Quality Control and Applications of Mass Spectrometry-based Metabolomics: From Cell Research to Large-scale Human Cohort Studies

By

YING ZHANG

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Chemistry

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

Oliver Fiehn, Chair

Carlito B. Lebrilla

John W. Newman

Committee in Charge

2021

Abstract

Metabolomics is an analytical approach for systematic profiling of metabolites in biofluids, cells and tissues. Gas chromatography - mass spectrometry (GC-MS) and liquid chromatography - mass spectrometry (LC-MS) are most popular techniques in metabolomics due to their high-throughput and high sensitivity. Common procedures in metabolomics include study design, sample collection, sample extraction, data acquisition, data normalization, statistical analysis, and biological interpretation. All procedures before data normalization may introduce analytical variations as a hurdle for statistical performance. With the purpose of improving quality assurance in metabolomics for human disease studies, my dissertation work started with applying internal standards and external quality controls for quality assurance in untargeted metabolomics as showed in chapter one. A new tool called Systematical Error Removal using Denoising Autoencoder (SERDA) was developed and the median relative standard deviations (RSD) of the training QC samples was reduced to 4.6% RSD after normalized by SERDA. Then, I performed untargeted metabolomics assays to explore the lipidomic alterations and risk prediction of Type 2 Diabetes (T2D) in the cohort of American Indians from the Strong Heart Family Study (SHFS). Multivariate analysis of lipidomics identified distinct lipidomic signatures that can differentiate high- from low-risk groups. Higher baseline level of 33 lipid species, including triacylglycerols, diacylglycerols, phosphoethanolamines, and phosphocholines, was significantly associated with increased risk of T2D at 5-year follow-up. Because studies in humans usually can only denote risk factors but not biochemical mechanisms, I continued my work on animal and cell studies in chapter three and four. In chapter three I showcase how metabolomics data can be integrated with phenotype information in mouse models to decipher the link between gene functions and metabolism and downstream diseases. Finally, I compared how different technologies can be used to measure isotope flux analyses in bacterial cell samples to detail specific differences in metabolic pathways under anaerobic or aerobic conditions. The comparison results revealed that all three instruments can provide similar biological conclusions with each instrument possessing their own advantages.

Introduction

Metabolomics has evolved from biomarker identification in cross-sectional analysis to predictor discovery in longitudinal research. Nowadays, large sample size studies have emerged for exploring innovative therapies of human diseases. Mass spectrometry (MS) in combination with chromatography has been widely used in metabolomics. Different instruments are required for specific studies based on their selectivity, sensitivity, mass resolution, and dynamic ranges. Both targeted and untargeted platforms are required to capture as complete metabolic profiles as possible for systematical studies of metabolic pathways and allow association with other features (*e.g.*, outcomes from other '*omics'*) from the same samples. To enable such studies, a high quality metabolomic profiling dataset must be obtained. Therefore, my dissertation has a leaning focus on quality assurance and quality controls in metabolomics, especially for large-scale sample size.

With more specific details, chapter one reports on my efforts to develop and validate normalization methods on GC-MS untargeted metabolomics in a very large-scale sample size analysis (> 4,000). Untargeted primary metabolomics on gas chromatography- mass spectrometry (GC-MS) suffers technical errors specifically due the need to increase the volatility of metabolites by chemical derivatizations. The development of advanced normalization method is therefore urgent for GC-MS based metabolomics. Due to the importance of branched-chain and aromatic amino acids in predicting T2D in Caucasian cohorts, I first tested internal standards (ISTD)-based normalization using 16 deuterated amino acids in small sample size experiments. I analyzed the repeatability and reproducibility by *N*-methyl-*N*-*tert*-butyldimethyl-silyl trifluoroacetamide (MTBSTFA) derivatization or *N*-methyltrimethylsilyl-trifluoroacetamide (MSTFA) derivatization on a nominal mass resolution GC-TOF MS, and propyl chloroformate (PCF) derivatization following a Ez:faast^{TM} Kit on a low resolution GC-SQ MS for targeted amino acids measurements. Then I applied ISTD-based normalization, sum-based and QC-based normalization to a total of 413 pooled quality control (QC) samples as training dataset and 413 BioIVT samples as extra validation dataset. Machine-learning tool Systematical Error Removal using Random Forest (SERRF) favors lipidomics

dataset than GC-MS primary metabolomics. SERRF algorithm gave 13.1% median precision based on the pooled quality control samples. However, when the resulting SERRF data model was applied to the BioIVT validation pools, the technical error increased to 34% relative standard deviation (RSD). Then I demonstrated that a newly developed machine learning tool Systematical Error Removal using Denoising Autoencoder (SERDA) method outperformed all other normalization strategies on my very large-scale GC-MS untargeted metabolomics data.

Chapter two depicts my quality assurance actions by incorporating quality controls into a very large-scale sample preparation and untargeted lipidomics assays for predicting the risk of type 2 diabetes (T2D) in American Indians. The International Diabetes Federation (IDF) estimates that 415 million people have diabetes worldwide, 91% of whom are diagnosed with type 2 diabetes mellitus. Diabetes is a metabolic disorder characterized by hyperglycemia resulting from defects in insulin secretion, insulin action, or both. Standard clinical diagnostics for T2D including fasting glucose, fasting insulin, BMI, age, or HbA1c are unsuitable for predicting the onset and development of pre-diabetes and insulin resistance. American Indians, an underrepresented cohort, suffer a much higher prevalence and incidence of T2D than other ethnic groups, and their metabolic changes in developing T2D may be different from other cohorts. I used two different sets of quality controls (QC, one set of commercially available plasma samples from BioIVT, one set of standardized reference materials from NIST) for untargeted lipidomics assay on UPLC-QTOF MS. SERRF was used for normalization as demonstrated before. Then the normalized lipidomics dataset was used for statistical analysis. We identified novel plasma lipids and lipidomic signatures that can predict onset and progression of T2D in American Indians beyond conventional risk factors including BMI, fasting glucose and insulin resistance. These newly identified molecular lipid species are perturbed years before the onset of T2D, and can help identify high-risk individuals who may benefit from early intervention.

Chapter three details my research on 30 mouse knockout lines with targeted and untargeted metabolomic assays to illustrate the sex-genotype interaction effects on plasma metabolites. Mouse models are a

critical tool in biomedical research; about 50% of all funded NIH projects use mice for at least one specific research aim. Understanding the effect of mouse gene dysfunction on metabolism will provide insights that may inform the function(s) of their human orthologs. I investigated the effect of loss-of-function mutations in 30 unique gene knockout (KO) lines on plasma metabolites. Steroids, bile acids, oxylipins, primary metabolites, biogenic amines, and complex lipids were analyzed with dedicated mass spectrometry platforms, yielding 1,035 identified metabolites in male and female KO mice and wildtype (WT) controls. Twenty-two percent of 23,698 KO versus WT comparison tests showed significant genotype effect on plasma metabolites. Fifty-six percent of identified metabolites were significantly different between the sexes in WT mice. Many of these metabolites were also found to have sexually dimorphic changes in KO lines. I also investigated the associations between plasma metabolites and IMPC phenotypes such as open field parameters and flinching response to an unexpected strong auditory stimulus. Strong sexual dimorphism was also found in metabolite-phenotype associations. Those associations may provide valuable insight for understanding gene functions and disease mechanisms. I demonstrate how to link metabolomics to genotypes and (disease) phenotypes. Sexual dimorphism must be considered as an important factor in interpreting the role(s) of a gene in metabolism and disease etiology.

Mechanic pathway was usually studied using *in vitro* models such as cell lines that permit a species-specific, simpler, more convenient, and more microscopic analysis. Chapter four shows my extension to stable isotope enrichment analysis using bacterial cells. Stable isotopic labeling uses non-radioactive isotopes such as ^{2}H, ^{13}C, ^{15}N, ^{18}O, and ^{34}S as labels for MS analysis for metabolite annotation and identification as well as understanding metabolite turnover and flux in organism. Simultaneous incorporation of stable isotopes into large numbers of metabolites can be accomplished using *in vitro* approaches with labeled nutrients as the source. Most often, such analyses are used with gas chromatography and mass spectrometry due to its ease of operation and reproducible mass spectral

databases. Metabolites require chemical derivatization before GC-MS analysis. The bacterium *Rothia mucilaginosa* belongs to the family of *Micrococcaceae* and is present and metabolically active in the airways and sputum of cystic fibrosis patients. In my study, *R. mucilaginosa* cultures were grown in triplicates in artificial-sputum medium spiked with 100 mM [U-$^{13}C_6$] d-glucose under anaerobic and aerobic conditions. I used *N*-methyl-*N*-*tert*-butyldimethyl-silyl trifluoroacetamide to derivatize the functional groups of metabolites to yield volatile TBDMS derivatives. Then I compared the performance of three commercially available instruments (low-resolution GC-SQ MS, low-resolution GC-TOF MS, and high-resolution GC-QTOF MS) on identical samples. Overall, all three GC-MS instruments can be used to perform stable isotope tracing studies for glycolytic intermediates, TCA metabolites, and amino acids, yielding similar biological results, with high-resolution GC-QTOF MS offering additional capabilities to identify chemical structures of unknown compounds that might show significant isotope enrichment in biological studies.

While this dissertation's work spreads across different species including human, mice and bacterium as well as multiple mass spectrometry instruments, all the studies can be taken as a whole to showcase that how metabolomics can be utilized from diverse aspects and how human diseases, or gene functions can be better understood using other *in vitro* or *in vivo* models. The most important side of this work is to address the importance of quality assurance and quality controls in metabolomics for optimal biological interpretation as also initiated by several authoritative organizations including the metabolomics standards initiative in toxicology (MERIT), the European centre for ecotoxicology and toxicology of chemicals (ECETOC), the metabolomics standards initiative (MSI), and the metabolomics quality assurance and quality control consortium (mQACC). The second purpose is to show that metabolome, aimed at the whole set of all metabolites in a biological sample, can only be comprehensively embraced using multiple techniques rather than a single platform/assay. Finally, animal and cell models are essential for human diseases/gene function research and biological differences exist not only between human and other animal

species but also between different human cohorts, metabolomics study must take into consideration of sexual dimorphism as a critical factor. This limited work may be beneficial to other future and related studies in metabolomics.

Acknowledgements

Foremost, I would like to express my deepest gratitude to my advisor Prof. Oliver Fiehn for his excellent guidance, trust and support to my Ph.D. study. He is a pioneer in the field of metabolomics with the attitude and substance of a genius. His enthusiasm in teaching and prudence in scientific research gave me great inspiration. I would like to thank Dr. Tobias Kind and Dr. Tong Shen for their motivation and affirmation at the very beginning of my Ph.D. study/research. Thanks Dr. Dinesh K. Barupal for your kindness and Dr. Brian DeFelice and my undergraduate team members in 2017 for assisting me in my project research. I would also like to express my appreciation to my dissertation committee members Dr. Carlito B. Lebrilla and Dr. John W. Newman, whose work demonstrated to me multiple applications of mass spectrometry. Thank all the members of the Fiehn laboratory, I enjoyed the lively discussions at group meetings very much. Thank you for all your care and mutual help over the past five years. Last but not least, I am fully grateful to my parents, family and friends for their love and encouragement throughout my life. This is the strength of my continuous progress in my future life.

Contents

Chapter 1 Normalization Strategies for Large-scale Untargeted Metabolomics by Gas Chromatography-Mass Spectrometry Analysis

Reproduced from unpublished manuscript: Ying Zhang, Sili Fan, Gert Wohlgemuth, Oliver Fiehn. "Normalization Strategies for Large-scale Untargeted Metabolomics by Gas Chromatography-Mass Spectrometry Analysis". ready for submission to scientific journal.

1.1 Abstract

Large-scale metabolomics assays are widely used in epidemiology for biomarker discovery and risk assessments. However, systematic errors introduced by instrumental signal drifting pose a big challenge in large-scale assays, especially for derivatization-based gas chromatography-mass spectrometry (GC-MS). Here, we compare the results of different normalization methods for a study with more than 4,000 human plasma samples involved in a type 2 diabetes cohort study, in addition to 413 pooled quality control (QC) samples, 413 commercial pooled plasma samples and a set of 25 stable isotope labeled internal standards used for every sample. Data acquisition was conducted across 1.2 years including seven column changes. In total 413 pooled QC (training) and 413 BioIVT samples (validation) were used for normalization comparisons. Surprisingly, neither internal standards nor sum-based normalizations yielded median precision of less than 30% across all 563 metabolite annotations. Yet, the machine-learning based SERRF algorithm gave 13.1% median precision based on the pooled quality control samples. However, when the resulting SERRF data model was applied to the BioIVT validation pools, the technical error increased to 34% relative standard deviation (RSD). We therefore used a newly developed denoising autoencoder method (SERDA). SERDA lowered the median standard deviations of the training QC samples down to 4.6% RSD, yielding an overall error of 19.5% RSD when applied to the independent BioIVT validation QC

samples. This is the largest study on GC-MS metabolomics ever reported, demonstrating that technical errors can be normalized and handled effectively for this assay.

1.2 Introduction

Metabolome is defined as the complete set of low molecular mass compounds (<1500 Da) synthesized or modified by a living cell or organism. Metabolomics is the simultaneous measurement of all small molecular metabolites that participate as substrates, reactants, signaling agents, intermediates, and products of enzyme-mediated reactions.[1] Mass spectrometry-based metabolomics has matured as a high-throughput, high-resolution, high-dimensional technique that identifies multiple metabolite markers present in significantly different abundances between different conditions in large human cohort studies, enabling the discovery of diagnostic and predictive metabolite levels for disease.[2-3]

Metabolomics can be integrated with transcriptomics and proteomics to find biomarkers of diseases or to elucidate biological mechanisms. For both goals, high-quality data mining is needed that removes unwanted (technical) variance. Such technical variance, is impacted by various forms of unwanted variations in conducting laboratory experiments, from batch-to-batch differences, variation between different instruments, inter-person variation, and drifts in instrument sensitivity across a specific sequence of samples.[4] To extract the biologically relevant information, such technical variance needs to be efficiently removed by data normalization methods after raw-data acquisition. Classic quantification strategies in analytical chemistry employ exogenous chemical surrogates as quality controls and for normalization against matrix effects, using either stable isotope labeled chemicals (deuterium or ^{13}C labeled) or structural analogs of target molecules. Because metabolomics aims to analyze ‘all’ metabolites, the use of internal

standards certainly faces limitations due to the complexity and differences of metabolomics mixtures. Overall Metabolomics normalization have evolved in the past two decades from scaling normalizations,[5-6] use of housekeeping metabolites,[4] normalization based on internal or external standards,[7-9] and quality control samples (QC)-based normalizations.[10-13]

Specifically, QC-based normalization methods are favored today.[14] Systematic error removal using random forest (SERRF) normalization has been shown to outperform classic QC-normalizations such as locally estimated scatterplot smoothing (LOESS) in large-scale untargeted lipidomics.[13, 15] However, no such analysis has been conducted for GC-MS based untargeted metabolomics. Interestingly, GC-MS based metabolomics studies typically are much smaller in size than LC-MS based studies, usually with fewer than 1,500 samples.[7, 16-19] Untargeted primary metabolomics on gas chromatography- mass spectrometry (GC-MS) suffers technical errors specifically due the need to increase the volatility of metabolites by chemical derivatizations. In addition, involatile materials may accumulate in the GC-injection liners and the beginning of chromatography columns. Such deposits may alter the local catalysis environment for the delicate balance of derivatization products.[20]

GC-MS is an ideal platform to detect volatile compounds. For primary metabolites with higher boiling points, a derivatization step reduces boiling points by exchanging acidic hydrogens against derivatization groups. For chemical derivatizations, a wide array of strategies and reagents can be employed, ranging from alkylations, acylations to silylations and others.[21-23] For example, alkylations use boron-trifluoride/butanol or dimethylformamide dimethylacetals,[24-26] silylation by *N,O*-Bis(trimethylsilyl)-trifluoroacetamide, *N*-methyltrimethylsilyl-trifluoroacetamide (MSTFA) or *N*-Methyl-*N*-*tert*-

butyldimethylsilyltrifluoroacetamide (MTBSTFA), acylation by propyl- or ethylchlorofomate, acetic anhydride or fluorinated anhydrides, or chiral derivatization reactions. Among silylating agents, trimethylsilylations are most frequently used in metabolomics, with MSTFA being the most widely utilized agent [27] due to its ease in handling and wide range of substrates encompassing hydroxyl-, carboxyl-, amino- or thiol- functional groups. In contrast, other derivatization agents are hampered by less convenient operation and narrower metabolite ranges. For example, boron-trifluoride in alkylation and anhydrides in acylation are corrosive, flammable and highly toxic.

We here investigate and compare different quality control strategies for metabolomics of human plasma, including derivatization agents, internal standards, external quality control samples and computational modeling (**Figure 1.1**). We compared three derivatization agents with deuterated internal standards in three different trials across three months. We then applied two different external quality controls for a type 2 diabetes study of > 4,000 human plasma samples, a QC pool made of extracts of the cohort samples and another QC pool that was commercially available. A new modeling tool called systematic error removal using denoising autoencoder (SERDA) showed an overwhelmingly better performance than SERRF in large-scale GC-MS based metabolome dataset. We further compared SERDA with other traditional normalization methods (e.g., mTIC, fTIC, iTIC, metabolite-ISTD ratio) and investigated the performance by combining different normalization methods.

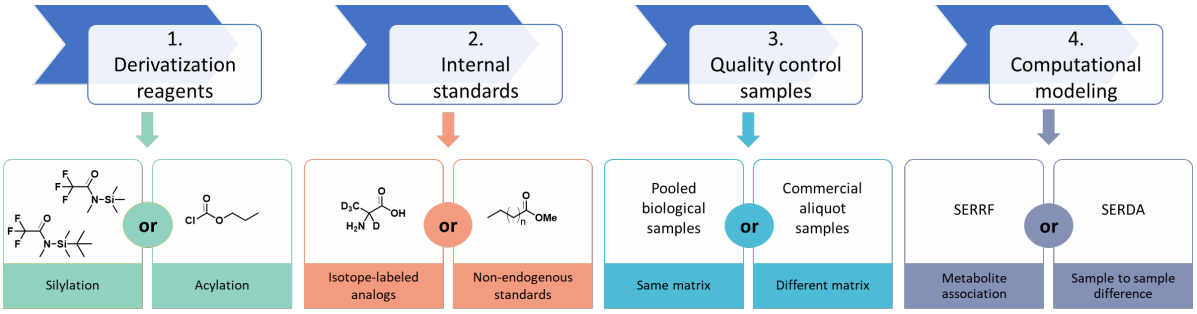


Figure 1.1. Overview of study plan for normalization in a very large-scale human cohort plasma samples.

1.3 Methods

1.3.1 Reagents

Pooled disodium EDTA plasma was purchased from BioIVT (Westbury, NY), aliquoted into portions of 30 µL and stored at -80 °C freezer until extraction. The EZ:faast^{TM} Amino Acid Analysis sample testing kit for propyl-chloroformate (PCF) derivatization was purchased from Phenomenex Inc (Torrance, CA, USA). 4,104 dipotassium EDTA plasma samples were obtained from study participants of a large-scale human cohort for diabetes risk factor analysis. Samples were extracted as published previously [28-29] and aliquoted into analytical samples and backup extracts. 1,032 backup extracts were merged, homogenized and aliquoted as QC samples. To match the cohort plasma matrix, dipotassium EDTA plasma was purchased from BioIVT (Westbury, NY) and used as validation sample set.

HPLC grade extraction solvents methanol, methyl-tertiary butyl ether (MTBE) and water were obtained from Sigma-Aldrich (Dorset, U.K.). Twenty-five deuterium labeled amino acids were purchased from Cambridge Isotope and were used as internal standards in the extraction solutions for the human cohort study. The following concentrations were added to plasma: alanine-d4 (400 mM), arginine-d7 (110 mM), asparagine-d3 (100 mM, aspartic acid-d3 (50 mM, glutamic acid-d5 (150 mM), glutamine-d5 (600 mM), glycine-d5 (400 mM), histidine-d5 (150 mM), homocysteine-d4 (100 mM), isoleucine-d10 (100 mM), leucine-d10 (250 mM), lysine-d8 (200 mM), methionine-d5 (60 mM, ornithine-d2 (100 mM), phenylalanine-d8 (100 mM), proline-d7 (200 mM), serine-d3 (150 mM), threonine-d5 (200 mM),

tryptophan-d8 (80 mM), tyrosine-d7 (100 mM), valine-d8 (400 mM), 2-aminobutyric acid-d6 (40 mM), 2-hydroxybutyric acid-d3 (60 mM), 3-hydroxybutyric acid-d4 (100 mM), sorbitol-d8 (50 mM). Only 16 of these amino acids were used in the initial derivatization normalization tests.

1.3.2 Sample preparations for GC-MS

For untargeted analyses, plasma samples were extracted using Matyash liquid-liquid extraction method with cold methanol/ MTBE/water.[28] 40 µL of aliquoted plasma was thawed to room temperature and kept on ice during the following steps. Samples were vortexed for 10s with 225 µL of ice-cold methanol, followed by adding 750 µL ice-cold MTBE. Samples were shaken for 6 min at 4°C. 188 µL of room temperature water containing the internal standards given above was added. Samples were vortexed for 20 s followed by centrifugation at 12,210 × *g* for 2 min. The lipophilic phase was decanted. The remaining hydrophilic phase was transferred to a new Eppendorf tube, dried down under vacuum and used for derivatization.

For silylations, derivatization started using 10 µL of methoxyamine hydrochloride in pyridine (40 mg/mL, with 5 ug/mL sorbitol-d8) and shaken at 30°C for 90 min. Trimethylsilylation was performed by 90 µL N-methyl-N-(trimethylsilyl) trifluoroacetamide (MSTFA) containing C8-C30 fatty acid methyl esters (FAMEs) at 37°C for 30 min. For derivatization with tertiary butyl-dimethylsilylation, 90 µL of MTBSTFA containing C8-C30 fatty acid methyl esters (FAMEs) was used at 80°C for 30 min. Samples were centrifuged at 12,210 × *g* for 2 min and transferred to crimp top vials for GC-TOF MS detection.

For targeted derivatization of amino acids in 100 µL plasma sample using propyl-chloroformate (PCF), samples were prepared by solid phase extraction method as described previously[30] following manufactures instructions for the EZ:faast™ Amino Acid Analysis sample testing kit.

1.3.3 Gas chromatography / mass spectrometry conditions

Each mass spectrometer was coupled to an Agilent 7890 GC system (Santa Clara, CA). For silylated samples, a Restek (Bellefonte, PA) RTX-5Sil MS column was used (30m length, 0.25 mm i.d, 0.25 µm df,

95% dimethyl/5%diphenyl polysilox-ane film) with an additional 10m guard column. The oven temperature was held at initial temperature at 50°C for 1 min, increased at 20°C/min to 330°C, and kept isothermal for 5 min. The injection temperature was 275°C. The injection volume was 0.5 μL in splitless mode. Silylated samples were measured on a LECO Pegasus IV TOF MS (St. Joseph, MI, USA) at +70 eV, source temperature 250 °C; scan range 85-700 *m/z*; sampling rate 17 Hz.

For PCF-derivatized amino acids, a Zebron™ ZB-AAA GC column (10m length, 0.25mm i.d.) was used. Carrier gas (helium) flow rate was kept constant at 1.5 mL/min (60kPa). The initial oven temperature was held at 110°C, and then ramped at 30°C/min from 110° to 320°C with no final hold. The injection temperature was 250°C. The injection volume was 2.0μL at a split ratio of 1:15. PCF derivatized amino acids were analyzed by a low-resolution Agilent 5977 single quadrupole MSD (Santa Clara, CA, USA) at +70 eV, source temperature 240 °C, quadrupole temperature 180 °C, scan range 45-450 *m/z*, sampling rate 4 Hz.

1.3.4 Data processing

For silylated samples, raw data were deconvoluted by the Leco instrument software ChromaTOF version 4.5. For silylated samples, deconvoluted data were submitted to the BinBase database for alignment and compound identification. Data files for PCF derivatized amino acids were processed using MassHunter Quantitative Analysis B.07.00 version.

1.3.5 Data normalization

Data of PCF-derivatized amino acids were normalized to corresponding internal standards as described previously[30] and are named ISTD normalization. For MTBSTFA-derivatized amino acids, the same ISTD normalization method was used by internal standards. In addition, three sum-normalization methods were compared: (a) raw amino acid peak intensities were normalized to the sum of all deuterated internal standards, called iTIC. (b) Secondly, data were normalized to the sum of all retention time marker compounds (fatty acid methyl esters), called fTIC. (c) Thirdly, data were normalized to the sum of all

identified metabolites, here: amino acids, hydroxyl acids and related compounds, called mTIC. The same methods were used to compare the normalization of trimethylsilylated samples in the initial comparison of derivatization methods. In addition, human cohort samples that underwent trimethylsilylation derivatization were normalized by two methods using quality control samples (QC): (1) SERRF (Systematic Error Removal using Random Forest)[13] and a new method that we present here (2) SERDA (Systematic Error Removal using Denoising Autoencoder). Statistical analyses were performed by Friedman nonparametric paired tests with adjusted Dunn's significance thresholds of $p < 0.0332$ in GraphPad Prism 8.4.3.

1.4 Results

1.4.1 GC-MS based metabolomics: data normalization for small sample sets

We first tested the two most common silylation reactions, trimethylsilylation (TMS) and tertiary-butyl dimethylsilylation (TBDMS), and compared these broad-range, untargeted reagents against a commercially available targeted assay for amino acid quantifications by chloroformate reaction. The broad-range silylation agents produced products with unstable ratios for primary amines, introducing unwanted variances in untargeted metabolomics. For example (**Figure 1.2**), trimethylsilylation usually generates two trimethylsilylated valine products, valine 1TMS with only the carboxyl acidic hydrogen replaced by TMS (**Figure 1.2a**, *m/z* 156), or valine 2TMS with one hydrogen of amine group and carboxyl proton replaced by TMS (**Figure 1.2b**, *m/z* 144). In principle, isotope labeled internal standards should correct for such difficulties in stabilizing reaction conditions and yield exactly the same TMS-derivatization ratios for amino acids. We used 16 isotope-labeled metabolite analogs and spiked them into the extraction solution to correct for all technical variations as their corresponding metabolites, from extraction, to derivatization, injection to the gas chromatograph and mass spectrometry. We found that the product ratios of *N,O*-TMS derivatized

amino acids to only O-TMS derivatized amino acids varied between pooled QC plasma samples despite all measures of pre-analytical quality controls such as regularly cutting columns, cleaning injectors, or exchanging injector needles.[20] As expected, internal stable isotope standards reduced this technical error. For example (**Figure 1.2a**), the two trimethylsilylated products valine-d8 1TMS (*m/z* 164) and valine-d8 2TMS (*m/z* 152) displayed similar ratios between two QC samples as the endogenous valine 1TMS and valine 2TMS products (**Figure 1.2a,b**). However, this control for unwanted technical variation did not completely eliminate technical errors when we compared trimethylsilylation to tertiary-butyl dimethylsilylation and targeted chloroformate derivatization (**Figure 1.2c**). For this initial comparison, we used 30 commercial plasma samples that were extracted in three independent replicate studies, each conducted on month apart (**Figure 1.3, Supplementary Table 1.1**). We limited the analysis to 16 amino acids that were detectable in all three derivatization methods. For examples, arginine was not amenable to any of the methods, while MSTFA was not yielding detectable signals for histidine and cysteine in the plasma samples analyzed here, due to lower sensitivity. Before normalization, the raw data of all three

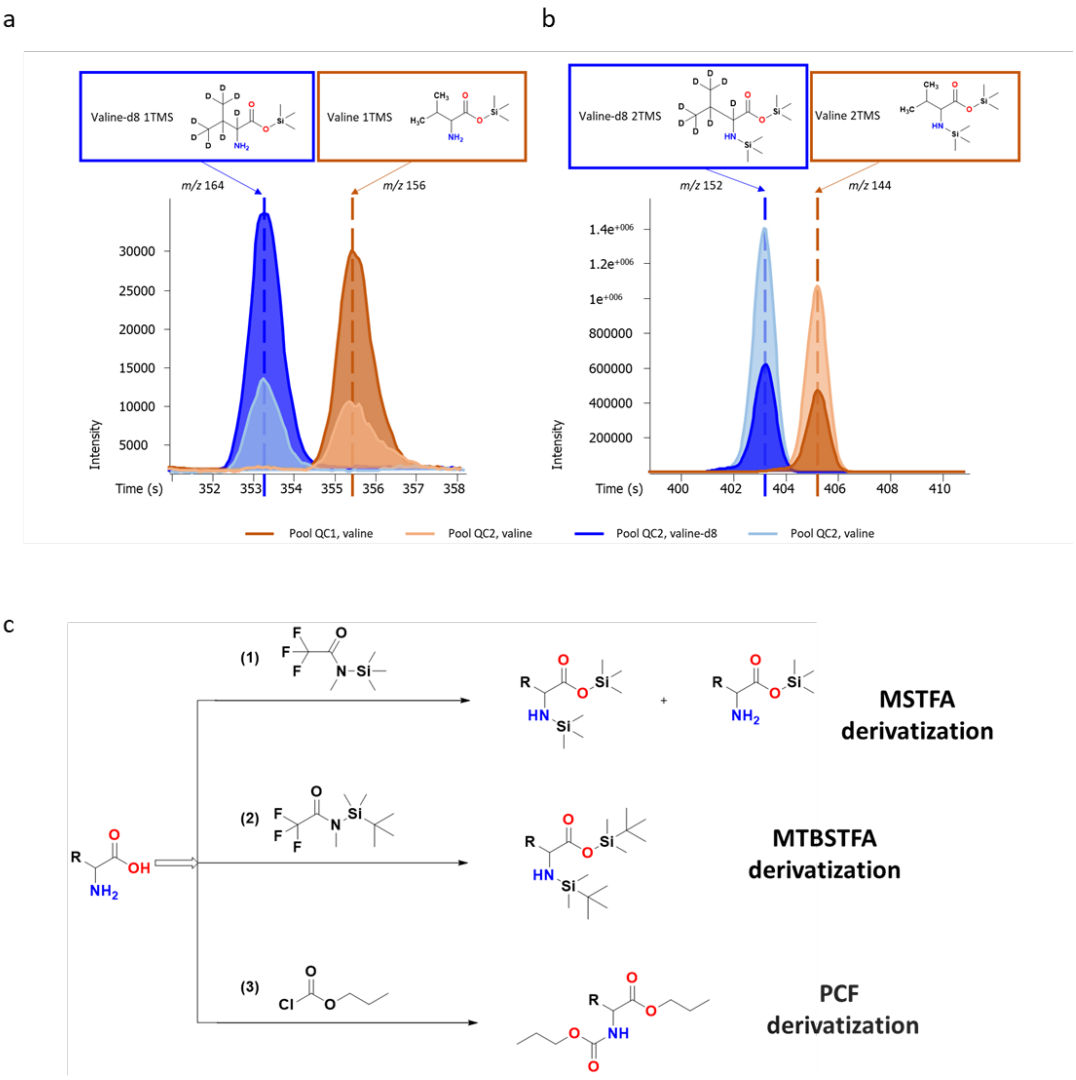


Figure 1.2. Reaction schemes of MSTFA, MTBSTFA and PCF derivatization, and the chromatography of valine derivatized with MSTFA. a) valine-1TMS and valine-d8-1TMS products. b) valine-2TMS and valine-d8-2TMS products. c) reaction schemes for MSTFA, MTBSTFA and PCF derivatization of amino acids.

derivatization methods showed significant variance between the three independent analyses (**Figure 1.3**). Yet, average raw data precision worsened from PCF to MTBSTFA to MSTFA, possibly due to the removal of matrix effects when using PCF derivatization under the Ez:faast protocol that uses a solid phase extraction method. Even after normalization to each individual stable-isotope labeled amino acids, PCF

derivatizations gave the lowest precision with 2.7% average coefficient of variance (CV), followed by MTBSTFA at 8.9% CV and MSTFA at 9.6% CV. Although using internal standards for normalization reduced overall systematical errors for three tested derivatization agents, residual variance was found across all amino acids, likely due to a random combination of all analytical errors, ranging from pipetting to extraction, moisture during derivatization and instrument performance.

In metabolomics, such precision values are regarded as acceptable. However, metabolomics aims at analyzing a wide range of compounds, with the coverage of compound classes decreasing from MSTFA to MTBSTFA to PCF derivatization. One cannot include internal standards for all possible small molecule identifications in GC-MS based metabolomics. Therefore, we tested this data set whether other normalization methods might yield acceptable results for MSTFA or MTBSTFA derivatization. To this end we used three sum-based normalizations: (a) the sum of all internal isotope labeled standards (total ion chromatogram, iTIC), (b) the sum of all 13 fatty acid methyl esters that are added as retention index markers in our protocol (fTIC) and (c) the sum of all identified metabolites (mTIC) (**Supplementary Table 1.1**). Interestingly, the mTIC normalization worked better than iTIC or fTIC for correcting errors for the 16 amino acids for both methods, with 13.5% CV for MSTFA and 8.3%CV for MTBSTFA (**Figure 1.3**). Both derivatization methods showed little improvement in precision when using fTIC normalization in comparison to the raw data, possibly because fatty acid methyl esters did not undergo any derivatization but only account for random errors during injection. In comparison, iTIC normalizations yielded slightly better precisions for both MSTFA and MTBSTFA than fTIC because the individual amino acids showed similar error trend as all amino acids as a group. Nevertheless, mTIC should be regarded as best sum

normalization method for untargeted GC-MS analyses for small data sets like this, especially for MTBSTFA.

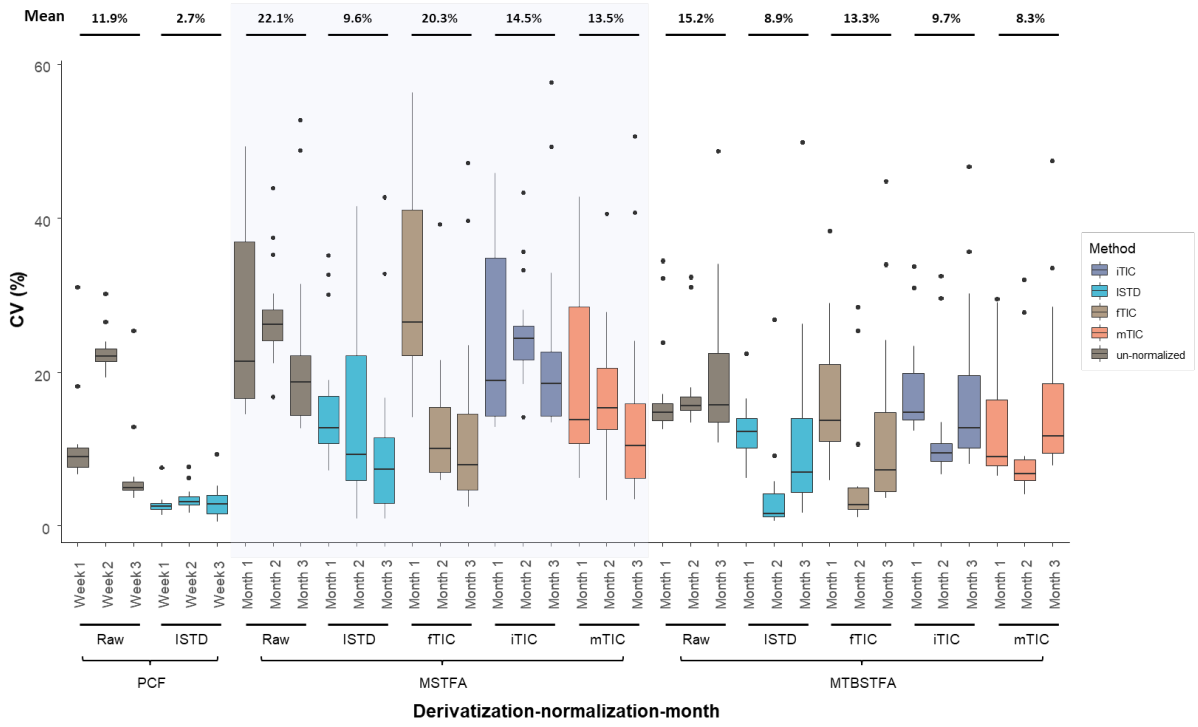


Figure 1.3. Coefficient of variance of 16 amino acids by three derivatization reagents.

1.4.2 GC-MS based metabolomics: data normalization for large and very large sample sets

Most published GC-MS based metabolomics studies use fewer than 100 samples. Only a single study has been published with almost 1,200 samples[16] using relatively matrix-poor tobacco leaf extracts. Apart from instrument drifts, differences in the types and amounts of involatile residues in biological matrices (such as complex lipids or incomplete removal of proteins) may cause additional technical errors in GC-MS based metabolomics. We here used human K_2EDTA plasma samples as example of a matrix that is highly

enriched in fat and protein contents. Such plasma samples are most often used in very large clinical and epidemiological cohort studies that makes this sample type very relevant to be studied with respect to residual technical errors. We used 25 stable-isotope labeled metabolites during the extraction to investigate if such classic internal standards could be used beyond their corresponding unlabeled endogenous metabolites, to correct for drifts during data acquisition and reduce technical (random) errors for the metabolome at-large. In addition, we employed four further types of quality control samples to improve analytical precision: (a) From a cohort of 4,104 human K_2EDTA plasma samples, we pooled half of the extracts obtained from the first 1,032 study samples and aliquoted this pool into 413 cohort-derived quality control (QC) samples. These QC samples were used for data normalization for MSTFA-derivatization based GC-TOF MS metabolomics that possess the widest range of metabolite coverage including amino acids, bioorganic acids, sugars, hydroxyl acids and fatty acids. Pool QC samples were added after each subset of 10 clinical cohort samples. (b) Secondly, we added one method blank and one commercial BioIVT K_2EDTA plasma sample as independent secondary quality controls for validation purposes. (c) Third, NIST SRM1950 human plasma QC samples[31-35] were added after each set of 40 human cohort samples. (d) Fourth, we further analyzed a total of 102 technical replicate samples that were used within a single set of 80 samples (**Figure 1.4a**).

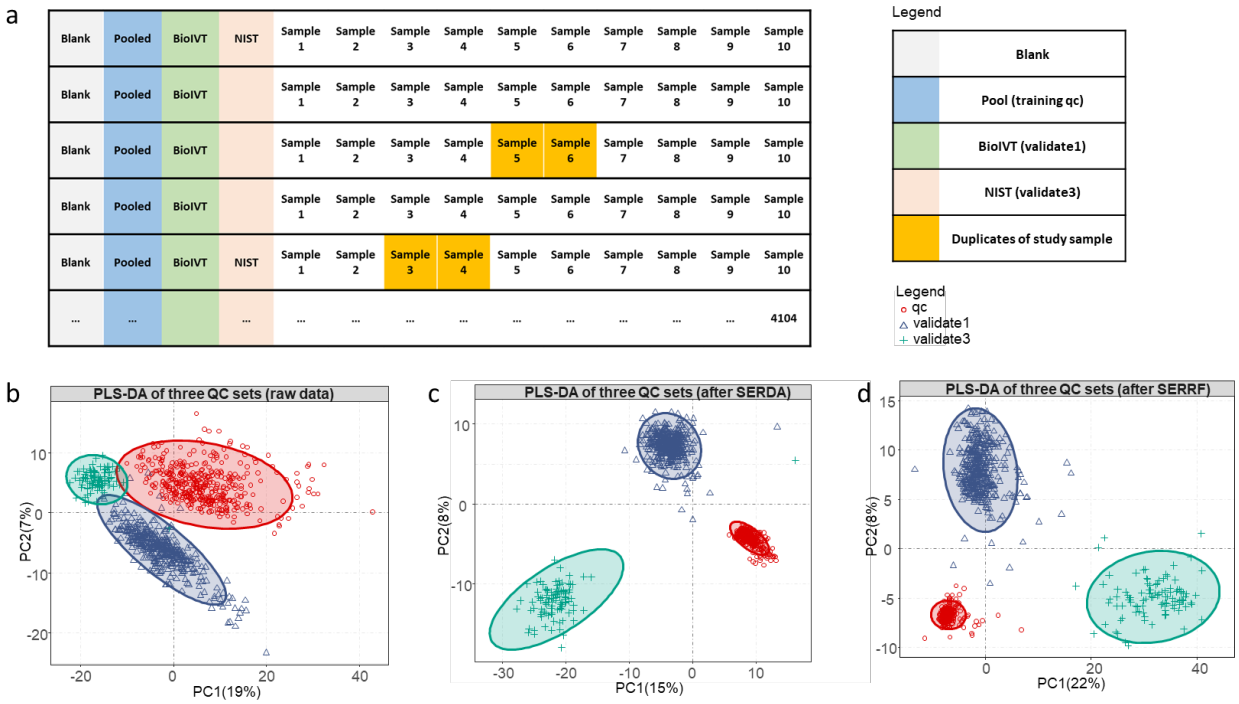


Figure 1.4. Sequence of sample acquisition and distribution of three QC sets after GC-MS metabolomics.

Table 1.1. Comparison of different normalization methods using pooled extracts as training QC set and BioIVT as validation QC set.

Statistics for %CV	Raw		SERDA Pool every 10		SERRF Pool every 10		fTIC		iTIC		mTIC		SERDA Pool every 10 + fTIC		SERDA Pool every 10 + mTIC		SERDA Pool every 10 + iTIC	
	Pool QC	BioIVT validate1	Pool QC	BioIVT validate1	Pool QC	BioIVT validate1	Pool QC	BioIVT validate1	Pool QC	BioIVT validate1	Pool QC	BioIVT validate1	Pool QC	BioIVT validate1	Pool QC	BioIVT validate1	Pool QC	BioIVT validate1
Stdev	122.3%	89.8%	14.3%	15.1%	11.2%	94.6%	111.9%	91.6%	116.3%	89.5%	114.3%	101.7%	14.3%	15.2%	14.2%	16.0%	14.2%	15.0%
Mean	83.7%	73.5%	6.8%	23.3%	15.2%	53.4%	74.2%	67.3%	82.5%	79.7%	75.0%	83.1%	7.0%	23.6%	7.2%	25.1%	6.9%	24.8%
Median	59.7%	56.3%	4.6%	19.6%	13.1%	34.0%	53.1%	51.1%	59.1%	62.8%	53.1%	59.5%	4.7%	19.9%	4.9%	21.2%	4.7%	21.3%

Data acquisition was conducted across 1.2 years in 7 batches with many column cuts and >60 injection liner exchanges in addition to 8 column changes and instrument autotunings following the detailed recommendations published earlier.[20] Due to these frequent, but necessary interventions, raw pooled QC data showed large technical variations. We first tested the three sum-normalization methods used in the small amino-acid derivatization method sets above (fTIC, mTIC and iTIC). As expected, none of the classic

sum-based normalization methods yielded acceptable precisions for such large-scale studies, with unacceptably high median technical errors between 53-63% for both cohort pool QC and commercial plasma QC samples (**Table 1.1**). Next, we used a machine-learning based data normalization method that we previously successfully used for large-scale lipidomics data with more than 5,000 samples (Systematic Error Removal by Random Forest, SERRF).[13] SERRF uses correlation patterns of signal drifts of multiple metabolites in QC samples to determine correction factors that are then applied to the biological samples. SERRF avoids overcorrection by using any single metabolite, unlike the classic LOESS algorithm ('locally estimated scatterplot smoothing).[36] When applied to GC-TOF MS cohort pool QC samples, SERRF indeed greatly reduced the median technical error to only 13.1% CV (**Table 1.1**), clearly below the margin of 30% CV that had been proposed for metabolomics.[37] Correspondingly, supervised classification of the cohort pool QC, commercial pool QC and NIST plasma pool QCs showed a large shrinkage of the data dispersion for the training data (cohort pool QC) compared to the raw data (**Figure 1.4b,d**). However, when the SERRF model was applied to the primary validation BioIVT commercial plasma QC samples, data still showed considerable dispersion (**Figure 1.4d**) and a median 34% CV (**Table 1.1**). In comparison to the success of SERRF in lipidomics, this diminished normalization power in GC-TOF MS metabolomics may be due to a higher random effect on absolute intensities (trimethylsilylation ratios, **Figure 1.2a,b**). In contrast, no chemical derivatization is required in lipidomics that therefore only has to be corrected for signal drift patterns due to systematic errors that occur gradually across hundreds of samples in a continuous way. To better correct for random effects that may be caused by derivatizations or the less controllable splitless injection procedure in gas chromatography, we developed and applied a new normalization method, Systematic Error Removal by Denoising Autoencoder (SERDA). Denoising autoencoders are used

to recognize signals despite large but random noise. Denoising autoencoders first randomly hide some features from input data and automatically generate a new dataset that is similar to the input data.[38] After capturing the useful information interrupted by noise signals by iterative neural network machine learning, the tool generates a reconstructed output with the same shape as the input data in the decoder stage. This initially graph-structured domain knowledge can also be used for developing drug combinations therapy for disease treatment,[39] showing that this algorithm can be adopted for extracting information in a complicated chemical or biological combinatorial space. Hence, we assumed that autoencoders may be better used for data normalization in untargeted GC-MS metabolomics that show high random variance in addition to systematic drifts. When we applied SERDA to the GC-TOF MS untargeted metabolomics samples, we found this new normalization technique to greatly outperform the SERRF algorithm. For training cohort QC samples, a residual median technical error of 4.6% CV was achieved (**Table 1.1**) along with low dispersion in multivariate clustering of the three types of QC samples (Figure 4c). Even more importantly, when the SERDA model was applied to the independent commercial BioIVT QC samples, the data dispersion in multivariate cluster was smaller than with SERRF data (**Figure 1.4c,d**) and overall median errors after SERDA modeling were found at 19.6% CV, well within the acceptable limits in untargeted metabolomics. When we tested the effect of sum-normalization on top of SERDA modeled data, neither mTIC, fTIC nor iTIC normalizations further improved the residual technical errors of SERDA data (**Table 1.1**).

