MULTISEGMENT INJECTION-NONAQUEOUS CAPILLARY ELECTROPHORESIS-MASS SPECTROMETRY BASED METABOLOMICS FOR RAPID ANALYSIS OF FATTY ACIDS FOR DIETARY AND CARDIOMETABOLIC RISK ASSESSMENT

MULTISEGMENT INJECTION-NONAQUEOUS CAPILLARY ELECTROPHORESIS-MASS SPECTROMETRY BASED METABOLOMICS FOR RAPID ANALYSIS OF FATTY ACIDS FOR DIETARY AND CARDIOMETABOLIC RISK ASSESSMENT

By Sandi M. Azab, M.Sc.

A thesis submitted to the School of Graduate Studies in Partial Fulfilment of the Requirements

for the Degree

Doctor of Philosophy

August 2020

DOCTOR OF PHILOSOPHY (2020)
(Chemistry and Chemical Biology)

McMaster University
Hamilton, ON

TITLE:	Multisegment injection–nonaqueous capillary electrophoresis–mass spectrometry based metabolomics for rapid analysis of fatty acids for dietary and cardiometabolic risk assessment
AUTHOR:	Sandi M. Azab, M.Sc. (Alexandria University)
SUPERVISOR:	Professor Dr. Philip Britz-McKibbin
PAGES	xix, 214

Abstract

Biomarkers play a key role in human health for early disease screening for public health, improved diagnosis of human diseases, and monitoring of treatment responses on an individual level. However, there is urgent need for high throughput technologies to accelerate biomarker discovery based on comprehensive analyses of metabolites and lipids in complex human biofluids. This thesis contributes to new advances in metabolomics research and biomarker discovery in three major areas. In specific, this thesis describes (1) the development and validation of a multiplexed separation platform based on multisegment injection-nonaqueous capillary electrophoresis-mass spectrometry (MSI-NACE-MS) for the rapid and accurate determination of fatty acids and synthetic environmental lipids in blood specimens, (2) the application of this technique in support of nutritional epidemiology for objective assessment of dietary fat intake in women that can be correlated with self-reported food frequency questionnaires, and (3) the discovery of a panel of serum biomarkers for differentiation of peripheral artery disease (PAD) in older persons at high risk for limb amputation and death. *Chapter I* provides an introduction in biomarkers and metabolomics, including an overview of the data workflow when performing comprehensive metabolite profiling with emphasis on methods applicable to comprehensive fatty acid analysis. *Chapter II* introduces a novel assay based on MSI-NACE-MS for high throughput analyses of nonesterified or total hydrolyzed fatty acids in serum/plasma extracts following a rigorous method optimization and an inter-method comparison to conventional gas chromatography (GC) to demonstrate good mutual agreement. MSI-NACE-MS enables multiplexed analyses of seven samples within a single run with stringent quality control, robust inter-batch correction, and accurate electromigration modeling for lipid identification without pre-column chemical derivatization and complicated sample workup procedures. *Chapter III* further expands

concentration sensitivity when using MSI-NACE-MS/MS for rapid biomonitoring of low nanomolar levels of perfluoroalkyl substance (PFAS) exposures as they are a class of ubiquitous environmental pollutants with endocrine-disruption functions. Perfluorooctanoic acid and perfluorooctanesulfonic acid were quantified in a subset of serum extracts from pregnant women before and after 2009 with lower exposures measured in accordance with enforced PFAS regulation. *Chapter IV* demonstrates the clinical utility of circulating NEFA for accurate nutritional assessment of fat intake in women. MSI-NACE-MS was applied in observational and intervention studies to demonstrate that polyunsaturated omega-3 and saturated odd-chain NEFAs serve as promising dietary biomarkers to monitor intake of oily fish/fish oil supplement and full-fat dairy intake, respectively. *Chapter V* describes an untargeted characterization of the serum metabolome of nondiabetic PAD patients for differentiation of chronic limb-threatening ischemia from intermittent claudication as compared to the ankle brachial index. A panel of serum metabolites, including several amino acids and fatty acids were identified as promising clinical biomarkers for early diagnosis and/or prognostication of PAD that also provides insights into its underlying pathophysiology. Lastly, *Chapter VI* provides an overview of the major contributions derived from this thesis, as well as a perspective on future research initiatives.

Acknowledgments

"Thank you, Lord, my one defense, for you have sustained me with your righteous right hand and your presence has never left me."

I would like to express my deepest gratitude to my supervisor, Dr. Philip Britz-McKibbin, for taking a chance on me by accepting me into his group while I was thousands of miles away; I will always be indebted. I would like to thank him for his supervision, teaching and motivation throughout the years as well as his ideas, his time and attention throughout our many discussions that amplified my passion for metabolomics research. I thank Dr. Britz for his constructive feedback, useful pressure and sometimes criticism that would always push me for more and for my best. Moreover, I am truly grateful for his ability to reassure me whenever I was panicking, having faith in me and for giving me the flexibility I needed as a grad student and mother. I would also like to thank my committee members Dr. Giuseppe Melacini and Dr. Russell de Souza for their feedback and suggestions during my committee meetings, including unexpected questions that helped me think wider and more comprehensively. I want to further thank Dr. de Souza for introducing me to nutritional epidemiology and for all that he had taught me, always with patience, encouragement and support.

I am truly thankful for all Britz group members; a unique body of support, mentoring, laughter, friendship and fun. I want to especially thank my colleague, friend, brainstorming-buddy and "partner in crime" Ritchie Ly for his continuous support as we ventured together the world of MSI-NACE-MS with stubbornness and vigour in the face of many technical and instrumental challenges, and his constant support that "we've got this". Dr. Meera Shanmuganathan, thank you for your support, optimism and friendship. Zach Kroezen, Stellena Mathiaparanam, Rebecca Hum, Alicia Mell, Biban Gill, Erick Helmezci, Jordan Holzschuher and Megan Magee; I am truly grateful for the good times we had in the lab. Former lab members, Dr. Mai Yamamoto, Dr.

Adriana N. de Macedo, Dr. Alicia DiBattista, Dr. Michelle Saoi and Dr. Nadine Wellington; I am indebted to having you when I first started always answering my many questions.

I am honoured to have learned so much from our collaborators, Dr. Sonia Anand, Dr. Russell de Souza, Dr. Koon Teo, Dr. Stephanie Atkinson, Dr. Stuart M. Philips and Dr. Mohammad Qadura. Thank you for your time, precious advice and the chance to be part of your exciting and impactful research.

I am sincerely indebted to several special academic mentors throughout my life including Herr. Hans-Jürgen Köhl, my Abitur chemistry teacher who first ignited my passion for chemistry and biology, and my Master's supervisor Dr. Mohamed Abou-Shoer, who always believed in me and fostered my love for research, chemistry and teaching. I also gratefully acknowledge funding support from the Egyptian Ministry of Higher Education. If I could, I would dedicate this thesis to my country Egypt; a place that is living within me and that I don't overcome missing, and to Canada; a country that has given me exceptional opportunities and where I learned how to ride a bike.

My parents, Eman and Mokbel, thank you for always believing in me, praying for me, and pushing me forward. My sister, Dr. Carol Azab, thank you for your love, friendship and babysitting help, but also for your academic advice calming me down and giving me guidance anytime I needed. Lastly, I am beyond grateful for my husband, Moushir, for his constant, untiring love and support on the good and the tough days. My sons, Alexander and Adam, you are the sunshine in my life and I love you beyond words; whenever you are thinking of giving up, persevere just a little bit more and again some more!

Table of Contents

List of Figures

Supporting Figures

List of Tables

Supporting Tables

List of Abbreviations and Symbols

$[M+H]^+$	Protonated molecular ion
$[M-H]^-$	Deprotonated molecular ion
μ_{EOF}	Electroosmotic mobility
μ_{ep}	Electrophoretic mobility
r	Pearson correlation coefficient
Vcap	Capillary voltage
18:0/C0	Stearic acid/carnitine ratio
ABI	Ankle brachial index
ACN	Acetonitrile
ADP	Adenosine diphosphate
ANOVA	Analysis of variance
Arg	Arginine
Arg/C3	Arginine/propionylcarnitine ratio
ATP	Adenosine triphosphate
AUC	Area under the curve
BEST	Biomarkers, Endpoints, and other Tools
BFIs	Biomarkers of food intake
BGE	Background electrolyte
BHT	Butylated hydroxytoluene
BMI	Body mass index
BRCA1/2	BReast CAncer genes 1 and 2
C0	Carnitine
C3	Propionylcarnitine
CAD	Coronary artery disease
CAPS	3-cyclohexylamino-1-propanesulfonic acid
CDC	Centers for Disease Control and Prevention
CE	Capillary electrophoresis
CE	Cholesterylesters
CE-MS	Capillary electrophoresis-mass spectrometry
CHD	Coronary heart disease
CI	Confidence interval
Cl-PFESAs	Chlorinated polyfluoroalkyl ether sulfonic acids
Cl-Tyr	3-Chloro-*L*-tyrosine
CLTI	Chronic limb-threatening ischemia
CMA	Correlation matrix analysis
CON	Controls
$COPSAC_{2010}$	Copenhagen Prospective Studies on Asthma in Childhood
Crt K	Creatine kinase
Crt Tr	Creatine transporter
CV	Coefficient of variance
DCM	Dichloromethane
DHA	Docosahexaenoic acid
DI-MS	Direct injection-mass spectrometry

DLW	Doubly labeled water
DOHaD	Developmental Origins of Health and Disease
DPBs	Dietary pattern biomarkers
DQI	Diet quality index
EDTA	Ethylenediaminetetraacetic acid
EFA	Esterified fatty acids
EI-MS	Electron impact ionization-mass spectrometry
EIE	Extracted ion electropherogram
EOF	Electroosmotic flow
EPA	Eicosapentaenoic acid
EPIC	European Prospective Investigation into Cancer and Nutrition
ESI	Electrospray ionization
ESI-MS	Electrospray ionization-mass spectrometry
F-Phe	4-fluoro-*L*-phenylalanine
FA	Fatty acids
FAMEs	Fatty acid methyl esters
FAMILY	Family Atherosclerosis Monitoring In earLY life
FC	Fold change
FCIBs	Food component intake biomarkers
FDA	Food and Drug Administration
FDR	False discovery rate
FFA	Free fatty acids
FFQs	Food frequency questionnaires
FN	False negatives
FO	Fish oil
FP	False positives
GC	Gas chromatography
GC-EI-MS	Gas chromatography-electron impact ionization-mass spectrometry
GC-FID	Gas chromatography-flame ionization detection
GC-MS	Gas chromatography-mass spectrometry
GDM	Gestational diabetes mellitus
GFR	Glomerular filtration rate
GSH	Glutathione
HCA	Hierarchical cluster analysis
HILIC	Hydrophilic interaction liquid chromatography
His	Histidine
HMDB	Human Metabolome Data Base
HP-921	Hexakis(2,2,3,3-tetrafluoropropoxy)phosphazine
HRMS	High resolution MS
IC	Intermittent claudication
ID	Inner diameter
IMS	Ion mobility spectrometry
INR	International normalized ratio
iPrOH	Isopropanol
IQR	Interquartile range
K-S	Kolmogorov-Smirnov

