, R.; Johnson, E.; Ziegler-Graham, K.; Arrighi, H. M. Forecasting the Global Burden of Alzheimer's Disease. *Alzheimers Dement. J. Alzheimers Assoc.* **2007**, *3* (3), 186–191. https://doi.org/10.1016/j.jalz.2007.04.381.

(10) Selkoe, D. J.; Hardy, J. The Amyloid Hypothesis of Alzheimer's Disease at 25 Years. *EMBO Mol. Med.* **2016**, *8* (6), 595–608. https://doi.org/10.15252/emmm.201606210.

(11) O'Brien, R. J.; Wong, P. C. Amyloid Precursor Protein Processing and Alzheimer's Disease. *Annu. Rev. Neurosci.* **2011**, *34*, 185–204. https://doi.org/10.1146/annurev-neuro-061010-113613.

(12) Cramer, P. E.; Cirrito, J. R.; Wesson, D. W.; Lee, C. Y. D.; Karlo, J. C.; Zinn, A. E.; Casali, B. T.; Restivo, J. L.; Goebel, W. D.; James, M. J.; Brunden, K. R.; Wilson, D. A.; Landreth, G. E. ApoE-Directed Therapeutics Rapidly Clear β-Amyloid and Reverse Deficits in AD Mouse Models. *Science* **2012**, *335* (6075), 1503–1506. https://doi.org/10.1126/science.1217697.

(13) Gu, L.; Guo, Z. Alzheimer's Aβ42 and Aβ40 Peptides Form Interlaced Amyloid Fibrils. *J. Neurochem.* **2013**, *126* (3), 305–311. https://doi.org/10.1111/jnc.12202.

(14) Borchelt, D. R.; Thinakaran, G.; Eckman, C. B.; Lee, M. K.; Davenport, F.; Ratovitsky, T.; Prada, C.-M.; Kim, G.; Seekins, S.; Yager, D.; Slunt, H. H.; Wang, R.; Seeger, M.; Levey, A. I.; Gandy, S. E.; Copeland, N. G.; Jenkins, N. A.; Price, D. L.; Younkin, S. G.; Sisodia, S. S. Familial Alzheimer's Disease–Linked Presenilin 1 Variants Elevate Aβ1–42/1–40 Ratio In Vitro and In Vivo. *Neuron* **1996**, *17* (5), 1005–1013. https://doi.org/10.1016/S0896-6273(00)80230-5.

(15) Cruchaga, C.; Chakraverty, S.; Mayo, K.; Vallania, F. L. M.; Mitra, R. D.; Faber, K.; Williamson, J.; Bird, T.; Diaz-Arrastia, R.; Foroud, T. M.; Boeve, B. F.; Graff-Radford, N. R.; Jean, P. S.; Lawson, M.; Ehm, M. G.; Mayeux, R.; Goate, A. M.; Consortium, for the N.-L. F. S. Rare Variants in APP, PSEN1 and PSEN2 Increase Risk for AD in Late-Onset Alzheimer's Disease Families. *PLOS ONE* **2012**, *7* (2), e31039. https://doi.org/10.1371/journal.pone.0031039.

(16) Mawuenyega, K. G.; Sigurdson, W.; Ovod, V.; Munsell, L.; Kasten, T.; Morris, J. C.; Yarasheski, K. E.; Bateman, R. J. Decreased Clearance of CNS Beta-Amyloid in Alzheimer's Disease. *Science* **2010**, *330* (6012), 1774. https://doi.org/10.1126/science.1197623.

(17) LaFerla, F. M.; Green, K. N. Animal Models of Alzheimer Disease. *Cold Spring Harb. Perspect. Med.* **2012**, *2* (11), a006320. https://doi.org/10.1101/cshperspect.a006320.

(18) Farris, W.; Mansourian, S.; Chang, Y.; Lindsley, L.; Eckman, E. A.; Frosch, M. P.; Eckman, C. B.; Tanzi, R. E.; Selkoe, D. J.; Guénette, S. Insulin-Degrading Enzyme Regulates the Levels of Insulin, Amyloid β-Protein, and the β-Amyloid Precursor Protein Intracellular Domain in Vivo. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100* (7), 4162–4167. https://doi.org/10.1073/pnas.0230450100.

(19) Hampel, H.; Schneider, L. S.; Giacobini, E.; Kivipelto, M.; Sindi, S.; Dubois, B.; Broich, K.; Nisticò, R.; Aisen, P. S.; Lista, S. Advances in the Therapy of Alzheimer's Disease: Targeting Amyloid Beta and Tau and Perspectives for the Future. *Expert Rev. Neurother.* **2015**, *15* (1), 83–105. https://doi.org/10.1586/14737175.2015.995637.

(20) Sperling, R. A.; Aisen, P. S.; Beckett, L. A.; Bennett, D. A.; Craft, S.; Fagan, A. M.; Iwatsubo, T.; Jack, C. R.; Kaye, J.; Montine, T. J.; Park, D. C.; Reiman, E. M.; Rowe, C. C.; Siemers, E.; Stern, Y.; Yaffe, K.; Carrillo, M. C.; Thies, B.; Morrison-Bogorad, M.; Wagster, M. V.; Phelps, C. H. Toward Defining the Preclinical Stages of Alzheimer's Disease: Recommendations from the National Institute on Aging-Alzheimer's Association Workgroups on Diagnostic Guidelines for Alzheimer's Disease. *Alzheimers Dement. J. Alzheimers Assoc.* **2011**, *7* (3), 280–292. https://doi.org/10.1016/j.jalz.2011.03.003.