Next, we analyzed the effect of SERDA on 102 biological duplicate samples that were interspersed into the data acquisition of the total of 4,104 cohort study samples (**Figure 1.5, Supplementary Table 1.2**). 48

biological replicates were measured in adjacent positions within GC autosamplers (Figure 4a) while 54 additional biological replicates were measured apart within a set of 80 samples. Across all metabolites, correlation coefficients for both adjacent and non-adjacent biological replicates were found at excellent r_{xy} = 0.98 with ranges of r_{xy} = 0.81 to 1.0 for non-adjacent replicates that were only slightly worse than adjacent replicates with ranges of r_{xy} = 0.90 to 1.0 (**Figure 1.5, Supplementary Table 1.2**). This data showed that the SERDA algorithm correctly normalized metabolite intensities in biological replicates for both high- and low-abundant compounds.

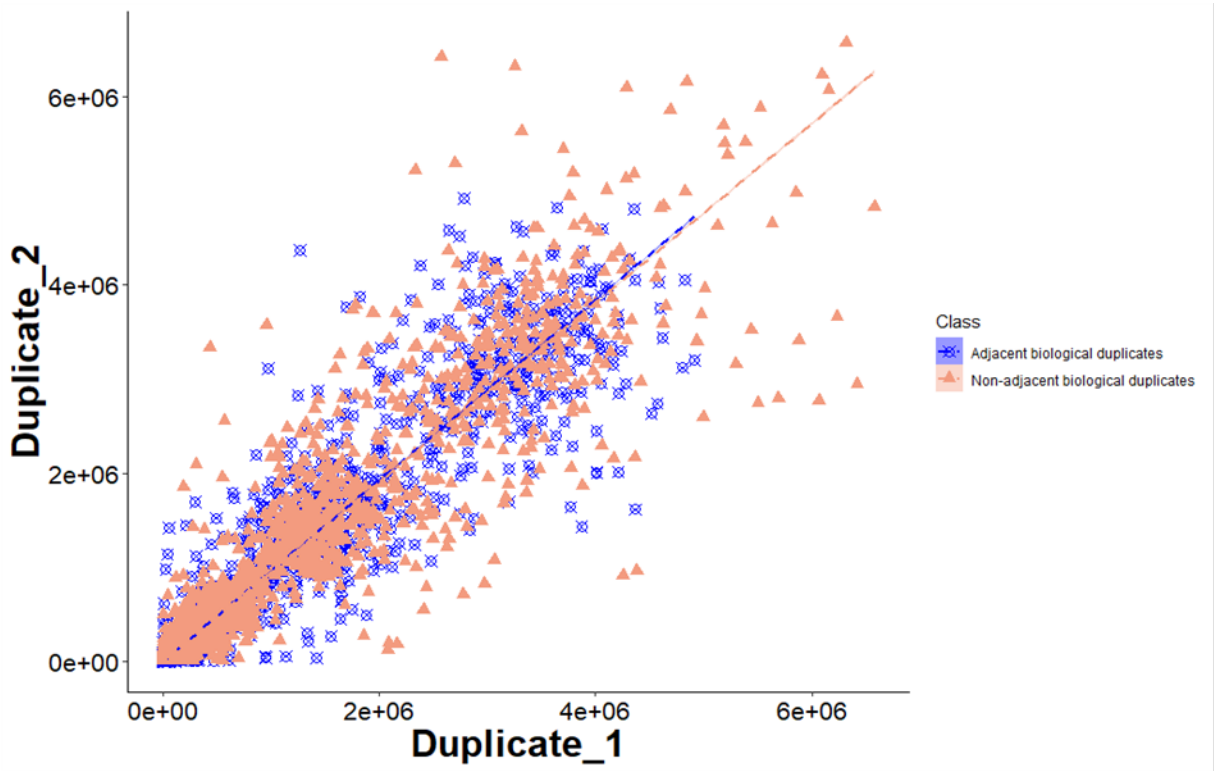


Figure 1.5. Correlation of extra biological duplicates after normalized by SERDA.

Table 1.2. Reduction of QC numbers in SERDA normalization.

Statistics for %CV	SERDA Pool every 10			SERDA Pool every 20			SERDA Pool every 40			SERDA Pool every 80		
	Pool QC	BioIVT validate1	Pool validate2	Pool QC	BioIVT validate1	Pool validate2	Pool QC	BioIVT validate1	Pool validate2	Pool QC	BioIVT validate1	Pool validate2
Stdev	14.3%	15.1%	*NA*	5.0%	19.5%	35.2%	1.0%	31.6%	58.3%	0.8%	46.5%	65.8%
Mean	6.8%	23.3%	*NA*	4.6%	27.5%	25.0%	2.9%	33.8%	33.4%	2.6%	39.8%	41.4%
Median	4.6%	19.6%	*NA*	3.8%	22.8%	19.1%	2.7%	26.2%	23.1%	2.4%	29.1%	27.0%

We then investigated how the number of training cohort QC samples affected the overall efficiency of the SERDA algorithm. To this end, we used fewer cohort QC samples for training and then applied the SERDA models on the remaining cohort QC pool samples and the commercial BioIVT QC pools. Median technical errors for commercial BioIVT pools worsened from 19.6% CV when using training QC samples after each set of 10 biological samples to 22.8% CV, 26.2% CV and 29.1% CV when using training QC samples after each set of 20, 40 or 80 biological samples (**Table 1.2**). Technical errors for the remaining cohort QC pool followed the same trend when using SERDA models for each set of 20, 40 or 80 biological samples (**Table 1.2**). Interestingly, technical errors for cohort QC pool samples remained 2.1-3.7% CV lower when used as testing samples than errors obtained by the BioIVT commercial pool samples. This observation can be explained by the slight differences in biological matrix compositions between the plasma QC pool of a specific age and ethnicity composition of a human cohort study, and the composition of commercial BioIVT samples. Based on this data, we recommend using one cohort pool QC sample for every 10 biological samples as it yielded significantly better precision than using fewer cohort QC pool sample. We also

advocate for using secondary test QC samples as shown here by commercial BioIVT QC plasma samples to independently test for the overall effect of the SERDA models.

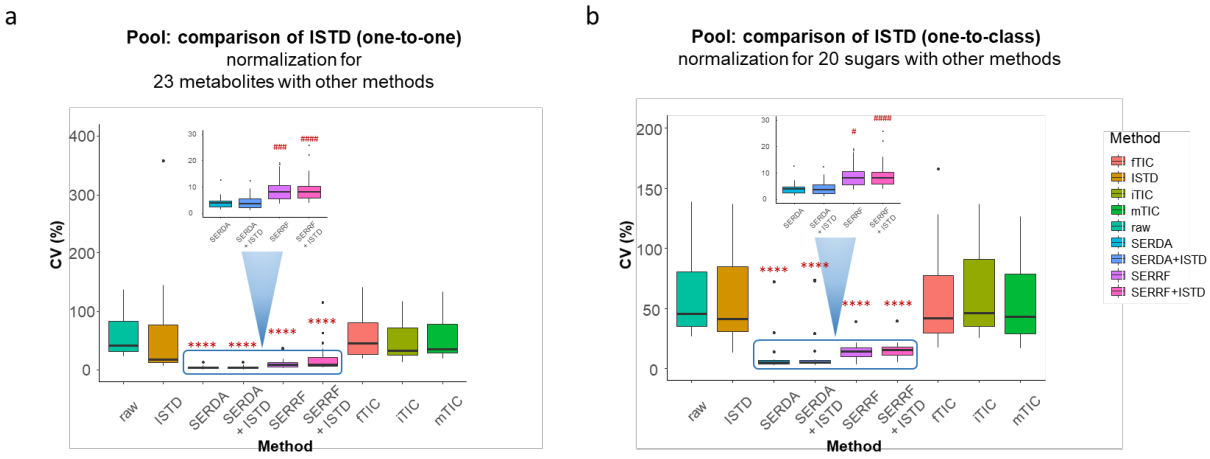


Figure 1.6. Comparison of ISTD absolute ratio normalization with QC-based and TIC-based normalization methods. Friedman nonparametric test was used for significance comparison with raw. P value threshold: 0.1234 (ns), 0.0332 (*), 0.0021 (**), 0.0002(***), 0.0001(****). Or Friedman nonparametric test was used for significance testing comparing to SERDA. P value threshold: 0.1234 (ns), 0.0332 (#), 0.0021 (##), 0.0002(###), 0.0001(####). One-to-one: absolute ratio was calculated by dividing the peak intensity of endogenous metabolite by corresponding deuterated ISTD. One-to-class: absolute ratio was calculated by dividing the peak intensity of endogenous metabolite by a single deuterated compound as an analog ISTD for the entire class.

Last, we investigated how SERDA normalization compared to the use of 25 internal standards (ISTD) that were spiked into the plasma extraction solution. For 23 of these compounds, the corresponding unlabeled endogenous plasma metabolites levels were detected by GC-TOF MS metabolomics; two additional internal standards (homocysteine-d4 and sorbitol-d8) were structurally similar to endogenous metabolites (cysteine and blood sugar alcohols). When using the 23 ISTDs to normalize one-to-one to their endogenous plasma counterparts, a median 19.1 %CV was found with proline and histidine as outliers (**Figure 1.6a**). For the same 23 endogenous compounds, SERDA normalization achieved a much lower technical error of

3.8% CV, and even SERRF outperformed the use of internal standards with 5% CV (**Figure 1.6a**). The same trends were observed for both cohort pool QC samples and commercial BioIVT plasma QC samples (**Figure 1.6a, Supplementary Figure 1.1a**). Use of sum normalization methods (mTIC, fTIC, iTIC) worsened the performance of the SERRF or SERDA methods alone, and did also not outperform the one-on-one use of internal standards (**Figure 1.6a**). When we used sorbitol-d8 as sole internal standard surrogate for a class of 20 detected sugars, sugar alcohols, or disaccharide, this 'one-to-class' normalization failed to improve the technical errors for these 20 sugars whereas SERDA still outperformed SERRF for carbohydrates for both cohort pool QCs and commercial BioIVT QCs (**Figure 1.6b, Supplementary Figure 1.1b**).

1.5 Conclusions

In summary, we here present the results of different normalization methods for GC-MS based metabolomics studies in human plasma. While MSTFA derivatization showed higher technical errors for amino acids than alternative MTBSTFA or PCF derivatizations, due to the much broader coverage of plasma metabolite annotations for trimethylsilylated compounds, MSTFA is still the reagent of choice for GC-MS profiling. Yet, due to both random and systematic errors in large-scale human plasma cohort studies, the technical errors in GC-MS based screening of primary (polar) metabolites are harder to control than in LC-MS based lipidomics.[13] Neither classic internal standards nor sum-based normalizations were sufficient to correct for GC-MS drifts, and even the machine-learning based SERRF method did not yield satisfying results. Yet, using denoising autoencoder algorithms implemented in the SERDA random neural networks presented here reduced the median residual errors to less than 20% CV, making the data useable for epidemiological statistics. To match blood matrix effects in the best way possible, we strongly recommend using quality control samples pooled from the exact same samples as the human cohort study. Commercial plasma QC or NIST SRM1950 QC samples should only be used as independent test samples to estimate the overall residual technical errors. The use of internal standards should be limited to estimate absolute concentration of specific plasma metabolites but should not be used for broad-scale normalization schemes.

1.6 References

1. Sakaguchi, C. A.; Nieman, D. C.; Signini, E. F.; Abreu, R. M.; Catai, A. M., Metabolomics-Based Studies Assessing Exercise-Induced Alterations of the Human Metabolome: A Systematic Review. *Metabolites* **2019,** *9* (8).

2. Wishart, D. S.; Tzur, D.; Knox, C.; Eisner, R.; Guo, A. C.; Young, N.; Cheng, D.; Jewell, K.; Arndt, D.; Sawhney, S.; Fung, C.; Nikolai, L.; Lewis, M.; Coutouly, M. A.; Forsythe, I.; Tang, P.; Shrivastava, S.; Jeroncic, K.; Stothard, P.; Amegbey, G.; Block, D.; Hau, D. D.; Wagner, J.; Miniaci, J.; Clements, M.; Gebremedhin, M.; Guo, N.; Zhang, Y.; Duggan, G. E.; Macinnis, G. D.; Weljie, A. M.; Dowlatabadi, R.; Bamforth, F.; Clive, D.; Greiner, R.; Li, L.; Marrie, T.; Sykes, B. D.; Vogel, H. J.; Querengesser, L., HMDB: the Human Metabolome Database. *Nucleic Acids Res* **2007,** *35* (Database issue), D521-6.

3. Zeki, O. C.; Eylem, C. C.; Recber, T.; Kir, S.; Nemutlu, E., Integration of GC-MS and LC-MS for untargeted metabolomics profiling. *J Pharm Biomed Anal* **2020,** *190*, 113509.

4. De Livera, A. M.; Dias, D. A.; De Souza, D.; Rupasinghe, T.; Pyke, J.; Tull, D.; Roessner, U.; McConville, M.; Speed, T. P., Normalizing and integrating metabolomics data. *Anal Chem* **2012,** *84* (24), 10768-76.

5. Scholz, M.; Gatzek, S.; Sterling, A.; Fiehn, O.; Selbig, J., Metabolite fingerprinting: detecting biological features by independent component analysis. *Bioinformatics* **2004,** *20* (15), 2447-54.

6. Borrego, S. L.; Fahrmann, J.; Datta, R.; Stringari, C.; Grapov, D.; Zeller, M.; Chen, Y.; Wang, P.; Baldi, P.; Gratton, E.; Fiehn, O.; Kaiser, P., Metabolic changes associated with methionine stress sensitivity in MDA-MB-468 breast cancer cells. *Cancer Metab* **2016,** *4*, 9.

7. Redestig, H.; Fukushima, A.; Stenlund, H.; Moritz, T.; Arita, M.; Saito, K.; Kusano, M., Compensation for systematic cross-contribution improves normalization of mass spectrometry based metabolomics data. *Anal Chem* **2009,** *81* (19), 7974-80.

8. Sysi-Aho, M.; Katajamaa, M.; Yetukuri, L.; Oresic, M., Normalization method for metabolomics data using optimal selection of multiple internal standards. *BMC Bioinformatics* **2007,** *8*, 93.

9. Boysen, A. K.; Heal, K. R.; Carlson, L. T.; Ingalls, A. E., Best-Matched Internal Standard Normalization in Liquid Chromatography-Mass Spectrometry Metabolomics Applied to Environmental Samples. *Anal Chem* **2018,** *90* (2), 1363-1369.

10. Dunn, W. B.; Wilson, I. D.; Nicholls, A. W.; Broadhurst, D., The importance of experimental design and QC samples in large-scale and MS-driven untargeted metabolomic studies of humans. *Bioanalysis* **2012,** *4* (18), 2249-64.

11. Li, B.; Tang, J.; Yang, Q.; Li, S.; Cui, X.; Li, Y.; Chen, Y.; Xue, W.; Li, X.; Zhu, F., NOREVA: normalization and evaluation of MS-based metabolomics data. *Nucleic Acids Res* **2017,** *45* (W1), W162-W170.

12. De Livera, A. M.; Sysi-Aho, M.; Jacob, L.; Gagnon-Bartsch, J. A.; Castillo, S.; Simpson, J. A.; Speed, T. P., Statistical methods for handling unwanted variation in metabolomics data. *Anal Chem* **2015,** *87* (7), 3606-15.

13. Fan, S.; Kind, T.; Cajka, T.; Hazen, S. L.; Tang, W. H. W.; Kaddurah-Daouk, R.; Irvin, M. R.; Arnett, D. K.; Barupal, D. K.; Fiehn, O., Systematic Error Removal Using Random Forest for Normalizing Large-Scale Untargeted Lipidomics Data. *Anal Chem* **2019,** *91* (5), 3590-3596.

14. Viant, M. R.; Ebbels, T. M. D.; Beger, R. D.; Ekman, D. R.; Epps, D. J. T.; Kamp, H.; Leonards, P. E. G.; Loizou, G. D.; MacRae, J. I.; van Ravenzwaay, B.; Rocca-Serra, P.; Salek, R. M.; Walk, T.; Weber, R. J. M., Use cases, best practice and reporting standards for metabolomics in regulatory toxicology. *Nat Commun* **2019,** *10* (1), 3041.

15. Han, T. L.; Yang, Y.; Zhang, H.; Law, K. P., Analytical challenges of untargeted GC-MS-based metabolomics and the critical issues in selecting the data processing strategy. *F1000Res* **2017,** *6*, 967.

16. Zhao, Y.; Hao, Z.; Zhao, C.; Zhao, J.; Zhang, J.; Li, Y.; Li, L.; Huang, X.; Lin, X.; Zeng, Z.; Lu, X.; Xu, G., A Novel Strategy for Large-Scale Metabolomics Study by Calibrating Gross and Systematic Errors in Gas Chromatography-Mass Spectrometry. *Anal Chem* **2016,** *88* (4), 2234-42.

17. Duan, L.; Ma, A.; Meng, X.; Shen, G. A.; Qi, X., QPMASS: A parallel peak alignment and quantification software for the analysis of large-scale gas chromatography-mass spectrometry (GC-MS)-based metabolomics datasets. *J Chromatogr A* **2020,** *1620*, 460999.

18. Bijlsma, S.; Bobeldijk, I.; Verheij, E. R.; Ramaker, R.; Kochhar, S.; Macdonald, I. A.; van Ommen, B.; Smilde, A. K., Large-scale human metabolomics studies: a strategy for data (pre-) processing and validation. *Anal Chem* **2006,** *78* (2), 567-74.

19. Adeola, H. A.; Papagerakis, S.; Papagerakis, P., Systems biology approaches and precision oral health: a circadian clock perspective. *Frontiers in physiology* **2019,** *10*, 399.

20. Fiehn, O., Metabolomics by Gas Chromatography-Mass Spectrometry: Combined Targeted and Untargeted Profiling. *Curr Protoc Mol Biol* **2016,** *114*, 30 4 1-30 4 32.

21. Beale, D. J.; Pinu, F. R.; Kouremenos, K. A.; Poojary, M. M.; Narayana, V. K.; Boughton, B. A.; Kanojia, K.; Dayalan, S.; Jones, O. A. H.; Dias, D. A., Review of recent developments in GC-MS approaches to metabolomics-based research. *Metabolomics* **2018,** *14* (11), 152.

22. Khodadadi, M.; Pourfarzam, M., A review of strategies for untargeted urinary metabolomic analysis using gas chromatography-mass spectrometry. *Metabolomics* **2020,** *16* (6), 66.

23. Curtius, H. C.; Wolfensberger, M.; Steinmann, B.; Redweik, U.; Siegfried, J., Mass fragmentography of dopamine and 6-hydroxydopamine. Application to the determination of dopamine in human brain biopsies from the caudate nucleus. *J Chromatogr* **1974,** *99* (0), 529-40.

24. Stavova, J.; Beranek, J.; Nelson, E. P.; Diep, B. A.; Kubatova, A., Limits of detection for the determination of mono- and dicarboxylic acids using gas and liquid chromatographic methods coupled with mass spectrometry. *J Chromatogr B Analyt Technol Biomed Life Sci* **2011,** *879* (17-18), 1429-38.

25. Rahn, W., König, W. A.,, GC/MS investigations of the constituents in a diethyl ether extract of an acidified roast coffee infusion. *Journal of High Resolution Chromatography* **1978,** *1* (1), 69-71.

26. Lamoureux, G., Agüero, C., A comparison of several modern alkylating agents *ARKIVOC* **2009,** *i*, 14.

27. Liebeke, M.; Puskas, E., Drying Enhances Signal Intensities for Global GC(-)MS Metabolomics. *Metabolites* **2019,** *9* (4).

28. Barupal, D. K.; Zhang, Y.; Shen, T.; Fan, S.; Roberts, B. S.; Fitzgerald, P.; Wancewicz, B.; Valdiviez, L.; Wohlgemuth, G.; Byram, G.; Choy, Y. Y.; Haffner, B.; Showalter, M. R.; Vaniya, A.; Bloszies, C. S.; Folz, J. S.; Kind, T.; Flenniken, A. M.; McKerlie, C.; Nutter, L. M. J.; Lloyd, K. C.; Fiehn, O., A Comprehensive Plasma Metabolomics Dataset for a Cohort of Mouse Knockouts within the International Mouse Phenotyping Consortium. *Metabolites* **2019,** *9* (5).

29. Yu, S.; Fan, J.; Zhang, L.; Qin, X.; Li, Z., Assessment of Biphasic Extraction Methods of Mouse Fecal Metabolites for Liquid Chromatography-Mass Spectrometry-Based Metabolomic Studies. *J Proteome Res* **2021,** *20* (9), 4487-4494.

30. Badawy, A. A.; Morgan, C. J.; Turner, J. A., Application of the Phenomenex EZ:faasttrade mark amino acid analysis kit for rapid gas-chromatographic determination of concentrations of plasma tryptophan and its brain uptake competitors. *Amino Acids* **2008,** *34* (4), 587-96.

31. Lipkie, T. E.; Janasch, A.; Cooper, B. R.; Hohman, E. E.; Weaver, C. M.; Ferruzzi, M. G., Quantification of vitamin D and 25-hydroxyvitamin D in soft tissues by liquid chromatography-tandem mass spectrometry. *J Chromatogr B Analyt Technol Biomed Life Sci* **2013,** *932*, 6-11.

32. Huber, A. H.; Kleinfeld, A. M., Unbound free fatty acid profiles in human plasma and the unexpected absence of unbound palmitoleate. *J Lipid Res* **2017,** *58* (3), 578-585.

33. Zhu, P.; Zhang, J.; Chen, Y.; Yin, S.; Su, M.; Xie, G.; Brouwer, K. L. R.; Liu, C.; Lan, K.; Jia, W., Analysis of human C24 bile acids metabolome in serum and urine based on enzyme digestion of conjugated bile acids and LC-MS determination of unconjugated bile acids. *Anal Bioanal Chem* **2018,** *410* (21), 5287-5300.

34. Simon-Manso, Y.; Lowenthal, M. S.; Kilpatrick, L. E.; Sampson, M. L.; Telu, K. H.; Rudnick, P. A.; Mallard, W. G.; Bearden, D. W.; Schock, T. B.; Tchekhovskoi, D. V.; Blonder, N.; Yan, X.; Liang, Y.; Zheng, Y.; Wallace, W. E.; Neta, P.; Phinney, K. W.; Remaley, A. T.; Stein, S. E., Metabolite profiling of

a NIST Standard Reference Material for human plasma (SRM 1950): GC-MS, LC-MS, NMR, and clinical laboratory analyses, libraries, and web-based resources. *Anal Chem* **2013,** *85* (24), 11725-31.

35. Joyce, R.; Kuziene, V.; Zou, X.; Wang, X.; Pullen, F.; Loo, R. L., Development and validation of an ultra-performance liquid chromatography quadrupole time of flight mass spectrometry method for rapid quantification of free amino acids in human urine. *Amino Acids* **2016,** *48* (1), 219-34.

36. Ballman, K. V.; Grill, D. E.; Oberg, A. L.; Therneau, T. M., Faster cyclic loess: normalizing RNA arrays via linear models. *Bioinformatics* **2004,** *20* (16), 2778-86.

37. Dunn, W. B.; Broadhurst, D.; Begley, P.; Zelena, E.; Francis-McIntyre, S.; Anderson, N.; Brown, M.; Knowles, J. D.; Halsall, A.; Haselden, J. N.; Nicholls, A. W.; Wilson, I. D.; Kell, D. B.; Goodacre, R.; Human Serum Metabolome, C., Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. *Nat Protoc* **2011,** *6* (7), 1060-83.

38. Lange, S. R., Martin, Deep auto-encoder neural networks in reinforcement learning. *The 2010 International Joint Conference on Neural Networks* **2010**, 1-8.

39. Karimi, M.; Hasanzadeh, A.; Shen, Y., Network-principled deep generative models for designing drug combinations as graph sets. *Bioinformatics* **2020,** *36* (Suppl_1), i445-i454.

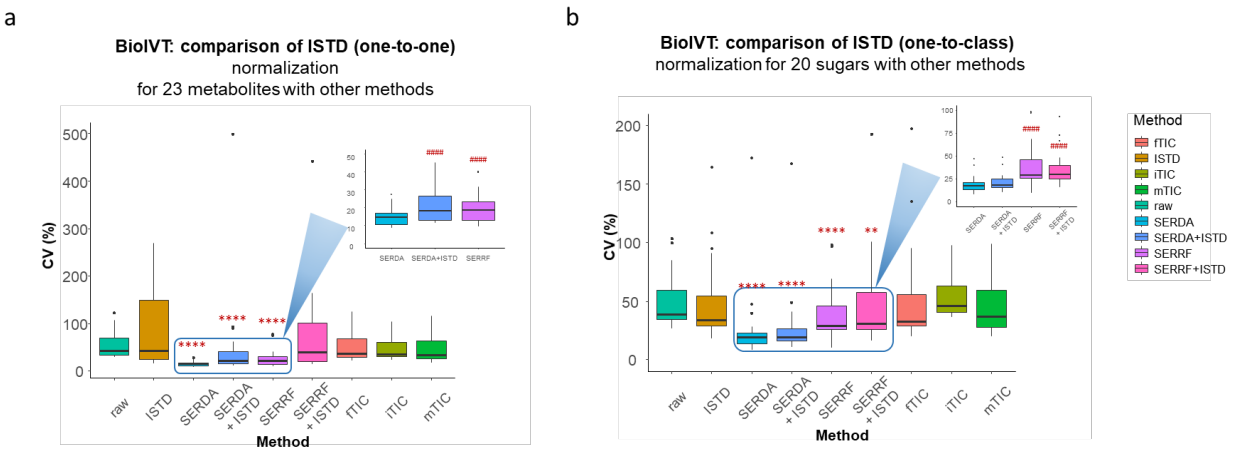
1.7 Supplementary Information

Supplementary Table 1.1. Average of median %CV of 16 amino acids by three derivatization reagents across three months

Reagents	Raw	ISTD	fTIC	iTIC	mTIC
PCF	11.9%	2.7%			
MSTFA	22.1%	9.6%	20.3%	14.5%	13.5%
MTBSTFA	15.2%	8.9%	13.3%	9.7%	8.3%

Supplementary Table 1.2. Correlation of extra biological duplicates after normalized by SERDA.

Statistics for correlation	Number of duplicates	Mean of correlation coefficient	S.D. of correlation coefficient	Max.	Min.
Adjacent biological duplicates	48	0.98	2.15%	1.00	0.90
Non-adjacent biological duplicates	54	0.98	3.03%	1.00	0.81
Total	102	0.98	2.64%	1.00	0.81



Supplementary Figure 1.1. Comparison of ISTD absolute ratio normalization with QC-based and sum-based normalization methods. Friedman nonparametric test was used for significance comparison with raw. P value threshold: 0.1234 (ns), 0.0332 (*), 0.0021 (**), 0.0002(***), 0.0001(****). Or Friedman nonparametric test was used for significance testing comparing to SERDA. P value threshold: 0.1234 (ns), 0.0332 (#), 0.0021 (##), 0.0002(###), 0.0001(####). One-to-one: absolute ratio was calculated by dividing the peak intensity of endogenous metabolite by corresponding deuterated ISTD. One-to-class: absolute ratio was calculated by dividing the peak intensity of endogenous metabolite by a single deuterated compound as an analog ISTD for the entire class.

Chapter 2 Longitudinal Plasma Lipidome of Risk for Type 2 Diabetes in a Large Sample of American Indians with Normal Fasting Glucose: The Strong Heart Family Study

Reproduced from Guanhong Miao, Ying Zhang (co-first author), Zhiguang Huo, Wenjie Zeng, Jianhui Zhu, Jason G. Umans, Gert Wohlgemuth, Diego Pedrosa, Brian DeFelice, Shelley A. Cole, Amanda M. Fretts, Elisa T. Lee, Barbara V. Howard, Oliver Fiehn, and Jinying Zhao. "Longitudinal Plasma Lipidome and Risk of Type 2 Diabetes in a Large Sample of American Indians with Normal Fasting Glucose: The Strong Heart Family Study". Diabetes Care, 2021.

2.1 Abstract

OBJECTIVE: Comprehensive assessment of alterations in lipid species preceding type 2 diabetes (T2D) is largely unknown. We aimed to identify plasma molecular lipids associated with risk of T2D in American Indians.

RESEARCH DESIGN AND METHODS: Using untargeted liquid chromatography–mass spectrometry, we repeatedly measured 3,907 fasting plasma samples from 1,958 participants who attended two examinations (~5.5 years apart) and were followed up to 16 years in the Strong Heart Family Study. Mixed-effects logistic regression was used to identify lipids associated with risk of T2D, adjusting for traditional risk factors. Repeated measurement analysis was performed to examine the association between change in lipidome and change in continuous measures of T2D, adjusting for baseline lipids. Multiple testing was controlled by false discovery rate at 0.05.

RESULTS: Higher baseline level of 33 lipid species, including triacylglycerols, diacylglycerols, phosphoethanolamines, and phosphocholines, was significantly associated with increased risk of T2D (odds ratio [OR] per SD increase in log2-transformed baseline lipids 1.50–2.85) at 5-year follow-up. Of these, 21 lipids were also associated with risk of T2D at 16-year follow-up. Aberrant lipid profiles were also observed in prediabetes (OR per SD increase in log2-transformed baseline lipids 1.30–2.19 for risk lipids and 0.70–0.78 for protective lipids). Longitudinal changes in 568 lipids were significantly associated with changes in continuous measures of T2D. Multivariate analysis identified distinct lipidomic signatures differentiating high- from low-

risk groups.

CONCLUSIONS: Lipid dysregulation occurs many years preceding T2D, and novel molecular lipids (both baseline level and longitudinal change over time) are significantly associated with risk of T2D beyond traditional risk factors. Our findings shed light on the mechanisms linking dyslipidemia to T2D and may yield novel therapeutic targets for early intervention tailored to American Indians.

2.2 Introduction

Dyslipidemia, including high triglycerides and low HDL cholesterol (HDL-c), represents a hallmark of type 2 diabetes (T2D) (1). American Indians are nearly three times more likely to be diagnosed with T2D than non-Hispanic whites (2). Routine biochemical tests cannot capture all molecular lipids (i.e., lipidome) and thus have limited value in detecting early lipid disturbances implicated in disease. Lipidomics can identify many individual lipids and is well suited for characterizing the perturbed lipids preceding T2D. This is important, because prevention or delay of T2D has proven to be effective via both lifestyle and pharmacological interventions.

Previous studies have reported associations of altered blood lipids with T2D (3–9), obesity (10,11), and insulin resistance (IR) (9,11–13). Altered plasma triacylglycerols (TGs), diacylglycerols (DAGs), sphingomyelins (SMs), phosphoethanolamines (PEs), phosphocholines (PCs), and cholesterol esters (CEs) have been associated with T2D in different populations (3–5,7). However, most existing studies only measured baseline plasma lipids, which did not reflect change in plasma lipidome during T2D development. We are aware of only one lipidomic study that measured 207 known lipids in plasma sample collected at two time points (1 year apart) in a high-risk population, but none was associated with T2D after adjusting for baseline levels (7). Moreover, previous studies included predominately participants of European descent or at-risk individuals (as a result of matching in nested case-control design or inclusion of prediabetes as control). We are aware of two metabolomic studies among normoglycemic individuals with Chinese (4) or

European ancestry (14), both of which only measured baseline metabolites and had low coverage of plasma lipidome.

Here we report findings from the first longitudinal lipidome profiling in 3,916 fasting plasma samples from 1,962 normoglycemic American Indians who attended two examinations (~5.5 year apart) and were followed up to 16 years in the Strong Heart Family Study (SHFS). Our goal is to identify novel molecular lipids and lipidomic signatures that can predict the onset and progression of T2D beyond traditional risk factors.

2.3 Research Design and Methods

2.3.1 Participants

All participants are American Indians in the SHFS (2001-ongoing), a family-based prospective study designed to identify genetic and metabolic factors for cardiovascular disease (CVD), T2D and related factors in American Indians (15). Briefly, 2,780 tribal members (aged 18 and older) from three geographic regions (Arizona, North/South Dakota, Oklahoma) were initially examined in 2001-2003 and re-examined in 2006-2009 (mean 5.5 years apart) using the same protocols. At each visit, participants completed questionnaires collecting information for demography, family history, medical records, and lifestyle. Biospecimens were also collected at each visit. Detailed descriptions for study design, laboratory methods (16) and phenotype collection of the SHFS were reported previously (15,16). More information on blood sample collection and sample handling is described in the Supplementary Materials. All participants provided informed consent. The SHFS protocols were approved by the institutional review board of each participating institution and tribe.

Participants in the current study met the following criteria: 1) attended clinical examinations and had available fasting plasma samples at both baseline (2001–2003) and 5-year follow-up (2006–2009); 2) were free of overt CVD at baseline. Participants with missing information for fasting glucose or hypoglycemia medications at either time point were excluded.

2.3.2 Definition of incident T2D and T2D-related traits

At baseline and 5-year follow-up, diabetes was defined as fasting plasma glucose (FPG) $7.0 mmol/L or receiving hypoglycemic medications. Impaired fasting glucose (IFG) was defined as FPG of 6.1–6.9 mmol/L and no hypoglycemic medications, and normal fasting glucose (NFG) was defined as FPG <6.1 mmol/L. Because SHFS did not collect biospecimens after the 5-year visit, T2D at 16-year follow-up was ascertained based on medical records abstracted by trained research staff. Incident T2D was defined as participants who had NFG at baseline (2001-2003) but developed T2D by end of 5-year (2006-2009) or 16-year follow-up (31 December 2018). Insulin resistance was assessed using homeostatic model assessment (HOMA): HOMA-IR = fasting glucose (mg/dL) × insulin (μU/mL)/405 (17). Pancreatic β-cell function (HOMA-β) was assessed using the formula: 360 × fasting insulin (μU/mL) / (fasting glucose (mg/dL) - 63) (16). Insulin sensitivity was estimated by calculating the Quantitative Insulin sensitivity check index (QUICKI =1/[log insulin (mU/L) + log baseline glucose (mg/dL)] (18).

2.3.3 Lipidomic data acquisition

Fasting plasma was extracted using a modified liquid-liquid extraction method (cold methanol/MTBE/water) (19), then subjected to liquid chromatography–mass spectrometry on both positive and negative ion modes. Raw data were processed using in-house cloud-based software (LCBinBase) with peak detection and deconvolution algorithms adapted from MSDIAL (20). Lipid peak intensity results were manually checked against raw data files. Features and peaks with 50% missing values across all samples were removed. Batch effect was corrected by SERRF (21). Details on sample extraction, laboratory protocols, and data preprocessing are described in the Supplementary Materials.

After preprocessing and quality control, we obtained 1,809 lipids (579 known, 1,230 unknown) in 3,916 samples. Coefficient variations of the Bioreclamation and National Institutes of Standards and Technology samples were 9% and 16%, respectively. Relative abundances of duplicated samples were highly correlated ($r > 0.95$).

2.3.4 Statistical analyses

Prior to statistical analysis, lipidomic data were $\log_2$ transformed, then standardized to zero mean and unit SD. Multimodal distribution was detected by the Hartigans-Dip test (22). Outlier samples were detected by principal component analysis, and those beyond mean ± 5 SD for any of the first three PCs were further removed. Our final analysis included 1,542 lipids (518 known, 1,024 unknown) in 3,907 samples (1,958 at baseline, 1,949 at 5-year follow-up). Of these, 1,945 participants had complete data at both time points.

Prospective association analyses. To identify novel molecular lipids associated with risk of T2D, we constructed a mixed-effects logistic regression model, in which level of baseline lipid was the independent variable and incident status of T2D was the dependent variable, adjusting for TRFs including age, sex, study site, BMI, fasting glucose, IR, HDL-c, and total triglycerides at baseline. The model tested the fixed effect of baseline lipids on risk of T2D and included random effect to account for relatedness among family members. Similar analysis was conducted to identify lipids predictive of risk of prediabetes (IFG). These analyses were performed among 1,161 NFG participants followed through 2006–2009 (mean 5.5-year follow-up).

Of 1,161 NFG participants at baseline, 989 individuals had information on T2D status through 31 December 2018 (mean follow-up 16 years), during which 176 participants developed incident T2D. The same statistical model was used to identify lipids associated with 16-year risk of T2D among these 989 participants. Multivariable adjusted odds ratios (ORs) and 95% CIs were calculated for each lipid. Multiple testing was controlled by false discovery rate (FDR) (23), and an FDR adjusted *P* value (i.e., q value) <0.05 was considered significant.

Repeated measurement analyses. To examine the longitudinal association between changes in plasma lipidome and changes in continuous measures of T2D, including fasting glucose, fasting insulin, IR (HOMA-IR), β-cell function (HOMA-β), and insulin sensitivity (QUICKI), we constructed a mixed-effects linear regression model, in which 5-year change in each continuous measure (i.e., difference between 5-

year follow-up and baseline, standardized to N [0,1]) was the outcome and change in each lipid was the independent variable, adjusting for age, sex, site, change in BMI, baseline lipid, and trait under study. The model tested the fixed effect of change in lipids on the change in T2D-related traits and included random effect to account for relatedness among family members. We also estimated to what extent the variation in 5-year change of each trait could be explained by change in plasma lipidome using the R package *r2glmm*. In addition, we tested the association of change in lipidome (focusing on known lipids whose 5-year changes were associated with changes in continuous measures of T2D at $q < 0.05$) with risk of T2D at 16-year follow-up, adjusting for TRFs and baseline lipids.

Multivariate analysis by partial least squares discriminant analysis. To identify discriminatory lipids and lipidomic signatures associated with risk of T2D, we conducted partial least squares discriminant analysis (PLS-DA) using the R package *mixOmics* (24). The model included all 1,542 lipids among 1,161 NFG participants at baseline, adjusting for TRFs. Optimal number of components and potential overfitting of the model were assessed with 10-fold cross validation (repeated 100 times). Lipids with variable importance in projection (VIP) scores ≥ 1.5 were considered crucial in identifying participants who developed incident T2D or IFG (cases) from those who remained NFG (controls) over 5 years of follow-up. Discriminant power of the validated model was assessed by area under the receiver-operating curve.

Sensitivity analyses. To examine whether diet quality (assessed by the Alternate Healthy Eating Index), physical activity (steps per day), or use of lipid-lowering medications (yes/no) affected our results, we additionally adjusted for baseline levels of these covariates in the above-described prospective analyses.

2.4 Results

After stringent quality control, we obtained data for 1,542 plasma lipids (both positive and negative ionization) in 3,907 samples from 1,958 participants at both baseline and 5-year follow-up.

Of the 1,161 NFG participants followed through 2006–2009 (mean follow-up 5.5 years), 205 participants developed incident IFG and 73 developed incident T2D. Of these 1,161 NFG participants, 989 were

followed through 31 December 2018 (mean follow-up 16 years), during which 176 participants developed incident T2D. **Table 2.1** presents the baseline characteristics of participants according to diabetic status by end of 5-year and 16-year follow-up. Compared with participants who did not develop incident T2D, those who developed incident T2D had higher levels of BMI, waist circumference, blood pressure, total triglycerides, fasting glucose, fasting insulin, IR, and β-cell function, but lower levels of HDL-c and insulin sensitivity, at baseline. **Supplementary Table 2.1** presents the differences in baseline characteristics of participants between incident T2D or IFG, as compared with those who remained NFG, by end of 5-year follow-up.

Table 2.1. Baseline characteristics of NFG participants according to diabetes status by end of 5-year and 16-year follow-up

	5-year follow-up (*n* = 1,161)			**16-year follow-up (*n* = 989)**		
Characteristic	Cases (*n* = 73)	Noncases (*n* = 1,088)	*P*	Cases (*n* = 176)	Noncases (*n* = 813)	*P*
Age (years)	35.8 ± 11.3	36.1 ± 13.0	0.84	36.7 ± 10.9	35.9 ± 12.9	0.40
Female (%)	44 (60)	693 (64)	0.54	110 (62)	539 (66)	0.347
BMI (kg/m2)	34.9 ± 7.6	29.6 ± 6.6	1.75×10^{-10}	33.9 ± 7.4	29.2 ± 6.2	4.96×10^{-12}
Waist (cm)	110.4 ± 18.4	97.3 ± 15.9	4.83×10^{-10}	106.8 ± 17.7	96.1 ± 15.0	1.07×10^{-11}
SBP (mmHg)	123.4 ± 14.1	119.4 ± 14.5	0.01	122.6 ± 14.3	118.6 ± 14.3	3.66×10^{-1}
3DBP (mmHg)	79.0 ± 9.9	76.0 ± 10.5	7.62×10^{-3}	78.6 ± 10.6	75.7 ± 10.4	1.37×10^{-3}
HDL (mg/dL)	49.1 ± 15.5	53.9 ± 14.9	0.02	48.4 ± 11.6	54.7 ± 15.1	8.01×10^{-9}
LDL (mg/dL)	101.8 ± 29.0	100.3 ± 29.2	0.66	99.4 ± 25.5	101.0 ± 29.4	0.5
Triglycerides (mg/dL)	177.9 ± 124.8	139.0 ± 78.9	8.93×10^{-4}	166.0 ± 118.2	136.7 ± 76.4	3.90×10^{-3}

Total cholesterol (mg/dL)	184.3 ± 35.1	181.8 ± 33.7	0.55	179.8 ± 31.0	182.9 ± 33.8	0.3
Fasting glucose (mg/dL)	91.3 ± 6.5	88.9 ± 6.6	9.36×10^{-3}	90.6 ± 6.4	88.7 ± 6.6	2.58×10^{-3}
Fasting insulin (mU/mL)	21.1 ± 16.0	12.4 ± 10.3	1.34×10^{-7}	17.3 ± 13.0	11.9 ± 10.0	1.56×10^{-5}
HOMA-IR	4.7 ± 3.5	2.8 ± 2.3	8.96×10^{-8}	3.9 ± 3.0	2.6 ± 2.3	1.12×10^{-5}
HOMA-b	295.8 ± 273.5	180.3 ± 155.6	1.39×10^{-7}	237.6 ± 203.3	174.8 ± 153.1	3.87×10^{-4}
QUICKI	0.14 ± 0.01	0.15 ± 0.02	1.40×10^{-8}	0.14 ± 0.01	0.15 ± 0.02	9.05×10^{-10}

Data are presented as mean ± SD or *n* (%). *P* values were obtained by generalized estimating equation to account for correlation among family members. Cases are participants who developed incident T2D by end of follow-up. Noncases are participants who did not develop incident T2D by end of follow-up.

2.4.1 Novel lipid species associated with risk of T2D over TRFs

After adjusting for TRFs (age, sex, site, BMI, fasting glucose, IR, total triglycerides, and HDL-c) and multiple testing ($q < 0.05$), we found that higher baseline levels of 56 lipids (33 known, 23 unknown) were significantly associated with risk of T2D at 5-year follow-up. The 33 known lipids included 19 glycerolipids (TGs and DAGs) and 14 glycerophospholipids (PCs and PEs) with ORs (per SD increase in log2-transformed baseline lipids) ranging from 1.50 to 2.85. Of these, 21 lipids were also associated with risk of T2D at 16-year follow-up (**Fig. 2.1**).

By contrast, 90 plasma lipids (49 known, 41 unknown) were positively or inversely associated with risk of future IFG at 5-year follow-up ($q < 0.05$). Of the 49 known lipids, higher baseline levels of 44 lipids, including glycerolipids (TGs and DAGs), glycerophospholipids (PCs, PEs, PIs, and PGs), SMs, and CEs, were positively (OR per SD increase in log2-transformed baseline lipids ranging from 1.30 to 2.19), while 5 lipids, including LPC(20:4), PC(17:1/22:5), PC(39:6), PC(p-18:1/20:4)/PC(o-18:2/20:4), and SM(d40:3), were inversely (OR per SD increase in log2-transformed baseline lipids ranging from 0.70 to 0.78)

associated with risk of future IFG. Twenty-four of 49 lipids were also significantly associated with 5-year risk of T2D. Besides the known lipids mentioned above, our untargeted lipidomics also identified multiple lipids with unknown structures associated with future risk of T2D/IFG. **Figure 2.2** displays the lipidome--wide associations of plasma lipidome with risk of T2D (**Fig. 2.2A**) and IFG (**Fig. 2.2B**) during 5-year follow-up.

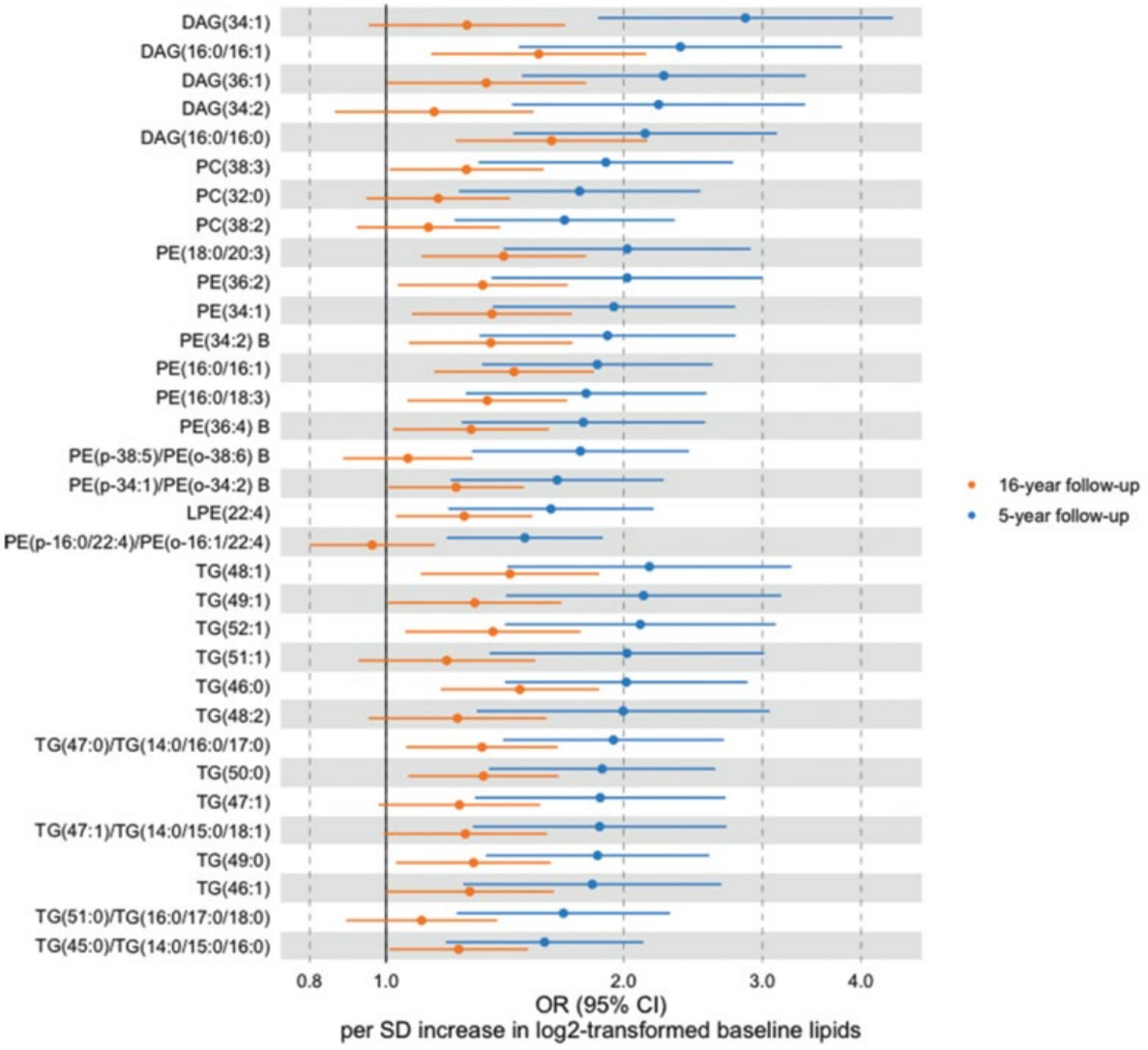


Figure 2.1. Baseline plasma lipids significantly associated with risk of T2D. Only known lipids are shown. The solid line represents OR = 1. Of the 33 known lipids associated with 5-year risk of T2D ($q < 0.05$), 21 lipids were also associated with 16-year risk of T2D. The mixed-effects logistic regression model adjusted for age, sex, BMI, fasting glucose, IR, HDL-c, and total triglycerides at baseline. Random effect was included in the model to account for relatedness among family members.