KNN	k-Nearest neighbor
LC	Liquid chromatography
LC-MS	Liquid chromatography-mass spectrometry
LC-MS/MS	Liquid chromatography coupled to tandem mass spectrometry
LIPID MAPS	Lipids Metabolites and Pathways Strategy
LOD	Limit of detection
LOQ	Limit of quantification
LPA	Linear polyacrylamide
Lys	Lysine
m/z	Mass-to-charge ratio
MEKC	Micellar electrokinetic capillary chromatography
MeOH	Methanol
MFE	Molecular Feature Extractor
MMA	Monomethylarginine
MRM	Multiple reaction monitoring
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
MSEA	Metabolite set enrichment analysis
MSI-CE-MS	Multisegment injection-capillary electrophoresis-mass spectrometry
MSI-NACE-MS	Multisegment injection-nonaqueous capillary electrophoresis-mass spectrometry
MSI-NACE-MS/MS	Multisegment injection-nonaqueous capillary electrophoresis-tandem mass spectrometry
MTBE	Methyl-*tert*-butyl ether
MUFA	Monounsaturated fatty acids
NACE	Non-aqueous capillary electrophoresis
NBS	Newborn screening
NEFAs	Non-esterified fatty acids
NIH	National Institutes of Health
NIST	National Institute of Standards and Technology
NMR	Nuclear magnetic resonance
NMS	2-napthalenesulfonic acid
OCFA	Odd-chain fatty acids
oxo-Pro	Oxo-proline
PAD	Peripheral artery disease
PAG	Phenylacetylglutamine
PARP	Poly ADP-ribose polymerase
PCA	Principal component analysis
PCs	Principal components
PFASs	Perfluoroalkyl substances
PFHxS	Perfluorohexane sulfonic acid
PFNA	Perfluorononanoic acid
PFOA	Perfluorooctanoic acid
PFOS	Perfluorooctanesulfonic acid
Phe	Phenylalanine
Phe:Tyr	Phenylalanine:tyrosine ratio

PKU	Phenylketonuria
PL	Phospholipid
PLS-DA	Partial least squares-discriminant analysis
Plsgen	Plasminogen
Plsn	Plasmin
POPs	Persistent organic pollutants
PQN	Probabilistic quotient normalization
PUFA	Polyunsaturated fatty acids
PURE	Prospective Urban and Rural Epidemiological
QC	Quality control
QQQ	Triple quadrupole
QTOF	Quadrupole time-of-flight
QTOF-MS	Quadrupole time-of-flight-mass spectrometry
R^2	Coefficient of determination
RCT	Randomized controlled trials
RMT	Relative migration time
ROC	Receiver operating characteristic
RPA	Relative peak area
RSD	Relative standard deviation
RT	Retention time
S-W	Shapiro-Wilk
SCF	Supercritical fluid extraction
SFA	Saturated fatty acids
SHARE	Study of Health and Risk in Ethnic Groups
SIM	Selected ion monitoring
SNR	Signal-to-noise ratio
SPE	Solid phase extraction
SRM	Standard reference material
START	SouTh Asian birth cohoRT
$t_{1/2}$	Half-life
T2DM	Type-2 diabetes mellitus
TAG	Triacylglycerides
TFA	Trans fatty acids
TG	Triglycerides
TIE	Total ion electropherogram
TLC	Thin layer chromatography
TMA	Trimethylamine
TMAO	Trimethylamine *N*-oxide
TN	True negatives
TOF	Time-of-flight
TP	True positives
Tyr	Tyrosine
UV	Ultraviolet absorbance
VIP	Variable importance in projection

Declaration of Academic Achievement

The following material has been previously published and is reprinted with written permission:

Chapter II. Reprinted and adapted from Azab SM, Ly R, Britz-McKibbin P. Robust method for high-throughput screening of fatty acids by multisegment injection-nonaqueous capillary electrophoresis–mass spectrometry with stringent quality control. *Anal. Chem.* 2019, *91*, 2329-2336. Copyright (2019) American Chemical Society.

Chapter III. Reprinted and adapted from Azab SM, Hum R, Britz-McKibbin P. Rapid biomonitoring of perfluoroalkyl substance exposures in serum by multisegment injection-nonaqueous capillary electrophoresis-tandem mass spectrometry. *Anal. Sci. Adv.* 2020, *1*, 1-10. https://doi.org/10.1002/ansa.202000053.

Chapter IV. Reprinted and adapted from Azab SM, de Souza RJ, Teo KK, Anand SS, Williams NC, Holzschuher J, McGlory C, Philips SM, Britz-McKibbin P. Serum non-esterified fatty acids have utility as dietary biomarkers of fat intake from fish, fish oil and dairy in women. *J. Lipid Res.* 2020, *61*, 933–944.

Chapter V. Reprinted and adapted from Azab SM, Zamzam A, Syed M, Abdin R, Qadura M, Britz-McKibbin P. Serum metabolic signatures of chronic limb-threatening ischemia in patients with peripheral artery disease. *J. Clin. Med.* 2020, *9*, 1877.

Chapter I: Introduction to Biomarkers, Metabolomics and Lipidomics for New Advances in Clinical Medicine

Chapter I: Introduction to Biomarkers, Metabolomics and Lipidomics for New Advances in Clinical Medicine

"An ounce of prevention is worth a pound of cure"

1.1 Introduction to biomarkers in clinical medicine

The term biomarker or biological marker, can be defined as any objectively measured characteristic reflecting a biological or physiological process relevant to human health, ranging from a genetic mutation in cell, a circulatory protein to a tissue imaging test.[1] Well-known examples of classical biomarkers are elevated blood glucose for diagnosis of diabetes, high blood pressure for evaluating hypertension, and excessive blood lactate as an indicator of tissue hypoxia.[2] In all cases, biomarkers require determination of normal reference ranges in large populations in order to determine critical thresholds to screen or diagnose human diseases with adequate sensitivity, specificity, and importantly clinical utility, such as improved patient outcomes, and cost savings for healthcare.[3,4] Historically, the Bence Jones protein (*i.e.,* light-chain immunoglobulin) in urine was the first quantitative laboratory test devised in 1847 for cancer.[2] However, elevated protein excretion in urine (*i.e.,* proteinuria/albuminuria) may also be caused by other non-malignant conditions, as well as dehydration, inflammation and kidney damage due to hypertension and late-stage diabetes. Only in the 1960s did the term "biomarker" emerge in the literature to describe discrete molecular abnormalities associated with disease.[2] Nowadays, universal newborn screening (NBS) programs for early detection of genetic disorders in the population is a prime example of the utility of biomarker screening for early treatment of rare diseases that otherwise suffer from high mortality/morbidity rates when relying on symptomatic diagnosis.[5] Currently in the province of Ontario, NBS allows for early detection of about 20 genetic diseases using high throughput direct infusion-tandem mass spectrometry (MS/MS) technology for multiplexed analysis of amino acids and acylcarnitines in dried blood spot extracts collected

from a heel prick shortly after birth. Mortality and morbidity are significantly reduced for early onset genetic disorders if newborn screening and confirmatory diagnosis is achieved promptly after birth prior to later clinical presentations. For example, phenylketonuria (PKU) is diagnosed by blood phenylalanine exceeding 360 μM, and is treated by lifelong restriction and therapeutic monitoring of dietary phenylalanine intake to avoid irreversible cognitive impairment and development disabilities.[6] As a result, diet plays critical roles in the effective treatment of most genetic diseases especially when introduced early in life.

In 2016, the FDA-NIH biomarker working group established the BEST (Biomarkers, Endpoints, and other Tools) resource, as a necessary tool to harmonize terms and definitions in the field of translational science for unambiguous communication based on 7 distinctive classes of clinical biomarkers.[7] A *diagnostic* biomarker is used to detect a disease or a disease subtype in individuals, such as use of glycosylated hemoglobin (HbA1c) for the diagnosis of type 2 diabetes, and glomerular filtration rate (GFR) for the diagnosis of chronic kidney disease. A *monitoring* biomarker is one that is measured *over time* to assess disease progression or response to an intervention based on the change of its trajectory in response to a defined endpoint of clinical significance, such as use of blood international normalized ratio (INR) with a target range of 2 to 3 to optimize anticoagulation response to warfarin therapy.[7] A *pharmacodynamic/response* biomarker is used as an indicator of a biological response to exposure to a certain (exogenous) drug or therapeutic agent, such as serum LDL cholesterol to assess response to prescribed cholesterol-lowering statins. A *predictive* biomarker is one that is able to identify those patients who might respond to a certain drug or agent whether favourably or unfavourably as in the case of BReast CAncer genes 1 and 2 (BRCA1/2) in women diagnosed with ovarian cancer to estimate likelihood of response to treatment with PARP (Poly ADP-ribose polymerase) inhibitors.[8]

BRCA1/2 mutations can be also utilized in risk assessment of cancer recurrence in women already diagnosed with breast cancer. This latter example is defined as *prognostic biomarker,* which assesses the probability of progression or deterioration of a pre-existing disease. On the other hand, *susceptibility or risk biomarker* would indicate the likelihood of developing a certain disease or condition in apparently healthy or asymptomatic individuals.[7] Similarly, BRCA 1/2 gene mutations are used to reveal women predisposed to breast malignancy without a previous cancer diagnosis.[7] Lastly, a *safety biomarker* measures the presence or extent of toxic or adverse effects of exposure to a certain drug, or environmental agent, such as use of neutrophil counts to estimate cytotoxic effect of chemotherapeutic agents and direct course of treatment accordingly.