(21) Ryan, N. S.; Rossor, M. N.; Fox, N. C. Alzheimer's Disease in the 100 Years since Alzheimer's Death. *Brain J. Neurol.* **2015**, *138* (Pt 12), 3816–3821. https://doi.org/10.1093/brain/awv316.

(22) Gordon, B. A.; Blazey, T. M.; Su, Y.; Hari-Raj, A.; Dincer, A.; Flores, S.; Christensen, J.; McDade, E.; Wang, G.; Xiong, P. C.; Cairns, P. N. J.; Hassenstab, J.; Marcus, D. S.; Fagan, P. A. M.; Jack, P. C. R.; Jr; Hornbeck, R. C.; Paumier, K. L.; Ances, P. B. M.; Berman, S. B.; Brickman, A. M.; Cash, D. M.; Chhatwal, J. P.; Correia, S.; Förster, S.; Fox, P. N. C.; Graff-Radford, P. N. R.; Fougère, P. C. la; Levin, J.; Masters, P. C. L.; Rossor, P. M. N.; Salloway, P. S.; Saykin, P. A. J.; Schofield, P. P. R.; Thompson, P. P. M.; Weiner, P. M. M.; Holtzman, P. D. M.; Raichle, P. M. E.; Morris, P. J. C.; Bateman, P. R. J.; Benzinger, T. L. S. Spatial Patterns of Neuroimaging Biomarker Change in Individuals from Families with Autosomal Dominant Alzheimer Disease: A Longitudinal Study. *Lancet Neurol.* **2018**, *17* (3), 241. https://doi.org/10.1016/S1474-4422(18)30028-0.

(23) Barthélemy, N. R.; Li, Y.; Joseph-Mathurin, N.; Gordon, B. A.; Hassenstab, J.; Benzinger, T. L. S.; Buckles, V.; Fagan, A. M.; Perrin, R. J.; Goate, A. M.; Morris, J. C.; Karch, C. M.; Xiong, C.; Allegri, R.; Mendez, P. C.; Berman, S. B.; Ikeuchi, T.; Mori, H.; Shimada, H.; Shoji, M.; Suzuki, K.; Noble, J.; Farlow, M.; Chhatwal, J.; Graff-Radford, N. R.; Salloway, S.; Schofield, P. R.; Masters, C. L.; Martins, R. N.; O'Connor, A.; Fox, N. C.; Levin, J.; Jucker, M.; Gabelle, A.; Lehmann, S.; Sato, C.; Bateman, R. J.; McDade, E. A Soluble Phosphorylated Tau Signature Links Tau, Amyloid and the Evolution of Stages of Dominantly Inherited Alzheimer's Disease. *Nat. Med.* **2020**, *26* (3), 398–407. https://doi.org/10.1038/s41591-020-0781-z.

(24) Kosicek, M.; Hecimovic, S. Phospholipids and Alzheimer's Disease: Alterations, Mechanisms and Potential Biomarkers. *Int. J. Mol. Sci.* **2013**, *14* (1), 1310–1322. https://doi.org/10.3390/ijms14011310.

(25) Huynh, K.; Lim, W. L. F.; Giles, C.; Jayawardana, K. S.; Salim, A.; Mellett, N. A.; Smith, A. A. T.; Olshansky, G.; Drew, B. G.; Chatterjee, P.; Martins, I.; Laws, S. M.; Bush, A. I.; Rowe, C. C.; Villemagne, V. L.; Ames, D.; Masters, C. L.; Arnold, M.; Nho, K.; Saykin, A. J.; Baillie, R.; Han, X.; Kaddurah-Daouk, R.; Martins, R. N.; Meikle, P. J. Concordant Peripheral Lipidome Signatures in Two Large Clinical Studies of Alzheimer's Disease. *Nat. Commun.* **2020**, *11* (1), 5698. https://doi.org/10.1038/s41467-020-19473-7.

(26) Meikle, P. J.; Wong, G.; Barlow, C. K.; Weir, J. M.; Greeve, M. A.; MacIntosh, G. L.; Almasy, L.; Comuzzie, A. G.; Mahaney, M. C.; Kowalczyk, A.; Haviv, I.; Grantham, N.; Magliano, D. J.; Jowett, J. B. M.; Zimmet, P.; Curran, J. E.; Blangero, J.; Shaw, J. Plasma Lipid Profiling Shows Similar Associations with Prediabetes and Type 2 Diabetes. *PLoS ONE* **2013**, *8* (9), e74341. https://doi.org/10.1371/journal.pone.0074341.

(27) Raffaitin, C.; Gin, H.; Empana, J.-P.; Helmer, C.; Berr, C.; Tzourio, C.; Portet, F.; Dartigues, J.-F.; Alpérovitch, A.; Barberger-Gateau, P. Metabolic Syndrome and Risk for Incident Alzheimer's Disease or Vascular Dementia. *Diabetes Care* **2009**, *32* (1), 169–174. https://doi.org/10.2337/dc08-0272.

(28) Gudala, K.; Bansal, D.; Schifano, F.; Bhansali, A. Diabetes Mellitus and Risk of Dementia: A Meta-analysis of Prospective Observational Studies. *J. Diabetes Investig.* **2013**, *4* (6), 640–650. https://doi.org/10.1111/jdi.12087.