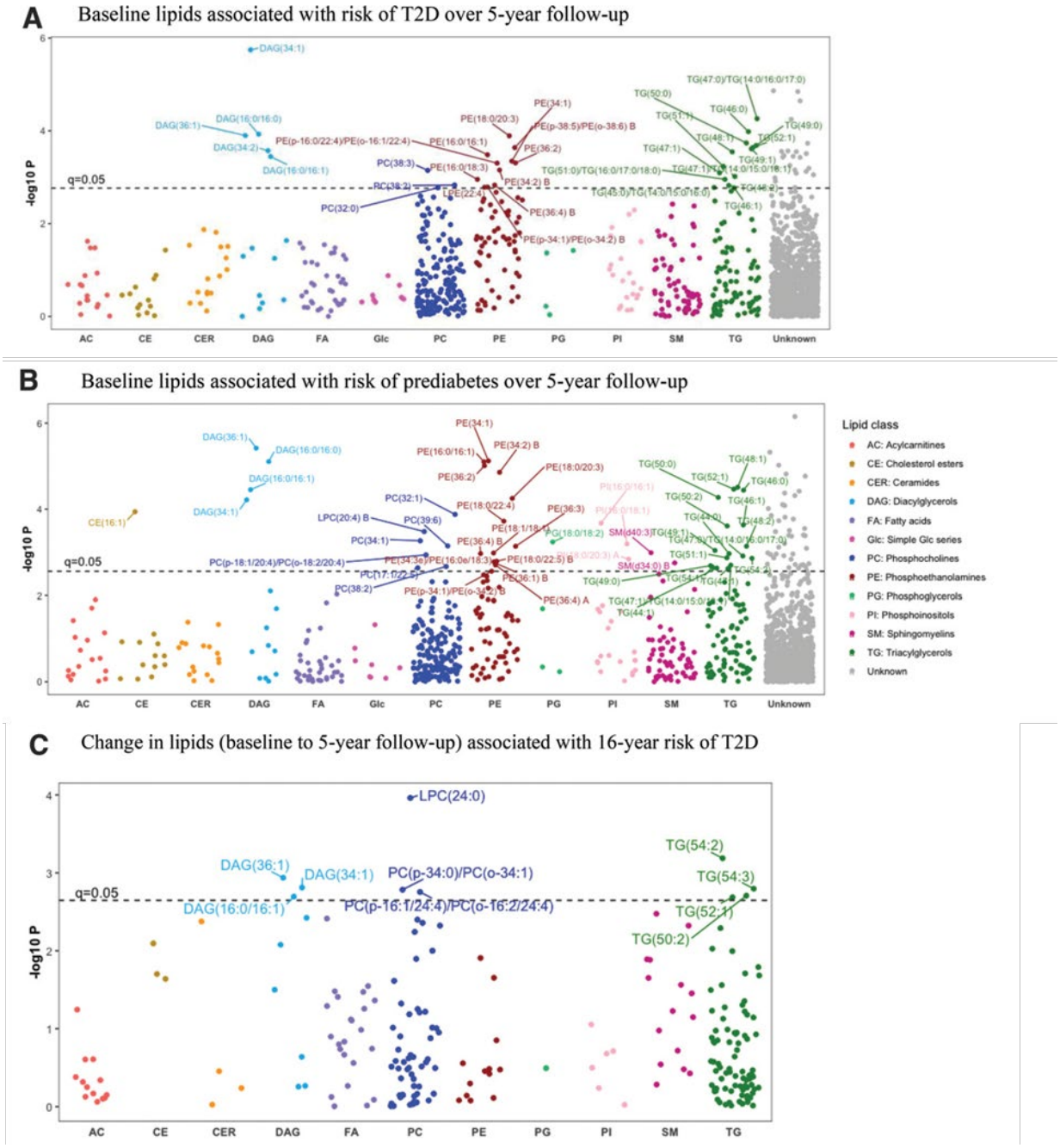


Figure 2.2. A and B: Manhattan plots showing the prospective associations of 1,542 baseline plasma lipid species with future risk of T2D (A) or prediabetes (B) over 5 years of follow-up. The mixed-effects logistic regression model adjusted for age, sex, study site, BMI, fasting glucose, IR, HDL-c, and total triglycerides at baseline. Random effect was included in the model to account for relatedness among family members. C: Manhattan plot showing the association between change in plasma lipidome (baseline to 5-year follow-up) and 16-year risk of T2D. x-axis indicates lipid class; y-axis indicates log10 P. The dashed line represents significance level at q = 0.05.

2.4.2 Longitudinal changes in plasma lipidome associated with changes in T2D traits

After correction for multiple testing ($q < 0.05$) and adjustments for covariates (age, sex, site, and change in BMI between baseline and follow-up) as well as baseline lipids and the trait under investigation, our repeated measurement analysis identified significant associations between 5-year changes in 568 lipids (230 known, 338 unknown) and changes in one or more T2D-related traits, including fasting glucose, fasting insulin, IR, pancreatic b-cell function, and insulin sensitivity. The 230 known lipids largely belong to TGs (n = 84), PCs (n = 57), fatty acids (FAs) (n = 23), SMs (n = 15), PEs (n = 13), acylcarnitines (ACs) (n = 13), and DAGs (n = 9). **Figure 2.3** illustrates the longitudinal association patterns between changes in lipidome and changes in T2D-related traits during 5-year follow-up (only lipids from seven main classes are shown). We observed a clear pattern that longitudinal changes in TGs, DAGs, and PEs were positively associated with changes in IR, b-cell function, and fasting insulin and inversely associated with changes in insulin sensitivity. In contrast, longitudinal changes in FAs and ACs were inversely associated with changes in IR, β-cell function, and fasting insulin and positively associated with changes in insulin sensitivity. In addition, changes in SMs were inversely associated with changes in fasting plasma glucose. To further evaluate the impact of altered plasma lipids on diabetes development, we estimated to what extent the variability in change of each T2D trait could be explained by the change in plasma lipids over 5-year follow-up period. After adjusting for covariates and baseline lipids and the trait under investigation, longitudinal changes in three major classes (TGs, PCs, and FAs) explained up to 8.4% variability in the change of T2D-related traits. These findings indicate a potential large impact of altered plasma lipidome on development of T2D. To examine the association between change in plasma lipidome and risk of T2D, we focused on the 230 known lipids whose longitudinal changes between baseline and 5-year follow-up were associated with one or more continuous measures of T2D and tested their associations with risk of T2D at 16-year follow-up. We found that 5-year changes in 10 lipids (7 TGs and DAGs, 3 PCs) were significantly associated with 16-year risk of T2D at $q < 0.05$ (**Fig. 2.2C**).

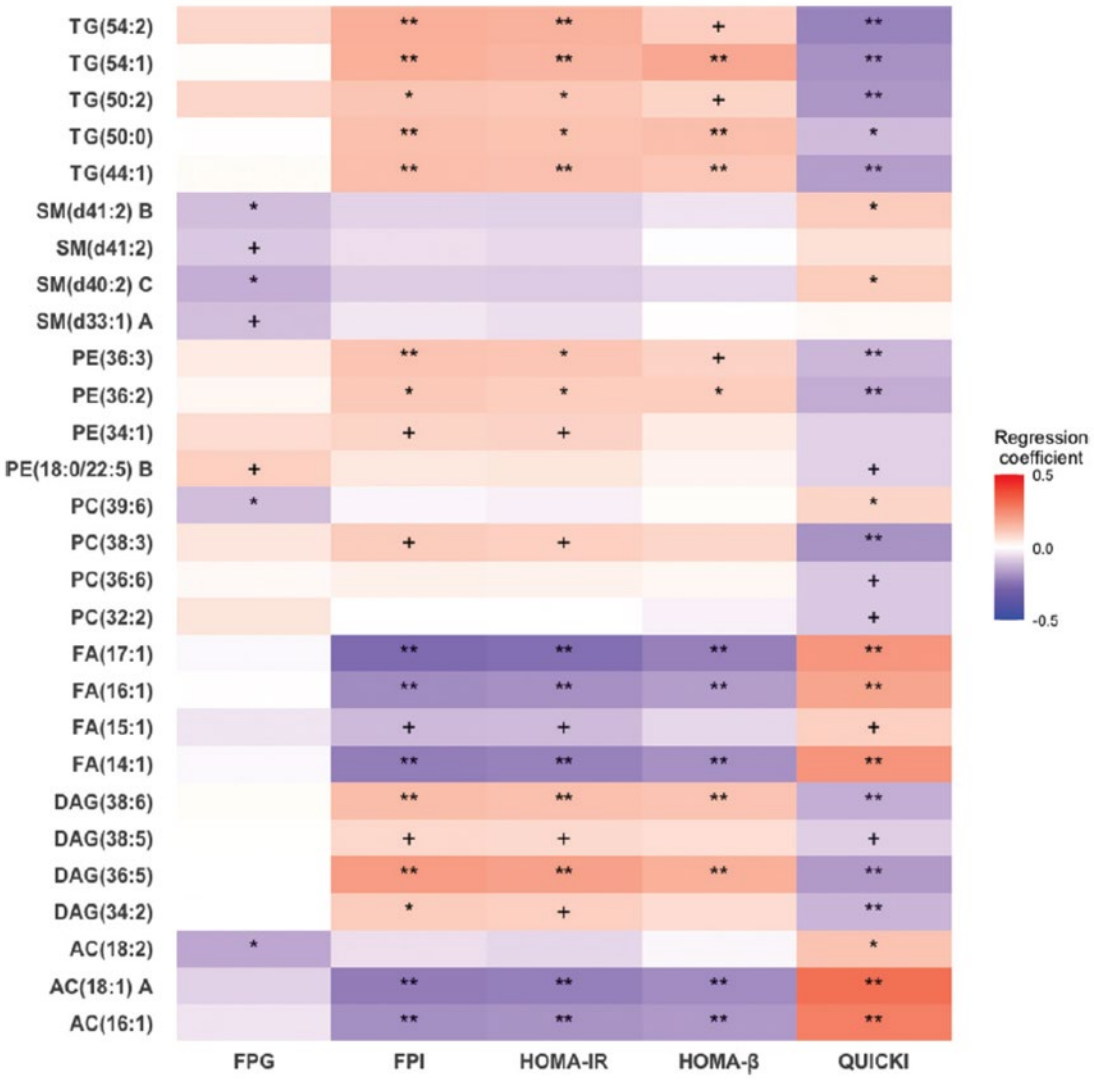


Figure 2.3. Heatmap illustrating the longitudinal associations between change in plasma lipidome and change in continuous diabetes traits during 5-year follow-up. Each row represents a lipid, and each column represents a phenotype. The heatmap is arranged based on lipid classes, and only top 10 known lipids in each class are shown. Color code is based on regression coefficients obtained from the linear mixed-effect model, in which change in lipid was the independent variable, and change in each of the T2D-related traits was the dependent variable. The model adjusted for age, sex, study site, change in BMI, and baseline levels of the specific lipid and the trait under investigation. Random effect was included in the model to account for relatedness among family members. FPG: fasting plasma glucose; FPI: fasting plasma insulin; HOMA-IR: insulin resistance; HOMA-b: b-cell function; QUICKI: insulin sensitivity. **P < 0.001, *P < 0.01, 1P < 0.05.

2.4.3 Discriminatory lipidomic signatures identified by PLS-DA

Our multivariate analysis revealed a distinct lipidomic signature that was able to separate cases (incident T2D or IFG) from controls (NFG) at 5-year follow-up. The validated PLS-DA model identified four clinical variables (site, baseline BMI, fasting glucose, and IR) and 128 discriminatory lipids with VIP score >1.5.

2.4.4 Results from sensitivity analyses

Additional adjustments for diet quality, physical activity, and lipid-lowering drugs did not change the associations between baseline lipids and 5-year risk of T2D (Supplementary Table 8). Of the 21 lipids associated with 16-year risk of T2D, 15 and 12 lipids remained significant after further adjustment for lipid-lowering drugs or physical activity, respectively. Four lipids remained significant after further adjustment for all three covariates (diet quality, physical activity, and lipid-lowering drugs).

2.5 Conclusions

Dyslipidemia may precede overt T2D by many years (24), but a full spectrum of plasma lipid species (i.e., lipidome) predicting the onset and progression of T2D among apparently healthy individuals remains largely unknown. Moreover, few studies have examined the longitudinal association of change in lipidome with T2D development in a large community sample of normoglycemic individuals. Here we report findings from a longitudinal profiling of plasma lipidome in nearly 4,000 fasting plasma samples collected at two time points from 2,000 apparently healthy American Indian men and women. Our study has several novel findings. First, we identified novel molecular lipids associated with risk of future T2D beyond TRFs, including age, sex, BMI, fasting glucose, IR, HDL-c, and total triglycerides. At baseline, individuals who later developed T2D exhibited elevated levels of TGs, DAGs, PCs, and PEs. Altered lipid profiles were also observed in prediabetes (i.e., IFG). These results demonstrate lipid dysregulation occurs years before symptoms appear, and if confirmed, the newly identified molecular lipids may serve as novel biomarkers for risk stratification and early intervention. Second, our repeated measurement analyses identified several lipid species (e.g., TGs, PCs, FAs, and SMs) whose changes over time were associated with changes in

T2D traits during 5 years of follow-up. Overall, changes in plasma lipidome explained up to 3.0%, 4.0%, and 8.4% variability in changes of fasting glucose, IR, and insulin sensitivity, respectively. Third, our multivariate analysis identified distinct lipidomic signatures associated with risk of T2D. To our knowledge, the current study represents the first comprehensive assessment of fasting plasma lipidome in relation to risk of T2D in a large community sample of apparently healthy adults among not only American Indians but other racial/ethnic groups as well.

In line with previous studies among European Caucasians (3,7,12,14,26), Mexican Americans (3), and Chinese individuals (4,27,28), we found that glycerolipids (e.g., TGs and DAGs) showed the most significant associations with risk of T2D in American Indians. In our study, higher baseline levels of multiple glycerolipids, including TAG(46:0), TAG (47:0)/TAG(14:0/16:0/17:0), TAG(48:1), TAG(49:1), TAG(49:0), TAG(50:0), TAG(52:1), DAG(36:1), DAG(16:0/16:1), and DAG(16:0/16:0), were consistently associated with risk of future T2D and prediabetes. Many lipids identified in ourstudy, such as TAG(48:1), TAG(48:2), TAG (49:1), TAG(50:0), TAG(51:0)/TAG(16:0/17:0/18:0), TAG(51:1), TAG(52:1), DAG(34:1), DAG(34:2), and DAG(16:0/16:0), were also reported to be associated with diabetes (in the same direction) in European Caucasians (3,12,26), Mexican Americans (3), and Chinese individuals (27,28). The mechanisms through which TGs and DAGs affect T2D are likely related to their roles in membrane fluidity, inflammation, oxidative stress, and insulin signaling, all of which may affect glucose metabolism and IR (29–32).

Besides TGs and DAGs, we found that higher baseline levels of PCs, including PC(32:0), PC(38:2), and PC(38:3), and PEs, including PE(16:0/16:1), PE(18:0/20:3), PE(34:1), and PE(36:2), were also associated with increased risk of T2D in American Indians. Many of these lipids, such as PC(32:0), PC(38:3), PE(34:1), and PE(36:2), were also associated with T2D in other cohorts (3). Moreover, previous studies have reported associations of different species of phospholipids with IR, T2D, and related traits (7,33,34), even though specific lipid molecules were not identical as a result of the use of different analytical platforms and/or sample heterogeneity across studies. PCs and PEs are the major components of cellular membranes,

and their roles in regulating glucose metabolism and insulin sensitivity have been reported in both animal models (35–37) and humans (33,38,39). Our results in American Indians, along with findings from other populations, demonstrate that altered plasma lipids precede T2D, and perturbed metabolism in TGs, DAGs, PCs, and PEs is implicated in T2D development.

In agreement with previous studies among European Caucasians and Mexican Americans (3,40), we found that lipid profiles of risk for prediabetes (IFG) were similar to those of T2D in American Indians. Specifically, of the 33 known lipids associated with risk of T2D, 24 lipids (largely TGs, DAGs, and PEs) were also associated with IFG in the same direction, indicating that aberrant lipid metabolism presents in prediabetes and persists during T2D development. Moreover, longitudinal changes (baseline to 5-year follow-up) in four of these 24 lipids, specifically DAG(16:0/16:1), DAG(34:1), DAG(36:1), and TAG(52:1), were also associated with risk of T2D at 16-year follow-up. These consistent findings lend support to an important role of these lipids in diabetes pathogenesis.

In line with previous studies reporting that SMs were either positively or inversely associated with T2D (7,9,12), we found that SM(d40:3) was inversely whereas SM(d34:0) was positively associated with risk of prediabetes. Our repeated measurement analysis showed that 5-year changes in SMs, such as SM(d40:2) and SM(d41:2), were inversely associated with changes in fasting glucose and positively associated with insulin sensitivity. In addition, we found that higher baseline levels of four PCs, including LPC (20:4), PC(39:6), PC(17:1/22:5), and PC(p-18:1/20:4)/PC(o-18:2/20:4), were inversely associated with risk of prediabetes. The observed bidirectional relationship of some SMs and PCs with T2D may be attributable to the different numbers of carbons and/or double bonds in these lipids. For example, previous research reported that lipids with lower numbers of carbons and double bonds were positively, whereas those with higher numbers of carbons and double bonds were inversely, associated with diabetes (12).

Our repeated measurement analysis identified a clear pattern between changes in lipidome and changes in T2D traits. Specifically, 5-year changes in TGs, DAGs, and PEs were positively associated with changes in fasting insulin, IR, and β-cell function and inversely associated with changes in insulin sensitivity. In

contrast, 5-year changes in ACs and FAs were inversely associated with changes in fasting insulin, IR, and β-cell function and positively associated with changes in insulin sensitivity. Moreover, we found that of the 230 known lipids whose longitudinal changes between baseline and 5-year follow-up were significantly associated with changes in continuous T2D traits, 37 lipids were also significantly associated with risk of T2D (both 5-year and 16-year follow-up) or IFG. These consistent results from different analyses support the robustness of our findings. Moreover, these novel and potentially important findings may provide mechanistic insights into the role of molecular lipids in diabetes development.

Our multivariate analysis identified distinct baseline lipidomic signatures able to classify participants into different risk groups. Top discriminatory lipids included TGs, DAGs, PCs, PEs, SMs, and ceramides. Of the 67 known discriminatory lipids, 30 were also identified in prospective and repeated measurement analyses. These results support important roles of these lipids in T2D development among American Indians.

Our study has several strengths. First, the repeatedly measured plasma lipidome in a large number of community-dwelling individuals represents a major strength of our study. To our knowledge, this is by far the largest longitudinal lipidomic study on risk of T2D in any racial/ethnic group. Second, unlike most previous studies that included a mixture of both normoglycemic individuals and individuals with prediabetes (i.e., those with IFG or IGT among whom early metabolic disturbances might have already been present) at baseline, our study focused on normoglycemic individuals who were also free of overt CVD at baseline. This makes our results much easier to interpret. Moreover, because we focused on apparently healthy individuals at baseline, and only a small proportion (1.1%) of participants received lipid-lowering drugs, our results are less likely to be confounded by drug use. Third, compared with previous research, our high-resolution liquid chromatography–mass spectrometry identified a larger number of lipid species spanning five lipid categories, including fatty acyls, glycerolipids, glycerophospholipids, sphingolipids, and sterol lipids. The high-coverage plasma lipidome allowed us to identify novel lipid species associated with T2D and provided unprecedented opportunities for future lipidomic research.

Fourth, we conducted comprehensive statistical analyses, including prospective association analyses, repeated measurement analyses, and multivariate analyses, to identify novel lipids and lipidomic signatures associated with T2D. Fifth, many lipids were consistently detected by different statistical models and were also reported in previous studies across different populations. Moreover, our analyses adjusted for a comprehensive list of clinical covariates, including age, sex, BMI, fasting glucose, IR, total triglycerides, HDL-c, diet, physical activity, and use of lipid-lowering medications, signifying the robustness of our findings.

Our study has several limitations. First, although we detected more than 1,500 lipids covering a wide range of molecular lipid species, we were unable to match many lipids to the current databases. We were also unable to distinguish and identify isomeric lipids. These unknown compounds and isomers need to be characterized using *de novo* identification or additional experiments if considered of interest. Second, because of our focus on normoglycemic participants who were also free of overt CVD at baseline, the number of incident T2D cases during the 5-year follow-up period was relatively small, and therefore, our power in detecting predictive lipids was limited. However, there was a large number of incident T2D cases during 16-year follow-up. Third, T2D ascertainment was based on one single blood test at baseline and 5-year follow-up, and the diagnosis of T2D at 16-year follow-up was based on medical records (because of lack of blood samples at 16-year follow-up). Fourth, although our analyses adjusted for many known risk factors, we cannot exclude the possibility of potential confounding by unknown or unmeasured factors (e.g., changes in lifestyle factors, insulin secretion, and resistance over time). Fifth, although our analysis included a large number of American Indians, we did not have an external validation because of the lack of comparable cohorts comprising normoglycemic individuals who developed T2D for whom longitudinal lipidomic data were also available. However, many lipids were consistently identified in different models in our own study and were also previously associated with diabetes (in the same direction) in different populations. This increases the confidence in and signifies the robustness of some of our findings. Finally, the observational nature of our study precludes any causal inference regarding the role of altered lipids in

diabetes etiology.

In summary, we identified novel lipid species (both baseline level and longitudinal change over time) and lipidomic signatures associated with risk of T2D beyond conventional risk factors. The newly identified lipids were altered years before the onset of T2D or prediabetes and could help identify high-risk individuals who may benefit from early intervention. Targeting metabolic pathways involving these newly identified lipids would help develop precision strategies tailored to American Indians.

ACKNOWLEDGMENTS. The authors thank the Strong Heart Study (SHS) participants, Indian Health Service facilities, and participating tribal communities for their extraordinary cooperation and involvement, which has contributed to the success of the SHS. The content expressed in this article is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or the Indian Health Service (IHS).

FUNDING. This study was supported by National Institutes of Health grants 1R01DK107532-01A1. The Strong Heart Study (SHS) has been funded in whole or in part with federal funds from the National Heart, Lung, and Blood Institute, National Institute of Health, Department of Health and Human Services, under contract numbers 75N92019D00027, 75N92019D00028, 75N92019D00029, and 75N92019D00030. The study was previously supported by research grants: R01HL109315, R01HL109301, R01HL109284, R01HL109282, and R01HL109319 and by cooperative agreements: U01HL41642, U01HL41652, U01HL41654, U01HL65520, and U01HL65521.

DUALITY OF INTEREST. No potential conflicts of interest relevant to this article were reported.

AUTHOR CONTRIBUTIONS. J.Z. conceived the study, supervised the statistical analyses, and wrote the manuscript. Z. H. and W.Z. conducted statistical analyses. Y. Z., G.W., D.P., B.deF, and O.F., collected the LC-MS data and carried out initial QC and data pre-processing for the lipidomic data. J.H., J.G.U. processed the plasma samples, S.A.C., E.T.L., and B.V.H. contributed to data interpretation, discussion and reviewed the manuscript. J.Z. is the guarantor of this work and, as such, had full access to all of the data in

the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

2.6 References

1. Hermans MP, Valensi P. Elevated triglycerides and low high-density lipoprotein cholesterol level as marker of very high risk in type 2 diabetes. Current Opinion in Endocrinology & Diabetes and Obesity. 2018;25(2):118–129.

2. Centers for Disease Control and Prevention. National Diabetes Statistics Report, 2020. Atlanta, GA, Centers for Disease Control and Prevention, U.S. Dept of Health and Human Services, 2020.

3. Meikle PJ, Wong G, Barlow CK, Weir JM, Greeve MA, MacIntosh GL, et al. Plasma lipid profiling shows similar associations with prediabetes and type 2 diabetes. PloS one. 2013;8(9).

4. Lu J, He J, Li M, Tang X, Hu R, Shi L, et al. Predictive value of fasting glucose, postload glucose, and hemoglobin A1c on risk of diabetes and complications in Chinese adults. Diabetes care. 2019;42(8):1539–1548.

5. Suvitaival T, Bondia-Pons I, Yetukuri L, Pöhö P, Nolan JJ, Hyötyläinen T, et al. Lipidome as a predictive tool in progression to type 2 diabetes in Finnish men. Metabolism. 2018;78:1–12.

6. Cummings DE, Overduin J, Foster-Schubert KE. Gastric bypass for obesity: mechanisms of weight loss and diabetes resolution. The Journal of Clinical Endocrinology & Metabolism. 2004;89(6):2608–2615.

7. Razquin C, Toledo E, Clish CB, Ruiz-Canela M, Dennis C, Corella D, et al. Plasma lipidomic profiling and risk of type 2 diabetes in the PREDIMED trial. Diabetes Care. 2018;41(12):2617–2624.

8. Wong G, Barlow CK, Weir JM, Jowett JBM, Magliano DJ, Zimmet P, et al. Inclusion of plasma lipid species improves classification of individuals at risk of type 2 diabetes. PLoS One. 2013;8(10).

9. Chew WS, Torta F, Ji S, Choi H, Begum H, Sim X, et al. Large-scale lipidomics identifies associations between plasma sphingolipids and T2DM incidence. JCI insight. 2019;4(13).

10. Mousa A, Naderpoor N, Mellett N, Wilson K, Plebanski M, Meikle PJ, et al. Lipidomic profiling reveals early-stage metabolic dysfunction in overweight or obese humans. Biochimica et Biophysica Acta (BBA)-Molecular and Cell Biology of Lipids. 2019;1864(3):335–343.

11. Yin X, Willinger CM, Keefe J, Liu J, Fernández-Ortiz A, Ibáñez B, et al. Lipidomic profiling identifies signatures of metabolic risk. EBioMedicine. 2020;51:102520.

12. Rhee EP, Cheng S, Larson MG, Walford GA, Lewis GD, McCabe E, et al. Lipid profiling identifies a triacylglycerol signature of insulin resistance and improves diabetes prediction in humans. The Journal of clinical investigation. 2011;121(4):1402–1411.

13. Kopprasch S, Dheban S, Schuhmann K, Xu A, Schulte K-M, Simeonovic CJ, et al. Detection of independent associations of plasma lipidomic parameters with insulin sensitivity indices using data mining methodology. PLoS One. 2016;11(10).

14. Merino J, Leong A, Liu C-T, Porneala B, Walford GA, von Grotthuss M, et al. Metabolomics insights into early type 2 diabetes pathogenesis and detection in individuals with normal fasting glucose. Diabetologia. 2018;61(6):1315–1324.

15. North KE, Howard B V, Welty TK, Best LG, Lee ET, Yeh JL, et al. Genetic and environmental contributions to cardiovascular disease risk in American Indians: the strong heart family study. American journal of epidemiology. 2003;157(4):303–314.

16. Lee ET, Welty TK, Fabsitz R, Cowan LD, Le N-A, Oopik AJ, et al. The Strong Heart Study A study of cardiovascular disease in American Indians: design and methods. American journal of epidemiology. 1990;132(6):1141–1155.

17. Matthews DR, Hosker JP, Rudenski AS, Naylor BA, Treacher DF, Turner RC. Homeostasis model assessment: insulin resistance and $β$-cell function from fasting plasma glucose and insulin concentrations in man. Diabetologia. 1985;28(7):412–419.

18. Katz A, Nambi SS, Mather K, Baron AD, Follmann DA, Sullivan G, et al. Quantitative insulin sensitivity check index: a simple, accurate method for assessing insulin sensitivity in humans. The Journal of Clinical Endocrinology & Metabolism. 2000;85(7):2402–2410.

19. Cajka T, Smilowitz JT, Fiehn O. Validating quantitative untargeted lipidomics across nine liquid chromatography-high-resolution mass spectrometry platforms. Anal Chem. 2017; 89:12360–12368.

20. Tsugawa H, Cajka T, Kind T, Ma Y, Higgins B, Ikeda K, et al. MS-DIAL: data-independent MS/MS deconvolution for comprehensive metabolome analysis. Nature methods. 2015;12(6):523.

21. Fan S, Kind T, Cajka T, Hazen SL, Tang WHW, Kaddurah-Daouk R, et al. Systematic error removal using random forest for normalizing large-scale untargeted lipidomics data. Analytical chemistry. 2019;91(5):3590–3596.

22. Hartigan JA, Hartigan PM, others. The dip test of unimodality. The annals of Statistics. 1985;13(1):70–84.

23. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the royal statistical society Series B (Methodological). 1995;289–300.

24. Rohart F, Gautier B, Singh A, Lê Cao K-A. mixOmics: An R package for 'omics feature selection and multiple data integration. PLoS computational biology. 2017;13(11):e1005752.

25. Schofield JD, Liu Y, Rao-Balakrishna P, Malik RA, Soran H. Diabetes dyslipidemia. Diabetes Therapy. 2016;7(2):203–219.

26. Fernandez C, Surma MA, Klose C, et al. Plasma lipidome and prediction of type 2 diabetes in the population-based Malmö diet and cancer cohort. Diabetes Care 2020; 43:366–373.

27. Lu J, Lam SM,Wan Q, et al. High-coverage targeted lipidomics reveals novel serum lipid predictors and lipid pathway dysregulation antecedent to type 2 diabetes onset in normoglycemic Chinese adults. Diabetes Care. 2019; 42:2117–2126.

28. Lu L, Koulman A, Petry CJ, et al. An unbiased lipidomics approach identifies early second trimester lipids predictive of maternal glycemic traits and gestational diabetes mellitus. Diabetes Care. 2016; 39:2232–2239.

29. Erion DM, Shulman GI. Diacylglycerol-mediated insulin resistance. Nature medicine. 2010;16(4):400–402.

30. Kraegen EW, Cooney GJ, Ye JM, Thompson AL, Furler SM. The role of lipids in the pathogenesis of muscle insulin resistance and beta cell failure in type II diabetes and obesity. Exp Clin Endocrinol Diabetes. 2001;109(Suppl. 2): S189–S201.

31. Castoldi A, Monteiro LB, van Teijlingen Bakker N, et al. Triacylglycerol synthesis enhances macrophage inflammatory function. Nat Commun. 2020; 11:4107.

32. Glass CK, Olefsky JM. Inflammation and lipid signaling in the etiology of insulin resistance. Cell Metab. 2012; 15:635–645.

33. Meikle PJ, Summers SA. Sphingolipids and phospholipids in insulin resistance and related metabolic disorders. Nature Reviews Endocrinology. 2017;13(2):79-91.

34. Imamura F, Fretts A, Marklund M, et al.; InterAct Consortium; Fatty Acids and Outcomes Research Consortium (FORCE). Fatty acid biomarkers of dairy fat consumption and incidence of type 2 diabetes: a pooled analysis of prospective cohort studies. PLoS Med. 2018; 15:e1002670.

35. Funai K, Lodhi IJ, Spears LD, et al. Skeletal muscle phospholipid metabolism regulates insulin sensitivity and contractile function. Diabetes. 2016; 65:358–370.

36. Funai K, Song H, Yin L, et al. Muscle lipogenesis balances insulin sensitivity and strength through calcium signaling. J Clin Invest. 2013; 123:1229–1240.

37. Selathurai A, Kowalski GM, Burch ML, et al. The CDP-ethanolamine pathway regulates skeletal muscle diacylglycerol content and mitochondrial biogenesis without altering insulin sensitivity. Cell Metab. 2015; 21:718–730.

38. Newsom SA, Brozinick JT, Kiseljak-Vassiliades K, et al. Skeletal muscle phosphatidylcholine and phosphatidylethanolamine are related to insulin sensitivity and respond to acute exercise in humans. J Appl Physiol (1985) 2016; 120:1355–1363.

39. Pilon M. Revisiting the membrane-centric view of diabetes. Lipids Health Dis. 2016; 15:167.

40. Mamtani M, Kulkarni H, Wong G, et al. Lipidomic risk score independently and cost-effectively predicts risk of future type 2 diabetes: results from diverse cohorts. Lipids Health Dis 2016; 15:67.

2.7 Supplementary Information

Supplementary Table 2.1. Baseline characteristics of NFG participants according to diabetes status by end of 5-year follow-up

Characteristics	T2D (N=73)	IFG (N=205)	NFG (N=883)	P-value† (T2D vs. NFG)	P-value† (IFG vs. NFG)
Age (years)	35.8 ± 11.3	36.5 ± 12.7	36.0 ± 13.1	0.90	0.57
Female, n (%)	44 (60%)	123 (60%)	570 (65%)	0.45	0.20
BMI (kg/m^2)	34.9 ± 7.6	31.9 ± 6.9	29.1 ± 6.4	6.25×10^{-12}	5.37×10^{-9}
Waist (cm)	110.4 ± 18.4	103.0 ± 16.2	96.0 ± 15.6	4.43×10^{-10}	3.12×10^{-8}
SBP (mmHg)	123.4 ± 14.1	121.9 ± 13.6	118.9 ± 14.7	4.98×10^{-3}	3.54×10^{-3}
DBP (mmHg)	79.0 ± 9.9	78.5 ± 9.7	75.4 ± 10.6	1.82×10^{-3}	4.24×10^{-5}
HDL (mg/dL)	49.1 ± 15.5	51.2 ± 12.7	54.5 ± 15.3	0.01	2.64×10^{-3}
LDL (mg/dL)	101.8 ± 29.0	103.5 ± 29.7	99.5 ± 29.1	0.50	0.08
Triglycerides (mg/dL)	177.9 ± 124.8	151.2 ± 73.3	136.1 ± 79.9	1.28×10^{-3}	0.02
Total cholesterol (mg/dL)	184.3 ± 35.1	184.8 ± 34.9	181.1 ± 33.5	0.45	0.16
Fasting glucose (mg/dL)	91.3 ± 6.5	91.0 ± 6.8	88.4 ± 6.4	2.34×10^{-3}	6.29×10^{-5}
Fasting insulin (μU/ml)	21.1 ± 16.0	14.5 ± 10.1	11.9 ± 10.3	1.74×10^{-6}	7.58×10^{-3}
HOMA-IR	4.7 ± 3.5	3.3 ± 2.3	2.6 ± 2.3	1.41×10^{-6}	3.88×10^{-3}
HOMA-β	295.8 ± 273.5	193.5 ± 153.7	177.2 ± 155.9	2.64×10^{-7}	0.18
QUICKI	0.14 ± 0.01	0.14 ± 0.01	0.15 ± 0.02	2.19×10^{-9}	1.56×10^{-5}

Data are mean ± SD unless otherwise indicated.

†P-values were obtained by generalized estimating equation (GEE) model accounting for correlation among family members.

2.7.1 Participants and sample collection

Participants. The Strong Heart Family Study (SHFS, 2001-ongoing), a component of the Strong Heart Study (SHS, 1989-ongoing), is a multicenter, family-based prospective study designed to identify genetic factors for cardiovascular disease (CVD), T2D and their risk factors in American Indians. A total of 2,780 tribal members (≥18 years old) from 12 tribes residing in Arizona, North Dakota, South Dakota, and Oklahoma were initially examined in 2001-2003 and re-examined in 2006-2009 (mean 5.5 years apart) using the same

protocols. At each visit, participants received a personal interview to collect data on demographic characteristics, medical history and lifestyle risk factors including smoking, alcohol consumption, diet and physical activity. A physical examination was given to each participant, including anthropometric and blood pressure measurements and an examination of the heart and lungs. Biospecimens including fasting plasma samples were collected at each visit. Laboratory methods were reported previously (1). All living participants are currently being followed through 2026 to collect information for disease morbidity and mortality. The SHS protocols were approved by the Indian Health Service, institutional review boards for each participating institution, and participating communities. All participants have given informed consent. Detailed descriptions of the SHS protocols for the collection of phenotype data have been described previously (2).

Participants in the current study met the following criteria: 1) attended clinical examinations and had available fasting plasma samples at both baseline (2001–2003) and 5-year follow-up (2006–2009); 2) were free of overt CVD at baseline. Participants with missing information for fasting glucose or hypoglycemia medications at either time point were excluded.

Definition of T2D. According to the American Diabetes Association 2003 criteria (3), diabetes was defined as fasting plasma glucose ≥7.0 mmol/L or using hypoglycemic medications. Impaired fasting glucose (IFG) was defined as a fasting glucose of 6.1–6.9 mmol/L and no hypoglycemic medications, and normal fasting glucose (NFG) was defined as fasting glucose <6.1 mmol/L. Incident diabetes was defined as participants who had NFG at baseline (2001-2003) but developed new diabetes by the end of 5-year follow-up (2006-2009) or 16-year follow-up (December 31, 2017).

Assessments of clinical factors. Fasting plasma glucose, insulin, lipids, and lipoproteins were measured by standard laboratory methods as previously described *(1)*. Body mass index (BMI) was calculated as body weight in kilograms divided by the square of height in meters. Hypertension was defined as blood pressure levels ≥140/90 mmHg or use of antihypertensive medications. Insulin resistance was assessed using h*omeostatic model assessment* (HOMA) according to the following formula: HOMA-IR = fasting glucose

(mg/dL) × insulin (μU/mL)/405.(4) Pancreatic *β*-cell function (HOMA-β) was assessed using the formula: 360 × fasting insulin (μU/mL) / (fasting glucose (mg/dL) - 63).(4) Insulin sensitivity was estimated *by* calculating the Quantitative Insulin sensitivity check index (QUICKI =1/[log insulin (mU/L) + log baseline glucose (mg/dL)] *(5).* Renal function was assessed using the estimated glomerular filtration rate (eGFR) calculated by the MDRD equation *(6)*. For cigarette smoking, participants were classified as current smokers, former smokers, and never smokers. Alcohol consumption was determined by self-reported history of alcohol intake, the type of alcoholic beverages consumed, frequency of alcohol consumption, and average quantity consumed per day and per week. Physical activity was assessed by the mean number of steps per day calculated by averaging the total number of steps recorded each day during a 7-day period (7). Dietary intake, including total protein intake, total calories intake, total fat intake was assessed using the Block Food Frequency questionnaire *(8)*. Information on use of medications including anti-hypertensive, hypoglycemic and lipid-lowering drugs was also collected at each visit.

Blood sample collection. Participants were instructed to fast overnight before their visit, and fasting blood sample was collected into 10ml EDTA tubes at the SHS field centers. The tubes were then gently inverted and placed on ice or refrigerated (- 4 °C) immediately. Plasma sample was obtained by centrifuging the tubes for 10 minutes at 3,000 rpm at - 4 °C and aliquots (0.5 ml) were immediately stored at -80 °C until further analysis. For the current study, 0.5 ml fasting plasma sample, which were never thawed before, was shipped to Dr. Fiehn's lab at the West Coast Metabolomics Center (UC-Davis) on dry ice via FedEx overnight, and stored at -80 ºC immediately on arrival until further analyses. Samples were randomized before shipping to the Fiehn's laboratory, where randomization was performed again before the lipidomics analysis as described below. As part of the QC procedures, 109 duplicated samples (55 at baseline and 54 at 5-year follow-up) were included to evaluate analytical or measurement precision. Lipid values of the duplicated samples were highly correlated (Spearman's correlation coefficient, $\rho = 0.95$, $P < 2.2 \times 10^{-16}$). Laboratory technicians were blinded to all clinical data throughout the assays.

2.7.2 Lipidomics data acquisition via liquid chromatograph-mass spectrometry (LC-MS)

Plasma samples were first extracted based on a modified liquid-liquid extraction method (cold methanol/ MTBE/water). The extracted samples were then subjected to lipidomics analysis by LC-MS in both positive and negative ionization modes.

Lipidomics ESI (+) on 6550 Agilent LC-QTOF MS**.** The injection volume on ESI (+) mode was 3 μL. The mobile phase compositions were: A) Acetonitrile: water (60:40, v/v) with 10 mM ammonium formate and 0.1% formic acid; and B) Isopropanol: acetonitrile (90:10, v/v) with 10 mM ammonium formate and 0.1% formic acid. The LC gradient consisted of the following elution conditions: 0 min 15% (B); 0−2 min 30% (B); 2−2.5 min 48% (B); 2.5−11 min 82% (B); 11−11.5 min 99% (B); 11.5−12min 99% (B); 12−12.1 min 15% (B); and 12.1−15 min 15% (B).

Lipidomics ESI (-) on 6550 Agilent LC-QTOF MS. The injection volume on ESI (-) mode was 5 μL. The mobile phase compositions were: A) Acetonitrile: water (60:40, v/v) with 10 mM ammonium acetate; and B) Isopropanol: acetonitrile (90:10, v/v) with 10 mM ammonium acetate. The LC gradient consisted of the following elution conditions: 0 min 15% (B); 0−2 min 30% (B); 2−2.5 min 48% (B); 2.5−11 min 82% (B); 11−11.5 min 99% (B); 11.5−12min 99% (B); 12−12.1 min 15% (B); and 12.1−15 min 15% (B).

The lipids were separated on Agilent 1290 Infinity LC system using an Acquity CSH C18 column (100 mm × 2.1 mm, 1.7 μm) with an Acquity CSH C18 guard column (5 mm × 2.1 mm, 1.7 μm) (Waters, Milford, MA). The column temperature was 65 °C with a flow rate of 0.6 mL/min. Sample temperature was maintained at 4 °C throughout the experiment. The guard column was changed every 300 samples, and a new column was replaced every 1,000 samples. We also measured a Bioreclamation plasma sample per 10 samples and a NIST plasma sample per 40 samples as quality controls for monitoring the instrumental drift. Detailed methods for sample extraction and lipidomic analysis on ESI (+) and ESI (-) were described in the online supplementary methods.

2.7.3 Lipidomics data pre-processing and quality control

The lipidomics data were pre-processed using a new in-house cloud-based software (LC-BinBase) with peak detection and deconvolution algorithms adapted from MS-DIAL (9). Raw files were automatically converted into correct formats and the LC-BinBase algorithms performed peak picking, retention time alignment using internal standards, and gap filling from raw data for missed peaks. Lipid peak intensity results were manually checked against raw data files. Adducts were combined into single features for statistical assessments. False negative features and peaks with 50% missing values across all samples were removed. The batch effect of reformatted dataset was normalized by SERRF software (Systematic Error Removal using Random Forest) (10), which dramatically reduced the raw data variance coefficient by 23% in positive mode data and 25% in negative mode to less than 10% in result files. After data pre-processing, we obtained 1,809 lipids (579 known and 1,230 unknown lipids) in both positive and negative ionizations (787 positive and 1,022 negative) from 1,983 samples at baseline and 1,994 samples at 5-year follow-up. *The lipid levels of the duplicated samples were highly correlated (Spearman's correlation coefficient* $\rho = 0.95, p < 2.2 \times 10^{-16}$ *).*

2.7.4 References

1. Lee ET, Welty TK, Fabsitz R, Cowan LD, Le N-A, Oopik AJ, et al. The Strong Heart Study A study of cardiovascular disease in American Indians: design and methods. American journal of epidemiology. 1990;132(6):1141–55.

2. North KE, Howard B V, Welty TK, Best LG, Lee ET, Yeh JL, et al. Genetic and environmental contributions to cardiovascular disease risk in American Indians: the strong heart family study. American journal of epidemiology. 2003;157(4):303–14.

3. Genuth S, Alberti KG, Bennett P, Buse J, Defronzo R, Kahn R, et al. Expert Committee on the Diagnosis and Classification of Diabetes Mellitus2, the Expert Committee on the Diagnosis and

Classification of Diabetes Mellitus. Follow-up report on the diagnosis of diabetes mellitus. Diabetes Care. 2003;26:3160–7.

4. Matthews DR, Hosker JP, Rudenski AS, Naylor BA, Treacher DF, Turner RC. Homeostasis model assessment: insulin resistance and β-cell function from fasting plasma glucose and insulin concentrations in man. Diabetologia. 1985;28(7):412–9.

5. Katz A, Nambi SS, Mather K, Baron AD, Follmann DA, Sullivan G, et al. Quantitative insulin sensitivity check index: a simple, accurate method for assessing insulin sensitivity in humans. The Journal of Clinical Endocrinology & Metabolism. 2000;85(7):2402–10.

6. Froissart M, Rossert J, Jacquot C, Paillard M, Houillier P. Predictive performance of the modification of diet in renal disease and Cockcroft-Gault equations for estimating renal function. Journal of the American Society of Nephrology. 2005;16(3):763–73.

7. Tudor-Locke C, Ainsworth BE, Thompson RW, Matthews CE. Comparison of pedometer and accelerometer measures of free-living physical activity. Medicine & Science in Sports & Exercise. 2002;34(12):2045–51.

8. Willett W. Nutritional epidemiology. Vol. 40. Oxford university press; 2012.

9. Tsugawa H, Cajka T, Kind T, Ma Y, Higgins B, Ikeda K, et al. MS-DIAL: data-independent MS/MS deconvolution for comprehensive metabolome analysis. Nature methods. 2015;12(6):523–526.