The concept of biomarkers can also be extended to include dietary exposures with its own definitions in nutritional sciences and epidemiological studies. Well-established examples in the literature are the use of doubly labeled water (DLW) and 24 h urinary nitrogen for measurement of total energy expenditure and protein intake, respectively. A NIH-led initiative in 2018 to develop a common ontology for nutritional research classified dietary biomarkers into *exposure biomarkers*, which reflect (recent or long-term) intake of certain foods or nutrients, *susceptibility biomarkers* which signify risk to adverse effects caused by food (*e.g.,* iron overload from excessive red meat intake), and *outcome biomarkers* which assess the body's physiological responses to food (*e.g.,* changes in blood lipoprotein profiles).[9,10] It is increasingly evident that these dietary biomarker categories can overlap, where bioactive compounds from the consumption of certain foods modulates the metabolism/protein expression of host tissue, as well as gut microbiota activity. For instance, dietary choline is converted by the gut bacterial enzyme choline-TMA lyase (expressed mainly in anaerobic *Proteobacteria*) into trimethylamine (TMA) which is then converted by flavin monooxygenase in the liver into trimethylamine-*N*-oxide (TMAO) - a recently

identified pro-atherosclerotic metabolite.[11,12] Exposure biomarkers are further categorized into three distinctive groups. *Food component intake biomarkers* (FCIBs) refer to the chemical constituents found in different foods including classes of nutrients and non-nutrients (*e.g.,* fiber) relevant to food science and human health. *Biomarkers of food intake* (BFIs) are mostly non-nutrients or exogenous compounds associated with certain foods or food groups, such as proline betaine which is as a recently validated biomarker of recent citrus intake.[9] Lastly, *dietary pattern biomarkers* (DPBs) aim to characterize complex dietary patterns as a whole (*e.g.,* Mediterranean, Prudent or Western, ketogenic diets) in populations that diverges from conventional single nutrient research.[7,13] Growing interest in dietary biomarkers in nutritional epidemiology arises from "a radical need of reform" due to substantial measurement error, selective reporting, and bias associated with dietary self-report tools, such as food frequency questionnaires (FFQs), that impede the elucidation of reliable disease-diet relationships.[14,15]

Various '-omics' platforms have been utilized for biomarker discovery including genomics, epigenomics, transcriptomics, proteomics and metabolomics, that provide new molecular insights into the underlying pathophysiology of human diseases (**Figure 1.1**).[16] Where genomic biomarkers indicate "what might happen", metabolomic biomarkers indicate "what is happening" serving as a dynamic and sensitive downstream readout closest to phenotype, physiological responses, and importantly clinical outcomes.[11] Thus, metabolomics is an expanding field of functional genomics research that is aimed at global metabolite profiling of complex biological samples, including minimally invasive biofluids (*e.g.,* blood or urine). As compared to other "omics" disciplines, dynamic perturbations in metabolites as a result of environmental exposures can exceed changes in the proteome or transcriptome in concentration as well as speed (*i.e.,* flux)[17] However, there exists major technical hurdles in metabolomics research given the

underlying chemical diversity and dynamic range of the human metabolome, which includes a large fraction of detected yet unidentified compounds of biological significance. Moreover, metabolic phenotype changes are very dynamic and interdependent and thus the integration of multi-omic data sets within a systems biology approach is advantageous to better elucidate biological mechanisms.[16]

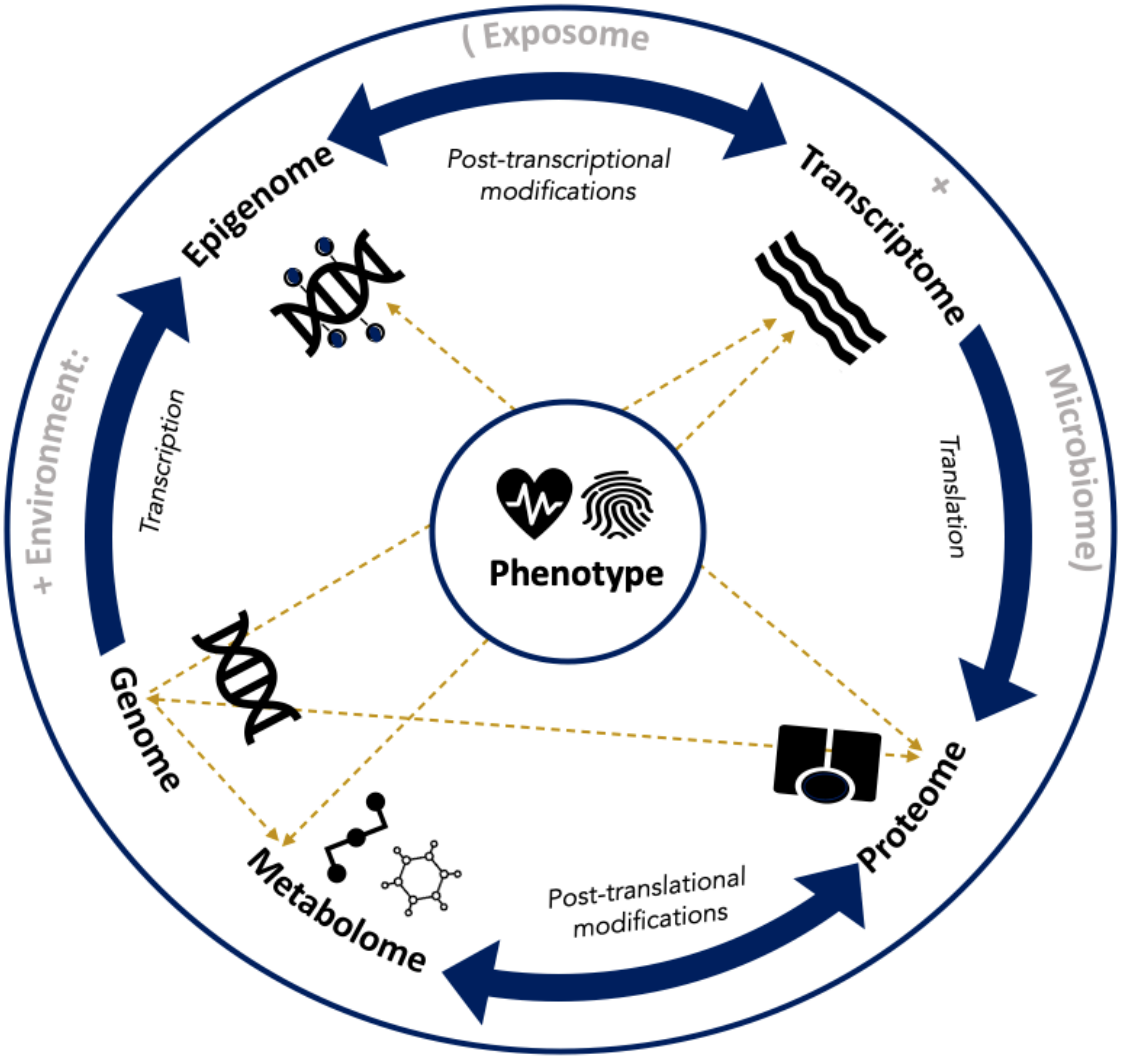


Figure 1.1. The '-omics' cascade from a systems biology viewpoint used to illustrate the etiology of the molecular phenotype from a central dogma framework that starts with the genome with a downstream direction towards the metabolome, where additional arrows depict other mechanistic pathways including reverse causation or feedback mechanisms. The influence of environmental exposures (*i.e.*, the exposome) in addition to host's microbiome, is also included given their direct influence on metabolic phenotype changes. Adapted from Chu et al.[16]

1.2 Metabolomics and the metabolomics data workflow

Metabolomics can be defined as the comprehensive characterization of all low molecular weight metabolites (< 1500 Da) in a biological sample using one or more analytical technologies in conjunction with complementary bioinformatic methods.[11] The metabolome can encompass thousands of metabolites, as in the case of human serum with over 25,000 putative compounds included in the Human Metabolome DataBase (HMDB), spanning nearly 11 orders of magnitude from millimolar (*e.g.,* glucose and urea) to picomolar concentration levels (*e.g.,* thyroxine and prostaglandin E1).[18] This vast collection of small molecules consists of both endogenous metabolites, such as amino acids, nucleic acids, small peptides, electrolytes, sugars, alcohols, organic acids and lipids, as well as exogenous/xenobiotic chemicals, including prescription and over-the-counter drugs, plant phytochemicals, food additives, pesticides, environmental contaminants, and persistent organic pollutants. The latter is often collectively referred to as the exposome, which reflects the totality of chemical exposures from conception to death.

To date, metabolomic experiments differ depending on the study question as well as the resolution, sensitivity and metabolome coverage of the instrumental method, and can be classified into four major sub-fields: 1. *targeted metabolomics* (*i.e.,* hypothesis-testing via quantitative analysis of sub-set of known metabolites), 2. *untargeted metabolomics* (*i.e.,* hypothesis-generating that includes the discovery of unknown metabolites of clinical significance), 3. *fluxomics* (*i.e.,* metabolic flux measurement using isotope labelling to assess metabolic pathway dynamics), and 4. *metabolite imaging* (*i.e.,* spatial characterization of metabolite distributions within tissues).[11] In general, the metabolomics workflow consists of series of interdependent steps (**Figure 1.2**) that need to be carefully implemented for meaningful and impactful research that avoids bias. As discussed in detail below, major hurdles to translational metabolomics research include poor study

designs, inadequate study power and replication, limited metabolome coverage and, importantly, unknown identification. These represent major bottlenecks at different stages of the metabolomics pipeline from pre-analytical (*e.g.*, inconsistent sample collection and storage), analytical (*e.g.*, lack of sample randomization and quality controls) to post-analytical (*e.g.*, inappropriate statistical methods not adjusted for confounding variables) steps that contribute to false discoveries.

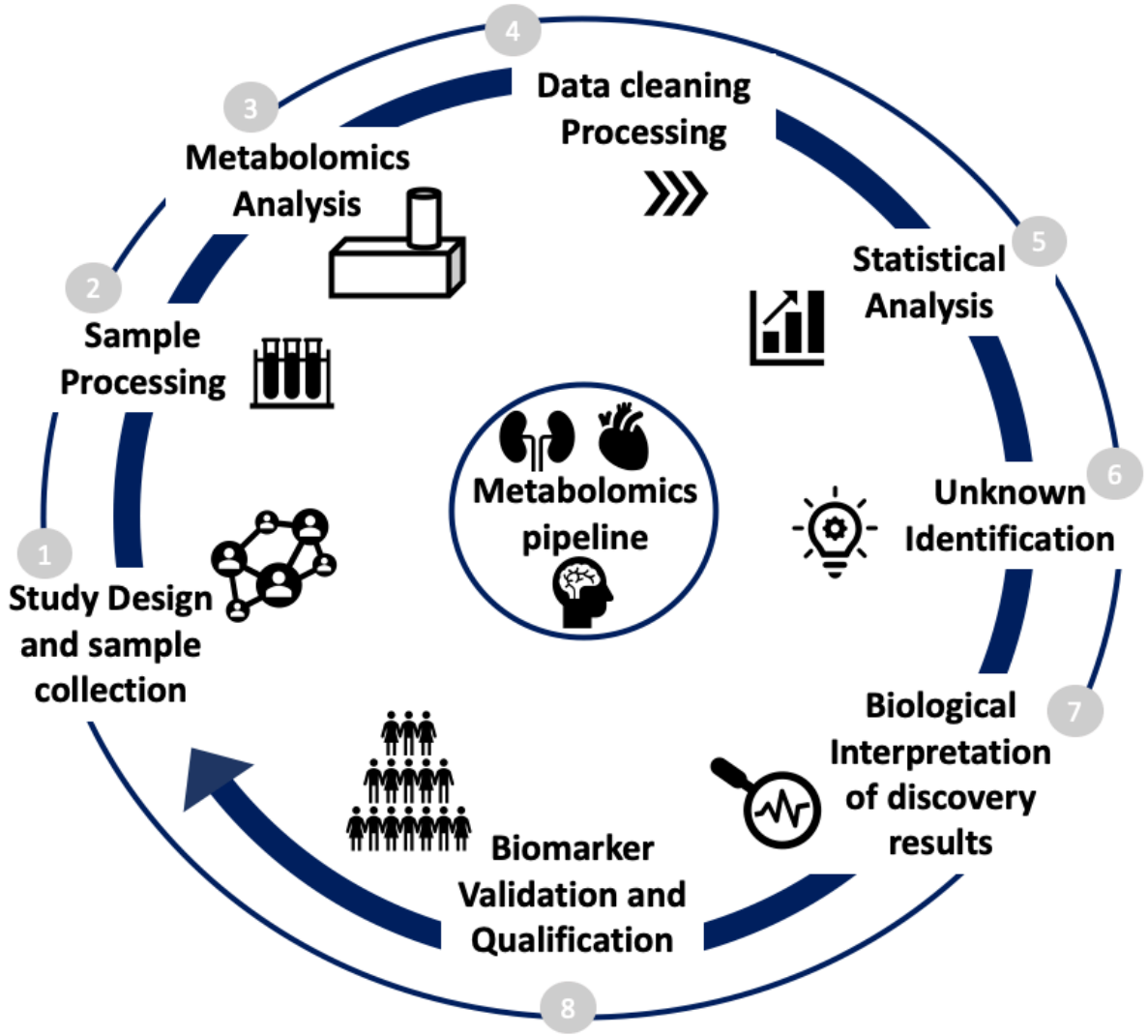


Figure 1.2. Scheme showing the major steps of the metabolomics workflow for biomarker discovery from study design and participant recruitment followed by sample pretreatment and chemical analysis. Also, data preprocessing when using high resolution MS platforms (*e.g.*, peak picking, filtering, scaling and normalization), and unknown metabolite identification and biological interpretation remain key bottlenecks in contemporary untargeted metabolomic workflows from biomarker discovery to clinical translation.