(29) Orešič, M.; Hyötyläinen, T.; Herukka, S.-K.; Sysi-Aho, M.; Mattila, I.; Seppänan-Laakso, T.; Julkunen, V.; Gopalacharyulu, P. V.; Hallikainen, M.; Koikkalainen, J.; Kivipelto, M.; Helisalmi, S.; Lötjönen, J.; Soininen, H. Metabolome in Progression to Alzheimer's Disease. *Transl. Psychiatry* **2011**, *1* (12), e57. https://doi.org/10.1038/tp.2011.55.

(30) Kim, W. S.; Li, H.; Ruberu, K.; Chan, S.; Elliott, D. A.; Low, J. K.; Cheng, D.; Karl, T.; Garner, B. Deletion of Abca7 Increases Cerebral Amyloid-β Accumulation in the J20 Mouse Model of Alzheimer's Disease. *J. Neurosci.* **2013**, *33* (10), 4387–4394. https://doi.org/10.1523/JNEUROSCI.4165-12.2013.

(31) Trushina, E.; Dutta, T.; Persson, X.-M. T.; Mielke, M. M.; Petersen, R. C. Identification of Altered Metabolic Pathways in Plasma and CSF in Mild Cognitive Impairment and Alzheimer's Disease Using Metabolomics. *PLoS ONE* **2013**, *8* (5), e63644. https://doi.org/10.1371/journal.pone.0063644.

(32) Puglielli, L.; Ellis, B. C.; Saunders, A. J.; Kovacs, D. M. Ceramide Stabilizes Beta-Site Amyloid Precursor Protein-Cleaving Enzyme 1 and Promotes Amyloid Beta-Peptide Biogenesis. *J. Biol. Chem.* **2003**, *278* (22), 19777–19783. https://doi.org/10.1074/jbc.M300466200.

(33) Mizuno, S.; Iijima, R.; Ogishima, S.; Kikuchi, M.; Matsuoka, Y.; Ghosh, S.; Miyamoto, T.; Miyashita, A.; Kuwano, R.; Tanaka, H. AlzPathway: A Comprehensive Map of Signaling Pathways of Alzheimer's Disease. *BMC Syst. Biol.* **2012**, *6*, 52. https://doi.org/10.1186/1752-0509-6-52.

(34) Casci, T. Personalized Omics Profiling. *Nat. Rev. Genet.* **2012**, *13* (5), 298–298. https://doi.org/10.1038/nrg3231.

(35) Nagana Gowda, G. A.; Zhang, S.; Gu, H.; Asiago, V.; Shanaiah, N.; Raftery, D. Metabolomics-Based Methods for Early Disease Diagnostics: A Review. *Expert Rev. Mol. Diagn.* **2008**, *8* (5), 617–633. https://doi.org/10.1586/14737159.8.5.617.

(36) Wood, P. L. Lipidomics of Alzheimer's Disease: Current Status. *Alzheimers Res. Ther.* **2012**, *4* (1), 5. https://doi.org/10.1186/alzrt103.

(37) Gallo Cantafio, M. E.; Grillone, K.; Caracciolo, D.; Scionti, F.; Arbitrio, M.; Barbieri, V.; Pensabene, L.; Guzzi, P. H.; Di Martino, M. T. From Single Level Analysis to Multi-Omics Integrative Approaches: A Powerful Strategy towards the Precision Oncology. *High-Throughput* **2018**, *7* (4), 33. https://doi.org/10.3390/ht7040033.

(38) *A Multiomics Approach to Heterogeneity in Alzheimer's Disease*. Medscape. http://www.medscape.com/viewarticle/936317 (accessed 2022-01-16).

(39) Labory, J.; Fierville, M.; Ait-El-Mkadem, S.; Bannwarth, S.; Paquis-Flucklinger, V.; Bottini, S. Multi-Omics Approaches to Improve Mitochondrial Disease Diagnosis: Challenges, Advances, and Perspectives. *Front. Mol. Biosci.* **2020**, *7*.

(40) Sitnikov, D. G.; Monnin, C. S.; Vuckovic, D. Systematic Assessment of Seven Solvent and Solid-Phase Extraction Methods for Metabolomics Analysis of Human Plasma by LC-MS. *Sci. Rep.* **2016**, *6* (1), 38885. https://doi.org/10.1038/srep38885.

(41) Folch, J.; Lees, M.; Stanley, G. H. S. A SIMPLE METHOD FOR THE ISOLATION AND PURIFICATION OF TOTAL LIPIDES FROM ANIMAL TISSUES. *J. Biol. Chem.* **1957**, *226* (1), 497–509. https://doi.org/10.1016/s0021-9258(18)64849-5.

(42) Bligh, E. G.; Dyer, W. J. A Rapid Method of Total Lipid Extraction and Purification. *Can. J. Biochem. Physiol.* **1959**, *37* (8), 911–917. https://doi.org/10.1139/o59-099.

(43) Reis, A.; Rudnitskaya, A.; Blackburn, G. J.; Fauzi, N. M.; Pitt, A. R.; Spickett, C. M. A Comparison of Five Lipid Extraction Solvent Systems for Lipidomic Studies of Human LDL. *J. Lipid Res.* **2013**, *54* (7), 1812-1824-1812–1824. https://doi.org/10.1194/jlr.M034330.

(44) Mubeen, U.; Giraldi, L. A.; Jüppner, J.; Giavalisco, P. A Multi-Omics Extraction Method for the In-Depth Analysis of Synchronized Cultures of the Green Alga Chlamydomonas Reinhardtii. *JoVE J. Vis. Exp.* **2019**, No. 150, e59547. https://doi.org/10.3791/59547.