10. Fan S, Kind T, Cajka T, Hazen SL, Tang WW, Kaddurah-Daouk R, et al. Systematic error removal using random forest for normalizing large-scale untargeted lipidomics data. Analytical chemistry. 2019;91(5):3590–3596.

Chapter 3 Sex-Dependent Plasma Metabolome and Metabolite Associations with Mouse Phenotypes

Reproduced from Ying Zhang, Dinesh K. Barupal, Sili Fan, Bei Gao, Chao Zhu, Ann M. Flenniken, Colin McKerlie, Lauryl M. J. Nutter, K. C. Kent Lloyd and Oliver Fiehn. "Sex-dependent plasma metabolome and metabolite associations with mouse phenotypes". Major revision, revised, resubmitted to Nature Communication.

3.1 Abstract

Although metabolic alterations are observed in many monogenic and complex genetic disorders, the impact of most mammalian genes on cellular metabolism remains unknown. Understanding the effect of mouse gene dysfunction on metabolism can inform the functions of their human orthologues. We investigated the effect of loss-of-function mutations in 30 unique gene knockout (KO) lines on plasma metabolites including genes coding for structural proteins (11 of 30), metabolic pathway enzymes (12 of 30) and protein kinases (7 of 30). Steroids, bile acids, oxylipins, primary metabolites, biogenic amines, and complex lipids were analyzed with dedicated mass spectrometry platforms, yielding 827 identified metabolites in male and female KO mice and wildtype (WT) controls. Twenty-two percent of 23,698 KO versus WT comparison tests showed significant genotype effects on plasma metabolites. Fifty-six percent of identified metabolites were significantly different between the sexes in WT mice. Many of these metabolites were also found to have sexually dimorphic changes in KO lines. We used plasma metabolites to complement phenotype information exemplified for *Dhfr*, *Idh1*, *Mfap4*, *Nek2*, *Npc2*, *Phyh* and *Sra1*. Association of plasma metabolites with IMPC phenotypes showed dramatic sexual dimorphism in wildtype mice. We demonstrate how to link metabolomics to genotypes and (disease) phenotypes. Sex must be considered as critical factor in biological interpretation of gene functions.

3.2 Introduction

As a result of greater efforts to include both male and female subjects in biomedical research, sexually dimorphic abnormal phenotypes are now being recognized in disease studies that analyze the

pathophysiological consequences of genetic shift/reproduction [1-11]. Historically, sex-bias in animal studies were extensive resulting in potentially misleading or incomplete conclusions [12-16]. In neuroscience, for example, studies in male rats, mice, monkeys, and other mammals outnumbered those in females by 5.5 to 1 [16]. In 2014 the U.S. National Institutes of Health issued a policy mandating the inclusion of female and male subjects in animal and cell research [12] and that sex be factored as a biological variable into research design, analysis, and reporting [17]. Sexual dimorphism may be established by the action of hormones that are expressed in different amounts between the sexes [18]. These hormones may act differently on peripheral organs and thus impact a multitude of systems, including cardiovascular physiology, the immune system, and nutrient absorption, and can be involved in complex diseases such as type 2 diabetes, obesity, and Alzheimer's disease [19-23].

In general, sexual dimorphism has been observed in various classes of metabolites, such as branch-chain amino acids, phosphocholines, sphingomyelin, and urea cycle metabolites [24-27]. Yet, the impact of gene dysfunction on metabolism has not been broadly studied. Specific genes may have pleiotropic effects on the metabolic systems of female and male cells, tissues, organs, and organ systems [28-31]. To study the *in vivo* effects using a mammalian model, we obtained samples from KO mice generated by the International Mouse Phenotyping Consortium (IMPC), a global program linking 16 different research centres in 12 countries on 5 continents engaged in identifying the *in vivo* function of all protein coding human gene homologs in the mouse genome. The IMPC uses standardized phenotyping protocols that generate ~1,200 parameters. All of the IMPC's material and data resources are publicly available for each of ~7,000 KO lines phenotyped to date (https://www.mousephenotype.org/) [32], including 208 classified as continuous variables, some of which are metabolic traits, such as glucose tolerance and total plasma triacylglyceride content. For continuous traits in 2,186 KO lines, 9% of phenotype parameters showed significant genotypic differences of which 14% were sexually dimorphic [33].

Metabolomics is increasingly used in epidemiology and clinical research [34,35] to understand the mechanism of disease development [36,37]. Mass spectrometry is the dominant technology in targeted and

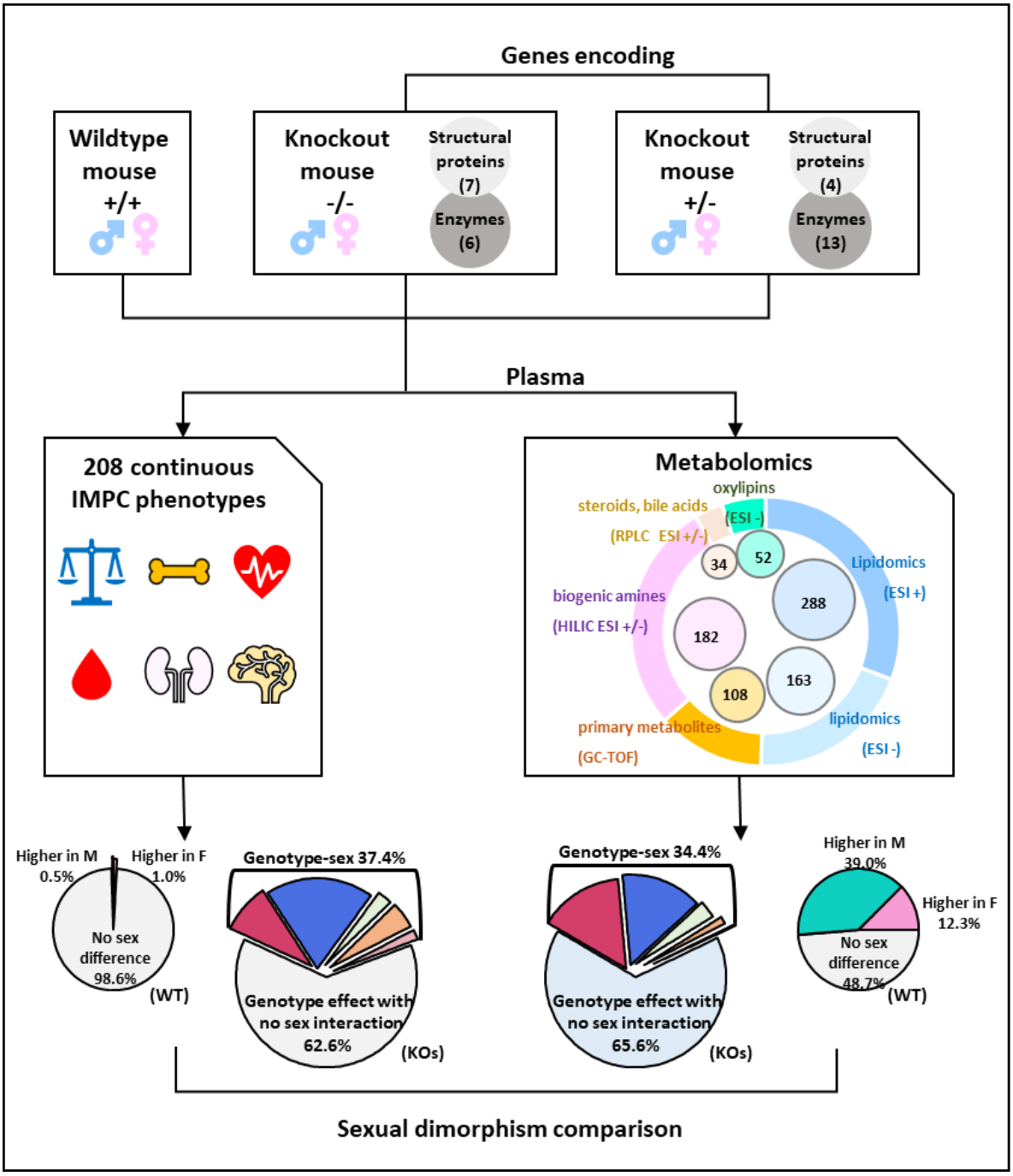


Figure 3.1. Graphic abstract of study design and data analysis. 30 gene knockout lines and corresponding wildtype controls were selected with phenotypic data available from IMPC. Plasma samples of 220 mice were analyzed using 5 assays including 3 untargeted metabolomic profiling and 2 targeted data acquisition methods.

untargeted metabolomics [38,39]. Version 4.0 of the Human Metabolome Database records more than 25,300 blood metabolites [40], but due to the chemical diversity of metabolites, individual analytic approaches can identify far fewer metabolites. Data sets from most individual metabolomic studies comprise 100─1,200 metabolites [41-43]. We used multiple analytical methods to interrogate several metabolic pathways, including hydrophilic interaction liquid chromatography-mass spectrometry (HILIC-MS), gas chromatography-time-of-flight mass spectrometry (GC-TOF MS), and reverse-phase liquid chromatography-mass spectrometry (RPLC-MS) [44]. Mice are the most often used *in vivo* model to study the functions of genes compared to human orthologues. We here showcase how metabolomic data can be used to complement the existing IMPC data repositories to link sex, genotypes and phenotypes.

3.3 Results

3.3.1 827 unique metabolites were detected in mouse plasma

Two-hundred and twenty mice (110 female and 110 male) on the C57BL/6NCrl background produced and phenotyped by The Center for Phenogenomics (TCP, Toronto, ON, Canada) as part of the IMPC project were used for this study. Approximately 30% of KO lines analyzed by the IMPC are embryonic lethal [45] for which test cohorts of heterozygous KO mice were used for adult phenotyping. Blood plasma was obtained at the end of the *in vivo* phenotyping pipeline. The metabolic impact of null mutations was tested on 17 adult heterozygous lines and 13 adult homozygous lines (**Fig. 3.1**) using 6 mice (3 female and 3 male) of each KO line. Control samples were collected from 40 (20 female and 20 male) co-housed sex- and age-matched C57BL/6NCrl WT mice over the same time period as the KO mice (**Supplementary Table 3.1**). KO lines of genes coding for structural proteins (11 of 30), metabolic pathway enzymes (12 of 30) and protein kinases (7 of 30) were selected to assess their impact on blood metabolic phenotypes. Among those genes, 16 gene KOs were used as models for human diseases (by similarities of annotation and orthology) (**Supplementary Table 3.1**). Using previously established IMPC criteria, phenotypes of these disease mouse models overlapped with human diseases that harbor mutations in the mouse orthologous genes [45]. We acquired comprehensive metabolomic data using a total of 20 µL lithium heparin plasma per mouse for

three untargeted metabolomics assays and an additional 50 µL lithium heparin plasma to target low abundant bile acids, steroids, and oxylipins [44]. In total, 827 unique metabolites identified using these five mass spectrometry assays were utilized for sexual dimorphism assessment, detecting 21 bile acids, 13 steroids, 52 oxylipins, 108 primary metabolites, 451 complex lipids, and 182 biogenic amines. The assays showed little overlap of compounds (**Fig. 3.1**), proving their complementary nature and utility.

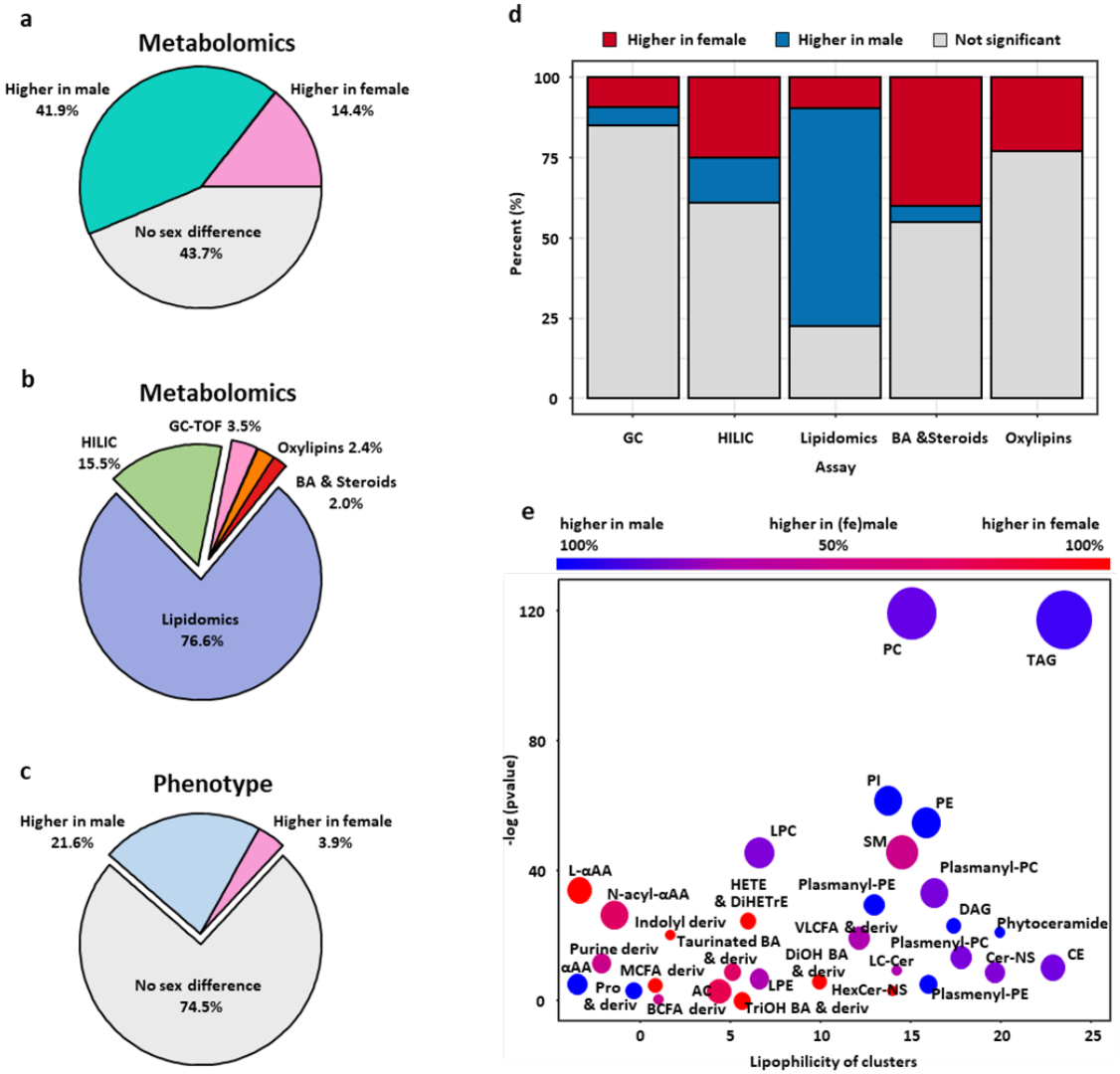
3.3.2 Sexual dimorphism in wildtype (WT) mice

3.3.2.1 Overview of sexual dimorphism of metabolomics and phenotype data in WT mice

The differences in metabolomic data were first assessed between 20 female and 20 male WT mice. Fifty-six percent of all plasma metabolites displayed sexual dimorphism using a generalized linear model at $P < 0.05$. The levels of 337 of 805 metabolites (41.9%) were increased in male mice compared to females, and 116 of 805 (14.4%) metabolites were more abundant in female than in male mice (**Fig. 3.2a**). With a more stringent criteria at FDR < 0.05, this analysis only slightly changed the proportion of significant metabolites from 56.3% to 51.8% with 39.3% plasma metabolites more abundant in male mice, and 12.6% in female mice (**Supplementary Fig. 3.2a**). Using chemical similarity enrichment statistics by ChemRICH impact plot [46], we found 31 classes of compounds that were enriched by direct comparison of the two sexes (**Fig. 3.2e**). Most complex lipids showed higher plasma levels in male mice than in females (**Fig. 3.2b, d, e**), for example, with higher abundances of triacylglycerides (TAG), phosphatidylcholines (PC), and phosphoinositides (PI) in WT males. Many other metabolites were found at higher levels in female mice than in males (**Fig. 3.2d**). Within the biogenic amines, some metabolite classes such as acylcarnitines, indole derivatives, and L-alpha-amino acids showed sexual dimorphism with higher levels detected in WT female mouse plasma (**Fig. 3.2e**). While the sexual dimorphism is dominated by complex lipids, the contribution of bile acids and oxylipins seem to exceed that of primary metabolites in terms of their proportions in each platform (**Fig. 3.2d**). When analyzing previously reported IMPC phenotype data [47], we also found a strong sexual dimorphism in *in vivo* phenotypes of these WT mice (**Fig. 3.2a, c**), with similarly more variables significantly greater in male mice compared to females. Overall plasma metabolites levels

were so different that perfect discrimination between the sexes was achieved by PLS-DA multivariate statistics (Supplementary Fig. 3.1).

Figure 3.2. Sex as a biological variable in wildtype mice. 2a) Metabolites that were significantly affected



by sex in 40 wildtype mice (n = 20 males & 20 females, generalized linear model was used at $P < 0.05$). 2b) The distribution of sex-affected metabolites among five metabolic assays ($P < 0.05$). 2c) IMPC phenotypes of continuous variables that were significantly affected by sex in 40 wildtype mice ($P < 0.05$). 2d) Metabolites affected by sex per metabolomics assay ($P < 0.05$). 2e) Chemical Similarity Enrichment Analysis between male and female wildtype mice by ChemRICH impact plot (larger dot size indicates greater number of metabolites in the cluster). PC: Phosphatidylcholines; TAG: Triacylglycerides; DAG: Diacylglycerides; PI: Phosphatidylinositols; PE: Phosphatidylethanolamines; plasmanyl-PC: plasmanyl-phosphocholines; plasmenyl-PC: plasmenyl-phosphocholines; plasmanyl-PE: plasmanyl-phosphoethanolamines; plasmenyl-PE: plasmenyl-phosphoethanolamines; SM: Sphingomyelins; LPC: Lyso-phosphocholines; LPE: Lysophosphatidylethanolamine; CE: Cholesteryl esters; Cer-NS: Ceramides; LC-Cer: Long Chain Ceramides; HexCer-NS: Glycosyl-N-acylsphingosines; Phytoceramide:

Phytoceramides; HETE & DiHETrE: Hydroxyeicosatetraenoic acids & Hydroxyeicosatrienoic acids; DiOH BA & deriv: Dihydroxy bile acids, alcohols and derivatives; TriOH BA & deriv: Trihydroxy bile acids, alcohols and derivatives; Taurinated BA & deriv: Taurinated bile acids and derivatives; VLCFA and deriv: Very long-chain fatty acids and derivatives; MCFA deriv: Medium-chain hydroxy acids and derivatives; BCFA deriv: Hydroxy fatty acids; AC: Acyl carnitines; L-αAA: Proteinogenic L-α-amino acids and derivatives; αAA: Non-proteinogenic L-α-amino acids and derivatives; N-acyl-αAA: N-acyl-alpha amino acids; Pro & deriv: Proline and derivatives: Indolyl deriv: Indolyl acids & derivatives; Purine deriv: Purine nucleosides.

Metabolite levels may be impacted by total body weight. Hence, sexual dimorphism of metabolites could be attributed to these known differences between the sexes. We therefore conducted a secondary analysis by adjusting metabolite levels to body weight at the time of blood collection. After adjusting to body weight at an FDR < 0.05, 51.3% of all metabolites showed sexual dimorphism (39.0% of metabolites higher in male mice and 12.3% higher in females, **Supplementary Fig. 3.2d**). For comparison, 0.5% of all 208 IMPC-measured continuous phenotypic variables were higher in wildtype male mice after body weight adjustment and FDR correction, and 1.0% of the phenotypes were increased in female mice (**Supplementary Fig. 3.2f**). These results showed that differences in body weight impacted the sex differences in phenotype data but not in plasma metabolite levels.

3.3.2.2 Example of metabolites with sexual dimorphism in WT mice

Table 3.1 shows 20 individual examples of metabolites that were significantly different between WT female and male mice. Metabolites were chosen based on metabolic classes, significance levels, and fold change. Most of these metabolites were also found to be affected by KO mutations, except for a single phosphatidylcholine membrane lipid (PC 40:4). Many of these compounds also showed sexual dimorphism in KO lines.

Table 3.1. Examples of top metabolites with sexual dimorphism in wildtype mice and their alterations in 30 knockout lines.

	Metabolite	Metabolite Class	Sex Effect in WT	Statistical Analysis in KO Lines	
				Genotype Effect	Genotype-sex Interaction
Higher in female wildtype mice	**Glycocholic acid**	Bile acids	**	*Ahcy, C8a, Cdk4, Dhfr, Galc, Idh1, Lmbrd1, Mfap4, Mmachc, Mvk, Phyh, Ptpn12, Rock1, Ulk3, Ywhaz*	*Ahcy, C8a, Cdk4, Dhfr, Lmbrd1, Mfap4, Mvk, Phyh, Ptpn12, Rock1, Ulk3, Ywhaz*
	TMAO	Biogenic amines	***	*Ahcy, Atp5b, Dhfr, Gnpda1, Lmbrd1, Npc2, Pebp1, Plk1*	*Plk1*
	Tauroursodeoxycholic acid	Bile acids	***	*Ahcy, C8a, Cdk4, Dhfr, Dync1li1, G6pd2, Gnpda1, Idh1, Lmbrd1, Mmachc, Mvk, Nek2, Npc2, Pebp1, Phyh, Ptpn12, Pttg1, Rock1, Ulk3, Ywhaz*	*Ahcy, C8a, Dhfr, Lmbrd1, Mvk, Nek2, Npc2, Pttg1*
	Chenodeoxycholic acid	Bile acids	**	*Atp5b, Dhfr, Pmm2, Pttg1*	/
	Progesterone	Steroids	*	*A2m, Atp6v0d1, Dync1li1, G6pd2, Galc, Mfap4, Phyh, Plk1, Sra1*	*A2m, Atp6v0d1, Mfap4, Plk1*
	2-Indolinone	Biogenic amines	***	*A2m, Ahcy, Atp5b, Atp6v0d1, C8a, Dhfr, Dync1li1, G6pd2, Galc, Iqgap1, Mmachc, Mvk, Nek2, Npc2, Pebp1, Phyh, Plk1, Pmm2, Pttg1, Rock1, Sra1*	/
	PGF3alpha	Oxylipins	***	*Lmbrd1, Mfap4, Mmachc, Pebp1, Phyh, Plk1, Ptpn12, Ywhaz*	*Lmbrd1, Mfap4, Pebp1, Phyh*
	11,12-EpETrE	Oxylipins	***	*Atp5b, Mvk, Sra1*	*Mvk*
	SM d36:2	Sphingomyelins	***	*Atp5b, Atp6v0d1, C8a, Pmm2, Ptpn12*	*C8a*
	HexCer-NS d18:1/16:0	Ceramides	***	*C8a, Sra1*	*C8a*
Higher in male wildtype mice	**Cer-NS d18:2/22:0**	Ceramides	***	*C8a, Dync1li1, G6pd2, Pebp1, Sra1*	*Dync1li1, G6pd2, Pebp1, Sra1*
	PC 40:4	Phosphocholines	***	/	/
	PC 38:2	Phosphocholines	***	*A2m, Atp5b, Dync1li1, Pebp1, Pmm2, Sra1*	*Sra1*
	Testosterone	Steroids	***	*Atp5b, Plk1*	/
	LPC 20:0	LysoPC lipids	***	*A2m, Cdk4, Dync1li1, Pebp1, Sra1*	/
	PI 16:0-16:1	Phosphatidylinositols	***	*A2m, Galc, Idh1, Nek2, Pebp1, Sra1, Ywhaz*	*Galc, Idh1, Nek2, Pebp1*
	PE 18:0-22:5	Phosphatidylethanolamines	***	*Ahcy, Galc*	*Ahcy*
	Cer (d18:1(4E)/22:0)	Ceramides	***	*G6pd2, Sra1*	*Sra1*
	PC 36:6	Phosphocholines	***	*Ulk3*	/
	TAG 14:0-16:0-18:2	Triacylglycerols	***	*A2m, Lmbrd1, Npc2, Pebp1*	*Npc2, Pebp1*

Note: Two-way ANOVA was used for statistical analysis. *, $P < 0.05$; **, $P < 0.01$, ***, $P < 0.001$.

Based on studies of human cohorts including men and women, blood metabolite levels usually do not show large differences [48], often displaying changes of less than 30% in human genome-wide association studies when analyzing the differences between sexes [49,50]. In this study we found more than 18.5% of 805 metabolites showed a sex difference greater than 2 folds in WT mice. Plasma levels of sex hormones were expected to be significantly different between the sexes, with testosterone only detected in the plasma of WT male mice. Progesterone was found at 2.5-fold higher levels in WT female mice due to their role in the estrous cycle, but it was also present in male mice as a crucial intermediate in the production of other endogenous steroids. A similar concentration ratio between sexes has also been reported in humans [51]. Yet, many other compounds were not well-known to be differentially present in the sexes. For example, the gut

microbial metabolite trimethylamine-N-oxide (TMAO) was found at 4-fold higher levels in female than in male mice (**Table 3.1**). Different levels of circulating TMAO have also been observed in humans [52]. Similarly, specific microbially transformed secondary bile acids like glycocholic acid and tauroursodeoxycholic acid were also found at higher levels in plasma from female mice, indicating a potentially differential impact of the gut microbiome on the sexes. A range of arachidonyl-lipid mediators such as the oxylipin 11,12-epoxyeicosa-5,8,14-trienoic acid (11,12-EpETrE) were detected in increased concentrations in female compared to male mice (**Table 3.1**), as well as 8,9-epoxyeicosatrienoic acid (8,9-EpETrE) and 8,9-dihydroxy-*5Z,11Z,14Z*-eicosatrienoic acid (8,9-DiHETrE). Conversely, the most significantly elevated plasma metabolites in male mice mainly belonged to polar and neutral lipid species (**Table 3.1, Fig. 3.2e**), including diacyl- and monoacylphosphatidylcholines (PC, LPC), phosphatidylethanolamines (PE), and neutral fats (triacylglycerides, TAG). Adjustment to whole body weight differences between the sexes did not change this finding. Higher plasma levels of both membrane phospholipids and fats are partly explained by higher levels of circulating lipoproteins in male mice. Indeed, high-density lipoprotein (HDL) cholesterol and non-HDL cholesterol were determined to be 37% and 7% higher in male mice, respectively (**Supplementary Table 3.3**), along with total plasma TAGs measured by the IMPC clinical chemistry protocols in concordance with our LC-MS/MS determination of significantly higher individual plasma TAG levels in male mice. Other selected IMPC phenotype differences are listed in **Supplementary Table 3.3**.

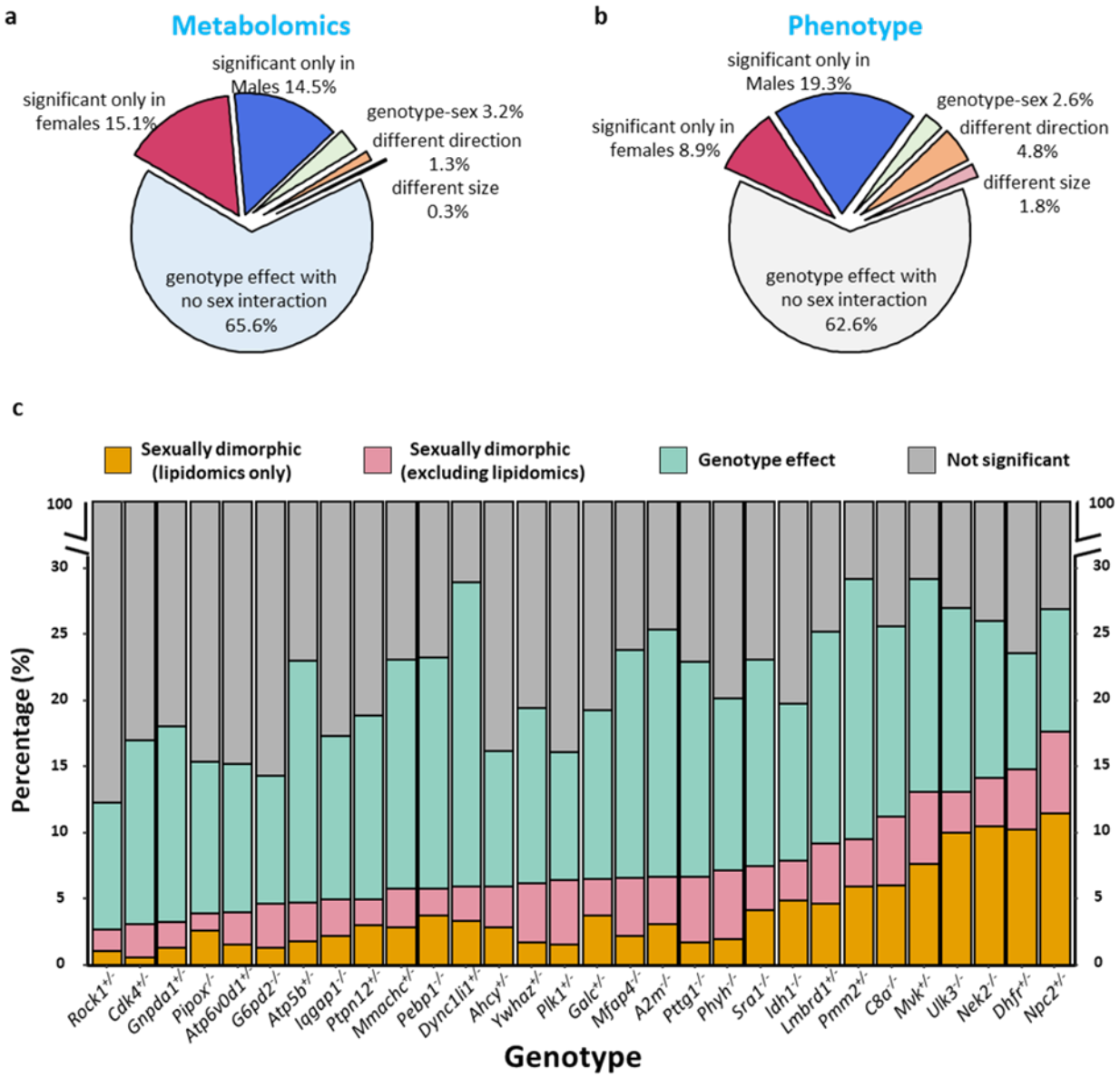


Figure 3.3. Overall genotype effect with sex interaction on metabolomics/phenotype data of 30 KO lines (n = 20 male & 20 female C57BL/6NCrl controls, n = 3 male & 3 female mice per KO line group. Two-way ANOVA at $P < 0.05$). (3a) Classification of significant genotype and sex interaction effect on the metabolomics dataset. (3b) Classification of significant genotype and sex interaction effect on the IMPC phenotype dataset. (3c) The proportion of metabolites that were altered by genotype effect and genotype-sex interaction effect for each KO line.

3.3.3 Sexual dimorphism in mouse KO lines

3.3.3.1 Two-way ANOVA revealed sexual dimorphism in 30 KO mouse lines

We found that more than 260 metabolites had significant genotype effects with more than 2-fold differences in plasma levels, such as a 3-fold change in plasma adenosine level in *Sra1*$^{-/-}$ (steroid receptor agonist 1) mice or 12(13)-Ep-9-KODE with greater than 3-fold change in *Npc2*$^{+/-}$ (NPC intracellular cholesterol transporter 2) mice compared to WT controls.

Next, we explored if plasma metabolite levels of KO lines were differentially influenced in male and female mice. All mice were housed at The Center for Phenogenomics under identical conditions using standard operating procedures and on the same inbred strain background, C57BL/6NCrl [45]. Hence, we could exclude environmental and husbandry effects that were unrelated to genotype-sex interactions. We conducted two-way ANOVA tests to assess the interaction effect of genotype and sex on all 208 continuous phenotypes reported by the IMPC and our metabolome data for the 30 KO lines in this study. IMPC phenotype data showed that 26.3% of all 4,756 comparisons across the 30 KO lines had significant phenotype differences (5.0% from body weight measured at different time points). Among those differences, 37.4% of the phenotypes were found to be sexually dimorphic (6.7% from body weight measured at different time points, **Fig. 3.3b**). Our metabolomics data identified 21.5% among 23,698 KO lines / WT comparison tests were significant at $P < 0.05$. Similarly, about one third of metabolomic changes showed significant genotype-sex interactions (**Fig. 3.3a**). Most changes were based on alterations in one sex but not the other. While both phenotype (**Fig. 3.3b**) and metabolome analyses (**Fig. 3.3a**) showed that a majority of the significant differences were true for both sexes and at similar magnitude and direction, phenotype analyses had a higher percent of traits that were significantly different from WT mice in both sexes but with changes in different directions or with different effect sizes between the two sexes (**Fig. 3.3b**). In comparison, metabolome analyses showed that the proportion of metabolites different from WT in only one sex was slightly higher with a lower proportion of metabolites showing opposite directions and different effect sizes in the two sexes (**Fig. 3.3a, b**). For metabolome data, the degree of sexual dimorphism was more pronounced in some

KO lines than in others (**Fig. 3.3c**). For example, 26.9% of all plasma metabolites were altered in *Npc2*$^{+/-}$ mice, but plasma levels of almost 64.8% of those compounds were sexually dimorphic (**Fig. 3.3c**). Similarly, 44.8% of all significantly genotype-affected metabolites in the *Mvk*$^{+/-}$ (metabolic enzyme mevalonate kinase) mice were found to be sexually dimorphic (**Fig. 3.3c**).

Initially we hypothesized that a number of genes might be so distant from enzymatic functions that no overall changes of metabolic phenotypes would be detected in such KO mouse plasma. However, we identified at least a few metabolites that were differently affected between the sexes compared to WT mice in each of the 30 KO lines (**Fig. 3.3c**). Each gene KO showed specific metabolites that were differentially regulated between male and female mice, ranging from 3-17% of the annotated metabolome (**Fig. 3.3c**). For example, the least affected KO mouse line was *Rock1*$^{+/-}$, coding for a gene involved in the regulation of cell motility, cell cycle, and cell adhesion, with about 12.3% of the metabolome found to be changed in *Rock1*$^{+/-}$ mice compared to WT controls of which 21.7% showed sexual dimorphism. A similarly small number of metabolic changes were found in *Ckd4*$^{+/-}$ mice, coding for another kinase involved in regulating cell cycle and tumorigenesis (**Fig. 3.3c**). As expected, genes that directly targeted metabolic enzymes or overall organ development had a stronger impact on the plasma metabolome with changes in up to 29% of all identified compounds. Examples of KO lines with genes targeting metabolism are *Pmm2*$^{+/-}$, coding for a phosphomannomutase involved in protein glycosylations, the cholesterol transporter, *Npc2*$^{+/-}$, and *Idh1*$^{-/-}$ that codes for isocitrate dehydrogenase 1. Similarly, a higher number of metabolic differences were found in gene KO lines involved in organ development such as the sonic hedgehog unc-51-like kinase 3 (*Ulk3*$^{-/-}$) that is involved in embryonic development (**Fig. 3.3c**).

3.3.3.2 Genotype-sex interactions on lipid metabolism in *Dhfr*$^{+/-}$, *Npc2*$^{+/-}$, *Nek2*$^{-/-}$, and *Sra1*$^{-/-}$ lines

Three genes, *Dhfr*, *Npc2*, and *Nek2,* had the highest proportion of lipids with genotype-sex interaction effects (**Fig. 3.3c**). Two of these genes code for enzymes that are known for their impact on cell division and therefore are actively studied in cancer research. DHFR

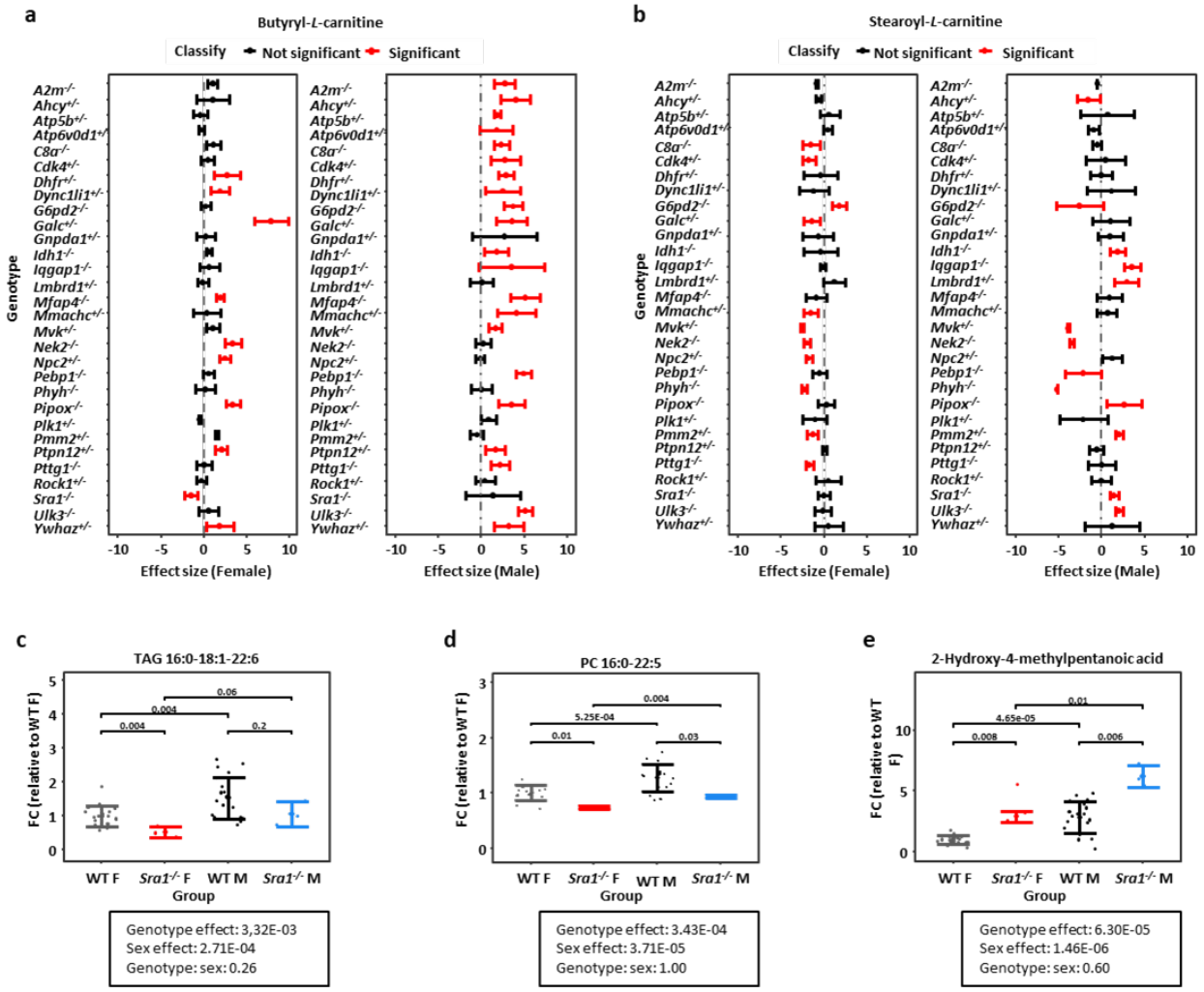


Figure 3.4. Examples of metabolites affected by genotype-sex interaction of 30 KO lines. Comparison of 20 male & 20 female C57BL/6NCrl control mice versus 3 male, 3 female mice per knockout line. Two-way ANOVA followed by individual comparisons using a generalized linear model, red= significant, black= not significant. Error bars represent standardized genotype effects with minimum and maximum values for each sex. (4a) Butyryl-L-carnitine. (4b) Stearyl-L-carnitine. (c–e) Plasma levels of three metabolites in *Sra1*$^{-/-}$. Boxplot represents standardized fold-changes normalized to that of wildtype female mice. Error bars stand for mean of fold-changes ±1 s.d.

(dihydrofolate reductase) is important in nucleoside biosynthesis [53] and NEK2 (NIMA-related kinase 2) is a centrosome kinase involved in *Wnt*-signaling pathways [54,55]. By altering the rate of cell division, both enzymes can act on lipid metabolism. More lipids were impacted in *Nek2*$^{-/-}$ females than in males while divergent changes in plasma lipids levels were observed in *Dhfr*$^{+/-}$ male and female mice.

The gene that had the highest fraction of sexually dimorphic plasma lipid levels was *Npc2*. Overall, 26.9% of all plasma metabolites were affected in *Npc2*$^{+/-}$ mice (**Fig. 3.3c**). Of these metabolites, 51.2% were detected by the lipidomics assay but were confined to *Npc2*$^{+/-}$ female mice. *Npc2* encodes the intracellular cholesterol transporter 2, regulating the transport of cholesterol to the perimeter membrane of late endosomes, to become available for transporting payloads to mitochondria, leading to cholesterol accumulation in lysosomes. In humans, a defect in NPC2-related cholesterol trafficking leads to the ultimately fatal Niemann-Pick Type C2 (NPC2) disease, an autosomal recessive complex lipid storage disorder [56]. Therefore, we expected to detect primary changes in lipid signatures. However, differences in the mitochondrial fatty acid transport molecules butyryl-carnitine and stearoyl-carnitine were only found in female *Npc2*$^{+/-}$ mice and not in male *Npc2*$^{+/-}$ mice (**Fig. 3.4a, b**).

Interestingly, both acylcarnitines were found to be dysregulated in many KO lines, often in a sexually dimorphic manner (**Fig. 3.4a, b**). Since mitochondrial fatty acid oxidation is a major contributor to overall energy usage, such data may enable a better understanding of the impact of gene dysfunction in human disease phenotypes. Lipids were also much more affected in *Sra1*$^{-/-}$ female than *Sra1*$^{-/-}$ male mice. For example, many triacylglycerides and phosphocholines were significantly down-regulated in *Sra1*$^{-/-}$ female mice, but not always in males (**Fig. 3.4c, d**). A bacterial leucine-derived metabolite [57,58] 2-hydroxy-4-methylpentanoic acid was higher in plasma for both *Sra1*$^{-/-}$ male and female mice (**Fig. 3.4e**). This observation shows again that gene KO mutations can have metabolic effects that involve the microbiome. The *Sra1* gene encodes the steroid receptor RNA activator protein and is involved in breast tumorigenesis and tumor progression. Our metabolomic results showed that triacylglycerides but also mitochondrial acylcarnitines, oxylipins, and sphingomyelins were differentially regulated in *Sra1*$^{-/-}$ female and male mice. Several oxylipins that can act as potent physiological mediators [59,60] were significantly upregulated including 8,9-DiHETrE and 11,12-EpETrE (in female *Sra1*$^{-/-}$ mice at $P < 0.05$), 11-HETE (in *Sra1*$^{-/-}$ male mice at $P < 0.01$), and 15-HETE (fold-change = 1.4 and 1.9 in *Sra1*$^{-/-}$ female and male mice, respectively)

while three other oxylipins were downregulated such as 15-HEPE and 15-KETE (in *Sra1*$^{-/-}$ female mice at $P < 0.05$) as well as 9,10-e-DiHO (in *Sra1*$^{-/-}$ male mice at $P < 0.01$).

3.3.3.3 *Phyh*$^{-/-}$, *Npc2*$^{+/-}$, and *Mfap4*$^{-/-}$ impact plasma lipid and peptide metabolism

Some genes are known for their implicit contribution to human diseases such as *Phyh*, *Pmm2*, and *Npc2*, while many others may not directly lead to a specific disease but may participate in the pathophysiology of some diseases. However, the influence of their dysfunction on metabolism is not fully understood. *Phyh*, a major player in Refsum disease, encodes phytanoyl-CoA hydroxylase responsible for breaking down phytanic acid in the alpha-oxidation pathway. Indeed, our metabolomic results showed differential regulation of plasma levels of 2-hydroxylated (branched) fatty acids such as 2-hydroxy-3-methylbutyric acid (fold-change = 1.6 in *Phyh*$^{-/-}$ male at $P < 0.001$) and 2-hydroxy-4-methylpentanoic acid (fold-change = 2.4 in *Phyh*$^{-/-}$ female at $P = 0.01$) and the carnitine transport forms of 3-methyl-fatty acids such as 3-hydroxyisovaleroylcarnitine (fold-change = 0.8 in *Phyh*$^{-/-}$ female at $P = 0.02$), isovaleryl-carnitine (fold-change = 0.8─0.9 in both sexes with $P = 0.2$) as well as derivatives of 3-methyl-fatty acid structures including valine (fold-change = 1.4 in *Phyh*$^{-/-}$ female at $P = 0.02$), isovaleryl-glycine (fold-change = 1.8 in *Phyh*$^{-/-}$ male at $P = 0.04$), and valine-dipeptides. Consequential genotype effect also exists for other metabolites such as decreased plasma levels of straight-chain fatty-acyl carnitines. Plasma prostanoids and prostaglandin were upregulated in *Phyh*$^{-/-}$ mice including progesterone (fold-change = 2.2 in *Phyh*$^{-/-}$ female at $P = 0.006$) and PGF3alpha (fold-change = 2.8 in *Phyh*$^{-/-}$ male at $P < 0.001$), and plasma bile acids were impacted by genotype-sex interaction including tauroursodeoxycholic acid (fold-change = 0.3 in *Phyh*$^{-/-}$ female at $P = 0.003$) and glycocholic acid (fold-change = 5.7 in *Phyh*$^{-/-}$ male at $P = 0.02$).