1.2.1 Study design and sample considerations for metabolomics

Study design encompasses the overarching experimental approach reflecting the objectives of the investigation, including a defined hypothesis, desired outcome, target population/control, and criteria used for participant selection/rejection. Epidemiological studies largely fall under observational, whether case-control, cross-sectional, cohort or longitudinal, whereas experimental designs are best exemplified by randomized controlled trials.[17] Clinical metabolomics studies often include "disease versus control" or paired "pre- and post-intervention" study designs to identify and quantify discriminating metabolites in specific biological samples. The initial stages of study design involving participant recruitment and biospecimen collection/storage are crucial for the quality of all downstream processes thereafter. This could be jeopardized in cases of inadequate sample sizes, non-uniform patient matching criteria, confounding, improper choice of specimens, and use of "specimens of convenience" for biomarker discovery based on mere availability from collaborating institutions.[19] "Garbage in, garbage out" remains an axiom that contributes to bias, false discoveries and poor reproducibility in metabolomics.[19] Various types of samples can be utilised for metabolomics studies ranging from cell cultures, tissues, and various biofluids.[20] The latter can be further classified into *metabolic fingerprinting* of circulating biofluids (*e.g.,* whole blood, serum or plasma) and *metabolic footprinting* of excreted biofluids (*e.g.,* urine, saliva, breath condensates, sweat or stool).[11] Careful selection of one or more biospecimen type(s) in accordance to the research question should be exercised. For instance, fatty acids analysis is a broad term that can encompass different lipid pools from adipose tissue, erythrocytes phospholipids, and circulating/protein-bound non-esterified fatty acids (NEFA) each reflecting a different turnover time, biochemical function/activity, and importantly clinical relevance.[21,22]

Further considerations are related to processes of sample collection, handling, transport and storage conditions, where failure to adhere to standardized operating protocols can inevitably contribute to bias with greater variability. Urine, cells or tissue samples are to be stored promptly at -80 °C immediately after collection whenever possible, whereas blood samples require processing prior to storage.[17] This should be done on ice and ideally within half an hour from blood sample retrieval followed by storage at -80 °C to halt incidental enzymatic activity, prevent hemolysis and ensure long-term chemical stability.[23] For preparation of plasma, clotting is prevented by collecting blood in anticoagulant tubes containing EDTA, citrate or heparin followed by centrifugation to fractionate plasma from cellular content in contrast to serum, where blood is allowed to naturally clot for a fixed time prior to fractionation.[23] Thus, serum lacks fibrinogen, prothrombin and other coagulation factors formed during fibrin clotting and removed along with blood cells, but has additional pro-inflammatory cytokines and lipids, such as sphingosine-1-phosphate and eicosanoids produced in response to clotting, that do not reflect physiological concentration levels.[18,24] Nevertheless, studies comparing serum and plasma have shown that both blood specimens are highly correlated and display similar results for the majority of metabolites/lipids with plasma being more reproducible and serum, in turn, exhibiting overall higher concentrations of metabolites.[18,24] Furthermore, fasting samples are superior to non-fasting samples as is 24 h urine collection as compared to single-spot/random urine samples. Yet, in practice, researchers do not necessarily have control over all factors given practical constraints and ethical issues involving patient recruitment/compliance, such as collection of fasting blood samples from children or pregnant women. Other uncontrollable confounders include between-subject differences of metabolism, co-morbidities, microbiome activity, environmental exposures, dietary/lifestyle patterns and food matrix effects, notably in the case of nutrition studies.[17,25]

1.2.2 Sample processing and chemical analysis

Prior to metabolomic analyses, sample processing is often required, depending on the specimen type and instrumental platform used for analysis. Sample workup is a critical step needed to concentrate low abundance metabolites, reduce background matrix interferences, or transform metabolites to facilitate their resolution and/or detection via chemical derivatization. For instance, gas chromatography (GC) methods typically require pre-column chemical derivatization (*e.g.*, transesterification, or oximation followed by trimethylsilylation), to convert sample analytes to a more volatile form that are transported through a wall-coated capillary by a gaseous mobile phase.[17] Similarly, for the quantitative analysis of trace levels of perfluoroalkyl substances (PFASs) in serum, liquid extraction and subsequent enrichment using solid phase extraction (SPE) cartridges is employed prior to liquid chromatography-tandem mass spectrometry (LC-MS/MS) to enable detection of the low nanomolar levels of these environmental contaminants.[26,27] In contrast, sample preparation for untargeted metabolomics should ensure wide selectivity for broad metabolome coverage with adequate sample clean-up as required, such as serum deproteination by organic solvent (*e.g.,* methanol or acetonitrile) or ultrafiltration followed by dilution before analysis by LC coupled to high resolution MS (HRMS).[28] Furthermore, it is important to keep freeze/thaw cycles at a minimum during sample processing to maximize metabolite stability by storing separate aliquots of samples for "one-time use".[29] For example, polyunsaturated fatty acids (PUFA) are prone to oxidation from heating, prolonged air exposure and/or presence of iron in case of hemolysis, thus careful sample handling is of utmost importance. Storage under nitrogen gas, at – 80 °C, or addition of lipid-soluble antioxidants, such as butylated hydroxytoluene (BHT) prior to storage, can further help to minimize losses during sample handling. [22,30,31]

At the same time of sample processing, a quality control (QC) is prepared to serve as a representative reference/pooled sample to assess technical variance and track variation in instrumental response. The same QC sample is analyzed at the start and end of analysis, as well as intermittently between batches of consecutive analytical runs after every 5-10 samples.[32,33] Ideally, the QC sample is a pool of aliquots from all study samples so that most of the detected metabolites are also present and monitored in the QC sample.[34] Alternatively, in large-scale studies or volume-restricted samples, the QC sample could also be of external or commercial source, such as a certified reference material from NIST.[35] In all cases, consistent use of the same QC samples allows for correcting of long-term signal drift that can inevitably occur in MS-based metabolomics despite maintaining standard operating procedures (*e.g.*, daily instrument calibration/tuning and ion source cleaning), via inter- and within-batch correction algorithms.[36,37]

In contrast to genomics or transcriptomics, where the entire DNA or RNA can be measured using a uniform analytical platform (*i.e.*, a gene sequencer), metabolites vary drastically in their chemical and physical properties belonging to hundreds of different chemical classes, so that no single analytical method can provide a complete measure of the metabolome.[11,20] Thus, it is common to apply several orthogonal platforms in parallel to maximize metabolome coverage provided there is adequate sample volume. The two main analytical techniques used in metabolomics are nuclear magnetic resonance (NMR) spectroscopy and high resolution mass spectrometry (HRMS); the latter is often coupled to various separation techniques to improve analytical performance, such as LC, GC, capillary electrophoresis (CE) or ion mobility spectrometry (IMS). These separation methods offer better selectivity to resolve isobaric/isomeric interferences while also reducing ion suppression/enhancement effects in nontargeted metabolomics when using direct infusion-MS without stable-isotope internal standards.

Table 1.1. Common instrumental techniques used in metabolomic studies.[17,38,39]

Platform	Advantages	Disadvantages
Nuclear Magnetic Resonance (NMR)	• Non-destructive method • Minimal sample workup • Quantitative • Unambiguous identification • Fast (2-10 min/sample) • Robust technology	• 10-100 X less sensitive than MS • High infrastructure/operating costs • Large sample volumes 100-500 μL • Complex deconvolution required due to spectral overlap in 1D NMR • Limited metabolome coverage
Direct Infusion Mass Spectrometry (DI-MS)	• Fast (1-3 min/sample) • Low sample volume • High sensitivity • No solvents • Ideal for targeted analysis	• Ion suppression/matrix effects • No separation of isomers/isobars • Stable-isotope internal standards crucial for quantification • Limited selectivity
Liquid Chromatography Mass Spectrometry (LC-MS)	• High sensitivity • Good retention time reproducibility (reversed-phase) • Wide selectivity/metabolome coverage • Simpler sample workup	• Slow (15-40 min/sample) • High solvent consumption/waste • Less robust, less reproducible retention times (HILIC) • Complicated separation mechanisms (HILIC)
Gas Chromatography Mass Spectrometry (GC-MS)	• Mature technology with extensive EI-MS library • High separation efficiency • Reproducible retention times • Ideal for volatile/thermally stable metabolites	• Complicated sample workup • Not for thermolabile compounds • Slow (20-40 min/sample) • Limited metabolome coverage • Unknown compound identification is difficult
Capillary Electrophoresis Mass Spectrometry (CE-MS)	• High separation efficiency for ionic metabolites • Effective desalting/minimal ion suppression • Low sample volume (< 5 μL) • Simpler sample workup • Minimal solvent/low operating costs	• Low sensitivity/limited coverage • Less reproducible migration times compared to reversed-phase LC • Technically challenging/less support from vendors • Few validated protocols for large-scale studies • Lack of robust CE-MS interfaces
Ion Mobility Mass Spectrometry (IM-MS)	• Ultra-fast (< 100 ms) • Separation of isobars/isomers • Accurate prediction of collisional cross-section area • High reproducibility	• Least mature technology • Prone to ion suppression • Poor orthogonality (high correlation) with MS • Limited peak capacity/resolution