(45) Kang, J.; David, L.; Li, Y.; Cang, J.; Chen, S. Three-in-One Simultaneous Extraction of Proteins, Metabolites and Lipids for Multi-Omics. *Front. Genet.* **2021**, *12*, 313. https://doi.org/10.3389/fgene.2021.635971.

(46) Löfgren, L.; Forsberg, G.-B.; Ståhlman, M. The BUME Method: A New Rapid and Simple Chloroform-Free Method for Total Lipid Extraction of Animal Tissue. *Sci. Rep.* **2016**, *6* (1), 27688. https://doi.org/10.1038/srep27688.

(47) 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.* **2008**, *49* (5), 1137–1146. https://doi.org/10.1194/jlr.D700041-JLR200.

(48) Reimand, J.; Isser, R.; Voisin, V.; Kucera, M.; Tannus-Lopes, C.; Rostamianfar, A.; Wadi, L.; Meyer, M.; Wong, J.; Xu, C.; Merico, D.; Bader, G. D. Pathway Enrichment Analysis and Visualization of Omics Data Using g:Profiler, GSEA, Cytoscape and EnrichmentMap. *Nat. Protoc.* **2019**, *14* (2), 482–517. https://doi.org/10.1038/s41596-018-0103-9.

(49) Li, S.; Park, Y.; Duraisingham, S.; Strobel, F. H.; Khan, N.; Soltow, Q. A.; Jones, D. P.; Pulendran, B. Predicting Network Activity from High Throughput Metabolomics. *PLoS Comput. Biol.* **2013**, *9* (7). https://doi.org/10.1371/journal.pcbi.1003123.

(50) Phelan, V. V. Feature-Based Molecular Networking for Metabolite Annotation. *Methods Mol. Biol. Clifton NJ* **2020**, *2104*, 227–243. https://doi.org/10.1007/978-1-0716-0239-3_13.

(51) Chawla, S.; Pund, A.; B., V.; Kulkarni, S.; Diwekar-Joshi, M.; Watve, M. Inferring Causal Pathways among Three or More Variables from Steady-State Correlations in a Homeostatic System. *PLoS ONE* **2018**, *13* (10), e0204755. https://doi.org/10.1371/journal.pone.0204755.

(52) Székely, G. J.; Rizzo, M. L.; Bakirov, N. K. Measuring and Testing Dependence by Correlation of Distances. *Ann. Stat.* **2007**, *35* (6), 2769–2794. https://doi.org/10.1214/009053607000000505.

(53) Heller, R.; Heller, Y.; Gorfine, M. A Consistent Multivariate Test of Association Based on Ranks of Distances. *Biometrika* **2013**, *100* (2), 503–510. https://doi.org/10.1093/biomet/ass070.

(54) Kinney, J. B.; Atwal, G. S. Equitability, Mutual Information, and the Maximal Information Coefficient. *Proc. Natl. Acad. Sci.* **2014**, *111* (9), 3354–3359. https://doi.org/10.1073/pnas.1309933111.

(55) Mooij, J.; Janzing, D.; Peters, J.; Schölkopf, B. Regression by Dependence Minimization and Its Application to Causal Inference in Additive Noise Models. In *Proceedings of the 26th Annual International Conference on Machine Learning*; ICML '09; Association for Computing Machinery: New York, NY, USA, 2009; pp 745–752. https://doi.org/10.1145/1553374.1553470.

(56) Glymour, C.; Zhang, K.; Spirtes, P. Review of Causal Discovery Methods Based on Graphical Models. *Front. Genet.* **2019**, *10*.

(57) Jahagirdar, S.; Suarez-Diez, M.; Saccenti, E. Simulation and Reconstruction of Metabolite–Metabolite Association Networks Using a Metabolic Dynamic Model and Correlation Based Algorithms. *J. Proteome Res.* **2019**, *18* (3), 1099–1113. https://doi.org/10.1021/acs.jproteome.8b00781.

(58) Krämer, A.; Green, J.; Pollard, J., Jr; Tugendreich, S. Causal Analysis Approaches in Ingenuity Pathway Analysis. *Bioinformatics* **2014**, *30* (4), 523–530. https://doi.org/10.1093/bioinformatics/btt703.

(59) *Causal integration of multi-omics data with prior knowledge to generate mechanistic hypotheses | Molecular Systems Biology*. https://www.embopress.org/doi/full/10.15252/msb.20209730 (accessed 2021-09-23).

(60) Jafarzadeh, S. Mediation Analysis: Understanding Causal Pathways to Disease Outcomes. *Osteoarthritis Cartilage* **2020**, *28*, S17. https://doi.org/10.1016/j.joca.2020.02.015.

(61) Ramanan, N.; Natarajan, S. Causal Learning From Predictive Modeling for Observational Data. *Front. Big Data* **2020**, *3*.

(62) Hyman, B. T.; Trojanowski, J. Q. Consensus Recommendations for the Postmortem Diagnosis of Alzheimer Disease from the National Institute on Aging and the Reagan Institute Working Group on Diagnostic Criteria for the Neuropathological Assessment of Alzheimer Disease. *J. Neuropathol. Exp. Neurol.* **1997**, *56* (10), 1095–1097. https://doi.org/10.1097/00005072-199710000-00002.

(63) Lejardi, D.; Technologies, A. Improving Coverage of the Plasma Lipidome Using Iterative MS/MS Data Acquisition Combined with Lipid Annotator Software and 6546 LC/Q-TOF. 10.