Because homozygous null mutations in *Npc2* are lethal, the viable heterozygotes were used for this study (**Supplementary Table 3.1**). Human NPC2 is mainly expressed in lung, thyroid, and gall bladder [61]. In addition to Niemann-Pick Type C2 (NPC2) disease, NPC2 is involved in chronic obstructive pulmonary disease (COPD)[62]. Microfibril-associated protein 4 coding gene *Mfap4* also has high expression in lung

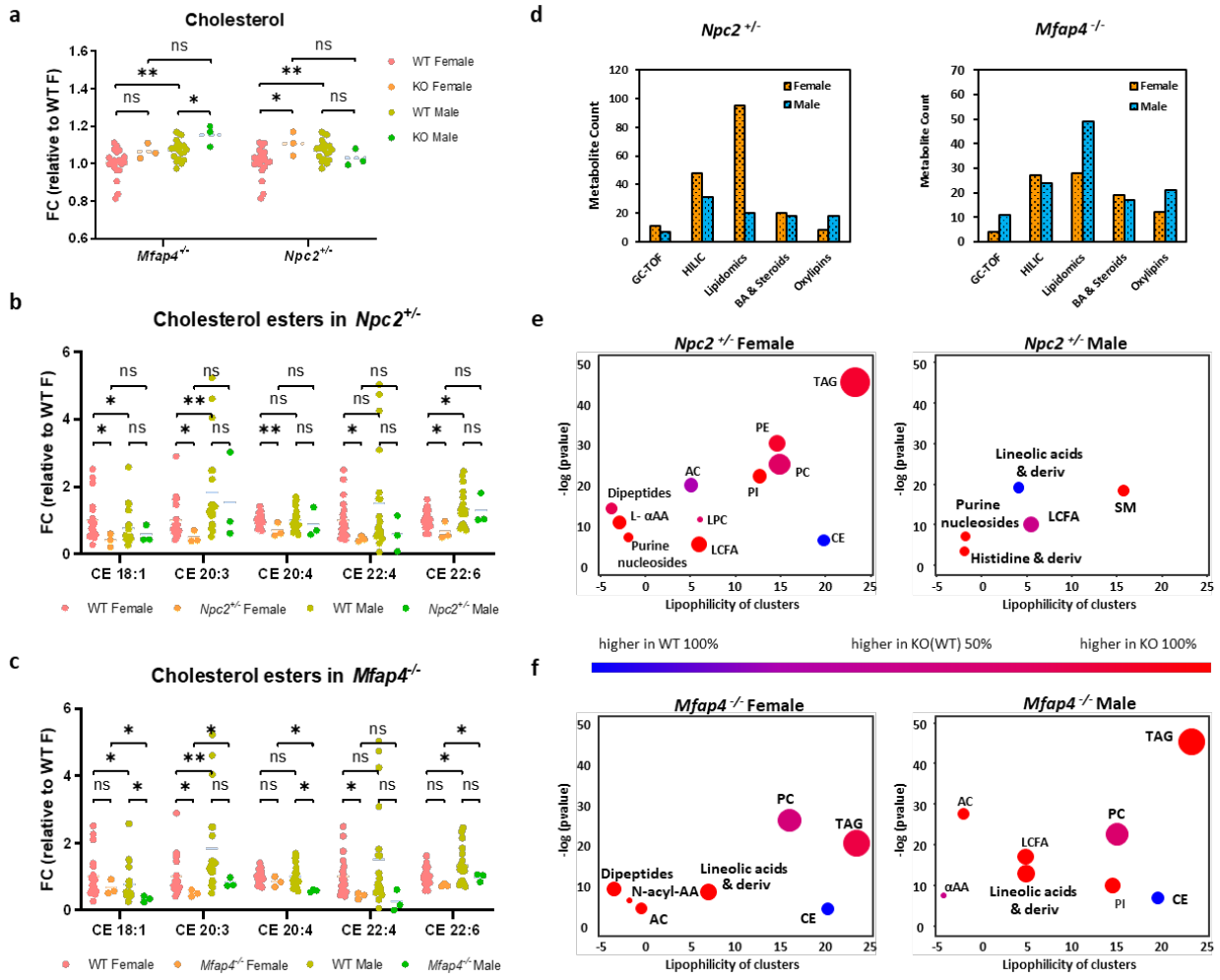


Figure 3.5. Comparison of sexually dimorphic alterations of plasma metabolites between *Npc2*$^{+/-}$ and *Mfap4*$^{-/-}$ mice. Comparison of 20 male & 20 female C57BL/6NCrl control mice versus 3 male, 3 female mice per knockout line. Significance levels given as *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$. (5a) Plasma cholesterol levels in *Npc2*$^{+/-}$ and *Mfap4*$^{-/-}$ mice versus controls. (5b) Plasma cholesterol ester levels in *Npc2*$^{+/-}$ mice versus controls. (5c) Plasma cholesterol ester levels in *Mfap4*$^{-/-}$ versus controls. (5d) Number of significant metabolites across five metabolomics assays for for *Npc2*$^{+/-}$ and *Mfap4*$^{-/-}$ versus controls. (5e) Chemical set enrichment plots of mouse plasma metabolites in *Npc2*$^{+/-}$ versus controls. (5f) Chemical set enrichment plots of mouse plasma metabolites in *Mfap4*$^{-/-}$ versus controls. For both bubble plots (e,f), larger bubbles indicate a higher number of metabolites per chemical set. Set names: TAG: Triacylglycerols; PC: Phosphatidylcholines; PE: Phosphatidylethanolamines; PI: Phosphatidylinositols; CE: Cholesteryl esters; LPC: Lyso-phosphocholines; AC: Acyl carnitines; LCFA: Long-chain fatty acids; MCFA: Medium-chain fatty acids; αAA: α-amino acids; N-acyl-αAA: N-acyl-α-amino acids; SM: Sphingomyelins; Linoleic acid & deriv: Linoleic acids & derivatives; His & deriv: Histidine & derivatives.

and gall bladder [61] and is upregulated in COPD [63]. The MFAP4 protein has binding specificities for both collagen and carbohydrates. We report here the first metabolomic effects of these two genes (**Fig. 3.5**). Because cholesterol trafficking is impaired in *Npc2*$^{+/-}$ mutants, we investigated if lower plasma levels of cholesterol esters (CE) were observed. Indeed, a range of both plasma cholesterol esters and free cholesterol were found significantly altered, specifically in female mice (**Fig. 3.5a, b**). Similarly, significant differences in both CE lipids and free cholesterol were found in *Mfap4*$^{-/-}$ mice, but male mice were more affected (**Fig. 3.5a, c**). For both genotypes, the largest number of metabolic changes were observed in lipid metabolism, extending to steroids, bile acids, oxylipins (**Fig. 3.5d**), and a range of phospholipids and neutral lipids that showed great effects in metabolite set enrichment statistics (**Fig. 3.5e─f**), with notable differences between *Npc2*$^{+/-}$ female and male mice. Effects on hydrophilic metabolites were found for amino acids in both sexes of *Mfap4*$^{-/-}$ and *Npc2*$^{+/-}$ genotypes, but not in other metabolic modules such as dipeptides, carbohydrates, or nucleosides. In comparison, sexually dimorphic metabolic alterations were much more prevalent in *Npc2*$^{+/-}$ mice (**Fig. 3.5e**) than that in *Mfap4*$^{-/-}$ mice (**Fig. 3.5f**). In combination, these results suggest that genes participating in the same disease (e.g., *Mfap4* and *Npc2* in COPD) may execute major effects on similar metabolic modules but, in addition, may also exert specific sexually dimorphic influence over other metabolic phenotypes.

3.3.3.4 Metabolic alterations in *Idh1*$^{-/-}$ mice

The enzyme isocitrate dehydrogenase 1 (IDH1) catalyzes the cytoplasmic oxidative decarboxylation of isocitrate to α-ketoglutarate with an identical reaction performed by IDH3 in the tricarboxylic acid cycle (TCA cycle, **Fig. 3.6a**). Therefore, the overall effect of the *Idh1*$^{-/-}$ allele might not be detectable on plasma concentrations of these metabolites due to a compensatory effect of the corresponding mitochondrial transporters and mitochondrial IDH3 enzyme (**Fig. 3.6a**). Plasma levels of the substrates of IDH1 catalyzed reactions, citrate and isocitrate, showed significant increases in *Idh1*$^{-/-}$ mice (isocitrate with 1.4─1.6 fold-changes at $P < 0.04$ for both sexes, citrate with 1.3─1.4 fold-changes at $P = 0.04$ in female mice and $P = 0.07$ in male mice, **Fig. 3.6b**). Consequently, plasma levels of the reaction product α-ketoglutarate were

downregulated in both sexes with 0.5─0.6 fold-changes at $P < 0.04$ (**Fig. 3.6b**). This KO line is therefore a prime example for a direct match of the immediate biochemical reaction in the cell to plasma metabolite levels. Downstream metabolite levels such as plasma glutamate were also

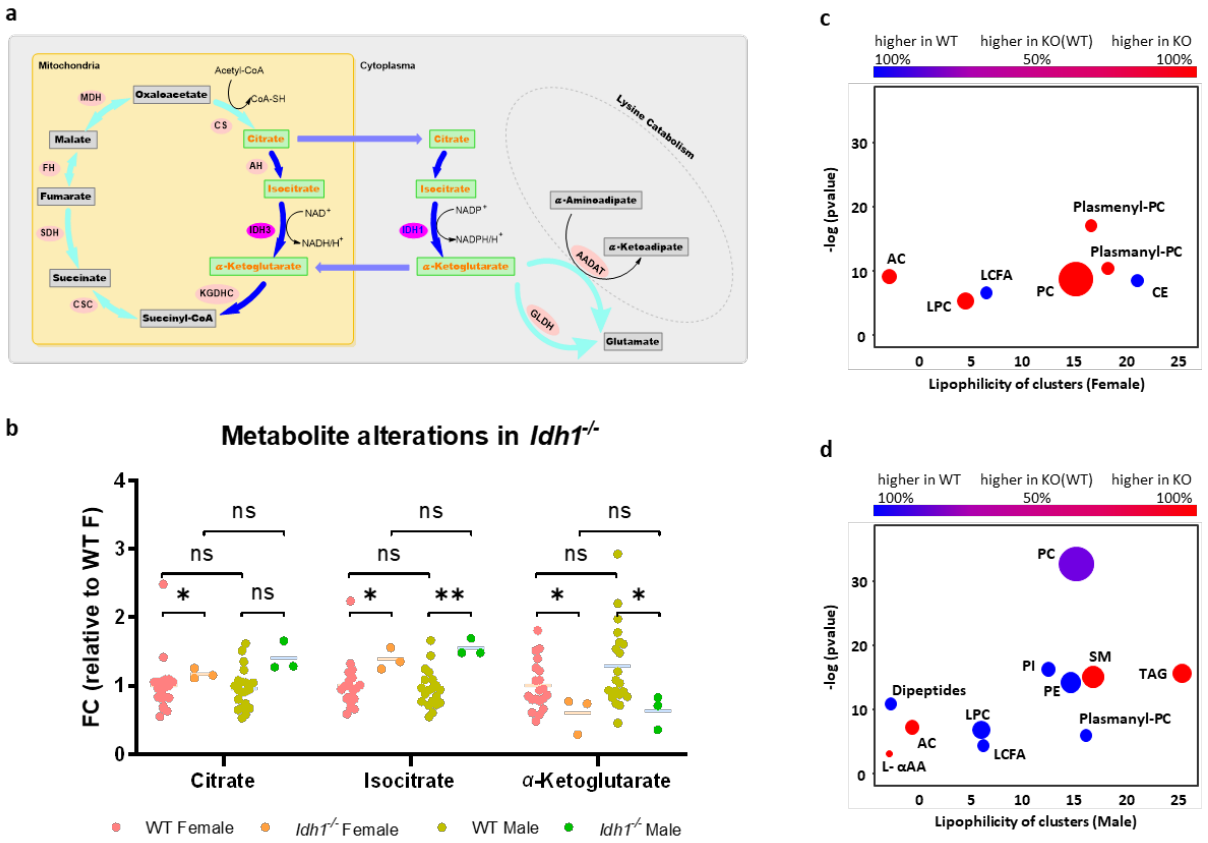
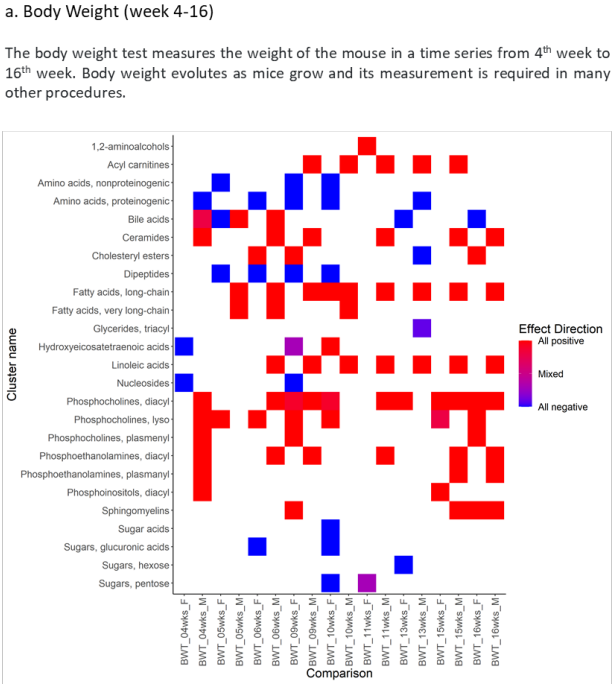


Figure 3.6. Sexual dimorphic alteration of plasma metabolites in *Idh1*$^{-/-}$ mice. Comparison of 20 male & 20 female C57BL/6NCrl control mice versus 3 male, 3 female Idh1 -/- mice. Significance levels given as *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$. (a) Representative metabolic map of TCA cycle metabolism in cells. CS: Citrate synthase; AH: Aconitase; IDH3: Isocitrate dehydrogenase (NAD), mitochondrial; IDH1: Isocitrate dehydrogenase (NADP+), cytosolic; KGDHC: α-Ketoglutarate dehydrogenase; CSC: Succinyl coenzyme A synthetase; SDH: Succinate dehydrogenase; FH: Fumarase; MDH: Malate dehydrogenase; GLDH: Glutamate dehydrogenase; AADAT: α-Aminoadipate aminotransferase. (b) Expected mouse plasma content changes of isocitrate, citrate, and α-ketoglutarate as the loss of IDH1 function. (c) Chemical set enrichment plots of mouse plasma metabolites in female *Idh1*$^{-/-}$ mice versus controls. (d) Chemical set enrichment plots of mouse plasma metabolites in male *Idh1*$^{-/-}$ mice versus controls. For both bubble plots (c,d), larger bubbles indicate a higher number of metabolites per chemical set. Set names: PC: Phosphatidylcholines; Plasmanyl-PC: Plasmanyl-phosphocholines; Plasmenyl-PC: plasmenyl-phosphocholines; SM: Sphingomyelins; CE: Cholesteryl esters; TAG: Triacylglycerols; PE: Phosphatidylethanolamines; LPC: Lyso-phosphocholines; AC: Acyl carnitines; PI: Phosphatidylinositols; LCFA: Long chain fatty acids; L-αAA: L-α-amino acids.

found significantly decreased in *Idh1*$^{-/-}$ male mice (fold-change = 0.76 at $P = 0.03$) but not in female mice. Further pleiotropic effects of *Idh1* KO mutation were found in both sexes alike as well as in a sexually dimorphic way. For example, in both sexes *Idh1*$^{-/-}$ mutants showed set enrichment increases in acylcarnitines and decreases in free long chain fatty acids. Yet, in *Idh1*$^{-/-}$ male mice, we found significantly decreased plasma levels of phosphatidylinositols, phosphatidylethanolamine, and phosphatidylcholines as well as increased levels of triacylglycerides (**Fig. 3.6d**), while in *Idh1*$^{-/-}$ female mice phosphatidylcholines levels were increased, and levels of cholesteryl esters were decreased (**Fig. 3.6c**).

3.3.4 Sexual dimorphism in metabolite-phenotype correlations

Sixteen out of thirty mouse KO lines are used as models for human diseases. Those genes showed overlap between phenotypes observed in human diseases and phenotypes observed in mouse KOs for orthologous genes (**Supplementary Table 3.1**). The relationships between plasma metabolite and phenotype parameters/parameter series may provide additional information to understand the gene functions and disease etiology. Due to limited number of mice per sex-per KO line, the associations were only performed in WT mice where strong sexual dimorphism was also observed.



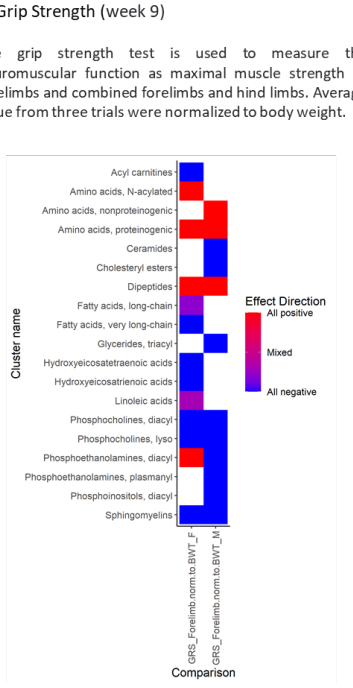


Figure 3.7. Heatmap of spearman correlation of metabolite clusters and IMPC open field phenotypes. (a) Correlation of metabolite clusters with body weight phenotype parameter series in WT mice. (b) Correlation of metabolite clusters with grip strength phenotype parameters in WT mice. Spearman correlation with $P < 0.05$, sample size in each metabolite and phenotype > 14, for each gender respectively. Red color, positive correlation; blue color, negative correlation. Axis labels: x-axis, phenotypes; y-axis, metabolite clusters.

To assess the magnitude of the associations between metabolites and IMPC phenotypes, spearman correlation analysis was performed in wildtype mice. 803 metabolites were correlated to 208 phenotypes, yielding 167,024 metabolite-phenotype spearman-rank correlations in wildtype mice. 20,435 correlations (12.2%) were significant at $P < 0.05$. Among those correlations, 9,136 were only significant in females and 10,392 were only significant in males. Only 742 correlations (3.6%) were significant in both sexes and in the same direction, indicating an overwhelming sexual dimorphism in metabolite-phenotype correlations. A small fraction of 165 metabolite-phenotype correlations were even statistically significant with opposite directions between the two sexes. For example, TAG 16:0-18:0-22:0, a triacylglyceride with three saturated free fatty acid chains, was negatively correlated with the phenotype 'center average speed' (one parameter in open field tests for anxiety) in wildtype females ($r = -0.71$ at $P = 0.0007$), but turned positive in wildtype male mice ($r = +0.74$ at $P = 0.0002$). Pre-pulse inhibition (PPI) is the suppression of an acoustic startle reflex (ASR) to an intense stimulus when a weak pre-pulse stimulus precedes the startle stimulus. A reduction of PPI is thought to reflect dysfunction of sensorimotor gating that normally suppresses excessive behavioral responses to disruptive stimuli [64]. Several clinical studies have shown that a number of human disorders have impaired PPI including: schizophrenia, Huntington's disease, fragile X syndrome, and autism. Phosphatidylcholine PC 40:4, was positively correlated with % Pre-pulse inhibition - PPI2 in female WT mice but was negatively correlated in male WT mice. Several other PCs/Plasmanyl-PCs were showed sexual dimorphism correlations with other PPI phenotype parameters.

Because every single phenotype/metabolite association might be caused by random chance, we accumulated all data in more robust chemical enrichments statistics (ChemRICH) analysis. Based on 10

phenotype pipelines, metabolites were clustered by ClassyFire [65] and subjected to ChemRICH analysis to associate clusters of metabolites with specific phenotypes. As expected, metabolomic measurements were correlated with blood chemistry data acquired by the IMPC (**Supplementary Fig. 3.5b**). For example, 'metabolomic' triacylglycerides and hexoses were positively correlated with IMPC triglycerides and glucose in both sex (**Supplementary Fig. 3.5b**). Dramatic sexual dimorphism in metabolite cluster-phenotype correlations were found in each of the 10 phenotype pipelines. For example, body weight (**Fig. 3.7a**) was highly positively associated with acylcarnitines and linoleic acid-derivatives as well as long-chain fatty acids-derivatives in males, but not in females, starting from week 5. Conversely, lysophosphatidylcholines were positively associate with body weight in females from 5 weeks onward, but not in males. Similarly, Dipeptides and proteinogenic amino acids were negatively associated with body weight in females from 5─10 weeks, but not in males (**Fig. 3.7a**). Since body weight includes muscle mass, we also studied the association of plasma metabolites with grip strength, normalized to body weight (**Fig. 3.7b**). While several compound classes (proteinogenic amino acids, dipeptides, sphingomyelins and phosphocholines) were found correlated in both sexes with forelimb grip strength, there was also remarkable sexual dimorphism. For example, plasma nonproteinogenic amino acids in males were positively associated with forelimb grip strength but not in females. Conversely, in females, N-acylated amino acids were found to be positively associated with forelimb grip strength but not in males. Similarly, sexual dimorphism was found for mitochondrial acylcarnitines and in linoleic acid-derivatives and other compound classes.

3.4 Discussion

We found that overt phenotypes measured by the IMPC showed ~25% of all phenotypic measures to be sexually dimorphic, while more than 56% of all plasma metabolites were significantly different between males and females. This finding indicates that sex has a major impact on many different metabolic pathways in mice [66,67]. Moreover, FDR correction and body weight adjustment did not influence the statistical results of sexual dimorphism in metabolomics data. We therefore conclude that sex differences in plasma

metabolite levels were not simply due to differences in body weight. Oxylipins, indolyl derivatives, bile acids, amino acids, and sphingomyelins were all increased more in female mice than males, pointing to differences in the use of metabolites with regulatory roles that are well-known for sphingomyelins [68], oxylipins [69], or indoles [70] and bile acids [71]. Even metabolites like TMAO that are clearly formed by gut intestinal microbes showed different levels in male and female mice, the level of which was also reported for rats and humans [25,72,73]. TMAO might act by stabilizing proteins as a "chemical chaperone" in the endoplasmic reticulum (ER) [74]. Levels of TMAO are associated with cardiovascular risk [74,75] and likely other diseases [76-81]. This finding shows that levels of metabolites produced by many routes (including by different microbiomes) are sex-dependent suggesting that metabolomic data might best be interpreted differently between the sexes. Similar to TMAO, we found EpETrEs (also called EETs) to be sexually dimorphic. EETs are derived from arachidonate and have opposite effects to TMAO with respect to vasodilatory impact and other cardioprotective effects [82-84]. Lipid mediators including EETs are important actors in the regulation of a range of physiological parameters such as blood pressure [69,85] but are not generally known to have sexually dimorphic levels in humans. Hence, such sexually dimorphic metabolites might contribute to the well-known differences in cardiovascular risk in men compared to women [86]. In addition, we found many lipid classes to be elevated in WT male mice such as membrane lipids, cholesteryl esters, and triacylglycerides, whereas signaling lipids like phospho-sphingolipids were elevated in female mice. Similar to TMAO and EETs, such findings have implications for health effects as excessive accumulation of TAG are known to be associated with a range of diseases including hepatic steatosis and non-alcoholic fatty liver disease [87]. Indeed, in humans, men also have higher TAG blood levels [88] and higher VLDL lipoprotein particle levels than women [89], which is highly correlated with higher risk of atherosclerotic cardiovascular disease in men [90]. On the other hand, though sexual dimorphism was also observed in many human healthy cohort studies, the sex effect on metabolism can differ greatly between humans and mice as well as between different human cohorts [49,50,91-93]. Sexual dimorphism is expected to be different between species. On average, the protein-coding regions of the mouse and human genomes are 85 percent identical. Some regions are highly evolutionarily conserved because they are required for

function. In contrast, the non-coding regions are much less similar (only 50 percent or less). Since mice (or other animal models) have different life span and maturational rate from humans, the findings from animal models may only serve as a reference for metabolism and pathophysiology in human diseases, but the major impacts caused by gene homologue mutations can definitely exceed the differences in sexual dimorphism between different species as exemplified for gene *Dhfr*, *Idh1*, *Mfap4*, *Nek2*, *Npc2*, *Phyh* and *Sra1*.

We also found that the magnitude of overall genotype effects on phenotypic traits and cellular metabolites was strikingly similar, with 26.3% of IMPC phenotypes (21.3% when excluding body weight measured at different time points) in KO lines differing from WT, and 21.5% of all metabolites being significantly affected by genotype effect or genotype-sex interactions. This finding supports the concept that cellular metabolites provide additional information on gene function. Of these overall gene KO effects, 37.4% of the phenotypes (30.7% when excluding body weight measured at different time points) and 34.4 % of the metabolites were found to be sexually dimorphic. This proportion of differences between sexes in phenotypes and metabolites caused by KO mutations reinforces the importance of using both sexes when probing gene function. Differentiation of gene function by sex should be considered as an important factor in human disease etiology. Each of the 30 tested gene KO lines had clear and significant genotype or genotype-sex interaction effects on metabolism. This was true even for genes that had no direct impact on metabolic enzymes such as *C8a*, a gene involved in immune response, or *Dync1li1*, a gene involved in the intracellular protein transport and assembly. The mechanistic explanation for metabolomic sex differences remains to be investigated in detail. While we restricted analyses to plasma metabolite levels in order to study potential translation of mouse screens for clinical use, it is clear that plasma levels serve only as indirect footprints of cellular, tissue, and organ mechanisms underlying metabolic changes. Future studies using these and other gene KO lines will shed light on the target-specific effects of gene dysfunction and consequent impact on the metabolome. This, along with the metabolic data presented here, should in turn provide the basis for developing diagnostic tests for cell-to-organ system dysfunction and diseases.

Nevertheless, it was interesting that for several enzyme KOs, the effect of the perturbed biochemical reaction was directly detectable in plasma levels of the product metabolites. For example, IDH1 is one of three isoforms in mammals that catalyze the conversion of isocitrate to α-ketoglutarate. The activity of the other two isoforms could have had compensatory effects that might have led to undetectable changes on blood metabolites. Yet, both *Idh1*$^{-/-}$ male and female mice showed clear upregulation of plasma isocitrate and downregulation of plasma α-ketoglutarate with the same effect size. This initial perturbation of a major metabolic enzyme then caused additional downstream differences, specifically in lipid metabolites such as reduced long-chain fatty acid and increased mitochondrial acylcarnitine levels in both sexes. Yet, even for IDH1, important sexually dimorphic levels of plasma metabolites were observed, such as diametrically opposed changes of lyso-phosphatidylcholine and plasmanyl-phosphatidylcholine levels.

Previously Karp et al analyzed 234 phenotype parameters from 2,186 KO mouse lines across 10 IMPC centers and showed that phenotypic sexual dimorphism varied between centers [33]. For example, continuous phenotype trait parameters from TCP showed less sexual dimorphism than the average across all centers. The reason for this may be that the genes analyzed in one center inherently have less sexual dimorphism than the other centers. Indeed, sexually dimorphic phenotypes were higher in KO lines of genes expressing proteins that influence hormonal effects on behavior and physiology [33]. Several of the KO lines analyzed in this study included genes expressing proteins known to be involved in the regulation of sexual development (e.g. *Sra1*) [94]. The steroid receptor RNA activator protein SRAP, encoded by *Sra1,* regulates estrogen and androgen receptor signaling pathways. *Sra1* is an estrogen and androgen-dependent gene that contributes to the progression of breast cancer in women [95]. Mitochondrial dysfunction and increased fatty acid oxidation were shown to be associated with breast cancer [96,97]. Twenty-three percent of all metabolites were affected in the *Sra1*$^{-/-}$ mice, and more than twice as many metabolites including lipids were affected in *Sra1*$^{-/-}$ female mice compared to *Sra1*$^{-/-}$ male mice. In addition to lipids and acylcarnitines that showed sexual dimorphism in *Sra1*$^{-/-}$ female and male mice, several oxylipins such as 11,12-EpETrE, 11-HETE and 15-HETE were also differentially affected between the two sexes of *Sra1*$^{-/-}$ mice. Epoxyeicosatrienoic acids

(EETs, or EpETrEs) induce angiogenesis and initiate cancer cell migration [98]. On the other hand, oxylipins like 11-HETE are reported to have anti-mitogenic and anti-tumor activity [99,100]. Dysregulated plasma oxylipin levels in breast cancer patients indicated that these metabolites may become therapeutic target candidates [59,97,101], indicating such gene-metabolite functional data may provide better understanding of the role(s) of a gene in metabolic regulation and its involvement in human diseases including breast cancer. It also proved how data from mouse gene KOs could inform translational research into human diseases.

Several of the KO mouse lines analyzed have been shown previously to be associated with human disease phenotypes. For instance, mutations in human *PHYH* causes Refsum Disease [102] which includes visual impairment and hearing loss. *PHYH* and its mouse orthologue *Phyh* encode a peroxisomal protein that is involved in the alpha-oxidation of 3-methyl branched fatty acids. Alpha-oxidation is a process in which fatty acids are shortened by one carbon atom, producing 2-hydroxylated intermediates in the process [102,103]. Our metabolomic results indicated the deletion of *Phyh* differentially affected peroxisomal alpha-oxidation related metabolites of 2-hydroxylated (branched) fatty acids and their derivatives. *Phyh*$^{-/-}$ mutants also showed impact on metabolites that are related to generic peroxisomal functions [103] including downregulation of straight-chain fatty-acyl carnitine transport molecules. Other metabolites involved in peroxisomal oxidation were differentially altered such as prostanoids (progesterone, PGD2, PGF3alpha), and bile acids (tauroursodeoxycholic acid, glycocholic acid). Plasma metabolomic changes in *Phyh*$^{-/-}$ mice ranged from differential regulation of methylated- and acetylated amino acids to oxylipins including 13-KODE, 15,16-DiHODE, 9,10-DiHODE, and 9,10-DiHOME that were found in a sexually dimorphic manner. Therefore, our metabolic phenotype results confirm the primary role of *Phyh* but also adds valuable additional data to inform secondary mechanisms that link disease phenotypes to the underlying metabolic function(s). Another example is *Npc2* that causes Niemann-Pick disease C2, a hereditary neurovisceral lysosomal lipid storage disorder due to mutations in *NPC2*. Respiratory distress and lung disease were uniformly observed in patients with Niemann-Pick type C2 in early infancy [62,104,105]. *Mfap4* encoding microfibrillar-associated protein 4 (MFAP4) contributes to mature elastic fiber homeostasis and stability in

connective tissues such as lung, skin, and aorta [106]. Plasma MFAP4 is associated with chronic obstructive pulmonary disease (COPD) severity and may serve as a COPD biomarker [107]. With similar localization in tissues and implication in diseases of *MFAP4* and *NPC2*, we were prompted to investigate what extent they impact plasma metabolomics. Indeed, cholesterol and cholesterol esters showed similar alteration in plasma from *Mfap4*$^{-/-}$ and *Npc2*$^{+/-}$ mice. But they also had very different influences on the majority of other metabolites with diverse genotype-sex interaction effects, indicating that genes involved in the same disease may exhibit effects on similar metabolic classes while affecting other metabolic phenotypes differently.

In addition, we found numerous interesting correlations between metabolomics and IMPC phenotypes. Correlation analysis (like Spearman-rank calculations here) rely on multiple data points to associate different variables. While the number of samples per KO genotype was too small to confidently score associations between IMPC phenotypes and plasma metabolites, the strength, number and the degree of sexual dimorphism of such correlations in female and male wild type mice was astounding. While no one should mistake statistical associations with causal effects, we pose that associating visible IMPC phenotypes with molecular and metabolic variables may open the doors to better mechanistic understanding of mouse phenotypes, and eventually their links to corresponding human diseases. For example, center average speed is a common parameter used in mouse behavior and neurological and nerve system research. The blood triacylglyceride (16:0-18:0-22:0) consisting only of saturated fatty acyl groups showed completely opposite correlations with center average speed in wildtype female and male mice. The corresponding free palmitic acid was found to be involved in sexual dimorphisms of microglia (resident brain immune cells) [108] while its hydroxylated fatty acid ester derivative 5-PAHSA was involved in antioxidant response in mice and PC12 neuronal model cells [108,109]. While few studies have investigated how specific triacylglycerides are related to disease phenotypes[110], these studies show possible ways to mechanistically relate lipids to neurological differences, at least in cell and animal models.

In summary, the evidence for significant sexual dimorphism in both 30 KO lines and WT controls demonstrates that sex must be considered as an important factor in interpreting the role(s) of a gene in

metabolism and disease etiology. Such observations may serve to help understand the physiological consequences of genetic alterations underlying human diseases that manifest differently in men and women. Plasma metabolite alterations detected by metabolomics techniques can provide insight into, and contribute to unraveling the many complex links between gene functions and the etiology of complex diseases. With an overview of 30 mouse KO lines we here showed how comprehensive metabolomics data may inform such links, and how often metabolic effects are different between sexes. This report exemplifies the power for unraveling the links between gene functions and mouse models of human diseases once metabolomics data are combined with mouse phenotypes on the complement of >7,000 available IMPC mouse KO lines.

3.5 Methods

3.5.1 Mouse KO production and selection

All KO mouse lines were produced from C57BL/6N-derived targeted embryonic stem (ES) cells from the International Knockout Mouse Consortium (IKMC) using standard protocols at The Centre for Phenogenomics (TCP) in Toronto, Canada [111]. All procedures involving mice were performed in compliance with the Animals for Research Act of Ontario and the Guidelines of the Canadian Council on Animal Care. TCP's Institutional Animal Care Committee also reviewed and approved all procedures conducted on mice at TCP. Mice were subjected to phenotyping at TCP as part of the IMPC project using standardized protocols (IMPReSS, https://www.mousephenotype.org/impress). Lithium heparin plasma of 30 KO lines (three mice per sex and gene) were selected from TCP's bioarchive along with 40 corresponding C57BL/6NCrl WT mice (20 female, 20 male) [44]. All phenotype data are available from the IMPC website (https://www.mousephenotype.org/) [47]. All metabolomics data were obtained at the West Coast Metabolomics Center (WCMC) at UC Davis.

3.5.2 Data acquisition

Plasma samples were prepared by liquid-liquid extractions as published previously [44,112]. Pooled human plasma (BioIVT) samples were used as quality control (QC) samples in this study for each of the five platforms. Blanks and QC samples were prepared at the same time of mouse plasma sample extraction. All

samples were randomized via miniX study design software [39] with one blank and one quality control (QC) sample between every 10 mouse plasma samples. In total 22 QC samples were used for calculating the RSDs for each platform. Mass spectrometry data were acquired by using three untargeted assays: primary metabolites covering carbohydrates, amino acids, hydroxyl acids, and related compounds were analyzed on gas chromatography-time-of-flight mass spectrometry (GC-TOF MS, Leco Corporation, St. Joseph, MI, USA) [39,113,114], complex lipids such as phosphatidylcholines, ceramides, sphingomyelins, and triacylglycerides were acquired by Vanquish UHPLC system with Q-Exactive HF mass spectrometry (Thermo Scientific, Waltham, MA, USA) [114], and biogenic amines such as carnitines, dipeptides, nucleosides, and modified amino acids were analyzed by hydrophilic interaction chromatography-orbital ion trap mass spectrometry (HILIC-Q-Exactive MS/MS, Thermo Scientific, Waltham, MA, USA) [115]. Data on bile acids, steroids, and oxylipins were acquired by reversed phase LC separation followed by targeted analysis on a Sciex quadrupole-linear ion trap mass spectrometer (6500+ QTRAP MS/MS) [116,117]. Detailed methods about metabolite annotations have been previously published [44]. The median RSD for compounds in QC samples was less than 20%. After removing deuterated internal standards, metabolites with at least 70% missing values, repeatedly identified metabolites between platforms and metabolites with failed QC criteria (RSD_{QC} > 50%), 827 metabolites remained. HILIC-(ESI) MS data were normalized by the median value for each batch to remove batch effects. GC-TOF MS data was normalized using SERRF method [118]. Data from other assays were not normalized because no batch effects were observed. Identified compounds were used for statistics if they were positively detected in more than 30% of all samples.

3.5.3 Missing value treatment

For GC-TOF MS, the peak intensity values were automatically extracted from raw data (of the target m/z and target retention time) during the data processing procedure by GC BinBase [39], and missing values were automatically replaced by local noise.

For other metabolomic assays, missing values were replaced in R by half of the minimum of the non-missing values in each mouse genotype. Metabolites or phenotype data were discarded from analyses when

the percentage of missing values was > 70% of total number of mice in each sex group. Otherwise, missing phenotype data were replaced by the minimum of the non-missing values in each genotype mice.

3.5.4 Statistical analysis

Because all mice were housed at TCP under identical standard operating procedures and were based on the same WT strain C57BL/6NCrl, and plasma were collected from mice at similar ages (**Supplementary Table 3.1**), factors such as age, environment, and husbandry were considered to be constant. Statistical analysis was performed using R 3.6.0. An overview on statistical workflows is given in **Supplementary Fig. 3.4**.

Wildtype (WT) mice. Analysis methods for testing sexual dimorphism in WT mice were adapted from Karp et al [119]. and summarized in **Supplementary Fig. 3.4a** and **3.4b**. Briefly, generalized linear regressions were used to test the role of sex as a variable (Y ~ sex) or as a covariate with adjusting body weight (Y ~ Sex + Weight). Compounds are deemed to be statistically significantly affected by sex at $P < 0.05$ or FDR < 0.05 after Benjamini-Hochberg false discovery rate adjustments [120]. The same methods were used to analyze 208 IMPC phenotypes classified as continuous data parameters as extracted from the IMPC database. For graphic representation of results in boxplots or fold-change calculations, arithmetic means ± standard deviations were used.

KO mice. We adopted a previously published method that investigated gene-phenotype effects to assess statistical impacts in KO mice [119]. Briefly, we first studied the role of genotype alone, and subsequently we assessed the impact of genotype-sex interaction on plasma metabolite levels. Batch effect was not considered in analyses because all mice were raised and analyzed under identical conditions and metabolomics data were acquired within one batch. For both investigations, two-way ANOVA was used for testing the statistical significance of genotype-sex interaction effect on each metabolite ($P < 0.05$). Body weight was not accounted for due to the very small difference between sex effect adjusted to body weight and sex effect without adjusting to body weight in WT mice, as well as the low number of mice per KO genotype and sex. False discovery rate adjustment was not carried out on individual metabolite levels

because the purpose of this study was to generate novel biochemical hypotheses rather than the discovery of biomarkers for diagnostic purposes. Also, evidence showed that there was only a small difference between sex effect with FDR correction and sex effect without FDR correction in wildtype. Instead, we combined all metabolites into classes to conduct set enrichment analysis with FDR correction using ChemRICH software (see below). To test the genotype effect, a full model (Y ~ Genotype + Sex + Genotype: Sex) was compared with a null model (Y ~ Sex). For the genotype-sex interaction effect, the regression analysis compared the full model (Y ~ Genotype + Sex + Genotype: Sex) against the null model (Y ~ Genotype + Sex). Individual comparison of genotype / WT for each sex were used to classify the genotype effect (**Supplementary Fig. 3.4c**). If the plasma level of a specific metabolite was significantly different by the genotype effect and also by genotype-sex interaction, then we specified whether effects were found only in females or only in males, or in both sexes but in different directions of change, or in both sexes but with different effect sizes. However, if a metabolite level was significant at the genotype level alone but not for genotype-sex interaction effect, then the effect was classified as 'genotype effect with no sex difference' in **Fig. 3.3a** and **Fig. 3.3b**. For **Fig. 3.4a** and **Fig. 3.4b**, effect sizes were calculated by standardizing the associated arithmetic average estimate of each KO line to the mean of the corresponding WT mice. For other graphic representations in boxplots (**Fig. 3.4c─4e, Fig. 3.5,** and **Fig. 3.6**) or fold-change calculations, arithmetic means ± standard deviations were used. For continuous phenotypic traits extracted from the IMPC database, the same methods were used by comparing a full model (Y ~ Genotype + Sex + Genotype: Sex) to a null model (Y ~ Sex) for assessing the genotype effect and by comparing the full model (Y ~ Genotype + Sex + Genotype: Sex) to a null model (Y ~ Genotype + Sex) for assessing the genotype-sex interaction effect where the dependent variable is an adult or embryo phenotype parameter from IMPC.

Metabolite set enrichment statistics**.** Chemical Similarity Enrichment Analysis (ChemRICH) was used for finding differentially regulated clusters of metabolites [46]. ChemRICH defines different sets of molecules in metabolomics assays based on the Medical Subject Headings (MeSH) ontology term annotations and

Tanimoto chemical similarity calculations. The PubChem identifier of each identified metabolite was mapped to MeSH term in ChemRICH database. For compounds that were not contained in ChemRICH database, their MeSH terms were estimated using Tanimoto chemical similarity coefficients. According to PubChem CIDs, names, SMILES codes, and MeSH terms, metabolites were separated into non-overlapping chemical clusters. Then, Kolmogorov–Smirnov test was conducted on each cluster to assess the significance alteration, compared with random distribution. A *P*-value less than 0.05 or FDR *P*-value less than 0.05 were used as criterion to determine the cluster significance.

Metabolite-phenotype correlation statistics. For WT mice, Spearman-rank correlations were used to calculate associations between 803 metabolites and 208 IMPC phenotypes. ChemRICH was used for assessing metabolite clusters that were correlated with phenotypes[46]. Only metabolite-phenotype correlations with > 14 data points per variable and per sex were subjected to ChemRICH analysis. *P*-values < 0.05 were used to determine cluster significance levels. Correlation results can be found in **Fig. 3.7** and **Supplementary Fig. 3.5**.

Code availability. R code for reproducing the analyses shown in the main figures is available at https://github.com/ythzhang/KOMP_test. Additional code related to extended data and supplementary figures is available upon request from the corresponding author.

Acknowledgements

We thank staff members of the UC Davis West Coast Metabolomics Center, The Centre for Phenogenomics (Toronto, Canada) and the UC Davis Mouse Metabolic Phenotyping Center for their support in implementing the project.

Author contributions

Y. Z. and O. F. wrote the manuscript. O. F., K. C. L. and D. K. B. conceptualized the study. A. M. F., L. M. J. N., C. M., and K. C. K. L. generated the mouse strains, provided the plasma samples, and reviewed and edited the manuscript. D. K. B. generated the consolidated metabolomics dataset. Y. Z. and D. K. B.

retrieved the phenotype data. Y. Z., S. F., and B. G. performed statistical analysis. Y. Z. and O. F. interpreted biological effects.

Competing interests

The authors declare no conflict of interest.

Funding

The study was funded by NIH U2C ES030158 (to OF) and U42 OD012210 (MMRRC), U2C DK092993 (MMPC), and UM1 OD023221 (KOMP2) grants (to KCKL), and by Genome Canada and Ontario Genomics under award number OGI-051 (to CM).

3.6 References

1. Mauvais-Jarvis, F., Arnold, A.P. & Reue, K. A Guide for the Design of Pre-clinical Studies on Sex Differences in Metabolism. *Cell Metab* **25**, 1216-1230 (2017).

2. Tam, A. *et al.* Sex Differences in Airway Remodeling in a Mouse Model of Chronic Obstructive Pulmonary Disease. *Am J Respir Crit Care Med* **193**, 825-34 (2016).

3. Hanamsagar, R. *et al.* Generation of a microglial developmental index in mice and in humans reveals a sex difference in maturation and immune reactivity. *Glia* **65**, 1504-1520 (2017).

4. Mauvais-Jarvis, F. Sex differences in metabolic homeostasis, diabetes, and obesity. *Biology of Sex Differences2015* **6**, 14 (2015).

5. Arnold, A.P., Cassis, L.A., Eghbali, M., Reue, K. & Sandberg, K. Sex Hormones and Sex Chromosomes Cause Sex Differences in the Development of Cardiovascular Diseases. *Arterioscler Thromb Vasc Biol* **37**, 746-756 (2017).

6. Ratnu, V.S., Emami, M.R. & Bredy, T.W. Genetic and epigenetic factors underlying sex differences in the regulation of gene expression in the brain. *J Neurosci Res* **95**, 301-310 (2017).

7. Klein, S.L. & Flanagan, K.L. Sex differences in immune responses. *Nat Rev Immunol* **16**, 626-38 (2016).

8. Regitz-Zagrosek, V. & Kararigas, G. Mechanistic Pathways of Sex Differences in Cardiovascular Disease. *Physiol Rev* **97**, 1-37 (2017).

9. Labonte, B. *et al.* Sex-specific transcriptional signatures in human depression. *Nat Med* **23**, 1102-1111 (2017).

10. McCarthy, M.M. & Nugent, B.M. At the frontier of epigenetics of brain sex differences. *Front Behav Neurosci* **9**, 221 (2015).

11. Pollitzer, E. Biology: Cell sex matters. *Nature* **500**, 23-4 (2013).

12. Clayton, J.A. & Collins, F.S. Policy: NIH to balance sex in cell and animal studies. *Nature* **509**, 282-3 (2014).

13. Zucker, I. & Beery, A.K. Males still dominate animal studies. *Nature* **465**, 690 (2010).

14. Palanza, P. & Parmigiani, S. How does sex matter? Behavior, stress and animal models of neurobehavioral disorders. *Neurosci Biobehav Rev* **76**, 134-143 (2017).

15. Florez-Vargas, O. *et al.* Bias in the reporting of sex and age in biomedical research on mouse models. *Elife* **5**(2016).

16. Beery, A.K. & Zucker, I. Sex bias in neuroscience and biomedical research. *Neurosci Biobehav Rev* **35**, 565-72 (2011).

17. National Institutes of Health. NIH Policy on Sex as a Biological Variable. Available at: https://orwh.od.nih.gov/sex-gender/nih-policy-sex-biological-variable.

18. Quinn, M., Ramamoorthy, S. & Cidlowski, J.A. Sexually dimorphic actions of glucocorticoids: beyond chromosomes and sex hormones. *Ann N Y Acad Sci* **1317**, 1-6 (2014).

19. Lauretta, R., Sansone, M., Sansone, A., Romanelli, F. & Appetecchia, M. Gender in Endocrine Diseases: Role of Sex Gonadal Hormones. *Int J Endocrinol* **2018**, 4847376 (2018).

20. Imtiaz, B. *et al.* Postmenopausal hormone therapy and Alzheimer disease: A prospective cohort study. *Neurology* **88**, 1062-1068 (2017).

21. Jabbar, A. *et al.* Thyroid hormones and cardiovascular disease. *Nat Rev Cardiol* **14**, 39-55 (2017).

22. Nauck, M.A. & Meier, J.J. Incretin hormones: Their role in health and disease. *Diabetes Obes Metab* **20 Suppl 1**, 5-21 (2018).

23. Gribble, F.M. & Reimann, F. Function and mechanisms of enteroendocrine cells and gut hormones in metabolism. *Nat Rev Endocrinol* **15**, 226-237 (2019).

24. Audano, M., Maldini, M., De Fabiani, E., Mitro, N. & Caruso, D. Gender-related metabolomics and lipidomics: From experimental animal models to clinical evidence. *J Proteomics* **178**, 82-91 (2018).