The advantages and disadvantages of each technique is summarized in **Table 1.1** with LC-MS being the most widely used instrumental configuration in metabolomic studies based on reversed-phase and/or hydrophilic interaction (HILIC) modes of chromatography. CE-MS is a high efficiency microseparation technique based on the separation of charged ions in free solution under an electric field, which is optimal for resolution of polar/ionic metabolites ranging from phosphorylated sugars, organic acids, amino acids, nucleotides and weakly acidic/basic drugs.[40-43] In fact, CE-MS has recently been applied in large-scale serum metabolomic cohort studies with good long-term reproducibility.[44,45] There are two fundamental electrokinetic principles that govern CE separations, namely the electroosmotic flow (EOF) and the intrinsic electrophoretic mobility of an ion (*i.e.,* metabolite). The EOF serves as a natural electrokinetic pumping mechanism to transport of all ions towards the MS detector with low flow rates (nL/min).[46] The silanol groups on the inner wall of a fused-silica capillary are deprotonated at pH > 2, and thus form an overall negative surface charge, which attracts electrolytes of the opposite charge in the background electrolyte (BGE) as shown in **Figure 1.3**. This process results in the formation of an electric double layer of electrolytes in solution. The inner fixed layer (Stern layer) exhibits a strong adsorption of largely immobile ions while the outer (Helmholtz) layer of electrolytes are more diffuse/mobile in solution that generates a potential difference (*i.e.,* zeta potential) from the capillary surface to bulk solution.[47] The application of an external voltage across both ends of the capillary, causes transport of bulk solution towards the cathode (*i.e.,* EOF) that is mediated by the mobility of excess of solvated cations within the diffuse electric double layer (**Figure 1.3**).[48] The EOF or electroosmotic mobility (μ_{EOF}) is largely dependent on the composition of the capillary surface, as well as properties of the buffer solution, including pH, ionic strength, temperature and viscosity, and impacts the apparent migration time of an analyte. In contrast, the electrophoretic

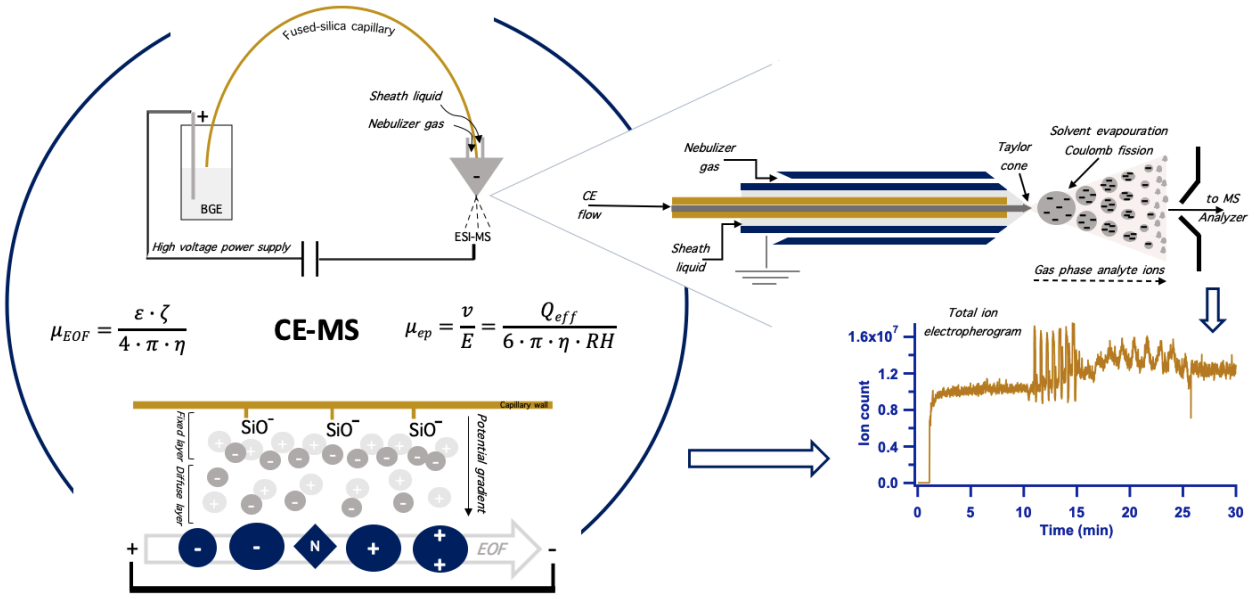


Figure 1.3 Schematic of fundamental CE separation principles when using a coaxial sheath liquid CE-MS interface (grounded) under negative ion mode conditions. Upper panel depicts CE-MS setup with a fused-silica capillary connected at the inlet (cathode) to a high voltage power supply when using a nonaqueous BGE for the separation of acidic lipids (*e.g.*, long-chain fatty acids) prior to their ion desorption in ESI-MS. Lower panel models Stern's model of the double layer charge distribution at the negatively charged silica wall leading to the formation of the EOF. Upon voltage application, the apparent migration time for acidic lipids are determined by the net superimposition of the electroosmotic (μ_{EOF}) and electrophoretic mobilities (μ_{ep}), where ε = buffer dielectric constant: ζ = zeta potential; η = viscosity; v = ion migration velocity; E = electric field; Q_{eff} = effective charge; R_H = solute's hydrodynamic radius.

mobility (μ_{ep}) directly impacts the selectivity of a separation and is dependent on differences in the intrinsic physicochemical properties of an ion, namely pK_a and molecular volume (*i.e.*, effective charge density). Therefore, in CE under normal polarity, small electrolytes (*e.g.*, Na^+ or K^+) will migrate first due to their high positive mobilities that allows for desalting of highly saline samples prior to electrospray ionization (ESI), followed by the migration of bulkier/minimally charged cationic/basic metabolites (*e.g.*, amines, amino acids, acylcarnitines). Neutral compounds are unresolved and co-migrate with the EOF, whereas acidic/anionic metabolites migrate in opposite direction towards the anode, but are transported to the detector due to the higher velocity of the EOF under most buffer conditions (**Figure 1.3**).[49] Low concentration sensitivity is a major limitation of CE-MS due to nanoliter volumes of samples introduced hydrodynamically on-

capillary (~ 5-10 nL), as well as post-capillary dilution effects when using a conventional coaxial sheath liquid interface for spray formation in CE-MS.[50] Alternatively, various sheathless interface or low-flow (*i.e.*, nanospray) interface designs have been developed to boost concentration sensitivity in CE-MS, such as a porous hydrofluoric-etched capillary tip, stainless-steel liner with treated/restricted capillary tips, or electrokinetically-driven sheath liquid flow through tapered glass emitter.[51-55] A unique advantage of CE-MS however is the ability to perform multiplexed separations to greatly improve sample throughput via serial injection of seven or more samples within a single analytical run.[56] Recent advances of multisegment injection (MSI)-CE-MS enable high throughput screening (< 3 min/sample) for comprehensive surveillance of drugs of abuse, as well as nontargeted metabolomics to identify serum biomarkers associated with physical inactivity in older persons.[33,41,49] Importantly, MSI-CE-MS also offers superior data fidelity when performing large-scale metabolomic studies that accelerates biomarker discovery based on temporal signal pattern recognition.[5]

CE separations are typically performed under aqueous buffer conditions for polar/ionic metabolites, which limits its applicability to a wide range of ionic yet lipophilic compounds, such as long-chain fatty acids. In this case, non-aqueous capillary electrophoresis (NACE) is an alternative strategy that makes use of a background electrolyte (BGE) comprised largely of an organic solvent solution.[57] In addition to greatly expanding separations for lipid analyses, NACE is also associated with lower currents due to the reduced conductivity of most organic solvents (*e.g.*, methanol) that allow use of higher voltages and larger bore capillaries with lower detection limits.[58] Furthermore, NACE is ideal for coupling to electrospray ionization (ESI)-MS for better droplet desolvation and ionization efficiency due to the lower surface tension and higher volatility of many organic solvents as compared to aqueous solutions. However, analyses by NACE-MS is

not necessarily without challenges, which may explain the limited number of reported studies to date and its lack of application in the field of lipidomics. For instance, acetonitrile, the most common solvent used in NACE can cause swelling of the outer polyimide capillary coating leading to incidental capillary breakage and poor long-term reproducibility.[59] Furthermore, as organic solvents used in NACE typically have lower boiling points and surface tension than water, this can create higher suctioning by the nebulizer gas used to facilitate spray generation when using coaxial sheath liquid CE-MS interfaces, causing air to be drawn within the capillary and leading to current instabilities.[60] Strategies to overcome these major technical challenges have been successfully implemented in the development of a robust method for rapid determination of NEFA from serum extracts by MSI-NACE-MS,[61] which is one of the major contributions described in this thesis.

1.2.3 Data processing and statistical analysis

Following data acquisition, deconvolution, pre-processing and pre-treatment of the raw data is necessary to produce clean data fit for multivariate or univariate statistical analysis in the form of a data matrix of samples (paired or unpaired) in columns and molecular features/metabolites (metabolite name, accurate mass, chemical shift bin) in rows as a function of their response (concentrations, integrated peak areas).[62] This phase of the nontargeted metabolomics data workflow remains a major bottleneck, especially for HRMS-based experiments.[63] It starts with conversion of the instrumental output into numeric values by peak integration and further includes data filtering, peak picking, and time alignment of molecular features if separations are coupled to MS. Data filtering is the removal of background noise, such as spurious signals corresponding to instrumental or random noise, whereas peak picking involves identification of reproducible and authentic metabolites.[64] ESI ionization technique commonly used in LC-MS and CE-MS metabolomics, is especially prone to formation of "untrue" molecular features. These can be

background contaminants (*e.g.,* solvent impurities), artifacts (informatic error such as baseline fluctuations), as well as redundant isotopic signatures (^{13}C, ^{15}N), in-source fragments, adducts or dimers – the latter features derive from a single metabolite and is referred to as signal degeneracy.[65,66] Hence, removal of such non-informative and error-prone signals is important for subsequent multivariate statistical analysis to reduce data overfitting and false discoveries. Indeed, stringent approaches to data filtering in LC-MS based metabolomics can lead to up to 90% reduction of data from 25,000 total molecular features to less than 1,000 authentic metabolites that are measurable with adequate precision and detection frequency.[63,65] Time alignment is the process of peak clustering between runs to correct for variation of retention or migration times due to instrumental drift, which offers an orthogonal descriptor when annotating metabolites besides their accurate mass (*m/z*), such as adjusted retention times or relative migration times. Efforts for efficient data cleaning include open-access web-based tools (*e.g.,* XCMS, MZmine), proprietary vendor software (Agilent MassHunter, Waters MassLynx), or in-house protocols, such as temporal signal pattern recognition using a dilution trend filter when using MSI-CE-MS.[5,67,68] Moreover, data generated over time are unavoidably prone to long-term signal drift in intensity and mass accuracy. Here, consistent QC samples, randomly and regularly analysed throughout a metabolomics study can be used to correct for such variance between samples both between and within batches via open-source mathematical algorithms, especially important in the case of large-scale MS-based metabolomics studies.[36,37]