(64) 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* **2015**, *12* (6), 523–526. https://doi.org/10.1038/nmeth.3393.

(65) Seabold, S.; Perktold, J. Statsmodels: Econometric and Statistical Modeling with Python; Austin, Texas, 2010; pp 92–96. https://doi.org/10.25080/Majora-92bf1922-011.

(66) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, É. Scikit-Learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12* (85), 2825–2830.

(67) Doran, S.; Arif, M.; Lam, S.; Bayraktar, A.; Turkez, H.; Uhlen, M.; Boren, J.; Mardinoglu, A. Multi-Omics Approaches for Revealing the Complexity of Cardiovascular Disease. *Brief. Bioinform.* **2021**, *22* (5), bbab061. https://doi.org/10.1093/bib/bbab061.

(68) Su, M.-W.; Chang, C.; Lin, C.-W.; Ling, S.-J.; Hsiung, C.-N.; Chu, H.-W.; Wu, P.-E.; Shen, C.-Y. Blood Multiomics Reveal Insights into Population Clusters with Low Prevalence of Diabetes, Dyslipidemia and Hypertension. *PLOS ONE* **2020**, *15* (3), e0229922. https://doi.org/10.1371/journal.pone.0229922.

(69) Schelbert, K. B. Comorbidities of Obesity. *Prim. Care Clin. Off. Pract.* **2009**, *36* (2), 271–285. https://doi.org/10.1016/j.pop.2009.01.009.

(70) Verdile, G.; Keane, K. N.; Cruzat, V. F.; Medic, S.; Sabale, M.; Rowles, J.; Wijesekara, N.; Martins, R. N.; Fraser, P. E.; Newsholme, P. Inflammation and Oxidative Stress: The Molecular Connectivity between Insulin Resistance Obesity, and Alzheimer's Disease. *Mediators Inflamm.* **2015**, *2015*, 1–17. https://doi.org/10.1155/2015/105828.

(71) Annema, W.; Eckardstein, A. von. Dysfunctional High-Density Lipoproteins in Coronary Heart Disease: Implications for Diagnostics and Therapy. *Transl. Res.* **2016**, *173*, 30–57. https://doi.org/10.1016/j.trsl.2016.02.008.

(72) Müller, C. P.; Reichel, M.; Mühle, C.; Rhein, C.; Gulbins, E.; Kornhuber, J. Brain Membrane Lipids in Major Depression and Anxiety Disorders. *Biochim. Biophys. Acta BBA - Mol. Cell Biol. Lipids* **2015**, *1851* (8), 1052–1065. https://doi.org/10.1016/j.bbalip.2014.12.014.

(73) Fabelo, N.; Martín, V.; Marín, R.; Moreno, D.; Ferrer, I.; Díaz, M. Altered Lipid Composition in Cortical Lipid Rafts Occurs at Early Stages of Sporadic Alzheimer's Disease and Facilitates APP/BACE1 Interactions. *Neurobiol. Aging* **2014**, *35* (8), 1801–1812. https://doi.org/10.1016/j.neurobiolaging.2014.02.005.

(74) Egawa, J.; Pearn, M. L.; Lemkuil, B. P.; Patel, P. M.; Head, B. P. Membrane Lipid Rafts and Neurobiology: Age-Related Changes in Membrane Lipids and Loss of Neuronal Function. *J. Physiol.* **2015**, *594* (16), 4565–4579. https://doi.org/10.1113/jp270590.

(75) Welti, R.; Li, W.; Li, M.; Sang, Y.; Biesiada, H.; Zhou, H.-E.; Rajashekar, C. B.; Williams, T. D.; Wang, X. Profiling Membrane Lipids in Plant Stress Responses. *J. Biol. Chem.* **2002**, *277* (35), 31994–32002. https://doi.org/10.1074/jbc.m205375200.

(76) Braverman, N. E.; Moser, A. B. Functions of Plasmalogen Lipids in Health and Disease. *Biochim. Biophys. Acta BBA - Mol. Basis Dis.* **2012**, *1822* (9), 1442–1452. https://doi.org/10.1016/j.bbadis.2012.05.008.

(77) Want, E. J.; Masson, P.; Michopoulos, F.; Wilson, I. D.; Theodoridis, G.; Plumb, R. S.; Shockcor, J.; Loftus, N.; Holmes, E.; Nicholson, J. K. Global Metabolic Profiling of Animal and Human Tissues via UPLC-MS. *Nat. Protoc.* **2012**, *8* (1), 17–32. https://doi.org/10.1038/nprot.2012.135.

(78) Shaner, R. L.; Allegood, J. C.; Park, H.; Wang, E.; Kelly, S.; Haynes, C. A.; Sullards, M. C.; Merrill, A. H. Quantitative Analysis of Sphingolipids for Lipidomics Using Triple Quadrupole and Quadrupole Linear Ion Trap Mass Spectrometers. *J. Lipid Res.* **2009**, *50* (8), 1692–1707. https://doi.org/10.1194/jlr.d800051-jlr200.

(79) 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 Analyt. Technol. Biomed. Life. Sci.* **2015**, *1002*, 260–266. https://doi.org/10.1016/j.jchromb.2015.08.018.

(80) Reichl, B.; Eichelberg, N.; Freytag, M.; Gojo, J.; Peyrl, A.; Buchberger, W. Evaluation and Optimization of Common Lipid Extraction Methods in Cerebrospinal Fluid Samples. *J. Chromatogr. B* **2020**, *1153*, 122271. https://doi.org/10.1016/j.jchromb.2020.122271.