25. Stanley, E.G. *et al.* Sexual dimorphism in urinary metabolite profiles of Han Wistar rats revealed by nuclear-magnetic-resonance-based metabonomics. *Anal Biochem* **343**, 195-202 (2005).

26. Honma, A. *et al.* Effect of acute total sleep deprivation on plasma melatonin, cortisol and metabolite rhythms in females. *Eur J Neurosci* **51**, 366-378 (2020).

27. Ellul, S. *et al.* Sex differences in infant blood metabolite profile in association with weight and adiposity measures. *Pediatr Res* (2020).

28. Wells, J.C. Sexual dimorphism of body composition. *Best Pract Res Clin Endocrinol Metab* **21**, 415-30 (2007).

29. Janssen, I., Heymsfield, S.B., Wang, Z.M. & Ross, R. Skeletal muscle mass and distribution in 468 men and women aged 18-88 yr. *J Appl Physiol (1985)* **89**, 81-8 (2000).

30. Greenman, A.C., Albrecht, D.M., Halberg, R.B. & Diffee, G.M. Sex differences in skeletal muscle alterations in a model of colorectal cancer. *Physiol Rep* **8**, e14391 (2020).

31. Karastergiou, K., Smith, S.R., Greenberg, A.S. & Fried, S.K. Sex differences in human adipose tissues - the biology of pear shape. *Biol Sex Differ* **3**, 13 (2012).

32. Ring, N. *et al.* A mouse informatics platform for phenotypic and translational discovery. *Mamm Genome* **26**, 413-21 (2015).

33. Karp, N.A. *et al.* Prevalence of sexual dimorphism in mammalian phenotypic traits. *Nat Commun* **8**, 15475 (2017).

34. Dunn, W.B., Broadhurst, D.I., Atherton, H.J., Goodacre, R. & Griffin, J.L. Systems level studies of mammalian metabolomes: the roles of mass spectrometry and nuclear magnetic resonance spectroscopy. *Chem Soc Rev* **40**, 387-426 (2011).

35. Pinu, F.R. *et al.* Systems Biology and Multi-Omics Integration: Viewpoints from the Metabolomics Research Community. *Metabolites* **9**(2019).

36. Menni, C., Zierer, J., Valdes, A.M. & Spector, T.D. Mixing omics: combining genetics and metabolomics to study rheumatic diseases. *Nat Rev Rheumatol* **13**, 174-181 (2017).

37. Johnson, C.H., Ivanisevic, J. & Siuzdak, G. Metabolomics: beyond biomarkers and towards mechanisms. *Nat Rev Mol Cell Biol* **17**, 451-9 (2016).

38. Cajka, T. & Fiehn, O. Toward Merging Untargeted and Targeted Methods in Mass Spectrometry-Based Metabolomics and Lipidomics. *Anal Chem* **88**, 524-45 (2016).

39. Fiehn, O. Metabolomics by Gas Chromatography-Mass Spectrometry: Combined Targeted and Untargeted Profiling. *Curr Protoc Mol Biol* **114**, 30 4 1-30 4 32 (2016).

40. Wishart, D.S. *et al.* HMDB 4.0: the human metabolome database for 2018. *Nucleic Acids Res* **46**, D608-D617 (2018).

41. Siskos, A.P. *et al.* Interlaboratory Reproducibility of a Targeted Metabolomics Platform for Analysis of Human Serum and Plasma. *Anal Chem* **89**, 656-665 (2017).

42. Miller, M.J. *et al.* Untargeted metabolomic analysis for the clinical screening of inborn errors of metabolism. *J Inherit Metab Dis* **38**, 1029-39 (2015).

43. Kennedy, A.D. *et al.* Metabolomics in the clinic: A review of the shared and unique features of untargeted metabolomics for clinical research and clinical testing. *J Mass Spectrom* **53**, 1143-1154 (2018).

44. Barupal, D.K. *et al.* A Comprehensive Plasma Metabolomics Dataset for a Cohort of Mouse Knockouts within the International Mouse Phenotyping Consortium. *Metabolites* **9**(2019).

45. Moore, B.A. *et al.* Genome-wide screening of mouse knockouts reveals novel genes required for normal integumentary and oculocutaneous structure and function. *Sci Rep* **9**, 11211 (2019).

46. Dinesh Kumar Barupal, O.F. Chemical Similarity Enrichment Analysis (ChemRICH) as alternative to biochemical pathway mapping for metabolomic datasets. *Scientific Reports* **7**, 14567 (2017).

47. Koscielny, G. *et al.* The International Mouse Phenotyping Consortium Web Portal, a unified point of access for knockout mice and related phenotyping data. *Nucleic Acids Res* **42**, D802-9 (2014).

48. Wang, T.J. *et al.* Metabolite profiles and the risk of developing diabetes. *Nat Med* **17**, 448-53 (2011).

49. Mittelstrass, K. *et al.* Discovery of sexual dimorphisms in metabolic and genetic biomarkers. *PLoS Genet* **7**, e1002215 (2011).

50. Krumsiek, J. *et al.* Gender-specific pathway differences in the human serum metabolome. *Metabolomics* **11**, 1815-1833 (2015).

51. van der Molen, H.J. & Groen, D. Determination of progesterone in human peripheral blood using gas-liquid chromatography with electron capture detection. *J Clin Endocrinol Metab* **25**, 1625-39 (1965).

52. Barrea, L. *et al.* Trimethylamine N-oxide, Mediterranean diet, and nutrition in healthy, normal-weight adults: also a matter of sex? *Nutrition* **62**, 7-17 (2019).

53. Rushworth, D., Mathews, A., Alpert, A. & Cooper, L.J. Dihydrofolate Reductase and Thymidylate Synthase Transgenes Resistant to Methotrexate Interact to Permit Novel Transgene Regulation. *J Biol Chem* **290**, 22970-6 (2015).

54. Fry, A.M. The Nek2 protein kinase: a novel regulator of centrosome structure. *Oncogene* **21**, 6184-94 (2002).

55. Mlodzik, M. The Dishevelled Protein Family: Still Rather a Mystery After Over 20 Years of Molecular Studies. *Curr Top Dev Biol* **117**, 75-91 (2016).

56. Hoglinger, D. *et al.* NPC1 regulates ER contacts with endocytic organelles to mediate cholesterol egress. *Nat Commun* **10**, 4276 (2019).

57. Robinson, J.I. *et al.* Metabolomic networks connect host-microbiome processes to human Clostridioides difficile infections. *J Clin Invest* **129**, 3792-3806 (2019).

58. Lytra, G., Miot-Sertier, C., Moine, V., Coulon, J. & Barbe, J.C. Influence of must yeast-assimilable nitrogen content on fruity aroma variation during malolactic fermentation in red wine. *Food Res Int* **135**, 109294 (2020).

59. Chocholouskova, M. *et al.* Reversed phase UHPLC/ESI-MS determination of oxylipins in human plasma: a case study of female breast cancer. *Anal Bioanal Chem* **411**, 1239-1251 (2019).

60. Zou, Z. *et al.* Inhibition of the HER2 pathway by n-3 polyunsaturated fatty acids prevents breast cancer in fat-1 transgenic mice. *J Lipid Res* **54**, 3453-63 (2013).

61. Fagerberg, L. *et al.* Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. *Mol Cell Proteomics* **13**, 397-406 (2014).

62. Guillemot, N., Troadec, C., de Villemeur, T.B., Clement, A. & Fauroux, B. Lung disease in Niemann-Pick disease. *Pediatr Pulmonol* **42**, 1207-14 (2007).

63. Brandsma, C.A. *et al.* A large lung gene expression study identifying fibulin-5 as a novel player in tissue repair in COPD. *Thorax* **70**, 21-32 (2015).

64. Jafari, Z., Kolb, B.E. & Mohajerani, M.H. Prepulse inhibition of the acoustic startle reflex and P50 gating in aging and alzheimer's disease. *Ageing Res Rev* **59**, 101028 (2020).

65. Djoumbou Feunang, Y. *et al.* ClassyFire: automated chemical classification with a comprehensive, computable taxonomy. *J Cheminform* **8**, 61 (2016).

66. Coll, A.P., Farooqi, I.S. & O'Rahilly, S. The hormonal control of food intake. *Cell* **129**, 251-62 (2007).

67. O'Shaughnessy, P.J. Hormonal control of germ cell development and spermatogenesis. *Semin Cell Dev Biol* **29**, 55-65 (2014).

68. Taniguchi, M. & Okazaki, T. The role of sphingomyelin and sphingomyelin synthases in cell death, proliferation and migration-from cell and animal models to human disorders. *Biochim Biophys Acta* **1841**, 692-703 (2014).

69. Nayeem, M.A. Role of oxylipins in cardiovascular diseases. *Acta Pharmacol Sin* **39**, 1142-1154 (2018).

70. Alexeev, E.E. *et al.* Microbiota-Derived Indole Metabolites Promote Human and Murine Intestinal Homeostasis through Regulation of Interleukin-10 Receptor. *Am J Pathol* **188**, 1183-1194 (2018).

71. Swann, J.R. *et al.* Systemic gut microbial modulation of bile acid metabolism in host tissue compartments. *Proc Natl Acad Sci U S A* **108 Suppl 1**, 4523-30 (2011).

72. Wang, Z. *et al.* Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. *Nature* **472**, 57-63 (2011).

73. Johnson, C., Prokopienko, A.J., West, R.E., 3rd, Nolin, T.D. & Stubbs, J.R. Decreased Kidney Function Is Associated with Enhanced Hepatic Flavin Monooxygenase Activity and Increased Circulating Trimethylamine N-Oxide Concentrations in Mice. *Drug Metab Dispos* **46**, 1304-1309 (2018).

74. Cho, C.E. & Caudill, M.A. Trimethylamine-N-Oxide: Friend, Foe, or Simply Caught in the Cross-Fire? *Trends Endocrinol Metab* **28**, 121-130 (2017).

75. Tang, W.H. *et al.* Intestinal microbial metabolism of phosphatidylcholine and cardiovascular risk. *N Engl J Med* **368**, 1575-84 (2013).

76. Kim, R.B. *et al.* Advanced chronic kidney disease populations have elevated trimethylamine N-oxide levels associated with increased cardiovascular events. *Kidney Int* **89**, 1144-1152 (2016).

77. Manor, O. *et al.* A Multi-omic Association Study of Trimethylamine N-Oxide. *Cell Rep* **24**, 935-946 (2018).

78. Organ, C.L. *et al.* Choline Diet and Its Gut Microbe-Derived Metabolite, Trimethylamine N-Oxide, Exacerbate Pressure Overload-Induced Heart Failure. *Circ Heart Fail* **9**, e002314 (2016).

79. Papandreou, C. *et al.* Plasma trimethylamine-N-oxide and related metabolites are associated with type 2 diabetes risk in the Prevencion con Dieta Mediterranea (PREDIMED) trial. *Am J Clin Nutr* **108**, 163-173 (2018).

80. Mondul, A.M. *et al.* Metabolomic analysis of prostate cancer risk in a prospective cohort: The alpha-tocolpherol, beta-carotene cancer prevention (ATBC) study. *Int J Cancer* **137**, 2124-32 (2015).

81. Guertin, K.A. *et al.* Serum Trimethylamine N-oxide, Carnitine, Choline, and Betaine in Relation to Colorectal Cancer Risk in the Alpha Tocopherol, Beta Carotene Cancer Prevention Study. *Cancer Epidemiol Biomarkers Prev* **26**, 945-952 (2017).

82. Caligiuri, S.P.B., Parikh, M., Stamenkovic, A., Pierce, G.N. & Aukema, H.M. Dietary modulation of oxylipins in cardiovascular disease and aging. *Am J Physiol Heart Circ Physiol* **313**, H903-H918 (2017).

83. Imig, J.D. *et al.* An orally active epoxide hydrolase inhibitor lowers blood pressure and provides renal protection in salt-sensitive hypertension. *Hypertension* **46**, 975-81 (2005).

84. Jenkins, C.M., Cedars, A. & Gross, R.W. Eicosanoid signalling pathways in the heart. *Cardiovasc Res* **82**, 240-9 (2009).

85. Caligiuri, S.P. *et al.* Dietary Flaxseed Reduces Central Aortic Blood Pressure Without Cardiac Involvement but Through Changes in Plasma Oxylipins. *Hypertension* **68**, 1031-8 (2016).

86. Kander, M.C., Cui, Y. & Liu, Z. Gender difference in oxidative stress: a new look at the mechanisms for cardiovascular diseases. *J Cell Mol Med* **21**, 1024-1032 (2017).

87. Haemmerle, G. & Lass, A. Genetically modified mouse models to study hepatic neutral lipid mobilization. *Biochim Biophys Acta Mol Basis Dis* **1865**, 879-894 (2019).

88. Yang, W. *et al.* Serum lipids and lipoproteins in Chinese men and women. *Circulation* **125**, 2212-21 (2012).

89. Wang, X., Magkos, F. & Mittendorfer, B. Sex differences in lipid and lipoprotein metabolism: it's not just about sex hormones. *J Clin Endocrinol Metab* **96**, 885-93 (2011).

90. Palmisano, B.T., Zhu, L., Eckel, R.H. & Stafford, J.M. Sex differences in lipid and lipoprotein metabolism. *Mol Metab* **15**, 45-55 (2018).

91. Barupal, D.K. *et al.* The circulating lipidome is largely defined by sex descriptors in the GOLDN, GeneBank and the ADNI studies. *bioRxiv*, 731448 (2019).

92. Rist, M.J. *et al.* Metabolite patterns predicting sex and age in participants of the Karlsruhe Metabolomics and Nutrition (KarMeN) study. *PLoS One* **12**, e0183228 (2017).

93. Ellul, S. *et al.* Metabolomics: population epidemiology and concordance in Australian children aged 11-12 years and their parents. *BMJ Open* **9**, 106-117 (2019).

94. Colley, S.M. & Leedman, P.J. Steroid Receptor RNA Activator - A nuclear receptor coregulator with multiple partners: Insights and challenges. *Biochimie* **93**, 1966-72 (2011).

95. Cooper, C. *et al.* Steroid Receptor RNA Activator bi-faceted genetic system: Heads or Tails? *Biochimie* **93**, 1973-80 (2011).

96. Martinez-Outschoorn, U.E., Sotgia, F. & Lisanti, M.P. Power surge: supporting cells "fuel" cancer cell mitochondria. *Cell Metab* **15**, 4-5 (2012).

97. Camarda, R. *et al.* Inhibition of fatty acid oxidation as a therapy for MYC-overexpressing triple-negative breast cancer. *Nat Med* **22**, 427-32 (2016).

98. Apaya, M.K., Chang, M.T. & Shyur, L.F. Phytomedicine polypharmacology: Cancer therapy through modulating the tumor microenvironment and oxylipin dynamics. *Pharmacol Ther* **162**, 58-68 (2016).

99. Schweiger, D., Furstenberger, G. & Krieg, P. Inducible expression of 15-lipoxygenase-2 and 8-lipoxygenase inhibits cell growth via common signaling pathways. *J Lipid Res* **48**, 553-64 (2007).

100. Greene, E.R., Huang, S., Serhan, C.N. & Panigrahy, D. Regulation of inflammation in cancer by eicosanoids. *Prostaglandins Other Lipid Mediat* **96**, 27-36 (2011).

101. Pakiet, A., Kobiela, J., Stepnowski, P., Sledzinski, T. & Mika, A. Changes in lipids composition and metabolism in colorectal cancer: a review. *Lipids Health Dis* **18**, 29 (2019).

102. Wanders, R.J. & Waterham, H.R. Peroxisomal disorders: the single peroxisomal enzyme deficiencies. *Biochim Biophys Acta* **1763**, 1707-20 (2006).

103. Waterham, H.R., Ferdinandusse, S. & Wanders, R.J. Human disorders of peroxisome metabolism and biogenesis. *Biochim Biophys Acta* **1863**, 922-33 (2016).

104. Griese, M. *et al.* Respiratory disease in Niemann-Pick type C2 is caused by pulmonary alveolar proteinosis. *Clin Genet* **77**, 119-30 (2010).

105. Bjurulf, B. *et al.* Niemann-Pick disease type C2 presenting as fatal pulmonary alveolar lipoproteinosis: morphological findings in lung and nervous tissue. *Med Sci Monit* **14**, CS71-5 (2008).

106. Pilecki, B. *et al.* Characterization of Microfibrillar-associated Protein 4 (MFAP4) as a Tropoelastin- and Fibrillin-binding Protein Involved in Elastic Fiber Formation. *J Biol Chem* **291**, 1103-14 (2016).

107. Johansson, S.L. *et al.* Microfibrillar-associated protein 4: a potential biomarker of chronic obstructive pulmonary disease. *Respir Med* **108**, 1336-44 (2014).

108. Wang, J.T. *et al.* A novel palmitic acid hydroxy stearic acid (5-PAHSA) plays a neuroprotective role by inhibiting phosphorylation of the m-TOR-ULK1 pathway and regulating autophagy. *CNS Neurosci Ther* **27**, 484-496 (2021).

109. Yanguas-Casas, N. *et al.* Sex differences in the phagocytic and migratory activity of microglia and their impairment by palmitic acid. *Glia* **66**, 522-537 (2018).

110. Kwon, H.M., Lim, J.S., Park, H.K. & Lee, Y.S. Hypertriglyceridemia as a possible predictor of early neurological deterioration in acute lacunar stroke. *J Neurol Sci* **309**, 128-30 (2011).

111. Gertsenstein, M. *et al.* Efficient generation of germ line transmitting chimeras from C57BL/6N ES cells by aggregation with outbred host embryos. *PLoS One* **5**, e11260 (2010).

112. Matyash, V., Liebisch, G., Kurzchalia, T.V., Shevchenko, A. & Schwudke, D. Lipid extraction by methyl-tert-butyl ether for high-throughput lipidomics. *J Lipid Res* **49**, 1137-46 (2008).

113. Showalter, M.R. *et al.* Obesogenic diets alter metabolism in mice. *PLoS One* **13**, e0190632 (2018).

114. Gao, B. *et al.* Multi-Omics Analyses Detail Metabolic Reprogramming in Lipids, Carnitines, and Use of Glycolytic Intermediates between Prostate Small Cell Neuroendocrine Carcinoma and Prostate Adenocarcinoma. *Metabolites* **9**(2019).

115. Blazenovic, I. *et al.* Structure Annotation of All Mass Spectra in Untargeted Metabolomics. *Anal Chem* **91**, 2155-2162 (2019).

116. Buckner, T. *et al.* Predictors of oxylipins in a healthy pediatric population. *Pediatr Res* (2020).

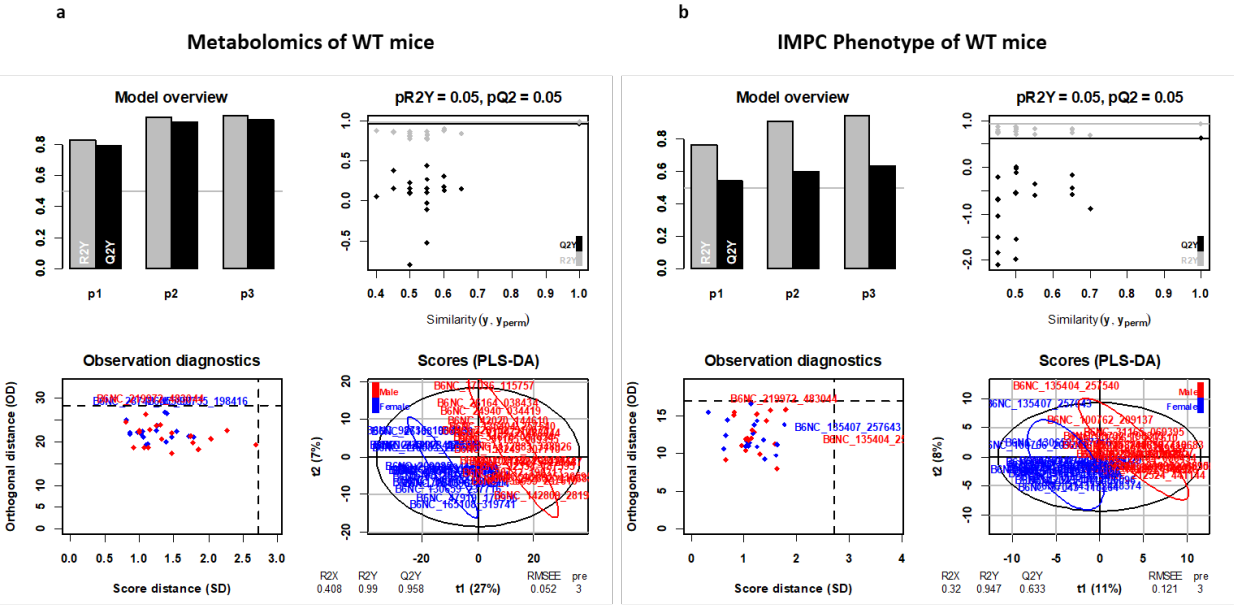
117. Pedersen, T.L. & Newman, J.W. Establishing and Performing Targeted Multi-residue Analysis for Lipid Mediators and Fatty Acids in Small Clinical Plasma Samples. *Methods Mol Biol* **1730**, 175-212 (2018).

118. Fan, S. *et al.* Systematic Error Removal Using Random Forest for Normalizing Large-Scale Untargeted Lipidomics Data. *Anal Chem* **91**, 3590-3596 (2019).

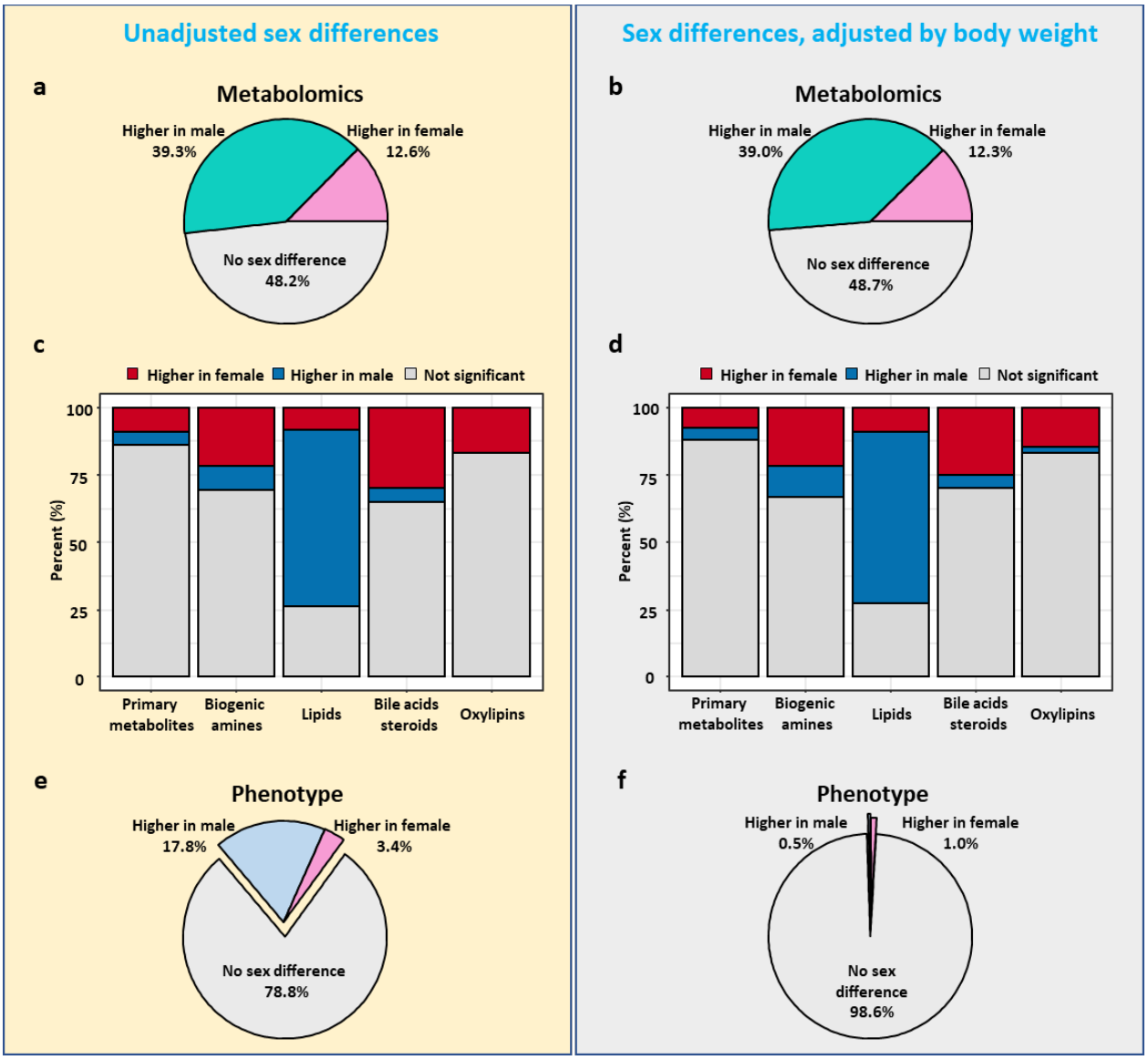
119. Karp, N.A., Heller, R., Yaacoby, S., White, J.K. & Benjamini, Y. Improving the Identification of Phenotypic Abnormalities and Sexual Dimorphism in Mice When Studying Rare Event Categorical Characteristics. *Genetics* **205**, 491-501 (2017).

120. Benjamini, Y.a.H., Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Statist. Soc.* **B**, 289-300 (1995).

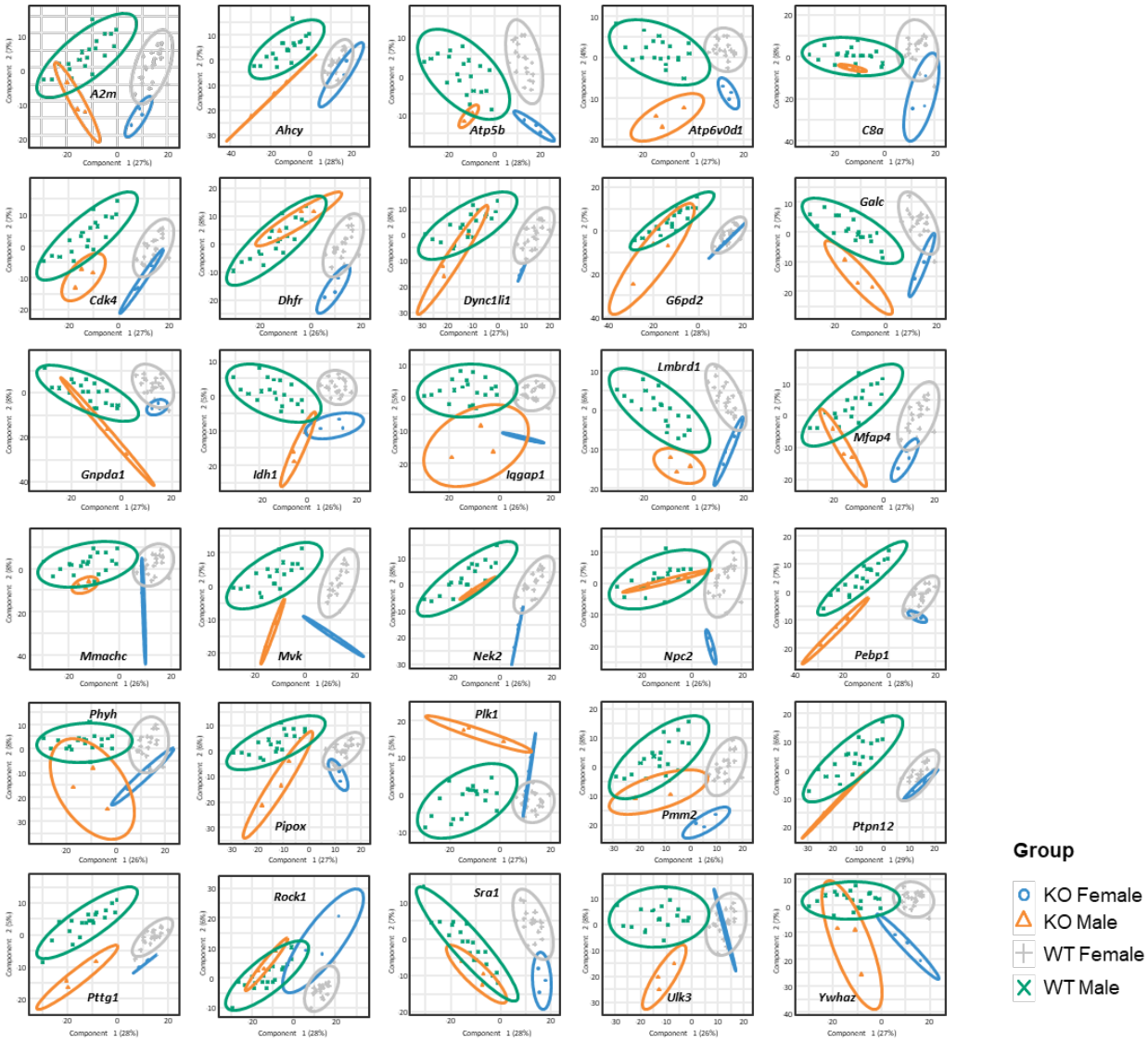
3.7 Supplementary Information



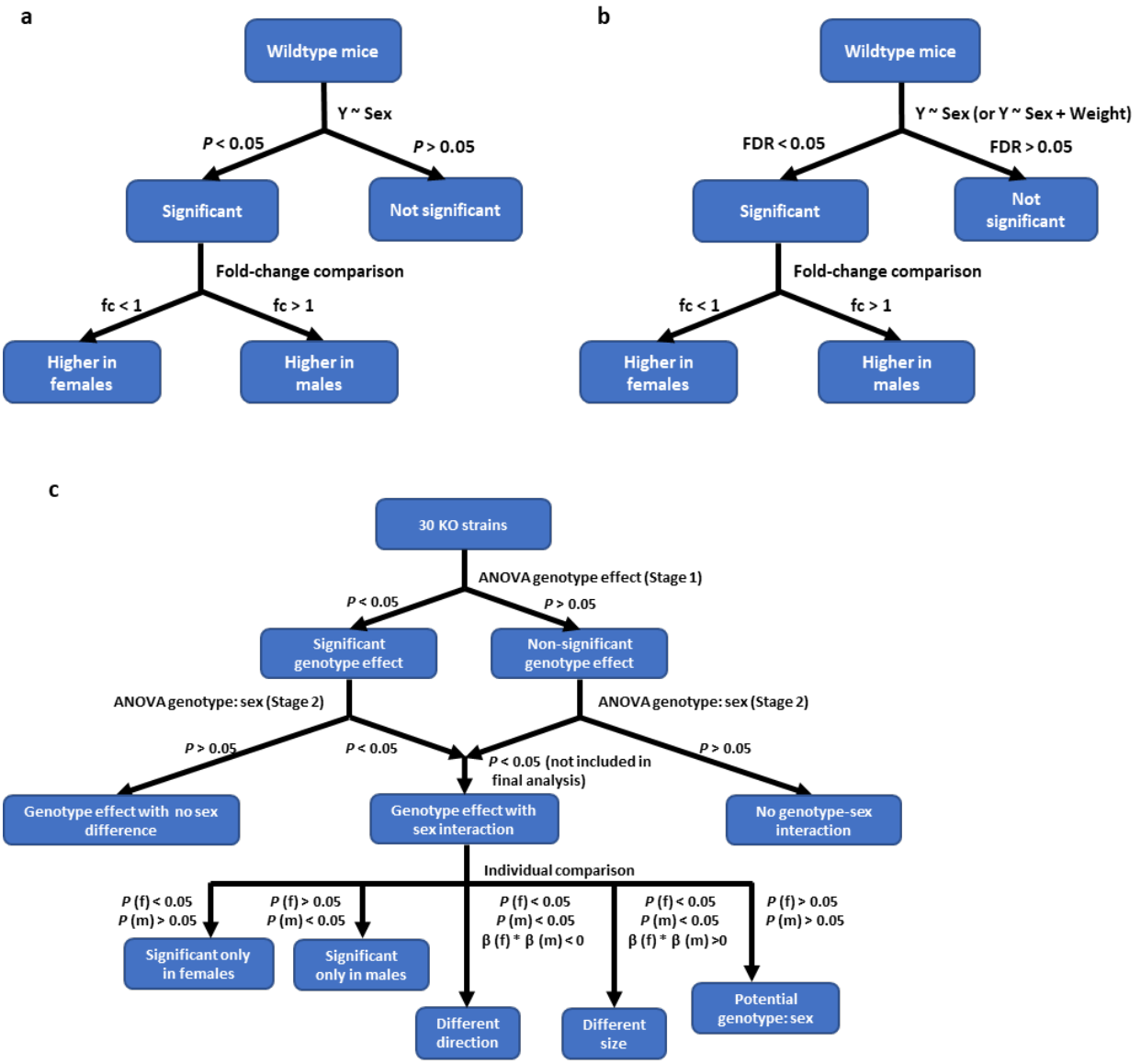
Supplementary Figure 3.1. Sexual dimorphism of metabolomics data and phenotype data in wildtype mice (n = 20 males & 20 females). (a) PLS-DA plot of metabolomics data in wildtype mice. (b) PLS-DA plot of IMPC phenotype dataset in wildtype mice.



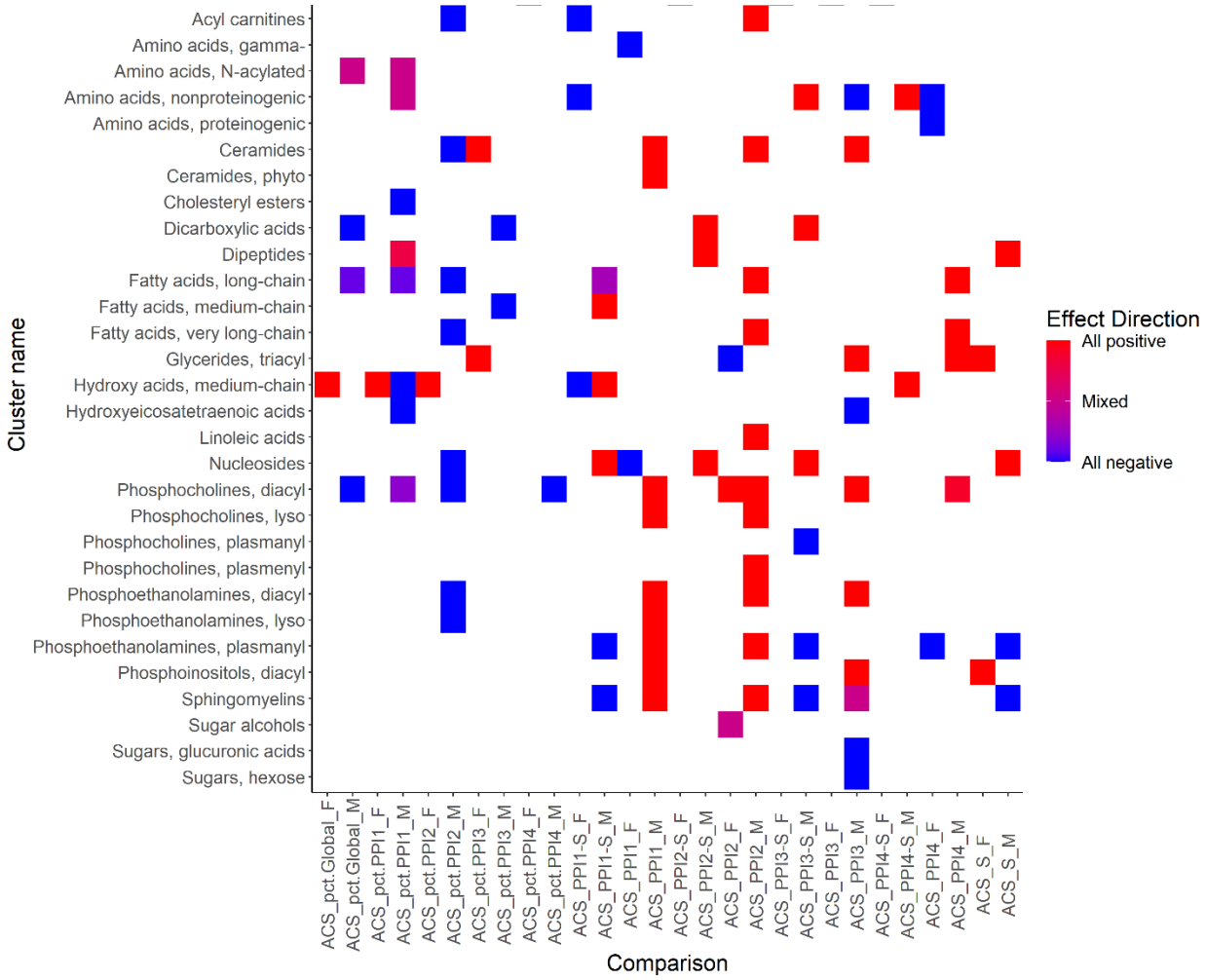
Supplementary Figure 3.2. Sex as a biological variable with a stricter significance threshold and normalized to body weight in wildtype C57BL/6NCrl mice (n = 20 males & 20 females, generalized linear model at FDR < 0.05). (a) Proportion of metabolites significantly affected by sex. (b) Proportion of metabolites significantly affected by sex, adjusted by body weight. (c) Assay-based proportion of sex-affected metabolites. (d) Assay-based proportion of sex-affected metabolites, adjusted by body weight. (e) Proportion of IMPC phenotypes of continuous traits significantly affected by sex. (f) Proportion of IMPC phenotypes of continuous traits significantly affected by sex; adjusted by body weight.



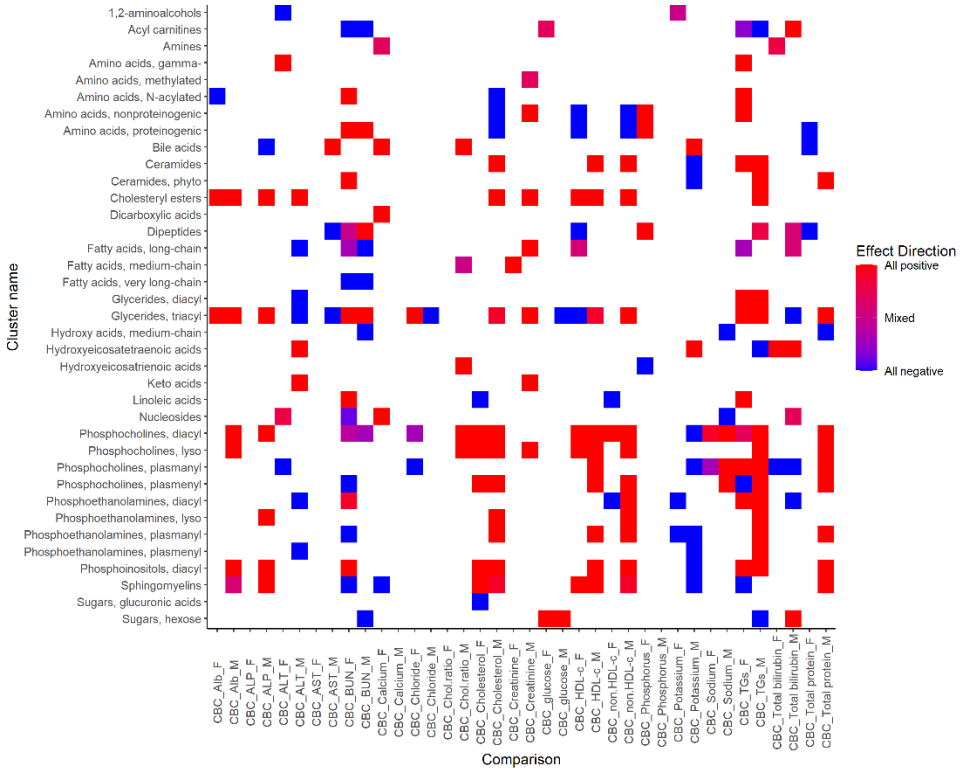
Supplementary Figure 3.3. PLS-DA plot of metabolomics data in 30 KO mouse lines compared to wildtype mice (n = 20 males & 20 females in C57BL/6NCrl wildtype mice, n = 3 male & 3 female mice in each KO mouse line). Circles: confidence intervals for plasma metabolic phenotype variance. Axes: Regression vectors with % of total explained variance.



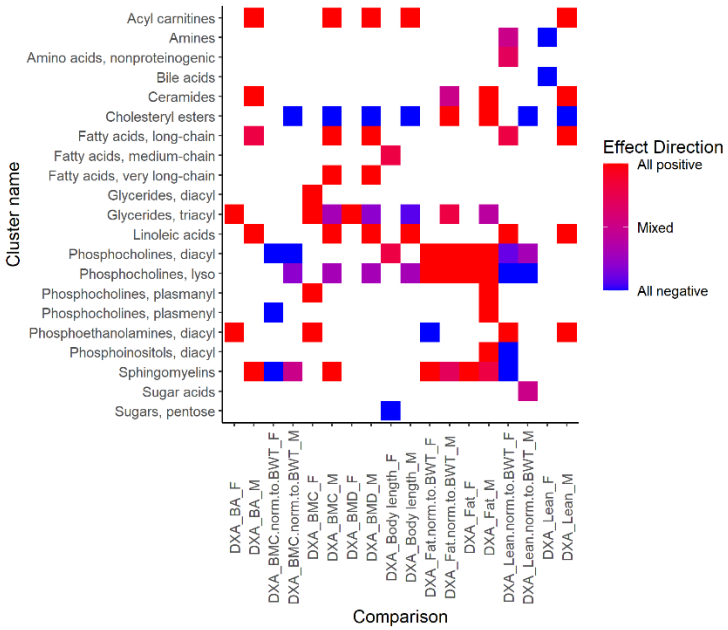
Supplementary Figure 3.4. Statistical analyses to study sexual dimorphism in wildtype mice and in knockout mice. (a) Statistical analysis in wildtype mice using generalized linear model at $P < 0.05$. (b) Statistical analysis in wildtype mice using generalized linear model at $P < 0.05$ with false discovery rate correction (FDR < 0.05) and body weight adjustment. (c) Statistical analysis for genotype-sex interaction effect in knockout lines compared to wildtype mice using two-way ANOVA followed by individual comparison by generalized linear model at $P < 0.05$.



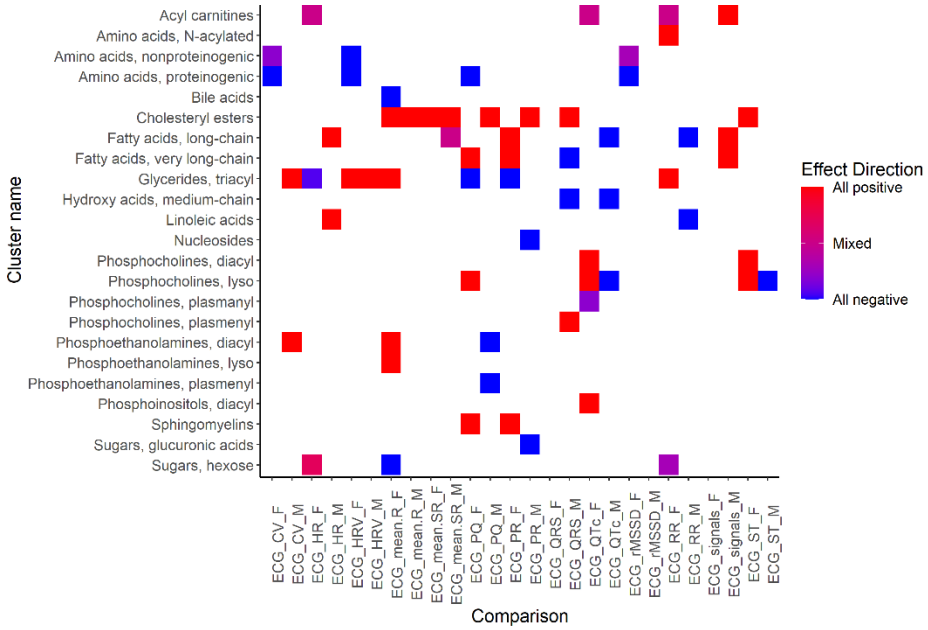
Supplementary Figure 3.5a. Acoustic Startle and Pre-pulse Inhibition (PPI, week 10). The acoustic startle response is characterized by an exaggerated flinching response to an unexpected strong auditory stimulus (pre-pulse). This response can be attenuated when it is preceded by a weaker stimulus (pre-pulse) and is the principle underlying pre-pulse inhibition (PPI). Several clinical studies have shown that a number of human disorders have impaired PPI including schizophrenia, Huntington's disease, fragile X syndrome, and autism.



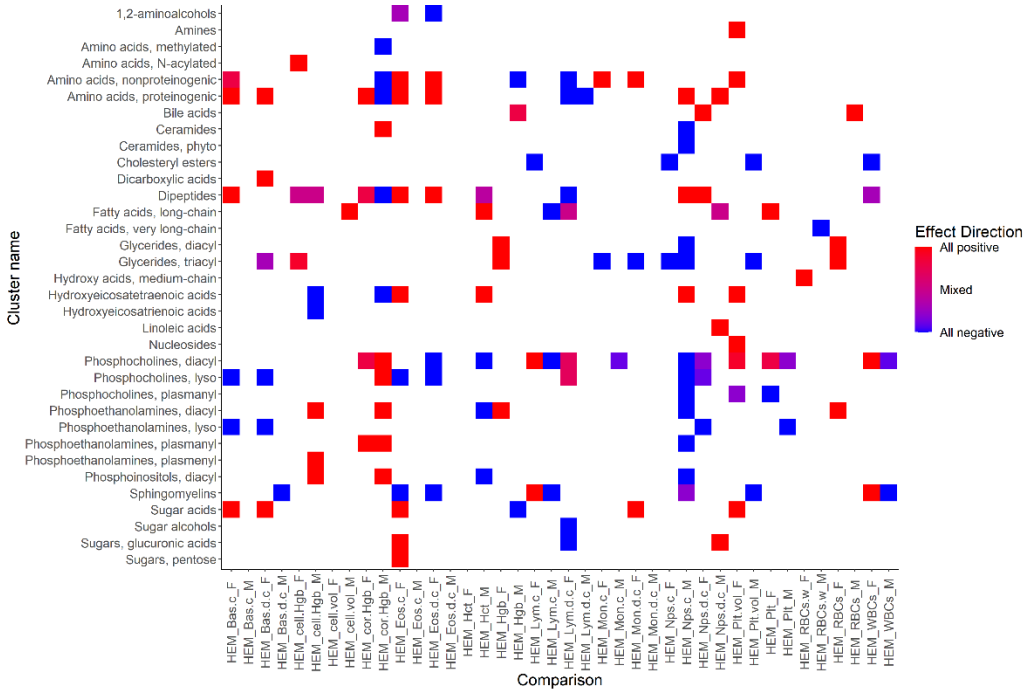
Supplementary Figure 3.5b. Clinical Chemistry (CBC, week 16). Clinical chemistry determines biochemical parameters in plasma including enzymatic activity, specific substrates and electrolytes.