Subsequently, QC samples ran repeatedly throughout the analysis are also processed, where molecular features are further excluded based on acceptable technical precision, as measured by the coefficient of variation (CV < 30-40%).[39,69] The US Food and Drug Administration (FDA) recommended guidelines of tolerance limits variation not to exceed 15%

for drugs, 20% for low abundance compounds with low signal to noise ($S/N < 10$), and 30% for biomarkers in discovery-based metabolomic studies.[29,39] Furthermore, molecular features not consistently detected (< 75% threshold) in the QCs or study samples are also excluded to prevent data skewing; otherwise, missing values imputation is performed by replacement with a small value, such as the concentration detection limit for metabolite (if known), the minimum response detected for metabolite/2 in all samples, or by a weighted k-nearest neighbour (KNN) algorithm.[69,70] QC samples also guide whether applying a batch-correction of a metabolomics data set is necessary as discussed earlier. Lastly, control charts of a recovery standard (*e.g.,* exogenous deuterated standard) added to all QC and study samples at a fixed concentration is an important method to assess long-term technical precision and flag sample outliers that may require exclusion or repeat analysis in case their response exceed action limits (± 3 s).[39]

Once completed, the resulting filtered data is then pretreated using normalization, transformation and scaling to generate data optimal for statistical analysis. Peak area normalization relative to internal standard peak area corrects for variation in sample injection volumes, where additional normalization processes may be required depending on sample type. This is less of a concern for hemodynamically controlled biofluids (*e.g.,* serum), unlike urine that requires adjustment to correct for variation in hydration status by normalizing to creatinine, osmolality, specific gravity, or using probabilistic quotient normalization (PQN).[71-73] Similarly, stool or wet tissue specimens are commonly normalized to their total dried mass after lyophilization and weighing. Autoscaling, pareto-scaling or range-scaling account for differences in metabolites concentration ranges and impart equal importance of all measured metabolites irrespective of their absolute abundances.[74] Moreover, data normality should be examined using histograms to study distribution patterns and either Shapiro-Wilk (S-W) or Kolmogorov-Smirnov (K-S) statistical

tests.[75] Following, *log*-transformation is helpful in making skewed metabolomic data more normally distributed while overcoming heteroscedasticity, which originates from the fact that technical and other sources of biological variation are not necessarily equally distributed around zero.[74] Choice of data pre-treatment or transformation can have considerable effect on overall data structure and statistical outcome, and thus should be carefully considered.[76]

Both univariate and multivariate statistical methods are routinely applied to metabolomics data in order to reveal biological findings relevant to a tested hypothesis or to generate a testable hypothesis when performing nontargeted metabolomics.[77] Univariate tests are often preferred due to the ease of direct interpretation (*e.g.,* student's t-test and ANOVA or their non-parametric equivalents) after exclusion of possible uninformative features or covariances.[78] These tests are often used to detect statistically significant differences between two (or more) groups, such as healthy control versus disease state. However, in case of metabolomics data, where hundreds of univariate tests (equivalent to total number of variables or metabolites analyzed) are conducted, the alpha-level criterion of significance (typically set at $p < 0.05$) increases with each additional test leading to a high probability of false discoveries as a result of mere chance (type I error).[79] Multiple hypothesis testing corrections are thus crucial and include the more conservative Bonferroni correction (alpha/number of tests) but can increase the risk for rejecting true discoveries (type II error).[80] Alternatively, false discovery rate (FDR) correction, such as the Benjamini-Hochberg procedure, is less strict as it controls the fraction of false positive results and thus is often preferred in case of pilot metabolomics studies at an early exploratory stage.[79] In either case, there is always a trade-off between these two errors of controlling false discoveries and loss of statistical power by rejecting true discoveries.[77]

In contrast to univariate statistics, some findings are often only detected collectively as individual metabolites can complement each other via coupled metabolic pathways.[78] A multivariate approach can also help minimize noise and biological variation within groups by the so called "effect of consistency at large".[78] Principal components analysis (PCA) is a widely used unsupervised data projection method that captures variance as a linear combination yet smaller set of latent variables or so called principal components (PCs).[74] Those PCs, which are new coordinate axes in the directions that confer a majority of the underlying variation found in the samples overall, aim to separate structure from noise. The first PC is the one that is in the direction of maximal variance between samples then the second is orthogonal to the first and so on thereafter. Subsequently, PCA is often used to investigate metabolomics data for overall patterns of variation and trends in the data, as well as detection of potential outliers in an unsupervised manner.[81] In contrast, partial least squares–discriminant analysis (PLS-DA) is a frequently used supervised multivariate method used in metabolomics for classification and regression,[82] which optimizes separation between groups by maximizing covariance based on prior knowledge of class membership (dependent variable Y) as linked to the raw data (independent variables X).[28] Unlike PCA, PLS-DA can detect subtle features/differences between groups in well balanced study designs without potential confounders.[81] Nevertheless, it should be carefully used since it is prone to data overfitting and misclassification with the number of variables exceeding the number of samples (*i.e.,* inadequate study power) especially when cross-validation/permutation testing of the training set is not performed.[83] In most cases, rigorous validation also requires an independent set of samples (*i.e.*, test set) to demonstrate predictive accuracy of model for group classification.

1.2.4 Unknown identification

The untargeted metabolomics approach includes all analytically captured features for hypothesis generation, and thus contains a large number of unidentified/unknown metabolites. Structure elucidation requires additional experiments using a MS system with high resolving power (> 10,000), mass accuracy (< 5 ppm), and sensitivity together with the ability to acquire MS/MS spectra via collisional-induced fragmentation of selected precursor ions at different voltages.[39] Though time and labour-intensive, unknown identification can lead to the discovery of novel metabolite classes of clinical or biological significance, such as oncometabolites and atherotoxins.[11,38] Thus, priority is given to those metabolites that show significance in the statistical analysis in contrast to data independent MS/MS acquisition modes. The challenge to identify unknown features is further compounded by peak degeneracy, where a single metabolite can generate ten or more molecular features/signals in ESI-MS with the same retention/migration time, but different *m*/*z*. For example, glutamine, a major blood amino acid was reported to generate 100 spectral features including complex adducts.[65] In 2007, the Chemical Analysis Working Group of the Metabolomics Standards Initiative established a 4-level system for metabolite identification aiming for standardized reporting and best practices as summarized in **Table 1.2**.[84] Level 1 exhibits the highest confidence level with unambiguous identification based on direct comparison to an authentic standard analyzed using the same instrument having the same accurate mass (*m*/*z*), retention/migration time, and MS/MS spectrum with high index of similarity (*i.e.*, *m*/*z* of specific product ions and their relative intensities). When a standard is unavailable, putative identification of the metabolite (level 2) or the metabolite class (level 3) is based on comparison of experimental MS/MS with reference spectra deposited in public databases, or *de novo* annotation or *in silico* prediction of MS/MS spectra for distinctive product ions/neutral losses (*i.e.*, functional groups or

Table 1.2. The four levels of metabolite identification confidence adapted from Dunn *et al*.[85]

Level	Confidence of Identity	Level of Evidence
1	Confidently identified compounds	Direct comparison to an authentic standard analysed under identical analytical conditions having the same two or more orthogonal properties i.e. *m/z*, retention time, MS/MS
2	Putatively identified compounds	Identification is based on similarity of physicochemical properties and MS/MS spectra with public/commercial spectral libraries, without reference to authentic chemical standards
3	Putatively annotated compound classes	Based on similarity of physicochemical properties and MS/MS spectra to a chemical class of compounds
4	Unknown compounds	Based on minimum annotation of accurate mass and retention/migration time, and most likely molecular formula, without known structural information.

structural motifs), respectively. Lastly, level 4 identification requires a minimum annotation of accurate mass and retention/migration time, and potentially most likely molecular formula, without known structural information, that, nevertheless imparts valuable information when reporting an unknown metabolite in a meaningful way.[84] Noteworthy, new reporting standards are currently being proposed by the Metabolomics Society for a 7-level system that also includes presence/absence of chirality, and stereochemical information, which is particularly challenging when classifying the exact identity of unknown lipids. Open-access databases, such as HMDB, LipidMaps, KEGG, PubChem and ChemSpider are typically used to search for a putative compound match based on experimental accurate mass or most likely molecular formula of an ion in large datasets of previously reported metabolites. However, in case no matches are found, *de novo* structural elucidation is required. Fragmentation spectra based on MS/MS experiments conducted at several collision energies (*e.g.,* 10, 20 and 40 V) when using ESI-MS (LC/CE-MS), as well as standardized (70 eV) electron impact ionization (EI)-MS spectra (GC-MS) play a

fundamental role in metabolite identification since a molecular formula is associated to dozens or hundreds of known chemical structures.[85] Other chemical or biological knowledge/information could be applied to confirm the identification of unknown metabolites, including selective chemical extraction or enzymatic reactivity/transformation, as well as *in silico* modeling of solute retention times or migration times that reflect the physicochemical properties of metabolites based on fundamental chromatographic or electrophoretic principles.[86] The latter strategy is effective at rejecting isobaric or isomeric candidate ions in cases when authentic standards are lacking.

1.2.5 Biological interpretations, biomarker validation and clinical translation

Candidate biomarkers identified following rigorous statistical analyses and structural elucidation in metabolomics are ideally linked to specific metabolic pathways and/or sources of exposure as required for biological or clinical interpretation of significance.[87] Biomarkers can also offer new insights into disease mechanisms and guide therapeutic choices, disease monitoring and even drug discovery.[87,88] Open-access metabolomics databases, such as the Human Metabolome Database (HMDB) provide a comprehensive resource for metabolite concentration ranges in specific biospecimens, and an overview on reported findings in the literature. Various computational tools for network modelling and pathway mapping ranging from metabolic pathway analysis, metabolite set enrichment analysis (MSEA) and mummichog can help reveal associated biological processes and their interaction with known enzymes and genes.[34,89-92] But biological interpretation is often not straightforward and is rather a long-term endeavor of multidisciplinary research, including follow-up studies in animal models, as well as human observational and randomized clinical trials to establish causation. In fact, metabolomics studies fall under two distinct objectives seeking either a broader understanding of disease pathogenesis or biomarker development; the latter is not targeted to explain biology.[93]