(81) Masson, P.; Alves, A. C.; Ebbels, T. M. D.; Nicholson, J. K.; Want, E. J. Optimization and Evaluation of Metabolite Extraction Protocols for Untargeted Metabolic Profiling of Liver Samples by UPLC-MS. *Anal. Chem.* **2010**, *82* (18), 7779–7786. https://doi.org/10.1021/ac101722e.

(82) Wu, H.; Southam, A. D.; Hines, A.; Viant, M. R. High-Throughput Tissue Extraction Protocol for NMR- and MS-Based Metabolomics. *Anal. Biochem.* **2008**, *372* (2), 204–212. https://doi.org/10.1016/j.ab.2007.10.002.

(83) Tajima, Y.; Ishikawa, M.; Maekawa, K.; Murayama, M.; Senoo, Y.; Nishimaki-Mogami, T.; Nakanishi, H.; Ikeda, K.; Arita, M.; Taguchi, R.; Okuno, A.; Mikawa,

R.; Niida, S.; Takikawa, O.; Saito, Y. Lipidomic Analysis of Brain Tissues and Plasma in a Mouse Model Expressing Mutated Human Amyloid Precursor Protein/Tau for Alzheimer's Disease. *Lipids Health Dis.* **2013**, *12*, 68. https://doi.org/10.1186/1476-511X-12-68.

(84) Chiu, H.-H.; Tsai, S.-J.; Tseng, Y. J.; Wu, M.-S.; Liao, W.-C.; Huang, C.-S.; Kuo, C.-H. An Efficient and Robust Fatty Acid Profiling Method for Plasma Metabolomic Studies by Gas Chromatography–Mass Spectrometry. *Clin. Chim. Acta* **2015**, *451*, 183–190. https://doi.org/10.1016/j.cca.2015.09.028.

(85) Masood, A.; Stark, K. D.; Salem, N. A Simplified and Efficient Method for the Analysis of Fatty Acid Methyl Esters Suitable for Large Clinical Studies. *J. Lipid Res.* **2005**, *46* (10), 2299–2305. https://doi.org/10.1194/jlr.d500022-jlr200.

(86) Carrapiso, A. I.; García, C. Development in Lipid Analysis: Some New Extraction Techniques and in Situ Transesterification. *Lipids* **2000**, *35* (11), 1167–1177. https://doi.org/10.1007/s11745-000-0633-8.

(87) Guckert, J. B.; Cooksey, K. E.; Jackson, L. L. Lipid Sovent Systems Are Not Equivalent for Analysis of Lipid Classes in the Microeukaryotic Green Alga Chlorella. *J. Microbiol. Methods* **1988**, *8* (3), 139–149. https://doi.org/10.1016/0167-7012(88)90015-2.

(88) Santos, R. R. dos; Moreira, D. M.; Kunigami, C. N.; Aranda, D. A. G.; Teixeira, C. M. L. L. Comparison between Several Methods of Total Lipid Extraction from Chlorella Vulgaris Biomass. *Ultrason. Sonochem.* **2015**, *22*, 95–99. https://doi.org/10.1016/j.ultsonch.2014.05.015.

(89) Tallino, S.; Duffy, M.; Ralle, M.; Cortés, M. P.; Latorre, M.; Burkhead, J. L. Nutrigenomics Analysis Reveals That Copper Deficiency and Dietary Sucrose Up-Regulate Inflammation Fibrosis and Lipogenic Pathways in a Mature Rat Model of Nonalcoholic Fatty Liver Disease. *J. Nutr. Biochem.* **2015**, *26* (10), 996–1006. https://doi.org/10.1016/j.jnutbio.2015.04.009.

(90) Cornett, D. S.; Reyzer, M. L.; Chaurand, P.; Caprioli, R. M. MALDI Imaging Mass Spectrometry: Molecular Snapshots of Biochemical Systems. *Nat. Methods* **2007**, *4* (10), 828–833. https://doi.org/10.1038/nmeth1094.

(91) Berry, K. A. Z.; Hankin, J. A.; Barkley, R. M.; Spraggins, J. M.; Caprioli, R. M.; Murphy, R. C. MALDI Imaging of Lipid Biochemistry in Tissues by Mass Spectrometry. *Chem. Rev.* **2011**, *111* (10), 6491–6512. https://doi.org/10.1021/cr200280p.

(92) 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* **2013**, *10* (8), 755–758. https://doi.org/10.1038/nmeth.2551.