Supplementary Figure 3.5c. Body Composition (DEXA lean/fat, week 14). Measure bone mineral content and density as well as body composition in mice using the DEXA (Dual Energy X-ray Absorptiometry) analyser.

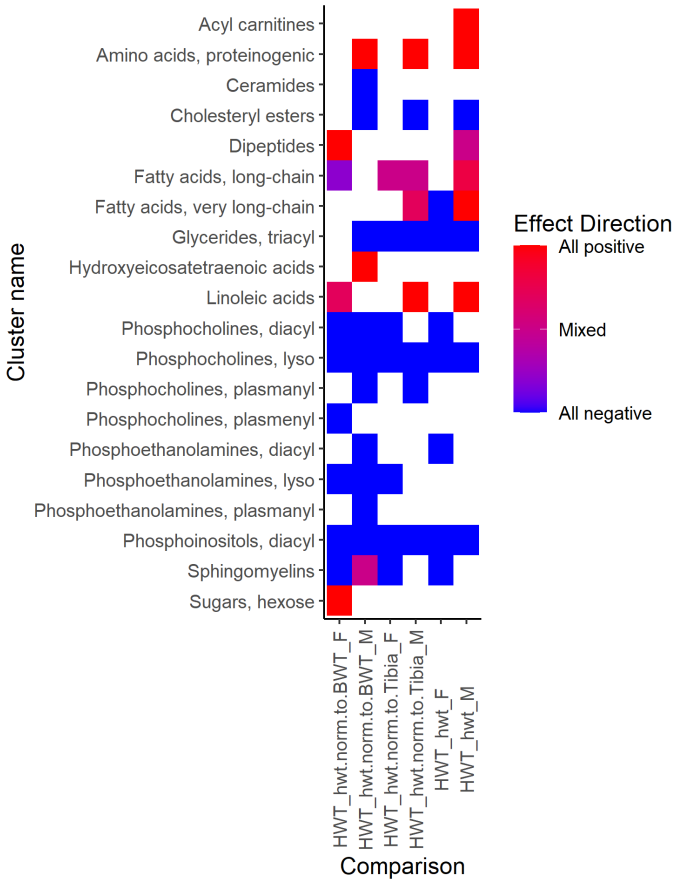


Supplementary Figure 3.5d. Electrocardiogram (ECG, week 12). Electrocardiogram provides a high throughput method to obtain Electrocardiograms in a conscious mouse or an anesthetized mouse. Electrocardiographic recordings usually help to detect abnormal myocardial action potential, conduction of impulse, disturbances in cardiac rate and rhythm, and altered autonomic activities.

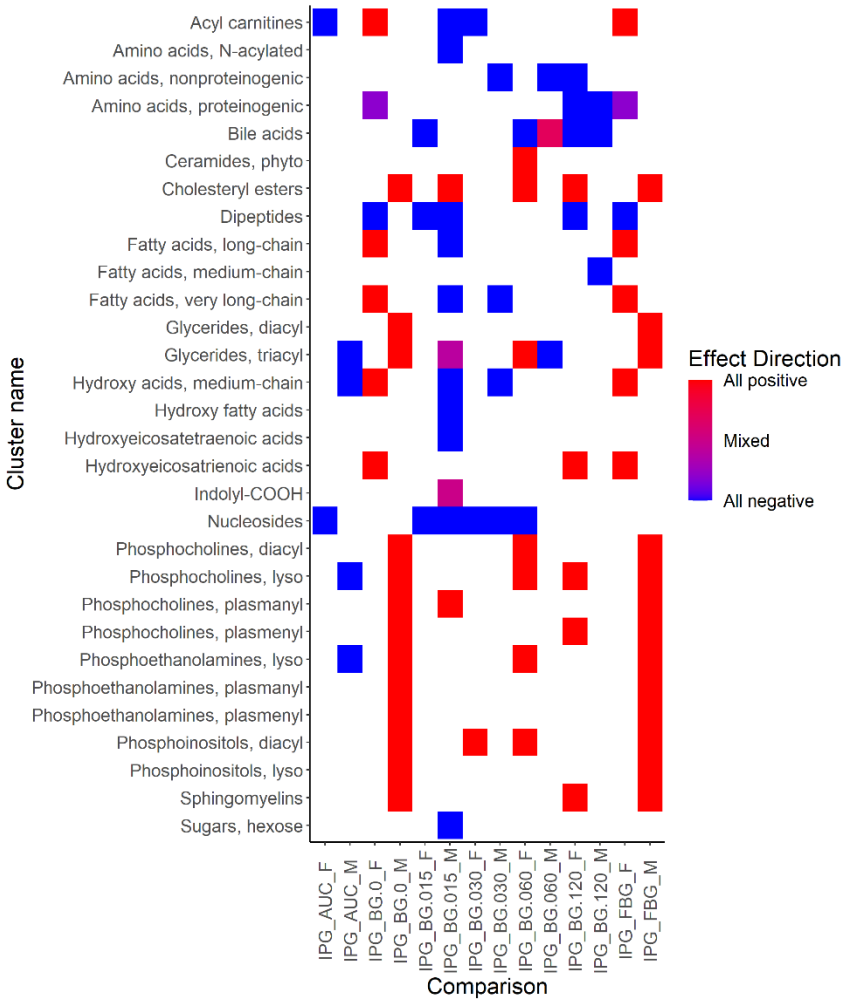


Supplementary Figure 3.5e. Hematology (week 16). Hematological assessment of blood determines blood cell counts (white blood cells, red blood cells, hemoglobin, and platelets) and additional hematological parameters (hematocrit, mean cell volume, mean corpuscular hemoglobin, mean cell hemoglobin

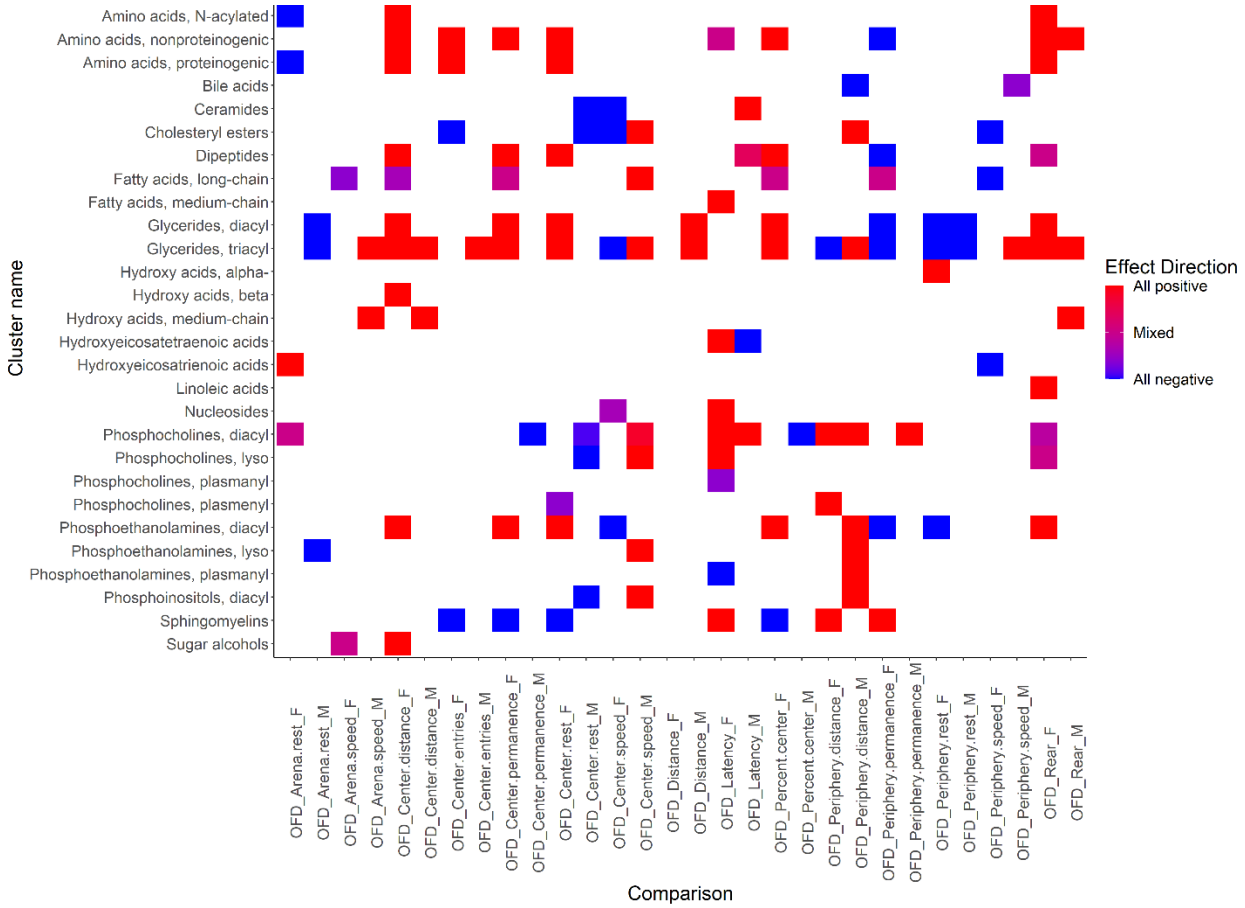
concentration) can be derived using these indices. These tests will indicate abnormalities in the production of blood and its components (blood cells and hemoglobin) as well as in the associated blood-forming organs.



Supplementary Figure 3.5f. Heart Weight (week 16). To evaluate cardiac size using heart weight and body weight.



Supplementary Figure 3.5g. Intraperitoneal glucose tolerance test (IPG, week 13). The glucose tolerance test measures the clearance of an intraperitoneally injected glucose load from the body. It is used to detect disturbances in glucose metabolism that can be linked to human conditions such as diabetes or metabolic syndrome.



Supplementary Figure 3.5h. Open Field (week 8). The Open Field test is used to assess anxiety and exploratory behaviors. It is based on the natural tendency of an animal to explore and to protect itself using avoidance which translates to a normal animal spending more time in the periphery of the Open Field arena than in the center (the most anxiogenic area).

Supplementary Table 3.1. Gene information of 30 knockout lines

Gene symbol	Name	Zygosity	Male		Female		New comments	Disease models (by annotation and orthology)
			Age (wks)	Number	Age (wks)	Number		
None (wildtype control)	Null	+/+	16.1 ± 0.3	20	16.1 ± 0.3	20	/	/
A2m	Alpha-2-macroglobulin	-/-	15.7 ± 0	3	15.7 ± 0	3	protein for proteinase inhibitor	Yes
Ahcy	S-adenosylhomocysteine hydrolase	+/-	16.1 ± 0.2	3	16.2 ± 0.3	3	Enzyme on pathway reactions (amino acid)	Yes
Atp5b	ATP synthase, H+ transporting mitochondrial F1 complex, beta subunit	+/-	16.3 ± 0	3	16 ± 0.5	3	Enzyme on ATP synthesis	/
Atp6v0d1	ATPase H+ Transporting lysosomal V0 Subunit D1	+/-	16.2 ± 0.1	3	15.9 ± 0	3	Enzyme on ATP synthesis	/
C8a	Complement component 8, alpha polypeptide	-/-	16 ± 0.2	3	16.2 ± 0.6	3	protein for immune response	Yes
Cdk4	Cyclin Dependent Kinase 4	+/-	15.9 ± 0	3	16 ± 0.4	3	Enzyme on protein residue	Yes
Dhfr	Dihydrofolate reductase	+/-	16.3 ± 0	3	16.1 ± 0.4	3	Enzyme on pathway reactions	Yes
Dync1li1	Dynein Cytoplasmic 1 Light Intermediate Chain 1	+/-	16 ± 0.3	3	16.4 ± 0	3	Protein on dynein function	/
G6pd2	Glucose 6-phosphate dehydrogenase 2	-/-	16.4 ± 0.3	3	15.9 ± 0.1	3	Enzyme on pathway reactions (pentose phosphate)	Yes
Galc	Galactosylceramidase	+/-	16.1 ± 0	3	16 ± 0	3	Enzyme on pathway reactions (galactosylceramide)	Yes
Gnpda1	Glucosamine-6-phosphate deaminase 1	+/-	16.3 ± 0.1	3	16.3 ± 0.1	3	Enzyme on pathway reactions (glucosamine)	/
Idh1	Isocitrate dehydrogenase	-/-	16 ± 0.2	3	16.1 ± 0.4	3	Enzyme on pathway reactions	Yes
Iqgap1	IQ motif containing GTPase activating protein 1	-/-	15.8 ± 0.1	3	15.9 ± 0.1	3	Protein on actin cytoskeleton	/
Lmbrd1	LMBR1 domain containing 1	+/-	16 ± 0.2	3	15.9 ± 0.2	3	Membrane protein	Yes
Mfap4	Microfibrillar-associated protein 4	-/-	16 ± 0.2	3	16.6 ± 0	3	Developmental protein	/
Mmachc	Methylmalonic aciduria cblC type, with homocystinuria	+/-	16 ± 0.4	3	16 ± 0.2	3	Enzyme on protein residue (cobalamin metabolic process)	Yes
Mvk	mevalonate kinase	+/-	16.4 ± 0.1	3	15.7 ± 0.1	3	Enzyme on pathway reactions (soprenoid and cholesterol biosynthesis)	Yes
Nek2	NIMA (never in mitosis gene a)-related expressed kinase 2	-/-	15.8 ± 0.2	3	16 ± 0.1	3	Enzyme on protein	Yes
Npc2	NPC intracellular cholesterol transporter 2	+/-	16 ± 0.2	3	16.2 ± 0.2	3	Protein on cholesterol efflux	Yes
Pebp1	Phosphatidylethanolamine binding protein 1	-/-	16.3 ± 0.2	3	16 ± 0.1	3	Protein for MAPK cascade	/
Phyh	Phytanoyl- CoA hydroxylase	-/-	16.1 ± 0.3	3	16 ± 0.3	3	Enzyme on pathway reactions (acy-CoAs)	Yes
Pipox	Pipecolic acid oxidase	-/-	15.8 ± 0.2	3	16.1 ± 0	3	Enzyme on pathway reactions	/
Plk1	Polo-like kinase 1, serine/threonine protein kinase	+/-	16.1 ± 0	3	16 ± 0.2	3	Enzyme on protein for cell cycle	/
Pmm2	Phosphomannomutase 2	+/-	16 ± 0.1	3	16.2 ± 0.1	3	Enzyme on pathway reactions	Yes
Ptpn12	Protein tyrosine phosphatase, non-receptor type 12	+/-	15.9 ± 0.1	3	15.7 ± 0.2	3	Enzyme on protein for cellular signaling cascades	Yes
Pttg1	Pituitary tumor-transforming gene 1	-/-	16.1 ± 0	3	15.8 ± 0.2	3	Protein on chromosome stability	/
Rock1	Rho-associated coiled-coil containing protein kinase 1	+/-	16 ± 0.2	3	16.1 ± 0.2	3	Enzyme on protein residue for apoptosis	/
Sra1	Steroid receptor RNA activator 1	-/-	16.2 ± 0.1	3	15.9 ± 0	3	Protein on apoptosis	/
Ulk3	Unc-51-like kinase 3	-/-	16.1 ± 0.2	3	16 ± 0.2	3	Enzyme on protein for autophagy	/
Ywhaz	Tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein	+/-	16.3 ± 0.1	3	16.4 ± 0.1	3	Protein on signaling pathways	/

Supplementary Table 3.2. Number of significantly changed metabolites by genotype effect and their alterations in 30 knockout strains.

Genotype	of significant metabolites	Significant clusters	FC > 2 or FC < 0.5	FC > 5 or FC < 0.2	FC > 10 or FC < 0.1
A2m$^{-/-}$	201	13	36	0	2
Ahcy$^{+/-}$	128	10	22	2	0
Atp5b$^{+/-}$	181	13	52	3	1
Atp6v0d1$^{+/-}$	120	10	29	2	0
C8a$^{-/-}$	203	14	50	1	3
Cdk4$^{+/-}$	135	16	29	3	1
Dhfr$^{+/-}$	184	13	43	5	1
Dync1li1$^{+/-}$	229	13	63	7	0
G6pd2$^{-/-}$	113	8	26	5	4
Galc$^{+/-}$	154	9	36	2	2
Gnpda1$^{+/-}$	143	9	29	2	2
Idh1$^{-/-}$	156	13	28	4	0
Iqgap1$^{-/-}$	137	13	17	1	2
Lmbrd1$^{+/-}$	198	13	53	2	1
Mfap4$^{-/-}$	192	10	34	4	1
Mmachc$^{+/-}$	182	13	35	2	3
Mvk$^{+/-}$	230	13	43	9	4
Nek2$^{-/-}$	205	15	47	1	3
Npc2$^{+/-}$	213	14	64	7	2
Pebp1$^{-/-}$	183	13	36	4	3
Phyh$^{-/-}$	159	8	39	5	5
Pipox$^{-/-}$	121	5	22	2	1
Plk1$^{+/-}$	127	10	27	3	2
Pmm2$^{+/-}$	232	18	59	6	7
Ptpn12$^{+/-}$	149	7	32	2	2
Pttg1$^{-/-}$	179	17	38	5	1
Rock1$^{+/-}$	97	7	26	2	1
Sra1$^{-/-}$	181	15	33	3	2
Ulk3$^{-/-}$	213	17	64	8	2
Ywhaz$^{+/-}$	154	12	36	7	7

Supplementary Table 3.3. Examples of sexual dimorphism of phenotypes (continuous traits) in wildtype mice and their alterations in 30 knockout lines.

Phenotype	Sex Effect in WT	Statistical Analysis in KO Lines	
		Genotype Effect	Genotype-sex Interaction
Heart weight, mg	***	A2m, Atp5b, Atp6v0d1, Cdk4, Gnpda1, Iqgap1	A2m, Atp6v0d1, Iqgap1
Body weight, g	***	A2m, Atp6v0d1, C8a, Cdk4, Dhfr, Dync1li1, Idh1, Iqgap1, Npc2, Pebp1, Plk1, Sra1	A2m, C8a, Dync1li1, Npc2, Pebp1, Sra1
Heart/Body weight, mg/g	ns	Atp5b, Atp6v0d1, C8a, Idh1, Mvk, Nek2, Pipox, Ulk3, Ywhaz	C8a, Pipox, Ywhaz
Alkaline phosphatase	***	Dhfr, Gnpda1, Plk1, Sra1, Ulk3	Dhfr, Sra1, Ulk3
Total cholesterol	***	Atp6v0d1, C8a, Dync1li1, Galc, Gnpda1, Idh1, Sra1, Ulk3	Galc
HDL cholesterol	***	A2m, Atp6v0d1, Galc, Gnpda1, Lmbrd1, Mfap4, Mmachc, Pebp1	Galc, Lmbrd1, Mmachc, Pebp1
Non-HDL cholesterol [a]	ns	Ahcy, C8a, Dhfr, Galc, Gnpda1, Idh1, Npc2, Rock1, Sra1, Ulk3	C8a, Galc, Idh1, Ulk3
Triglycerides	***	A2m, Ahcy, Dync1li1, Lmbrd1, Npc2 A2m, Ahcy, Dync1li1, Lmbrd1, Mfap4, Mvk, Npc2, Pipox	Ahcy, Npc2
Tibia length, mm	ns	Atp5b, Atp6v0d1, Cdk4, Dhfr, Dync1li1, G6pd2, Idh1, Iqgap1, Lmbrd1, Mfap4, Mmachc, Nek2, Phyh, Pipox, Plk1, Pmm2, Pttg1, Rock1, Sra1, Ywhaz	Atp5b, Dhfr, Dync1li1, Mfap4, Mmachc, Phyh, Pipox, Plk1, Pttg1, Sra1
Heart/Tibia length, mg/cm [a]	***	Ahcy, Atp5b, Atp6v0d1, Cdk4, Dhfr, G6pd2, Idh1, Iqgap1, Lmbrd1, Mfap4, Mmachc, Phyh, Pipox, Plk1, Pmm2, Ptpn12, Sra1, Ywhaz	Dhfr, Pipox, Pmm2, Sra1

[a] Values were additionally calculated from measured phenotypes. Two-way ANOVA was used for statistical analysis. ns, not significant; *, $P < 0.05$; **, $P < 0.01$, ***, $P < 0.001$.

Chapter 4 Comparing Stable Isotope Enrichment by Gas Chromatography with Time-of-Flight, Quadrupole Time-of-Flight and Quadrupole Mass Spectrometry

*Reproduced from Ying Zhang, Bei Gao, Luis Valdiviez, Chao Zhu, Tara Gallagher, Katrine Whiteson, Oliver Fiehn. "Comparing stable isotope enrichment by gas chromatography with time-of-flight, quadrupole time-of-flight, and quadrupole mass spectrometry," Anal. Chem., **2021**, 93 (4), 2174-2182.*

4.1 Abstract

Stable isotope tracers are applied in vivo and in vitro studies to reveal the activity of enzymes and intracellular metabolic pathways. Most often, such tracers are used with gas chromatography coupled to mass spectrometry (GC-MS) due to its ease of operation and reproducible mass spectral databases. Differences in isotope tracer performance of classic GC-quadrupole MS instrument and newer time-of-flight instruments are not well-studied. Here, we used three commercially available instruments for the analysis of identical samples from a stable isotope labeling study that used [U-$^{13}C_6$] d-glucose to investigate the metabolism of bacterium *Rothia mucilaginosa* with respect to 29 amino acids and hydroxyl acids involved in primary metabolism. The prokaryote *Rothia mucilaginosa* belongs to the family of *Micrococcaceae* and is present and metabolically active in the lungs and sputum of cystic fibrosis patients. Overall, all three GC-MS instruments (low-resolution GC-SQ MS, low-resolution GC-TOF MS, and high-resolution GC-QTOF MS) can be used to perform stable isotope tracing studies for glycolytic intermediates, TCA metabolites and amino acids, yielding similar biological results, with high-resolution GC-QTOF MS offering additional capabilities to identify chemical structures of unknown compounds that might show significant isotope enrichments in biological studies.

4.2 Introduction

Metabolomics has been widely used to study human diseases in the past 20 years.[1-7] While snapshot (steady-state) of metabolite profiles provides important information on cellular metabolism and physiology, such data do not necessarily reflect enzyme or pathway activity directly.[8] Instead, stable isotope tracing can be

used to track the fate of labeled atoms in metabolite substrates, quantify the rates of changes of these metabolites and elaborate the direction of pathway activities.[9-11] When isotope tracers are used along with absolute (molar) quantifications, fluxes across pathways can be discerned.[12-13] Isotope tracers can delineate the alternative use of substrates in uncontrolled cell growth in malignant tumors[14-16], determine the target pathways of novel drugs for disease therapy development,[17-19] investigate the metabolic cross-feeding between bacteria,[20-22] and assign/quantify the global plant metabolome that is riddled with heavy matrix interference.[23-24]

Mass spectrometry (MS) has played a central role in metabolomics and stable isotope tracing studies.[9] Gas chromatography with mass spectrometry (GC-MS) has been used for a long time for stable isotope labeling studies because many target compounds are directly amenable to GC-MS, including metabolites in the oxidative part of glycolysis, the TCA cycle and adjacent amino acids.[25-27] For many compounds including sugars and hydroxyl acids,[28-31] GC-MS offers an easier route to chromatographic separation than liquid chromatography (LC)-MS unless specific procedures like ion-pairing LC are used.[32] Small molecules analyzed by GC-MS require derivatization to decrease their polarity and to increase volatility, selectivity, and sensitivity, especially for metabolites bearing acidic protons.[30, 33] A typical silylation reagent used for derivatization in stable isotope tracing is *N*-Methyl-*N*-*tert*-butyldimethylsilyltrifluoroacetamide (MTBSTFA), which replaces non-sterically hindered active hydrogens (e.g., on -COOH, -OH, NH2, -HN, -SH groups) with *tert*-butyldimethylsilyl (TBDMS). MTBSTFA-derivatives produce characteristic fragmentation patterns with three major ions $[M]^+$, $[M-57]^+$, and $[M-131]^+$. The $[M-57]^+$ is generally the dominant ion and harbors the entire original metabolite[7, 34] and is therefore best suited for isotope tracer studies. In comparison, trimethylsilylation (TMS) reagents do not yield abundant molecular fragment ions[35] and are less stable and more sensitive to moisture. On one hand, due to steric hindrance, TBDMS derivatives only replace one active hydrogen per amino group, leading to better precision in quantitative analysis of amino-containing compounds than smaller TMS reagents[36-37] that generate several derivatives. On the other

hand, because of this effect, TBDMS-derivatization cannot derivatize sugars or sugar alcohols, limiting the use of TBDMS derivatization for general untargeted metabolomics compared to TMS reagents.

A wide range of GC-MS instruments are available varying in resolution and mass separation. However, comparison of the performance of different GC-MS instruments using stable isotope labeling technique is not well-studied. Isotope tracing studies investigate either isotopomer or isotopologue information of metabolites.[38-41] Isotopologue refers to molecules that differ only in their isotopic compositions. Isotopomers also consider the constitutional differences of isotopes within molecule structures.[42] Isotope enrichment studies are not often conducted together with absolute quantifications[43-44] which were used for the determination of absolute fluxes.[45] To date, GC-MS are widely used for isotopologue tracer studies for which we here compared on three types of GC-MS instruments.

We have applied [U-$^{13}C_6$] d-glucose to investigate the metabolism of *Rothia mucilaginosa* and reported the incorporation of glucose derived ^{13}C into glycolysis metabolites and some amino acid biosynthesis pathways using a low-resolution GC-single quadrupole-MS (GC-SQ MS).[20] Here, we investigate the same samples but used three different GC-MS instruments (**Figure 4.1, Figure S4.1**), with either low-resolution single quadrupole MS (GC-SQ MS), low-resolution time-of-flight MS (GC-TOF MS) or high-resolution GC-Quadrupole Time-of-Flight MS (GC-QTOF MS).

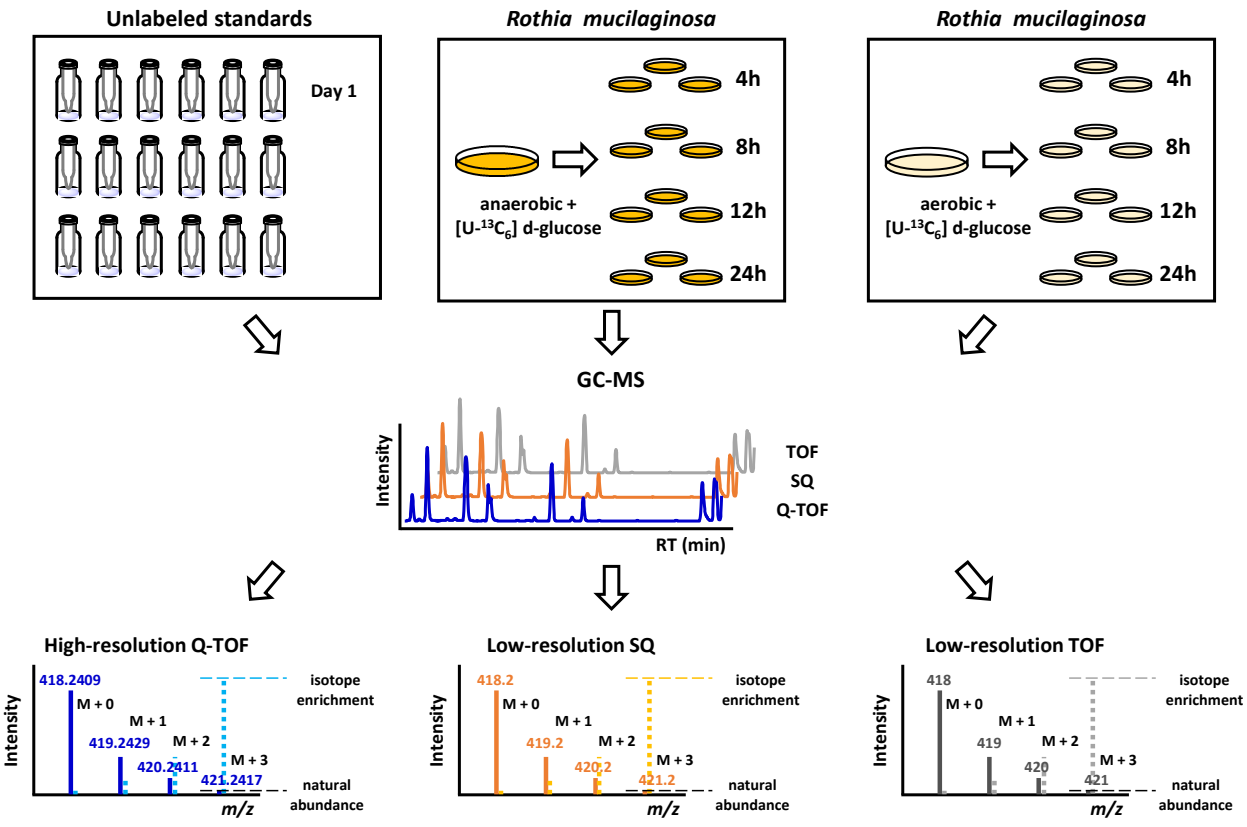


Figure 4.1. Evaluating the performance of isotope measurements for three GC-MS instruments using chemical standards and isotope tracer biological experiments.

4.3 Methods

4.3.1 Samples

A quality control mixture of 29 unlabeled metabolite standards (**Table S4.1**) was prepared to reach a final concentration of 1 mg/mL as stock solution. 50 µL aliquot from each stock solution was combined into a new tube and dried down. Then, the mixture was diluted to reach a final concentration of 50 µg/mL working solution.

Rothia mucilaginosa strain RmFLR01 was isolated from a cystic fibrosis (CF) patient at the UC San Diego Adult CF Clinic.[20, 46] *R. mucilaginosa* cultures were grown in triplicates in artificial-sputum medium[47] spiked with 100 mM [U-$^{13}C_6$] d-glucose (Cambridge Isotope Laboratory, Tewksbury, MA, USA) under

anaerobic and aerobic conditions (5% CO_2) at 37°C and harvested at 4, 8, 12, and 24 h for isotope tracer analyses.

4.3.2 Sample preparation

For the 29 unlabeled standards mixture, 10 µL of the final solution was dried down for GC-MS measurement. Derivatization and data acquisition of mixture aliquots by GC-MS were reproduced on three days for inter-day precision. Dried mixtures were derivatized by adding 10 µL of 40 mg/mL methoxyamine hydrochloride (Sigma-Aldrich, St. Louis, MO, USA) in pyridine (Sigma-Aldrich, St. Louis, MO, USA) and shaking at 30°C for 1.5 h. Subsequently, 90 µL MTBSTFA (Sigma-Aldrich, St. Louis, MO, USA) was added with 13 fatty acid methyl esters (FAMEs) as retention index markers and shaken at 80°C for 30 min. Samples were immediately transferred to crimp top vials and injected onto each GC-MS instrument.

Same samples of *R. mucilaginosa* cultures were extracted using published methods.[20] The extraction solvent was prepared by mixing 3 volume of acetonitrile (Fisher Scientific), 3 volume of isopropanol (Fisher Scientific), and 3 volume of water (Fisher Scientific) together and was then degassed by directing a gentle stream of nitrogen through the solvent for 5 min. The solvent was cooled to −20°C prior to extraction.[36, 48] Samples were added 1 mL pre-chilled, degassed acetonitrile: isopropanol: water (v/v/v 3:3:2, Fisher Scientific) followed by vortexing 30 s and shaking at 4°C for 5 min. Samples were centrifuged for 2 min at 12,210 × g to precipitate debris from extracts. Supernatants were collected and split into two equal portions. One aliquot was dried to completeness in a Labconco cold trap centrifuge evaporator and then resuspended in 0.5 mL degassed acetonitrile: water (v/v 1:1, Fisher Scientific) to remove triacylglycerides. Resuspension solutions were vortexed for 30s and centrifuged for 2 min. Supernatants were transferred into clean Eppendorf tubes and dried down completely. Dried extracts were derivatized as given above.

4.3.3 Gas chromatographic conditions

Each mass spectrometer was coupled to an Agilent 7890 GC system (Santa Clara, CA) installed with a Restek (Bellefonte, PA) RTX-5Sil MS column (30m length, 0.25 mm i.d, 0.25 µM d_f, 95%

dimethyl/5%diphenyl polysiloxane film) with an additional 10m guard column. For the quadrupole-MS and QTOF MS analyses, identical GC parameters were used by injecting 1 µL of derivatized sample into the GC in splitless mode at an injection temperature of 250°C and a constant flow of 1 mL/min. The initial oven temperature was held at 60°C for 0.5 min, and ramped at a rate of 10°C/min to 325°C that was maintained for 10 min for a total run time of 37 min.

For the low-resolution LECO TOF MS (St. Joseph, MI, USA) instrument, 1 µL of derivatized sample was injected in splitless mode with an injection temperature of 275°C. The oven temperature started at 50°C for 1 min, increased at 20°C/min to 330°C, and kept isothermal for 5 min.

Each standards mixture sample was injected 6 times to obtain intra-day precision data for each instrument. The comparison of these samples measured on three separate days gave data for inter-day precisions.

Each of the *R. mucilaginosa* biological triplicate samples was injected 10 times on each of the three GC-MS instruments to discriminate true biological isotope enrichment effects from the impact of technical imprecision.

4.3.4 MS analysis

We compared the high-resolution Agilent 7200 QTOF (Santa Clara, CA, USA), the low-resolution Agilent 5977 MSD (Santa Clara, CA, USA), and the low resolution LECO Pegasus IV TOF MS (St. Joseph, MI, USA) in this study. All three instruments used electron ionization (EI) mode at +70 eV electron voltage. Mass spectra were acquired from 85─700 m/z at 17 Hz and 250°C source temperature on the Pegasus IV TOF MS, 50─1050 m/z at 1.5 Hz and 240°C source temperature on the quadrupole-MS and from 50─1050 *m/z* at 5 Hz and 230°C source temperature on the QTOF MS instrument.

4.3.5 Data processing

Because of the chromatography parameter differences on three instruments, retention times (RT) were initially determined by running each metabolite as chemical standard on the high-resolution GC-QTOF MS. During derivatization, FAMEs (fatty-acid methyl esters) with 8 to 30 fatty acyl carbon lengths were added

into MTBSTFA. These FAMEs were used as internal standards for retention index markers. This FAME marker set yields a stable set system with fixed differences in retention times.[49] Despite other differences between the instruments, all three GC-MS instruments used the same electron ionization (EI) mode at +70 eV electron voltage and the spectra were very comparable among three instruments. An example is given in **Figure S4.1**.

LECO TOF MS raw data files (.CDF format) were converted to MassHunter formats (.D format) using Agilent GCMS translator software. MassHunter Quantitative Analysis B.07.00 version was used to process the data that acquired from all three GC-MS instruments to avoid potential differences caused by different data processing software parameters. We used identical S/N and RT tolerance settings for quantifying isotope tracers on the GC-QTOF and the GC-TOF MS instruments. Threshold of S/N was set to peak height ≥ 1% of the largest peak. RT tolerance was set to ± 0.08 min. For the GC-quadrupole MS, the S/N was set to peak height ≥ 0.2%. All peak integrations were manually validated within the vendor software to exclude potential artifacts based on possible differences in software settings. Theoretical isotope ratios were calculated for $[M{-}57]^+$ ion of each derivatized structure of the 29 standards using an online isotope calculator.[50] As this study is not for absolute quantification and absolute flux calculations, we normalized over-enrichment of isotopes to the unlabeled M^0 monoisotope ion, i.e., the $[M\text{-}57]^+$ fragment. Hence, normalization was performed as ratio M+1/M (or M+2/M or M+3/M) from peak heights of extracted ion chromatograms. Averages and standard deviations as well as technical errors and isotope enrichments were calculated in Microsoft Excel by correcting for natural isotope abundances. The data sets are available in the Metabolomics Workbench data repository with accession numbers 2231(ST001613) for low-resolution GC-TOF MS and 2278 (ST001608) for the other two GC-MS instruments.

4.4 Results and Discussion

4.4.1 Precision of isotope abundance measurements for three GC-MS instruments using 29 unlabeled metabolite standards

MTBSTFA derivatization replaces active hydrogens on polar functional groups and forms *tert*-butyldimethylsilyl (TBDMS) derivatives which are fragmented during electron impact ionization to generate the most abundant ion $[M-57]^+$ with *tert*-butyl cleavage (**Figure 4.2A**). TBDMS derivatization therefore has the advantage to yield a very abundant molecular fragment ion which contains the entire underivatized molecular skeleton with only one derivative product per metabolite. In addition, it had been previously shown that TBDMS derivatization yields better reproducibility than trimethylsilyl derivatization for amino acids which are important target compounds to be analyzed in metabolic isotope tracer studies. Using silylation reactions has a big advantage for isotope abundance analyses because the Si-atom has a high natural abundance of 5.1% for the M+1 isotope and 3.4% for the M+2 isotope, leading to easily detectable signals in GC-MS analyses even for low abundant metabolites in biological studies.

We measured 29 unlabeled metabolite standards that are typically investigated in metabolic flux studies including glycolysis intermediates, TCA metabolites, and amino acids on three GC-MS instruments. We first evaluated the overall performance of three instruments by calculating relative and absolute errors for the three most important isotope ratios (M+1/M, M+2/M and M+3/M). Our results showed that the low-resolution GC-TOF MS instrument yielded a significant systematic overrepresentation of the average isotope ratios for M+1/M, M+2/M, and M+3/M (**Figure 4.2C**) compared to GC-QTOF MS and GC-SQ MS instruments. The GC-TOF MS instrument also showed greater absolute errors (**Table 4.1**). These errors in the low-resolution GC-TOF MS instrument were clearly dependent on the molecular size of the chemicals (**Figure 4.2D**) as molecule $[M-57]^+$ fragments at < *m/z* 350 had isotopic ratios that were close to the theoretical values whereas at > *m/z* 350, systematic errors increased dramatically. These systematic errors are likely due to the way of this specific instrument accounting for ions by optimizing ion detection for speed, not for resolution, leading to non-optimal accounting for mass defects at higher *m/z* values. For

the single quadrupole, we found three outlier compounds, leucine and fumarate that showed detector saturation at $[M-57]^+$ monoisotopic ion peak heights, and mass spectral undersampling for pyruvate as very low abundant compound**.** The GC-QTOF MS showed the most accurate isotope ratios with only one compound (lysine) that had more than 5% absolute error for the M+1/M isotope ratio (**Figure 4.2D**). This excellent accuracy of the GC-QTOF MS (**Table 4.1**) might be due to the high mass resolving power that at least partly separates co-elution isobaric noise ions and therefore leads to better peak integration of low abundant isotope ions.

The mixture of 29 metabolite standards was injected six times each day to determine intra-day precision and reproduced for three days to test inter-day precision that were calculated as coefficient of variation (%CV). The low-resolution GC-TOF MS gave significantly better intra- and inter-day precisions than the other two instruments. For M+1/M (**Table 4.1**), GC-TOF MS instrument yielded only 1.1% CV on average for intra-day precision, compared to more than 2% CV for GC-quadrupole and GC-QTOF MS instruments. Similarly, for inter-day precision of M+1/M ratios, the low-resolution GC-TOF MS had best precision with 2.3% CV compared to 4─6% CV for the other two instruments. The same trends were observed for M+2/M with significantly better intra-day and inter-day precisions for the low-resolution GC-TOF MS instrument compared to the other two mass spectrometers. For M+3/M ratios, overall signal intensities were too low to obtain reliable data with 4─10% CV for intra-day and 5─13% CV for inter-day precisions. We verified the excellent precision but also the high absolute error for the low-resolution GC-TOF MS on four separate instruments (**Table S4.3**) for TMS-derivatized standards, analyzed on different days and by different personnel. The superior precision of the low-resolution GC-TOF MS instruments compared to the other two instruments was also verified by additional analysis of underivatized fatty acid methyl esters (**Table S4.4**). In combination, these findings indicate that isotope ratio measurements correspond to the physics of the ion optics and ion detection system of each instrument type, rather than to small differences between individual instruments of the same type or the type of chemical derivatizations.

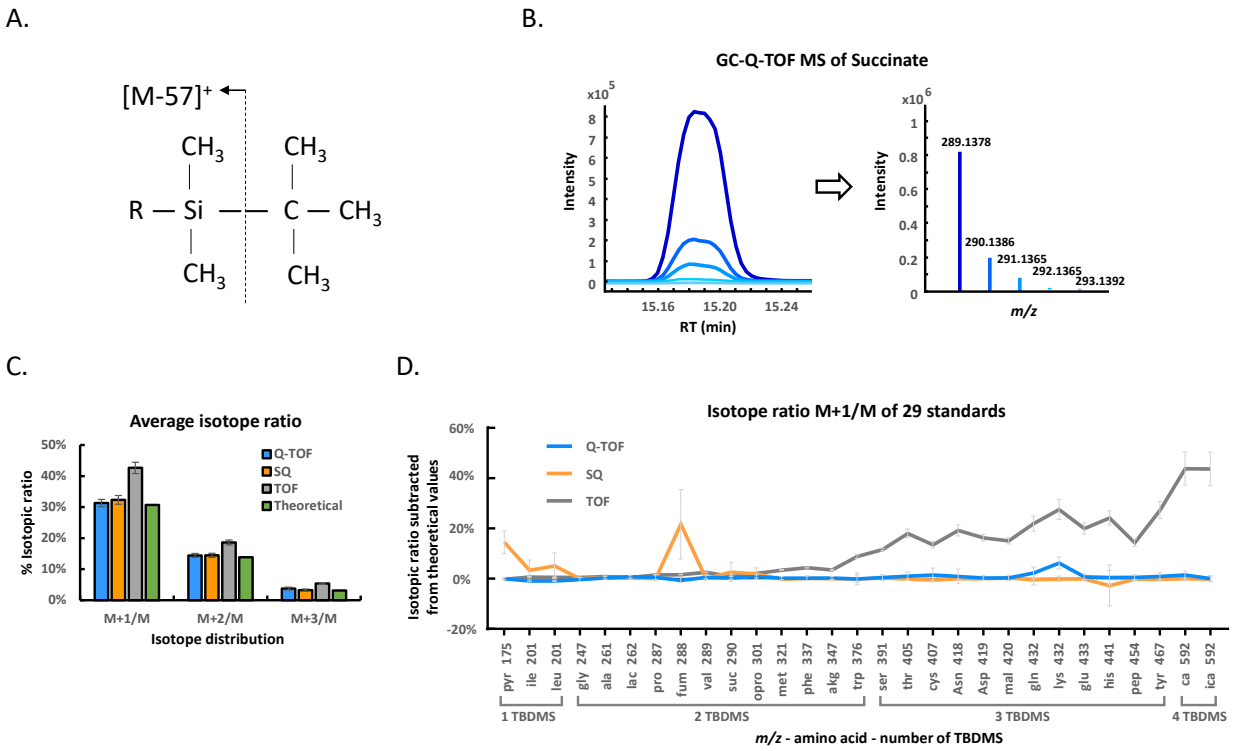


Figure 4.2. Overall performance of isotope detection using three instruments for 29 unlabeled standards. (A) MTBSTFA derivation leads to $[M-(tert\text{-butyl})]^+$ fragments that are used as M^+ surrogate ions. (B) Example of chromatographic ion traces and corresponding averaged mass spectrum for $[M-57]^+$ fragment of succinate.2TBDMS derivative at *m/z* 289 and its isotopologue intensities. (**C**) Average isotope ratios across all 29 metabolite standards and three GC-MS instruments. QTOF = high-resolution quadrupole time-of-flight MS, SQ = low-resolution single quadrupole MS, TOF = low-resolution time-of-flight MS. Isotope ratios are represented as arithmetic means ±1 σ. (**D**) Measured M+1/M ratios for 29 metabolites corrected by their theoretical abundances, sorted by number of TBDMS derivatizations and *m/z* values.

Table 4.1. Comparison of M+1/M and M+2/M isotope ratios of non-labeled standards from three instruments.
QTOF = high-resolution quadrupole time-of-flight MS, SQ = low-resolution single quadrupole MS, TOF = low-resolution time-of-flight MS.

	intra-day precision (%CV)			**inter-day precision (%CV)**			**absolute error (ratio %)**		
	QTOF	SQ	TOF	QTOF	SQ	TOF	QTOF	SQ	TOF
M+1/M	2.7%	3.3%	1.1%	3.7%	5.8%	2.3%	0.8%	3.1%	12.0%
M+2/M	3.4%	3.7%	1.5%	4.2%	6.1%	2.7%	0.6%	1.5%	4.7%

Interestingly, despite its excellent isotope accuracy, we did not find improved precision for the high-resolution QTOF MS instrument compared to classic unit-resolution quadrupole instrument. Instead, the better precision observed for the GC-TOF MS might be explained by its faster spectral acquisition rate of the. Using this instrument, spectra are acquired at 500 spectra/s and then accumulated to 17 spectra/s during initial raw data processing. In comparison, high-resolution GC-QTOF MS and classic quadrupole MS instruments are typically scanning at 2─5 spectra/s. At this spectral acquisition rate, data might be under-sampled for peaks that have band widths of 3 s (i.e. a total of 6-12 data points per peak) such as pyruvate measured on GC-SQ MS (**Figure 4.2D**), especially for lower-abundant M+1 and M+2 ions. Such undersampling of data points may certainly lead to lower precision.