Receiver operator characteristic (ROC) curve analysis is considered a standard approach to evaluate the performance of candidate biomarkers in clinical metabolomics research, which is robust to non-parametric data distribution, and has prediction power independent of disease prevalence, unlike prediction accuracy.[94,95] ROC curve models for one or more biomarkers or their ratios should be reported along with 95% confidence intervals (CI) towards a better harmonization with clinicians in order to facilitate biomarker translation "from the benchtop to the bedside".[93,95] ROC curves are based on the frequency of the biomarker or test under investigation of bringing forth true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN) when discriminating between disease cases and healthy controls.[3] Consequently, the graph comprises a visual representation of the sensitivity (TP/TP+FN) of a biomarker on the y-axis against 1 minus its specificity (TN/TN+FP) on the x-axis. Ideally the ROC curve would approach the top left corner of the graph with an area under the curve (AUC) or c-statistic equal to one, signifying a perfect classifier.[1,96] In contrast, AUC equal to 0.5 signifies that the test's prediction performs no better than mere chance. An AUC value from 0.9-1.0 is excellent, 0.8-0.9 is good, and 0.7-0.8 is generally considered fair.[93] The AUC is a neat combined visualization of sensitivity ("positivity in disease") and specificity ("negativity in health") that conveys the overall performance of a biomarker or biomarker panel test.[3] Noteworthy, ROC curves are fundamentally based on the prediction of a binary (*i.e.*, benign versus malignant) not a continuous outcome and they graphically display a continuum of the compromise between maximizing sensitivity (*i.e.*, the test not missing a true cancer diagnosis) and specificity (*i.e.*, test giving out a false cancer diagnosis); choice of the optimal threshold or critical cut-off concentration for a biomarker can be easily deduced from the curve in accordance to the decision makers' priority.[95,96] One example of ROC curve analysis is the composite Framingham risk score for coronary heart disease risk prediction, which combines

several biochemical (HDL, cholesterol and systolic blood pressure) and anthropometric (sex, age and smoking status) measurements with an AUC of 0.75.[1] An additional marker that would enhance this value to 0.77, though a modest contribution, would still be considered significant for increasing the accuracy of cardiovascular disease risk stratification in high-risk individuals.[1]

To move such biomarkers from exploratory to advanced phases, experiments must be repeated on independent/blinded samples drawn from the same population for validation.[95] Only when discovery and validation results are replicated, sufficient evidence can be drawn that the putative biomarker is worth pursuing in advanced clinical testing on a much larger scale.[93] At this point, not only study sample size increases that are collected from across multiple centres, but also more difficult clinical scenarios are included, such as cases with ambiguous diagnosis or comorbidities, so that results are more generalizable to a wider population.[95] Whenever possible, the performance of a new biomarker is compared to conventional symptomatic diagnosis or a currently accepted biomarker(s) lacking adequate specificity. At the advanced stage of biomarker translation, important statistical criteria for assessing biomarker performance include calibration (Hosmer-Lemeshow statistic) and reclassification, in addition to discrimination (ROC curve c-statistic).[1] Calibration describes the degree of agreement between the biomarker-predicted and actual event rates in longitudinal studies with clinical follow-up, whereas reclassification calculates the proportion of patients assigned a correct classification based on the merits of the new biomarker, such as a patient falsely assigned a pre-determined low-risk group would actually transition to a higher risk category.[1,97] Reclassification is of high clinical value in coronary heart disease as it may prompt early and aggressive LDL-lowering pharmacological interventions.[1,98] Unfortunately, many metabolomics biomarker studies stop at the discovery phase with "sometimes undeserved death" from promising early results in pilot studies as seen with many proteomics biomarkers.[87,99]

Furthermore, aside from the biological performance of a biomarker, other considerations must be regarded for successful translation to clinical settings, including the availability of an inexpensive and validated assay for routine screening or diagnostic testing within an accredited laboratory environment. **Table 1.3** summarizes characteristics of an ideal biomarker for peripheral artery disease (PAD).[100] PAD is a complex clinical syndrome categorized under cardiovascular disease, which is a form of atherosclerosis manifested in the lower extremities that leads to muscle pain, impaired walking, infections and tissue loss.[101] In fact, PAD remains poorly recognized by healthcare practitioners with late symptomatic diagnosis that requires confirmation with specialized ankle-brachial index (ABI) measurements.[102] Compared to coronary artery disease, there have been sparse metabolomic studies on PAD to date, where part of this thesis involves a comprehensive metabolomic and lipidomic study for improved differentiation of the two major subtypes of PAD based on a specific panel of serum biomarkers.

Table 1.3 Ideal biomarker for peripheral artery disease (PAD).

	Characteristics of the "ideal" PAD biomarker
1	Sensitivity to presence of PAD (positivity in disease)
2	Specificity to PAD (negativity in health)
3	Correlation to PAD prognosis
4	Correlation to disease-specific features e.g. walking time or the ankle-brachial index (ABI)
5	Minimal influence by confounding factors
6	High-throughput measurement
7	Reproducible and robust measurement
8	Cost effectiveness
9	Easy, non-invasive bio-sampling
10	Complementing current strategies
11	Endorsement by clinicians and patients

Adapted from Cook and Wilson 2010[100]

1.3 Fatty acids in metabolomics and biomarker discovery

1.3.1 Introduction to lipidomics

Lipidomics is a growing subset of metabolomics specifically focused on the comprehensive analysis of lipids.[103] The term was first introduced in 2003 by Spener *et al*. as a distinct '-omics' science defined as "the full characterization of lipid molecular species and of their biological roles with respect to expression of proteins involved in lipid metabolism and function, including gene regulation".[104] Lipids are essential for cellular function as they play critical roles in cell membrane formation, energy storage, cell growth and signaling augmented by their coupling to proteins and sugar moieties.[104,105] Indeed, in a global investigation of the human serum metabolome using 5 different analytical platforms for maximal coverage, the following statement highlights the importance of studying the lipidome: "Overall, the composition of human serum is dominated by diglycerides, triglycerides, phospholipids, fatty acids, steroids and steroid derivatives. This simply reinforces the fact that serum is a key carrier of lipoproteins, fats and hydrophobic nutrients".[18] What groups lipids together is their shared physicochemical properties rather than their distinctive chemical structure, unlike all other biochemical classes; they are insoluble in water and mostly soluble in organic solvents.[105] Therefore, a comprehensive classification system of lipids was necessary to capture the diversity of lipid chemical structures, which was introduced through LIPID MAPS (Lipids Metabolites and Pathways Strategy) database. According to their distinct hydrophobic and hydrophilic components, lipids have been classified into eight major categories, including fatty acids, glycerolipids, glycerophospholipids, sphingolipids, sterol lipids, prenol lipids, saccharolipids and polyketides.[106] Historically, lipids have also been broadly classified into simple and complex lipids, where the former yields two and the latter three or more products upon hydrolysis.[106]

1.3.2 Lipid extraction

Sample pretreatment is critical in lipidomic studies to extract lipophilic compounds from an aqueous biological matrix thereby eliminating interferences caused by water-soluble protein, electrolytes, carbohydrates or other hydrophilic/polar metabolites.[99,105] Ideally, this step should be fast, precise and efficient with high recovery to capture the wide range of lipids expanding over nearly 35 orders of magnitude on the octanol/water coefficient scale.[39,107] Folch extraction was developed in 1951 using chloroform/methanol (2:1 vol) solvent in 20-fold excess and has been the a widely utilized extraction method in lipidomics.[108] It was slightly modified by Bligh and Dyer in 1959 using smaller solvents volumes while maintaining equivalent extraction efficiency.[109] In either "gold standard" liquid extraction methods, the lipid-containing chloroform layer is more dense than the aqueous methanolic layer with precipitated protein matrix partitioning at the interface. In order to retrieve the lower layer, a careful aliquoting is required where minimal amounts of precipitate collected with the chloroform layer would interrupt or hamper the subsequent analysis.[110] In addition, chloroform is a known toxin and carcinogen and subsequently further research was prompted to develop simpler protocols using less harmful solvents for sample extraction in lipidomics.[107,111,112] In 2008, Matyash *et al*.[110] developed an alternative lipid extraction method yielding faster and cleaner extracts with similar or superior recoveries. It utilizes methyl-*tert*-butyl ether (MTBE) as the extraction organic solvent, which due to its lower density relative to water, forms the upper layer during phase separation, that can be collected with ease and with minimal losses due to dripping caused by the lower surface tension of most other organic solvents. Another protocol developed by Lofgren *et al*.[111] involves sample extraction in butanol/methanol followed by a two-phase separation using heptane/acidified ethylacetate. Moreover, one-phase extractions with just methanol or acetonitrile/methanol/water mixtures have also been used for

more selective extraction for certain lipid classes.[107,113] Finally, supercritical fluid extraction (SCF) has also been employed for fast lipidomics analysis, though not widely, due to high instrumentation costs.[114] In SCF, extraction is based on the increased solvation power of carbon dioxide at temperatures and pressures beyond its critical values and thus works as an excellent solvent for solubilizing non-polar compounds, where extraction of more polar lipids, such as fatty acids or phospholipids, requires addition of a solvent modifier (*e.g.,* methanol).[114]

1.3.3 Chemistry and biochemistry of fatty acids

Fatty acids are key molecules for living organisms as the major and fundamental building block of complex lipid and thus have an essential function in energy metabolism and storage, formation of cell membranes and cell mediators, such as prostaglandins and leukotrienes.[115] Fatty acyl structure is characterized by a hydrophobic chain formed of repeating methylene groups with a terminal methyl ($-CH_3$) group (ω-end) and a polar carboxylic head group and their synthesis occurs through malonyl-CoA elongation of an acetyl-CoA primer.[106] Fatty acids can be classified according to carbon chain length (short, medium, long and very long), degree of unsaturation (saturated, monounsaturated and polyunsaturated), and position of unsaturation (*e.g.*, n-3, n-6, and n-9). Their nomenclature commonly follows a "C number of carbon/number of double bond" scheme, where the position of the first double bond from the terminal methyl (x) is indicated by "ω-x" or by "n-x".[116] For instance, stearic acid is denoted as 18:0 while linoleic acid is denoted as 18:2 n-6 with a *cis* (Z) double bond geometry and its isomer linoelaidic acid with a *trans* (E) double bond geometry. Moreover, in the case of branched-chain fatty acids, the position of branching is indicated by "iso-" or "anteiso-" suffix for terminal isopropyl and isobutyl groups, respectively. Indeed, there are over 7 branched-chain fatty acids measured in meat, dairy and various fermented

food products (*e.g.,* sauerkraut, miso), as well as microflora from the newborn gastrointestinal tract, and human milk.[117]

Various biospecimens can be collected for determination of fatty acid composition and concentration, including adipose tissue from the buttock or abdomen, red blood cells (erythrocyte phospholipids), total plasma lipids or individual plasma lipid fractions ranging from non-esterified fatty acids (NEFA), phospholipids, cholesteryl esters, and triglycerides - in decreasing order of fatty acids turnover time from years to hours.[4,31,118] Less often measured biological samples include platelets, skin, breast milk, semen, buccal cells, neutrophils, monocytes and lymphocytes.[31] Moreover, some of these specimens are sensitive to fasting status, including plasma/serum NEFAs and triglycerides that undergo rapid changes in concentrations within hours of fat/meal consumption (*i.e.,* post-prandial).[31,119] Regrettably, detailed information of specimen collection and patient recruitment details (*e.g.,* fasting status, delays to storage, sample workup conditions, and analytical methodology) is often missing from original papers and sometimes difficult to deduce, causing significant ambiguity in reporting. Additionally, there are two formats used for reporting fatty acids concentrations, absolute molar concentrations, and more commonly, fatty acids profiling.[21] The latter involves reporting of individual fatty acids as a proportion relative to all other fatty acids measured in the specimen (*e.g.,* wt%) and is popular in nutritional studies.[21] Relative concentrations offer the advantage of measuring fatty acids in the context of each other as these molecules generally tend to compete for incorporation into various lipids, however this approach is subject to inter-laboratory variation with a lack of a unified agreement on total and unknown/unidentified fatty acids measured.[21] This poses a difficulty in comparing results of different studies and laboratories and could also mask significant biological differences in fatty acids between individuals.[4,120] Clearer reporting and greater harmonization of validated analytical

methods for reliable FA determination is critical to better guide nutritional sciences and evidence-based public health policies. Such efforts towards standardization include comprehensive fatty acids databases with absolute concentrations in large populations conducted for some lipid pools, paving the way to establish reference ranges for routine clinical testing of fatty acids for chronic disease prevention as done for LDL, HDL, total cholesterol and triglycerides.[4,121,122]