(93) Wang, M.; Carver, J. J.; Phelan, V. V.; Sanchez, L. M.; Garg, N.; Peng, Y.; Nguyen, D. D.; Watrous, J.; Kapono, C. A.; Luzzatto-Knaan, T.; Porto, C.; Bouslimani, A.; Melnik, A. V.; Meehan, M. J.; Liu, W.-T.; Crüsemann, M.; Boudreau, P. D.; Esquenazi, E.; Sandoval-Calderón, M.; Kersten, R. D.; Pace, L. A.; Quinn, R. A.; Duncan, K. R.; Hsu, C.-C.; Floros, D. J.; Gavilan, R. G.; Kleigrewe, K.; Northen, T.; Dutton, R. J.; Parrot, D.; Carlson, E. E.; Aigle, B.; Michelsen, C. F.; Jelsbak, L.; Sohlenkamp, C.; Pevzner, P.; Edlund, A.; McLean, J.; Piel, J.; Murphy, B. T.; Gerwick, L.; Liaw, C.-C.; Yang, Y.-L.; Humpf, H.-U.; Maansson, M.; Keyzers, R. A.; Sims, A. C.; Johnson, A. R.; Sidebottom, A. M.; Sedio, B. E.; Klitgaard, A.; Larson, C. B.; Boya P, C. A.; Torres-Mendoza, D.; Gonzalez, D. J.; Silva, D. B.; Marques, L. M.; Demarque, D. P.; Pociute, E.; O'Neill, E. C.; Briand, E.; Helfrich, E. J. N.; Granatosky, E. A.; Glukhov, E.; Ryffel, F.; Houson, H.; Mohimani, H.; Kharbush, J. J.; Zeng, Y.; Vorholt, J. A.; Kurita, K. L.; Charusanti, P.; McPhail, K. L.; Nielsen, K. F.; Vuong, L.; Elfeki, M.; Traxler, M. F.; Engene, N.; Koyama, N.; Vining, O. B.; Baric, R.; Silva, R. R.; Mascuch, S. J.; Tomasi, S.; Jenkins, S.; Macherla, V.; Hoffman, T.; Agarwal, V.; Williams, P. G.; Dai, J.; Neupane, R.; Gurr, J.; Rodríguez, A. M. C.; Lamsa, A.; Zhang, C.; Dorrestein, K.; Duggan, B. M.; Almaliti, J.; Allard, P.-M.; Phapale, P.; Nothias, L.-F.; Alexandrov, T.; Litaudon, M.; Wolfender, J.-L.; Kyle, J. E.; Metz, T. O.; Peryea, T.; Nguyen, D.-T.; VanLeer, D.; Shinn, P.; Jadhav, A.; Müller, R.; Waters, K. M.; Shi, W.; Liu, X.; Zhang, L.; Knight, R.; Jensen, P. R.; Palsson, B. Ø.; Pogliano, K.; Linington, R. G.; Gutiérrez, M.; Lopes, N. P.; Gerwick, W. H.; Moore, B. S.; Dorrestein, P. C.; Bandeira, N. Sharing and Community Curation of Mass Spectrometry Data with Global Natural Products Social Molecular Networking. *Nat. Biotechnol.* **2016**, *34* (8), 828–837. https://doi.org/10.1038/nbt.3597.

(94) Horai, H.; Arita, M.; Kanaya, S.; Nihei, Y.; Ikeda, T.; Suwa, K.; Ojima, Y.; Tanaka, K.; Tanaka, S.; Aoshima, K.; Oda, Y.; Kakazu, Y.; Kusano, M.; Tohge, T.; Matsuda, F.; Sawada, Y.; Hirai, M. Y.; Nakanishi, H.; Ikeda, K.; Akimoto, N.; Maoka, T.; Takahashi, H.; Ara, T.; Sakurai, N.; Suzuki, H.; Shibata, D.; Neumann, S.; Iida, T.; Tanaka, K.; Funatsu, K.; Matsuura, F.; Soga, T.; Taguchi, R.; Saito, K.; Nishioka, T. MassBank: A Public Repository for Sharing Mass Spectral Data for Life Sciences. *J. Mass Spectrom.* **2010**, *45* (7), 703–714. https://doi.org/10.1002/jms.1777.

(95) Tsugawa, H.; Ikeda, K.; Takahashi, M.; Satoh, A.; Mori, Y.; Uchino, H.; Okahashi, N.; Yamada, Y.; Tada, I.; Bonini, P.; Higashi, Y.; Okazaki, Y.; Zhou, Z.; Zhu, Z.-J.; Koelmel, J.; Cajka, T.; Fiehn, O.; Saito, K.; Arita, M.; Arita, M. A Lipidome Atlas in MS-DIAL 4. *Nat. Biotechnol.* **2020**, *38* (10), 1159–1163. https://doi.org/10.1038/s41587-020-0531-2.

(96) Martin Bland, J.; Altman, DouglasG. STATISTICAL METHODS FOR ASSESSING AGREEMENT BETWEEN TWO METHODS OF CLINICAL

MEASUREMENT. *The Lancet* **1986**, *327* (8476), 307–310. https://doi.org/10.1016/S0140-6736(86)90837-8.

(97) Gil, A.; Zhang, W.; Wolters, J. C.; Permentier, H.; Boer, T.; Horvatovich, P.; Heiner-Fokkema, M. R.; Reijngoud, D.-J.; Bischoff, R. One- vs Two-Phase Extraction: Re-Evaluation of Sample Preparation Procedures for Untargeted Lipidomics in Plasma Samples. *Anal. Bioanal. Chem.* **2018**, *410* (23), 5859–5870. https://doi.org/10.1007/s00216-018-1200-x.

(98) Fiehn, O. Metabolomics--the Link between Genotypes and Phenotypes. *Plant Mol. Biol.* **2002**, *48* (1–2), 155–171.

(99) Krumsiek, J.; Suhre, K.; Illig, T.; Adamski, J.; Theis, F. J. Bayesian Independent Component Analysis Recovers Pathway Signatures from Blood Metabolomics Data. *J. Proteome Res.* **2012**, *11* (8), 4120-4131-4120–4131. https://doi.org/10.1021/pr300231n.

(100) Rosato, A.; Tenori, L.; Cascante, M.; De Atauri Carulla, P. R.; Martins dos Santos, V. A. P.; Saccenti, E. From Correlation to Causation: Analysis of Metabolomics Data Using Systems Biology Approaches. *Metabolomics* **2018**, *14* (4), 37. https://doi.org/10.1007/s11306-018-1335-y.