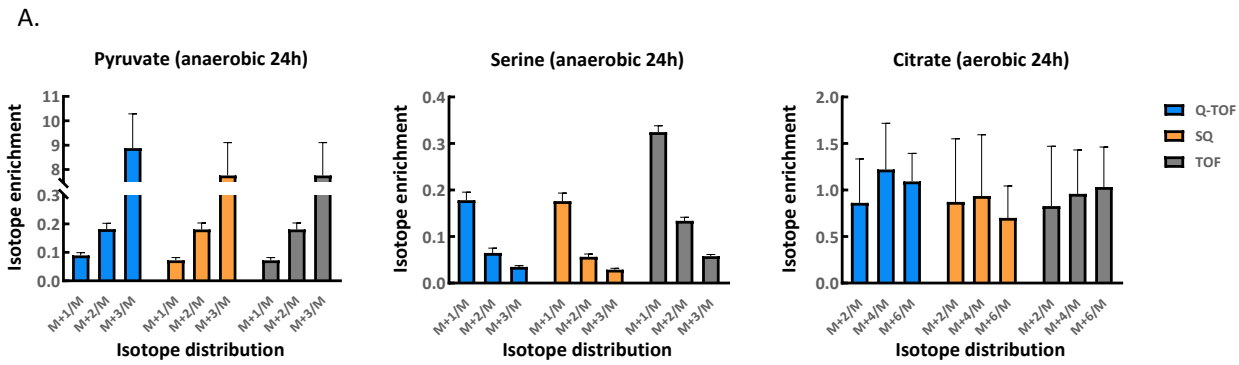


Figure 4.3. Isotope enrichment measured on three GC-MS instruments for selected metabolites in *R. mucilaginosa*. QTOF = high-resolution quadrupole time-of-flight MS, SQ = low-resolution single quadrupole MS, TOF = low-resolution time-of-flight MS. (3A─3C) Isotope enrichment of selected metabolites pyruvate, serine, and citrate in *R. mucilaginosa* grown for 24h under anaerobic or aerobic conditions depending on different $[M─57]^+$ fragment size. Isotope enrichments were corrected for natural abundance and are represented as arithmetic means ±1 σ. Calculation of σ was from 30 injections with a combination of both technical variation of 10 replicates and biological variation of 3 biological sample replicates.

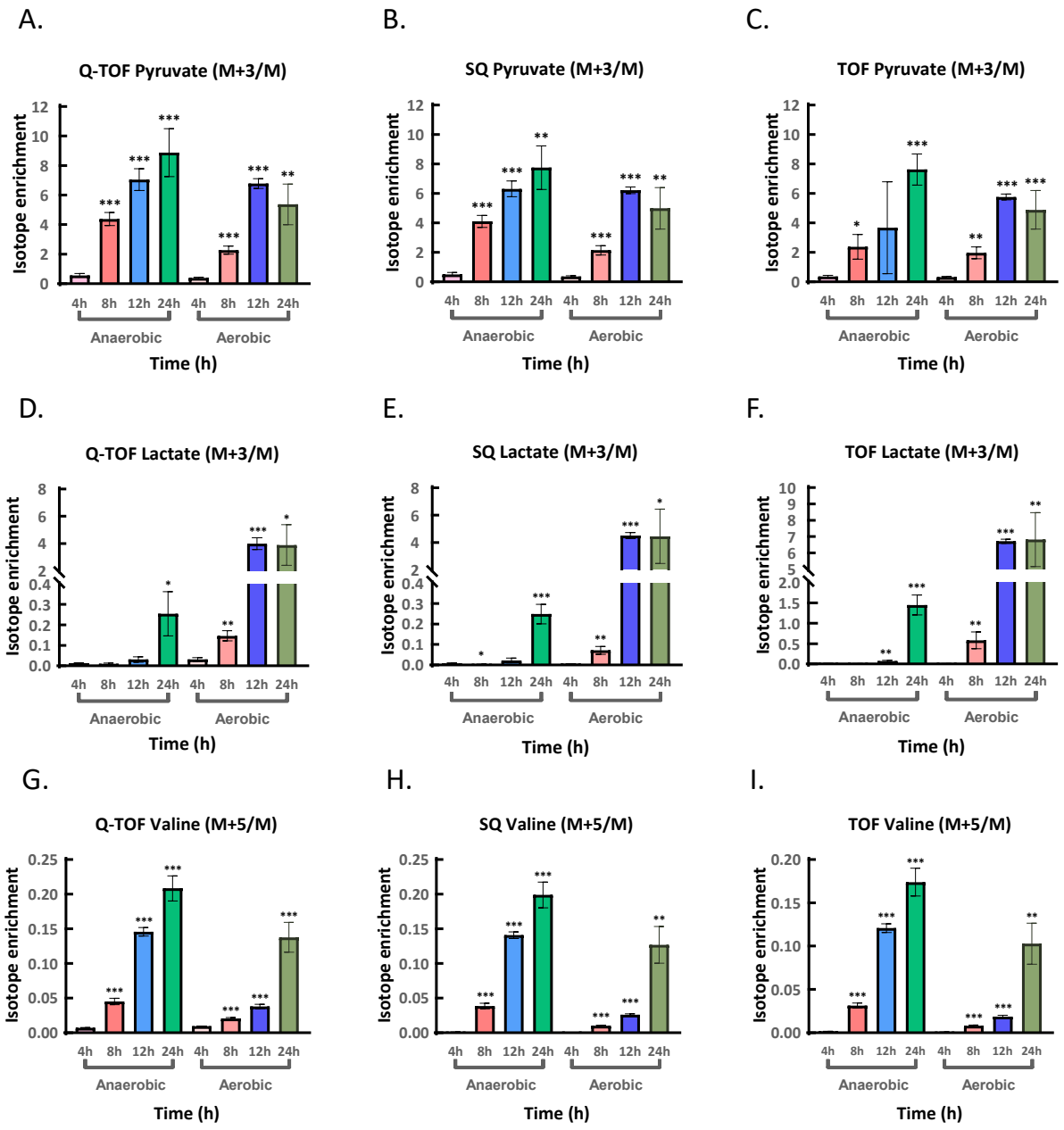


Figure 4.4. Increase in isotope enrichments measured on three GC-MS instruments for pyruvate, lactate and valine in *R. mucilaginosa.* QTOF = high-resolution quadrupole time-of-flight MS, SQ = low-resolution single quadrupole MS, TOF = low-resolution time-of-flight MS. Student *t*-test was used for assessing the significance levels by three biological replicates for enrichment at each time point compared to 4h time point in *R. mucilaginosa* grown for 4h─24h under anaerobic or aerobic conditions. **P* < 0.05, ***P* < 0.01, ****P* < 0.001 with values represented as arithmetic means ±1 σ. (4A─4C) Increase in isotope enrichments for pyruvate M+3/M. (4D─4F) Increase in isotope enrichments for lactate M+3/M. (4G─4I) Increase in isotope enrichments for valine M+5/M.

4.4.2 Performance of three GC-MS instruments on ^{13}C labeled *R. mucilaginosa* samples

Isotope tracing in biochemistry can highlight active enzymes and pathways that may differ between biological conditions. We therefore tested the performance of three GC-MS instruments to determine whether each of the instruments would yield the same biological conclusions. We here used a [U-$^{13}C_6$] d-glucose feeding experiment for which we previously published the overall biological insights using the GC-quadrupole-MS system,[20] and reproduced the study three times to compare GC-SQ MS to the GC-TOF MS and the GC-QTOF MS systems.

Rothia mucilaginosa (formerly *Stomatococcus mucilaginosus*) is a gram-positive coccus of the family *Micrococcaceae* and was found as a cause of oral, cutaneous, and central nervous system infections. All 29 metabolites were detected in *R. mucilaginosa* cultures grown under anaerobic and aerobic conditions from 4─24 hours. Three independent cultures per time point were harvested and measured repeatedly for ten times. Isotope enrichment of three primary metabolites, pyruvate, serine and citrate all showed clear incorporation of labeled carbon into the metabolite structures at 24h (**Figure 4.3A─C**). Isotope enrichment analysis showed that the 3-carbon molecule pyruvate significantly enriched its M+3 isotope (**Figure 4.3A**) with 8─9 folds higher abundance than the unlabeled M^+ monoisotopic parent molecule. Incomplete incorporation by one or two labeled carbons was much less pronounced for pyruvate. In comparison, the 3-carbon molecule serine (**Figure 4.3B**) showed only modest, but notable incorporation of labeled carbon into the M+1, M+2 and M+3 isotopologues. For the 6-carbon molecule citrate (under aerobic condition, **Figure 4.3C**) that is formed by addition of acetyl-groups using citrate synthase, we investigated the M+2, M+4 and M+6 isotopologues and also found small but detectable labeling. Citrate under aerobic conditions were still detected at low abundance, thus the isotope enrichment among three instruments were slightly different but gave overall similar ^{13}C incorporation patterns. These comparisons showed that all three GC-MS instruments correctly detected a large metabolic flux into the endpoint of glycolysis, pyruvate, with only small flux into off-stream biosynthesis of serine under anaerobic conditions. The instruments also all found an overall limited metabolic flux from pyruvate into the TCA cycle. Importantly, pyruvate did not

show a large difference between the GC-SQ and the other two instruments (**Figure 4.3A**), unlike for the unlabeled mixture experiment (**Figure 4.2D**). Moreover, despite the differences in reported isotope enrichment for each of the metabolites on the three GC-MS instruments, the biological conclusions would be the same, with high degree of ^{13}C incorporation in pyruvate but low ^{13}C enrichment for citrate and serine (**Figure 4.3A—C**). This biological conclusion would be even clearer if raw measurements were to be corrected for the systematic errors observed in the GC-TOF MS instrument by a post-hoc mass-dependent regression, which can be further explored in future studies.

The growth and metabolism of bacteria are sensitive to oxygen conditions. Thus, we examined the impact of oxygen condition on the metabolism of R. mucilaginosa and studied whether biological conclusions would differ between the three GC-MS instruments with respect to isotope enrichment over the course of one day under both aerobic and anaerobic conditions, correcting for natural isotope abundances (**Figure 4.4A—I**). The increase pattern in fully labeled pyruvate M+3 was similar for all three GC-MS instruments (**Figure 4.4A—C**) as well as for amino acids, citrate and other TCA cycle metabolites. All three instruments showed a decrease of fully labeled pyruvate M+3 under aerobic conditions at 24h compared to 12h, but not for anaerobic conditions (**Figure 4.4A—C**). Interestingly, all three instruments found much higher incorporation of labeled carbon into lactate under aerobic conditions than under anaerobic media between 8—24h growth (**Figure 4.4D—F**), reaching a plateau at 12h. Under anaerobic conditions, all three instruments showed only enrichment for lactate only at 24h (**Figure 4.4D—F**). For amino acid metabolite valine, greater isotope enrichment was observed under anaerobic conditions compared to aerobic growth conditions (**Figure 4.4G—I**). This indicated the incorporation rates of glucose derived 13C into valine were faster under anaerobic condition than aerobic environment. Similar biological results were observed by all three types of GC-MS instruments for tryptophan that showed clearly higher enrichment for tryptophan M+3 under aerobic conditions than anaerobic conditions (**Figure S4.2A—C**). Only when absolute ion intensities were too low, for example for phosphoenolpyruvate (**Figure S4.2D—F**), differences between the instruments might interfere with detection of biological differences. Specifically, temporal differences in

isotope enrichment in PEP from 4─24h growth were found only when using the quadrupole and QTOF instruments, but not the TOF instrument (**Figure S4.2D─F**). While we have not tested for biological ^{15}N incorporation,[51] similar performance of three GC-MS instruments as in ^{13}C labeling isotope enrichment analyses can be expected because such isotopologue enrichment analyses only consider the differences between full mass units, not the isotopic fine structures.[38, 52]

4.5 Conclusion

Our results show that the three tested types of GC-MS instruments were all capable of performing ^{13}C-based stable isotope ratio enrichments studies for primary metabolites. While the GC-QTOF MS showed better accuracy than the other two instruments, the low-resolution GC-TOF MS was superior in precision of ^{13}C labeled isotopologue analysis. It is important to note that none of these instruments enables isotopomer (^{13}C positional) analyses or studies combining ^{13}C, ^{18}O, and ^{15}N labeled substrates. Instruments to enable such studies for discerning the fine isotope structure of the labeled isotopes must have resolving power with $R > 250{,}000$.[38, 52] The nominal-mass GC-TOF MS showed systematic overrepresentation of isotope ratios for high *m/z* metabolites while the nominal-mass GC-single quadrupole yielded saturation effects at high metabolite concentrations. Nevertheless, when used in an example biological study on ^{13}C-glucose fed *Rothia* bacteria, all three GC-MS instruments yielded similar biological interpretations. While the low-resolution instruments are cheaper and easier to operate and maintain than high-resolution GC-MS instruments, accurate mass GC-QTOF MS instruments offer additional capabilities to identify chemical structures of unknown compounds.[53-54] This study exceeded the number of metabolites analyzed in the prior biology-focused study that only used GC-quadrupole MS instrumentation.[20] Interestingly, the inclusion of such a wider range of metabolic targets discovered differences in metabolic ^{13}C-enrichment for low abundant compounds that were not reported in the prior study.[20] For example, stark differences in isotope enrichment for low abundant tryptophan were observed with all three instruments between aerobic and anaerobic cultures. In addition, temporal ^{13}C-enrichment differences in some low abundant metabolites such as phosphoenolpyruvate were only detected by GC-QTOF MS and GC-SQ MS instruments, but not

by low-resolution GC-TOF mass spectrometry. Hence, this study shows how the number of metabolites in ^{13}C-enrichment analysis can be easily enlarged, offering the potential to extend the use of GC-MS to untargeted metabolome enrichment studies. If the aim of using ^{13}C-enrichment is to report absolute fluxes in biological studies, multiple time points would need to be analyzed in combination with absolute quantifications. Overall, we conclude that different GC-MS instruments showed to be highly useful for isotopic enrichment analysis in metabolomic studies.

Author Contributions

The experiment was conducted by Y. Z., B. G., L. V., and T. G. under the biological design of K.W. Data were analyzed and the manuscript was written by Y. Z., B. G., and O. F. All authors have given approval to the final version of the manuscript.

Acknowledge

This work was supported as a pilot project by NIH U24 DK097154 and NIH U2C ES030158. K.L.W. is supported by a Gilead CF Research Scholars Award (app_00b072). T.G. is supported through the National Science Foundation's Integrative Graduate Education and Research Traineeship (IGERT) program (grant DGE-1144901).

4.6 References

1. McCartney, A.; Vignoli, A.; Biganzoli, L.; Love, R.; Tenori, L.; Luchinat, C.; Di Leo, A., Metabolomics in breast cancer: A decade in review. *Cancer Treat Rev* **2018,** *67*, 88-96.

2. Yu, L.; Li, K.; Zhang, X., Next-generation metabolomics in lung cancer diagnosis, treatment and precision medicine: mini review. *Oncotarget* **2017,** *8* (70), 115774-115786.

3. Guasch-Ferre, M.; Hruby, A.; Toledo, E.; Clish, C. B.; Martinez-Gonzalez, M. A.; Salas-Salvado, J.; Hu, F. B., Metabolomics in Prediabetes and Diabetes: A Systematic Review and Meta-analysis. *Diabetes Care* **2016,** *39* (5), 833-46.

4. Rauschert, S.; Uhl, O.; Koletzko, B.; Hellmuth, C., Metabolomic biomarkers for obesity in humans: a short review. *Ann Nutr Metab* **2014,** *64* (3-4), 314-24.

5. Jacob, M.; Lopata, A. L.; Dasouki, M.; Abdel Rahman, A. M., Metabolomics toward personalized medicine. *Mass Spectrom Rev* **2019,** *38* (3), 221-238.

6. Misra, B. B.; Olivier, M., High Resolution GC-Orbitrap-MS Metabolomics Using Both Electron Ionization and Chemical Ionization for Analysis of Human Plasma. *J Proteome Res* **2020,** *19* (7), 2717-2731.

7. Higashi, R. M.; Fan, T. W.; Lorkiewicz, P. K.; Moseley, H. N.; Lane, A. N., Stable isotope-labeled tracers for metabolic pathway elucidation by GC-MS and FT-MS. *Methods Mol Biol* **2014,** *1198*, 147-67.

8. Zamboni, N.; Saghatelian, A.; Patti, G. J., Defining the metabolome: size, flux, and regulation. *Mol Cell* **2015,** *58* (4), 699-706.

9. Jang, C.; Chen, L.; Rabinowitz, J. D., Metabolomics and Isotope Tracing. *Cell* **2018,** *173* (4), 822-837.

10. Wilkinson, D. J., Historical and contemporary stable isotope tracer approaches to studying mammalian protein metabolism. *Mass Spectrom Rev* **2018,** *37* (1), 57-80.

11. Heuillet, M.; Bellvert, F.; Cahoreau, E.; Letisse, F.; Millard, P.; Portais, J. C., Methodology for the Validation of Isotopic Analyses by Mass Spectrometry in Stable-Isotope Labeling Experiments. *Anal Chem* **2018,** *90* (3), 1852-1860.

12. Liu, L.; Shah, S.; Fan, J.; Park, J. O.; Wellen, K. E.; Rabinowitz, J. D., Malic enzyme tracers reveal hypoxia-induced switch in adipocyte NADPH pathway usage. *Nat Chem Biol* **2016,** *12* (5), 345-52.

13. Wang, P.; Guo, L.; Jaini, R.; Klempien, A.; McCoy, R. M.; Morgan, J. A.; Dudareva, N.; Chapple, C., A (13)C isotope labeling method for the measurement of lignin metabolic flux in Arabidopsis stems. *Plant Methods* **2018,** *14*, 51.

14. Courtney, K. D.; Bezwada, D.; Mashimo, T.; Pichumani, K.; Vemireddy, V.; Funk, A. M.; Wimberly, J.; McNeil, S. S.; Kapur, P.; Lotan, Y.; Margulis, V.; Cadeddu, J. A.; Pedrosa, I.; DeBerardinis, R. J.; Malloy, C. R.; Bachoo, R. M.; Maher, E. A., Isotope Tracing of Human Clear Cell Renal Cell Carcinomas Demonstrates Suppressed Glucose Oxidation In Vivo. *Cell Metab* **2018,** *28* (5), 793-800 e2.

15. Roci, I.; Gallart-Ayala, H.; Schmidt, A.; Watrous, J.; Jain, M.; Wheelock, C. E.; Nilsson, R., Metabolite Profiling and Stable Isotope Tracing in Sorted Subpopulations of Mammalian Cells. *Anal Chem* **2016,** *88* (5), 2707-13.

16. Fan, T. W.; Lane, A. N.; Higashi, R. M.; Farag, M. A.; Gao, H.; Bousamra, M.; Miller, D. M., Altered regulation of metabolic pathways in human lung cancer discerned by (13)C stable isotope-resolved metabolomics (SIRM). *Mol Cancer* **2009,** *8*, 41.

17. Fan, T. W.; Lorkiewicz, P. K.; Sellers, K.; Moseley, H. N.; Higashi, R. M.; Lane, A. N., Stable isotope-resolved metabolomics and applications for drug development. *Pharmacol Ther* **2012,** *133* (3), 366-91.

18. Lane, A. N.; Higashi, R. M.; Fan, T. W., Preclinical models for interrogating drug action in human cancers using Stable Isotope Resolved Metabolomics (SIRM). *Metabolomics* **2016,** *12* (7).

19. Yuan, M.; Kremer, D. M.; Huang, H.; Breitkopf, S. B.; Ben-Sahra, I.; Manning, B. D.; Lyssiotis, C. A.; Asara, J. M., Ex vivo and in vivo stable isotope labelling of central carbon metabolism and related pathways with analysis by LC-MS/MS. *Nat Protoc* **2019,** *14* (2), 313-330.

20. Gao, B.; Gallagher, T.; Zhang, Y.; Elbadawi-Sidhu, M.; Lai, Z.; Fiehn, O.; Whiteson, K. L., Tracking Polymicrobial Metabolism in Cystic Fibrosis Airways: Pseudomonas aeruginosa Metabolism and Physiology Are Influenced by Rothia mucilaginosa-Derived Metabolites. *mSphere* **2018,** *3* (2).

21. Yuan, B. F.; Zhu, Q. F.; Guo, N.; Zheng, S. J.; Wang, Y. L.; Wang, J.; Xu, J.; Liu, S. J.; He, K.; Hu, T.; Zheng, Y. W.; Xu, F. Q.; Feng, Y. Q., Comprehensive Profiling of Fecal Metabolome of Mice by Integrated Chemical Isotope Labeling-Mass Spectrometry Analysis. *Anal Chem* **2018,** *90* (5), 3512-3520.

22. Hartl, J.; Kiefer, P.; Kaczmarczyk, A.; Mittelviefhaus, M.; Meyer, F.; Vonderach, T.; Hattendorf, B.; Jenal, U.; Vorholt, J. A., Untargeted metabolomics links glutathione to bacterial cell cycle progression. *Nat Metab* **2020,** *2* (2), 153-166.

23. Freund, D. M.; Hegeman, A. D., Recent advances in stable isotope-enabled mass spectrometry-based plant metabolomics. *Curr Opin Biotechnol* **2017,** *43*, 41-48.

24. Doppler, M.; Bueschl, C.; Kluger, B.; Koutnik, A.; Lemmens, M.; Buerstmayr, H.; Rechthaler, J.; Krska, R.; Adam, G.; Schuhmacher, R., Stable Isotope-Assisted Plant Metabolomics: Combination of Global and Tracer-Based Labeling for Enhanced Untargeted Profiling and Compound Annotation. *Front Plant Sci* **2019,** *10*, 1366.

25. Wittmann, C.; Heinzle, E., Modeling and experimental design for metabolic flux analysis of lysine-producing Corynebacteria by mass spectrometry. *Metab Eng* **2001,** *3* (2), 173-91.

26. Fischer, E.; Zamboni, N.; Sauer, U., High-throughput metabolic flux analysis based on gas chromatography-mass spectrometry derived 13C constraints. *Anal Biochem* **2004,** *325* (2), 308-16.

27. Beale, D. J.; Pinu, F. R.; Kouremenos, K. A.; Poojary, M. M.; Narayana, V. K.; Boughton, B. A.; Kanojia, K.; Dayalan, S.; Jones, O. A. H.; Dias, D. A., Review of recent developments in GC-MS approaches to metabolomics-based research. *Metabolomics* **2018,** *14* (11), 152.

28. Boets, E.; Gomand, S. V.; Deroover, L.; Preston, T.; Vermeulen, K.; De Preter, V.; Hamer, H. M.; Van den Mooter, G.; De Vuyst, L.; Courtin, C. M.; Annaert, P.; Delcour, J. A.; Verbeke, K. A., Systemic availability and metabolism of colonic-derived short-chain fatty acids in healthy subjects: a stable isotope study. *J Physiol* **2017,** *595* (2), 541-555.

29. Lima, V. F.; de Souza, L. P.; Williams, T. C. R.; Fernie, A. R.; Daloso, D. M., Gas Chromatography-Mass Spectrometry-Based (13)C-Labeling Studies in Plant Metabolomics. *Methods Mol Biol* **2018,** *1778*, 47-58.

30. Fiehn, O., Metabolomics by Gas Chromatography-Mass Spectrometry: Combined Targeted and Untargeted Profiling. *Curr Protoc Mol Biol* **2016,** *114*, 30 4 1-30 4 32.

31. McConnell, B. O.; Antoniewicz, M. R., Measuring the Composition and Stable-Isotope Labeling of Algal Biomass Carbohydrates via Gas Chromatography/Mass Spectrometry. *Anal Chem* **2016,** *88* (9), 4624-8.

32. Meissen, J. K.; Pirman, D. A.; Wan, M.; Miller, E.; Jatkar, A.; Miller, R.; Steenwyk, R. C.; Blatnik, M., Phenotyping hepatocellular metabolism using uniformly labeled carbon-13 molecular probes and LC-HRMS stable isotope tracing. *Anal Biochem* **2016,** *508*, 129-37.

33. Lai, Z.; Fiehn, O., Mass spectral fragmentation of trimethylsilylated small molecules. *Mass Spectrom Rev* **2018,** *37* (3), 245-257.

34. Schummer, C.; Delhomme, O.; Appenzeller, B. M.; Wennig, R.; Millet, M., Comparison of MTBSTFA and BSTFA in derivatization reactions of polar compounds prior to GC/MS analysis. *Talanta* **2009,** *77* (4), 1473-82.

35. Pietrogrande, M. C.; Bacco, D.; Mercuriali, M., GC-MS analysis of low-molecular-weight dicarboxylic acids in atmospheric aerosol: comparison between silylation and esterification derivatization procedures. *Anal Bioanal Chem* **2010,** *396* (2), 877-85.

36. Fiehn, O.; Wohlgemuth, G.; Scholz, M.; Kind, T.; Lee, D. Y.; Lu, Y.; Moon, S.; Nikolau, B., Quality control for plant metabolomics: reporting MSI-compliant studies. *Plant J* **2008,** *53* (4), 691-704.

37. Kind, T.; Wohlgemuth, G.; Lee, D. Y.; Lu, Y.; Palazoglu, M.; Shahbaz, S.; Fiehn, O., FiehnLib: mass spectral and retention index libraries for metabolomics based on quadrupole and time-of-flight gas chromatography/mass spectrometry. *Anal Chem* **2009,** *81* (24), 10038-48.

38. Rhoads, T. W.; Rose, C. M.; Bailey, D. J.; Riley, N. M.; Molden, R. C.; Nestler, A. J.; Merrill, A. E.; Smith, L. M.; Hebert, A. S.; Westphall, M. S.; Pagliarini, D. J.; Garcia, B. A.; Coon, J. J., Neutron-encoded mass signatures for quantitative top-down proteomics. *Anal Chem* **2014,** *86* (5), 2314-9.

39. Bequette, B. J.; Sunny, N. E.; El-Kadi, S. W.; Owens, S. L., Application of stable isotopes and mass isotopomer distribution analysis to the study of intermediary metabolism of nutrients. *J Anim Sci* **2006,** *84 Suppl*, E50-9.

40. Cline, G. W.; Lepine, R. L.; Papas, K. K.; Kibbey, R. G.; Shulman, G. I., 13C NMR isotopomer analysis of anaplerotic pathways in INS-1 cells. *J Biol Chem* **2004,** *279* (43), 44370-5.

41. Moseley, H. N.; Lane, A. N.; Belshoff, A. C.; Higashi, R. M.; Fan, T. W., A novel deconvolution method for modeling UDP-N-acetyl-D-glucosamine biosynthetic pathways based on (13)C mass isotopologue profiles under non-steady-state conditions. *BMC Biol* **2011,** *9*, 37.

42. Buescher, J. M.; Antoniewicz, M. R.; Boros, L. G.; Burgess, S. C.; Brunengraber, H.; Clish, C. B.; DeBerardinis, R. J.; Feron, O.; Frezza, C.; Ghesquiere, B.; Gottlieb, E.; Hiller, K.; Jones, R. G.; Kamphorst, J. J.; Kibbey, R. G.; Kimmelman, A. C.; Locasale, J. W.; Lunt, S. Y.; Maddocks, O. D.; Malloy, C.; Metallo, C. M.; Meuillet, E. J.; Munger, J.; Noh, K.; Rabinowitz, J. D.; Ralser, M.; Sauer, U.; Stephanopoulos, G.; St-Pierre, J.; Tennant, D. A.; Wittmann, C.; Vander Heiden, M. G.; Vazquez, A.; Vousden, K.; Young, J. D.; Zamboni, N.; Fendt, S. M., A roadmap for interpreting (13)C metabolite labeling patterns from cells. *Curr Opin Biotechnol* **2015,** *34*, 189-201.

43. Metallo, C. M.; Gameiro, P. A.; Bell, E. L.; Mattaini, K. R.; Yang, J.; Hiller, K.; Jewell, C. M.; Johnson, Z. R.; Irvine, D. J.; Guarente, L.; Kelleher, J. K.; Vander Heiden, M. G.; Iliopoulos, O.; Stephanopoulos, G., Reductive glutamine metabolism by IDH1 mediates lipogenesis under hypoxia. *Nature* **2011,** *481* (7381), 380-4.

44. Fan, J.; Kamphorst, J. J.; Rabinowitz, J. D.; Shlomi, T., Fatty acid labeling from glutamine in hypoxia can be explained by isotope exchange without net reductive isocitrate dehydrogenase (IDH) flux. *J Biol Chem* **2013,** *288* (43), 31363-9.

45. Zamboni, N.; Fendt, S. M.; Ruhl, M.; Sauer, U., (13)C-based metabolic flux analysis. *Nat Protoc* **2009,** *4* (6), 878-92.

46. Phan, J.; Meinardi, S.; Barletta, B.; Blake, D. R.; Whiteson, K., Stable isotope profiles reveal active production of VOCs from human-associated microbes. *J Breath Res* **2017,** *11* (1), 017101.

47. Quinn, R. A.; Whiteson, K.; Lim, Y. W.; Salamon, P.; Bailey, B.; Mienardi, S.; Sanchez, S. E.; Blake, D.; Conrad, D.; Rohwer, F., A Winogradsky-based culture system shows an association between microbial fermentation and cystic fibrosis exacerbation. *ISME J* **2015,** *9* (4), 1024-38.

48. Fiehn, O., Metabolite profiling in Arabidopsis. *Methods Mol Biol* **2006,** *323*, 439-47.

49. Skogerson, K.; Wohlgemuth, G.; Barupal, D. K.; Fiehn, O., The volatile compound BinBase mass spectral database. *BMC Bioinformatics* **2011,** *12*, 321.

50. Scientific Instrument Services. (2005, May 22). [Online]. Available: http://www.sisweb.com/referenc/tools/massdes.htm.

51. Lane, A. N.; Tan, J.; Wang, Y.; Yan, J.; Higashi, R. M.; Fan, T. W., Probing the metabolic phenotype of breast cancer cells by multiple tracer stable isotope resolved metabolomics. *Metab Eng* **2017,** *43* (Pt B), 125-136.

52. Frey, A. J.; Feldman, D. R.; Trefely, S.; Worth, A. J.; Basu, S. S.; Snyder, N. W., LC-quadrupole/Orbitrap high-resolution mass spectrometry enables stable isotope-resolved simultaneous quantification and (1)(3)C-isotopic labeling of acyl-coenzyme A thioesters. *Anal Bioanal Chem* **2016,** *408* (13), 3651-8.

53. Lai, Z.; Tsugawa, H.; Wohlgemuth, G.; Mehta, S.; Mueller, M.; Zheng, Y.; Ogiwara, A.; Meissen, J.; Showalter, M.; Takeuchi, K.; Kind, T.; Beal, P.; Arita, M.; Fiehn, O., Identifying metabolites by integrating metabolome databases with mass spectrometry cheminformatics. *Nat Methods* **2018,** *15* (1), 53-56.

54. Qiu, Y.; Moir, R.; Willis, I.; Beecher, C.; Tsai, Y. H.; Garrett, T. J.; Yost, R. A.; Kurland, I. J., Isotopic Ratio Outlier Analysis of the S. cerevisiae Metabolome Using Accurate Mass Gas

Chromatography/Time-of-Flight Mass Spectrometry: A New Method for Discovery. *Anal Chem* **2016,** *88* (5), 2747-54.

4.7 Supplementary Information

Table S4.1. Information or Standards name and provided vendor names.

Standards Name	CAS#	Catalog	Vendor	Abbreviation
Pyruvic acid	127-17-3	107360	Sigma-Aldrich	pyr
L-leucine	61-90-5	L8000	Sigma-Aldrich	leu
L-Isoleucine	73-32-5	I2752	Sigma-Aldrich	ile
Glycine	56-40-6	G7126	Sigma-Aldrich	gly
L-Alanine >98%	56-41-7	A7627	Sigma-Aldrich	ala
L(+)-Lactic acid	79-33-4	L1750	Sigma-Aldrich	lac
L-Proline	147-85-3	P0380	Sigma-Aldrich	pro
Fumaric acid	110-17-8	47910	Sigma-Aldrich	fum
L-Valine	72-18-4	V0500	Sigma-Aldrich	val
Succinic acid	110-15-6	398055	Sigma-Aldrich	suc
Oxoproline	98-79-3	83160	Sigma-Aldrich	opro
L-Methionine	63-68-3	M9625	Sigma-Aldrich	met
L-Phenylalanine	63-91-2	P2126	Sigma-Aldrich	phe
α-Ketoglutarate	328-50-7	75890	Sigma-Aldrich	akg
L-Tryptophan	73-22-3	T0254	Sigma-Aldrich	trp
D-Serine	312-84-5	S4250	Sigma-Aldrich	ser
L-Threonine	72-19-5	T8625	Sigma-Aldrich	thr
L-Cysteine	52-90-4	168149	Sigma-Aldrich	cys
L-Asparagine	70-47-3	A0884	Sigma-Aldrich	asn
L-Aspartic acid	56-84-8	A9256	Sigma-Aldrich	asp
L(-)-Malic acid	97-67-6	M1000	Sigma-Aldrich	mal
L-Glutamine	56-85-9	G3202	Sigma-Aldrich	gln
L-Lysine	56-87-1	L5501	Sigma-Aldrich	glu
L-Glutamic acid	56-86-0	G1251	Sigma-Aldrich	lys
L-Histidine	71-00-1	H8000	Sigma-Aldrich	his
D-Tyrosine	556-02-5	855456	Sigma-Aldrich	tyr
Phospho(enol)pyruvic acid monopotassium salt	4265-07-0	860077	Sigma-Aldrich	pep
Citric acid	77-92-9	251275	Sigma-Aldrich	ca
DL-Isocitric acid trisodium salt hydrate	1637-73-6	I1252	Sigma-Aldrich	ica

Table S4.2. InChiKeys, KEGG IDs, and PubChem CIDs for the 29 metabolites.

Metabolite Name	InChiKey	KEGG ID	PubChem CID
pyruvic acid	LCTONWCANYUPML-UHFFFAOYSA-N	C00022	1060
lactic acid	JVTAAEKCZFNVCJ-UHFFFAOYSA-N	C01432	612
alanine	QNAYBMKLOCPYGJ-REOHCLBHSA-N	C00041	5950
glycine	DHMQDGOQFOQNFH-UHFFFAOYSA-N	C00037	750
valine	KZSNJWFQEVHDMF-BYPYZUCNSA-N	C00183	6287
leucine	ROHFNLRQFUQHCH-YFKPBYRVSA-N	C00123	6106
isoleucine	AGPKZVBTJJNPAG-WHFBIAKZSA-N	C00407	6306
succinic acid	KDYFGRWQOYBRFD-UHFFFAOYSA-N	C00042	1110
proline	ONIBWKKTOPOVIA-BYPYZUCNSA-N	C00148	145742
fumaric acid	VZCYOOQTPOCHFL-OWOJBTEDSA-N	C00122	444972
oxoproline	ODHCTXKNWHHXJC-VKHMYHEASA-N	C01879	7405
methionine	FFEARJCKVFRZRR-BYPYZUCNSA-N	C00073	6137
serine	MTCFGRXMJLQNBG-REOHCLBHSA-N	C00065	5951
a-ketoglutaric acid	KPGXRSRHYNQIFN-UHFFFAOYSA-N	C00026	51
threonine	AYFVYJQAPQTCCC-GBXIJSLDSA-N	C00188	6288
phenylalanine	COLNVLDHVKWLRT-QMMMGPOBSA-N	C00079	6140
malic acid	BJEPYKJPYRNKOW-UHFFFAOYSA-N	C00711	525
aspartic acid	CKLJMWTZIZZHCS-REOHCLBHSA-N	C00049	5960
cysteine	XUJNEKJLAYXESH-REOHCLBHSA-N	C00097	5862
phosphoenolpyruvic acid	DTBNBXWJWCWCIK-UHFFFAOYSA-N	C00074	1005
glutamic acid	WHUUTDBJXJRKMK-VKHMYHEASA-N	C00025	33032
asparagine	DCXYFEDJOCDNAF-REOHCLBHSA-N	C00152	6267
lysine	KDXKERNSBIXSRK-YFKPBYRVSA-N	C00047	5962
glutamine	ZDXPYRJPNDTMRX-VKHMYHEASA-N	C00064	5961
histidine	HNDVDQJCIGZPNO-YFKPBYRVSA-N	C00135	6274
citric acid	KRKNYBCHXYNGOX-UHFFFAOYSA-N	C00158	311
isocitric acid	ODBLHEXUDAPZAU-ZAFYKAAXSA-N	C00311	1198
tyrosine	OUYCCCASQSFEME-QMMMGPOBSA-N	C00082	6057
tryptophan	QIVBCDIJIAJPQS-VIFPVBQESA-N	C00078	6305

Table S4.3. Comparison of M+1/M and M+2/M isotope ratios of 28 non-labeled QC mixture from four unit-resolution GC-TOF instruments.

Name	absolute error	within-instrument	A	B	C	D
M+1/M	2.8%	2.5%	1.7%	1.5%	3.1%	3.6%
M+2/M	4.3%	4.0%	4.9%	2.3%	3.5%	5.2%

Table S4.4. Comparison of M+1/M isotope ratios of non-labeled FAMES standards from three instruments. QTOF = high-resolution quadrupole time-of-flight MS, SQ = low-resolution single quadrupole MS, TOF = low-resolution time-of-flight MS

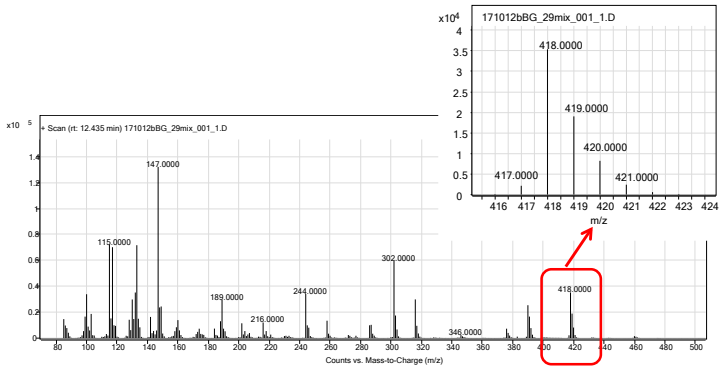
	M+1 / M	precision (% CV)	
FAMEs *m/z 87*	absolute error	inter-day	intra-day
SQ	4.4%	5.4%	2.8%
TOF	4.0%	3.7%	1.9%
QTOF	2.0%	6.4%	6.4%

We processed the data showed in **Table S4.4** using MassHunter Quantitative Analysis B07.00 and investigated the abundant *m/z* 87 fragment and its isotopes M+1, M+2 and M+3. *m/z* 87 is a typical ion fragment of FAMES. The results showed that all instrument showed a significant absolute error for this mass: the formula $C_4H_7O_2$ should have a ratio of M+1/M = 4.5%, but the measured ratios showed systematic errors ranging from measure values of 8.9% (quadrupole) to 6.5% (QTOF). These systematic errors lead to absolute errors of 2─4.5% and were larger than observed in our study for the *tert*-butyl-loss fragments $[M─57]^+$ of *tert*-butyldimethylsilylated metabolites. This effect may be due to the difficulties in peak integration for low abundant M+1 and M+2 peaks. This effect is less prominent in MTBSTFA derivatization because the silicon atom has significant abundance of its natural isotopes of M+1/M = 5.1% and M+2/M = 3.4%. For this reason, derivatized metabolites have higher total abundance in isotope ratio enrichment analyses in GC-MS chromatograms, and consequently, lower errors. In addition, analysis of *m/z* 87 cannot reflect the differences we observed for high *m/z* ions, specifically for the low-resolution GC-TOF MS instrument.

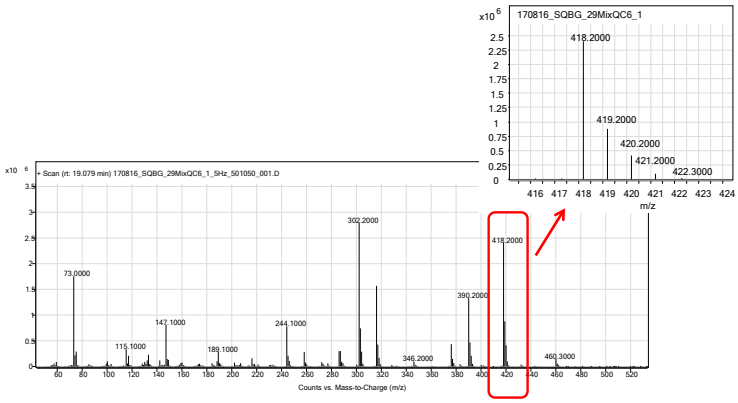
Figure S4.1. Full spectrum of aspartate 3TBS-derivatives for the three instruments.

The spectrum showed in **Figure 1** of the main manuscript was an example obtained by high-resolution GC-QTOF MS analyses. Only the relevant isotope data of the $[M-57]^+$ fragment were shown in **Figure 1**. Nominal *m/z* values obtained for identical compound are identical across instruments.

4.1A. Low-resolution GC-TOF MS: Aspartate



4.1B. Low-resolution GC-SQ MS: Aspartate



4.1C. High-resolution GC-QTOF MS: Aspartate

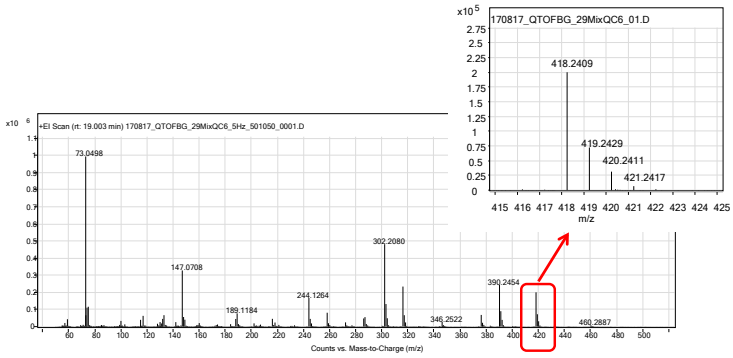


Figure S4.2. Increase in isotope enrichments measured on three GC-MS instruments for phosphoenolpyruvate (PEP) and tryptophan in *R. mucilaginosa*. (S4.2A-S4.2C) Isotope enrichment pattern for phosphoenolpyruvate, M+3/M. (S4.2D-S4.2F) Isotope enrichment pattern for tryptophan, M+3/M. QTOF = high-resolution quadrupole time-of-flight MS, SQ = low-resolution single quadrupole MS, TOF = low-resolution time-of-flight MS. Student *t*-test for three biological replicates was used for assessing significance levels for enrichment at each time point compared to 4h time point in *R. mucilaginosa* grown for 4h-24h under anaerobic or aerobic conditions. $*P < 0.05$, $**P < 0.01$, $***P < 0.001$, values were represented as mean ±1 S.D.

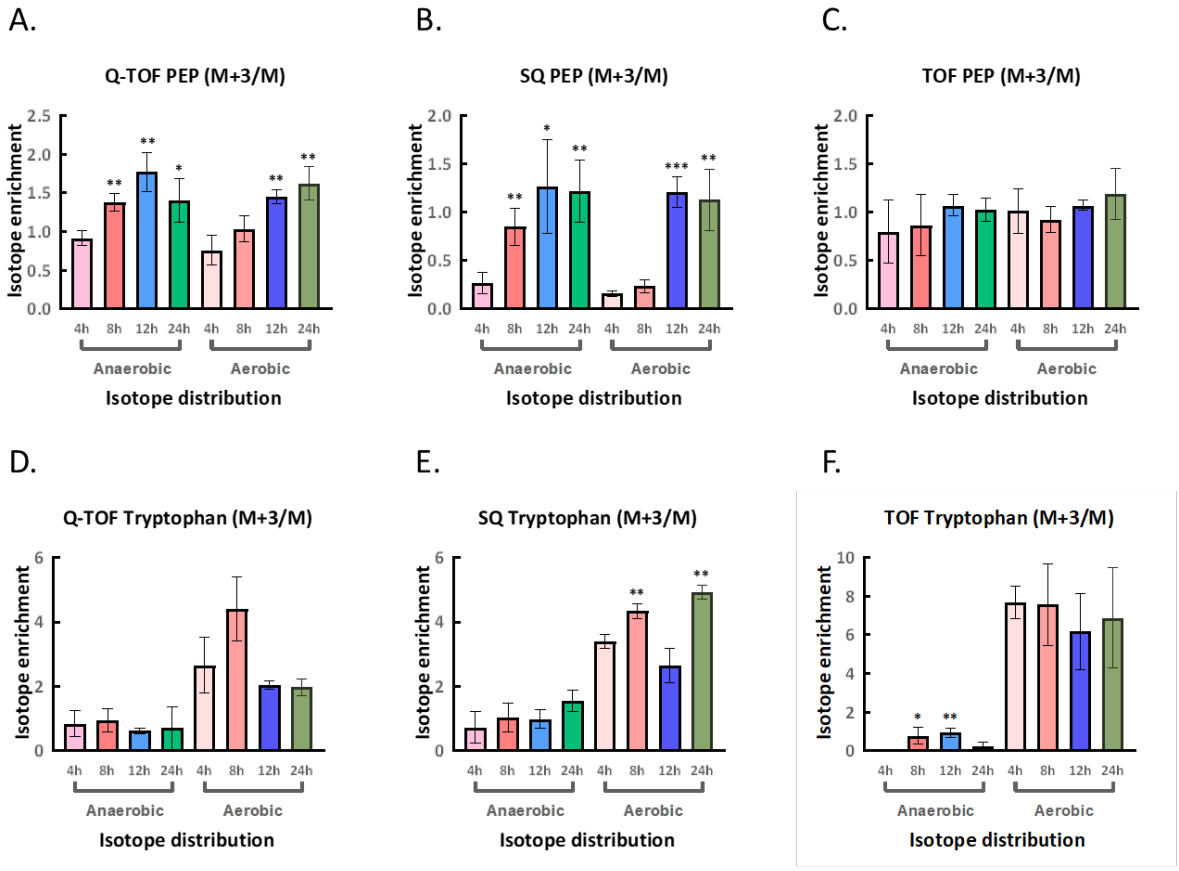


Figure S4.2 shows examples for two metabolites with low abundance in *R. mucilaginosa* samples, phosphoenolpyruvate (PEP) and tryptophan. QTOF MS and SQ MS showed PEP M+3/M started significant enrichment at 8h under anaerobic conditions and at 12h under aerobic conditions. On TOF MS, there is no significant enrichment but overrepresentation of enrichment across all time points. For tryptophan, the intensities are also relatively low on three GC-MS instruments, but all three instruments showed higher enrichment under aerobic conditions than anaerobic conditions.