(101) Jahagirdar, S.; Suarez-Diez, M.; Saccenti, E. Simulation and Reconstruction of Metabolite–Metabolite Association Networks Using a Metabolic Dynamic Model and Correlation Based Algorithms. *J. Proteome Res.* **2019**, *18* (3), 1099–1113. https://doi.org/10.1021/acs.jproteome.8b00781.

(102) Kotze, H. L.; Armitage, E. G.; Sharkey, K. J.; Allwood, J. W.; Dunn, W. B.; Williams, K. J.; Goodacre, R. A Novel Untargeted Metabolomics Correlation-Based Network Analysis Incorporating Human Metabolic Reconstructions. *BMC Syst. Biol.* **2013**, *7* (1), 107. https://doi.org/10.1186/1752-0509-7-107.

(103) Steuer, R. Review: On the Analysis and Interpretation of Correlations in Metabolomic Data. *Brief. Bioinform.* **2006**, *7* (2), 151–158. https://doi.org/10.1093/bib/bbl009.

(104) Horvath, S. *Weighted Network Analysis*; Springer: New York, NY, 2011. https://doi.org/10.1007/978-1-4419-8819-5.

(105) KENDALL, M. G. A NEW MEASURE OF RANK CORRELATION. *Biometrika* **1938**, *30* (1–2), 81–93. https://doi.org/10.1093/biomet/30.1-2.81.

(106) Vinh, N. X.; Epps, J.; Bailey, J. Information Theoretic Measures for Clusterings Comparison: Variants, Properties, Normalization and Correction for Chance. 18.

(107) Freedman, D.; Diaconis, P. On the Histogram as a Density Estimator:L2 Theory. *Z. Für Wahrscheinlichkeitstheorie Verwandte Geb.* **1981**, *57* (4), 453–476. https://doi.org/10.1007/BF01025868.

(108) Kinney, J. B.; Atwal, G. S. Equitability, Mutual Information, and the Maximal Information Coefficient. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111* (9), 3354-3359-3354–3359. https://doi.org/10.1073/pnas.1309933111.

(109) Liang, K.-C.; Wang, X. Gene Regulatory Network Reconstruction Using Conditional Mutual Information. *EURASIP J. Bioinforma. Syst. Biol.* **2008**, *2008* (1), 253894. https://doi.org/10.1155/2008/253894.

(110) Gretton, A.; Fukumizu, K.; Teo, C. H.; Song, L.; Schölkopf, B.; Smola, A. J. A Kernel Statistical Test of Independence. In *Proceedings of the 20th International Conference on Neural Information Processing Systems*; NIPS'07; Curran Associates Inc.: Red Hook, NY, USA, 2007; pp 585–592.

(111) Gretton, A.; Györfi, L. Consistent Nonparametric Tests of Independence. *J. Mach. Learn. Res.* **2010**, *11* (46), 1391–1423.

(112) Shen, C.; Vogelstein, J. T. The Exact Equivalence of Distance and Kernel Methods in Hypothesis Testing. *AStA Adv. Stat. Anal.* **2021**, *105* (3), 385–403. https://doi.org/10.1007/s10182-020-00378-1.

(113) Sejdinovic, D.; Sriperumbudur, B.; Gretton, A.; Fukumizu, K. Equivalence of Distance-Based and Rkhs-Based Statistics in Hypothesis Testing. *Ann. Stat.* **2013**, *41* (5), 2263–2291.

(114) Feizi, S.; Marbach, D.; Médard, M.; Kellis, M. Network Deconvolution as a General Method to Distinguish Direct Dependencies in Networks. *Nat. Biotechnol.* **2013**, *31* (8), 726–733. https://doi.org/10.1038/nbt.2635.

(115) Margolin, A. A.; Nemenman, I.; Basso, K.; Wiggins, C.; Stolovitzky, G.; Favera, R. D.; Califano, A. ARACNE: An Algorithm for the Reconstruction of Gene Regulatory Networks in a Mammalian Cellular Context. *BMC Bioinformatics* **2006**, *7* (S1), S7. https://doi.org/10.1186/1471-2105-7-S1-S7.

(116) Faith, J. J.; Hayete, B.; Thaden, J. T.; Mogno, I.; Wierzbowski, J.; Cottarel, G.; Kasif, S.; Collins, J. J.; Gardner, T. S. Large-Scale Mapping and Validation of Escherichia Coli Transcriptional Regulation from a Compendium of Expression Profiles. *PLoS Biol.* **2007**, *5* (1), e8. https://doi.org/10.1371/journal.pbio.0050008.

(117) Choi, K.; Medley, J. K.; König, M.; Stocking, K.; Smith, L.; Gu, S.; Sauro, H. M. Tellurium: An Extensible Python-Based Modeling Environment for Systems and Synthetic Biology. *Biosystems* **2018**, *171*, 74–79. https://doi.org/10.1016/j.biosystems.2018.07.006.

(118) Medley, J. K.; Choi, K.; König, M.; Smith, L.; Gu, S.; Hellerstein, J.; Sealfon, S. C.; Sauro, H. M. Tellurium Notebooks—An Environment for Reproducible Dynamical Modeling in Systems Biology. *PLOS Comput. Biol.* **2018**, *14* (6), e1006220. https://doi.org/10.1371/journal.pcbi.1006220.

(119) Fawcett, T. An Introduction to ROC Analysis. *Pattern Recognit. Lett.* **2006**, *27* (8), 861–874. https://doi.org/10.1016/j.patrec.2005.10.010.