1.3.4 Clinical significance/current status of fatty acids metabolomics studies

Consequently, measuring the composition of circulating fatty acids in blood would serve both as a readily accessible and objective biomarker of ingested fats reflecting nutrition status and complex dietary patterns, and may also serve as potential diagnostic, prognostic or risk biomarkers for chronic diseases associated with dyslipidemia and obesity. This can help bridge a current fundamental gap between epidemiology and nutrition sciences and demands a highly multidisciplinary and collaborative research approach of clinical and bioanalytical chemists, clinicians, dietitians, epidemiologists and statisticians. The degree of accuracy or truth of discoveries reported in observation studies relating diets/dietary fat to various diseases such as cancer, diabetes or cardiovascular disease, is dependent on the degree of accuracy and precision in measuring diet intake. However, a "heart-healthy diet" based on promotion of low saturated fat/cholesterol diets (plus high carbohydrate) introduced in the 1980s in the US and the UK impacting 276 million people was ultimately based on flawed results from secondary studies and not randomised controlled trials.[123] Consequently, current guidelines for reducing overall as well as saturated fat consumption to reduce CHD are being questioned and more research is undergoing to disentangle the effect of saturated fat on health while considering complex food matrices and overall macronutrient distribution.[124] **Table 1.4** is a summary of metabolomics studies that

Table 1.4 A review of published metabolomics studies focused on fatty acids analysis in relation to disease.

	Study	Subjects #	Biological fluid/fraction	Analytical methodology	Comments	Major Findings
1	Forouhi et al. (2014)[135]	27,296	Plasma: PL	Folch extraction; robotic SPE fractionation; GC-FID	3 parallel analytical systems	14:0, 16:0, 18:0 positively; 15:0, 17:0, 20:0, 22:0, 23:0, and 24:0 inversely associated with T2DM
2	Khaw et al (2012)[136]	7354	Plasma: PL Nested case con	Folch extraction; robotic SPE fractionation; GC-FID	3 parallel analytical systems	14:0, 16:0, 18:0 positively; n-6, 15:0, 17:0, inversely associated with CHD
3	Würtz et al. (2015)[134]	7256	Serum: Total	Minimal sample prep; NMR	combined MUFA signal	MUFA positively; n-6 FA and DHA negatively correlated with cardiovascular events
4	Bisgaard et al. (2016)[128]	736	Whole blood, breast milk	No prior extraction direct microwave transesterification with BF_3/methanol; GC-FID	significantly lower FA estimates by microwave transesterification	Third trimester DHA and EPA supplementation linked to reduced risk of asthma in offspring
5	Santaren et al. (1995)[137]	555	Serum: Total	Folch extraction; GC-FID	35 FA measured	14:0 and 16:0 positively; 15:0, 20:0, 22:0 inversely related to proinflammatory markers
6	Kurotani et al. (2012)[138]	437	Serum: PL, CE	Folch extraction; TLC fractionation; GC-FID	No isomeric resolution	18:0, 16:1, and 20:3 n-6 positively associated; 18:2 n-6 negatively associated with insulin resistance
7	Allalou et al. (2016)[139]	244	Plasma: NFFA	Hexane extraction; GC-MS	13 FA measured	16:1 significantly altered with incident T2DM
8	Yi et al. (2007)[140]	123	Plasma: EFA, NEFA	Methanolic KOH to methylate EFA, H_2SO_4-CH_3OH to methylate NEFA; GC-MS	Cis/trans isomers well resolved	Distinct differences in NEFA profiles between type 2 diabetic patients and controls

9	Suhre et al. (2010)[141]	100	Plasma: sample prep not reported	Metabolomics providers Biocrates Life Sciences and Metabolon Inc.	Isomeric resolution of 18:2 but not 18:3	Medium chain-length fatty acids and 20:4 decreased in T2DM
10	Jenkins et al. (2017)[142]	90	Plasma: NEFA	Direct infusion MS		15:0 directly correlated with dietary intake; 17:0 linked to metabolic disease
11	Dudzik et al. (2017)[32]	48	Plasma: NFFA	Methoxymation followed by silylation for derivatization; GC-MS		Second trimester C18:0 with discriminative power for T2DM post GDM
12	Oda et al. (2005)[143]	42	Serum: Total	Measurements conducted by a commercial laboratory center in Japan (SRL Inc.)	Measurement procedures not consistently defined	24:1 may have preventive effects on metabolic disorders
13	Manfredi et al. (2019)[144]	35	Serum: Total	Hexane extraction; GC-MS	37 FA measured	22:6, 20:5, 18:2 downregulated in IBD
14	Perreault et al. (2014)[145]	30	Serum: Total, PL, TAG	Folch extraction; TLC fractionation; GC-FID	28 FA analyzed Isomeric resolution reported for 18:1	14:0, 18:0 in MHO resembled that of LH 14:0 and 16:0 positively associated; 18:0 inversely associated with inflammation
15	Volk et al. (2014)[146]	16	Plasma: TAG, CE, PL	Methodology not mentioned; only a reference for GC	5 FA measured; isomeric resolution not reported	Increased dietary saturated fat not accumulated in plasma during carbohydrate restriction. 16:1 tracked incrementally with dietary carbohydrate
16	Weitkunat et al. (2017)[147]	16	Serum: PL	MTBE extraction; SPE fractionation; GC-MS	Superior isomeric resolution	OCFA: possible biomarker for fiber and dairy intakes; inverse relation to diabetes

specifically investigate the role of fatty acids in chronic diseases as well as their relationship to various health outcomes with some major studies discussed here. Firstly, ω-3 fatty acids are at the center of an on-going controversial debate on their putative health benefits and have been tested in several clinical intervention studies with ambiguous results.[125] For instance, a meta-analysis of 20 RCT studies found no association between ω-3 supplementation (fish oil) or dietary counseling and lower risk for cardiovascular events.[125] However, ω-3 dosage regimes tested were low and variable (0.2-1.8 g/day), most studies were targeted for secondary prevention, and circulating fatty acids concentrations were not directly measured to assess dietary compliance or differences in metabolism.[125] In contrast, the recently published ANCHOR trial reported a significant reduction of serum triglycerides in patients already on statins when supplemented with 4 g/day of a highly pure form of eicosapentaenoic acid (EPA).[126,127] On a different note, the Copenhagen Prospective Studies on Asthma in Childhood ($COPSAC_{2010}$) double-blind randomized controlled trial of 736 pregnant women found that 2.4 g/day ω-3 supplementation in the third trimester reduced asthma and wheeze in offspring followed until 5 years of age.[128,129] The greatest benefit was found in children whose mothers' EPA and docosahexaenoic acid (DHA) concentrations were in the lowest tertile at baseline.[128] Conversely, their findings suggested a maximum threshold for baseline ω-3 fatty acid plasma concentrations beyond which supplementation was ineffective.[128] Together, these studies suggest there may be notable benefits of DHA and EPA supplementation within specific populations (*i.e.,* nutrient deficient), when using optimal dosages and that quantitative determination of fatty acids can stratify health status and tailor dietary/supplemental recommendations while allowing for improved treatment monitoring, including dietary adherence. A similar debate relates to the association of saturated fat with cardiovascular disease, where meta-analysis rely on epidemiological studies that rely on subjective self-report tools without direct

analysis of circulating fatty acids concentrations.[130,131] In contrast, an outstanding meta-analysis used an individual-level pooled analysis of 30 cohort studies of participants with either circulating or adipose tissue relative concentrations of arachidonic acid and linoleic acid (essential ω-6 PUFA found in nuts, vegetable oils and whole grains) to investigate their association to cardiovascular disease and mortality with an editorial commentary titled "are we getting closer to the truth?".[132,133] Another NMR-based metabolomic study of 7256 subjects computed a risk score for predicting cardiovascular disease based on serum phenylalanine and monounsaturated fatty acids (positive correlation), as well as ω-6 fatty acids and DHA (negative correlation), which was equivalent to conventional risk factors following extensive validation in two independent cohorts (n = 6185).[134] The EPIC study of more than 25,000 participants, one of the largest with hydrolyzed serum fatty acids concentrations from the phospholipid fraction, revealed differential (opposing) directions of the relationship of even and odd-chain saturated fatty acids with type-2 diabetes mellitus and coronary heart disease.[135,136] In conclusion, methods that enable reliable quantitative determination of fatty acids have promising clinical applications that complement standard blood lipid panels. However, there is urgent need for the development of comprehensive and high throughput screening methods for fatty acids in support of large-scale epidemiological studies and clinical trials.

1.4 Thesis motivation: Metabolomics for dietary and cardiometabolic risk assessments

Metabolomics has an emerging role in the investigation of physiological and pathophysiological processes for clinical research by using advanced analytical methods to identify and quantify metabolites and lipids in complex biological samples. The validation and translation of biomarkers, in turn, play an important role in clinical decision-making and treatment algorithms that can also guide public health policies. However, challenges pertaining to biological sampling,

analytical reporting, biomarker clinical utility and method validation hamper effective translation of scientific discoveries into clinical medicine.

Fatty acids are key molecules for cell membranes, cellular fuel and signaling, which also may serve as specific biomarkers of dietary fat intake in women during pregnancy, particularly relevant to the Developmental Origins of Health and Disease (DOHaD). DOHaD is a field of research based on the hypothesis of "fetal origins of adult disease" where a mother's intrauterine milieu dictated by her dietary/environmental exposures during a critical window of development can impact the health of her child later in life.[148] Epidemiological observations have well supported the DOHaD theory, where nutritional deficiencies and environmental exposures are two major programming stimuli that modulate chronic disease susceptibility in the offspring.[149,150] Consequently, there is great need for new analytical platforms that allow for rapid biomonitoring and reliable assessment of maternal exposures. In this context, the work in this thesis aims to develop MSI-NACE-MS as a robust method for high throughput determination for serum fatty acids without fractionation, hydrolysis and chemical derivatization, thus, optimal for large-scale clinical or epidemiological studies (*Chapter II*). Furthermore, rigorous validation of MSI-NACE-MS in conjunction with multiple reaction monitoring was extended for targeted yet sensitive analysis of nanomolar levels of synthetic PFASs, a class of persistent organic pollutants and endocrine-disrupting chemicals linked to adverse health outcomes in children from prenatal exposures (*Chapter III*). Next, a methodologically underreported lipid fraction, serum NEFAs, was validated as a convenient lipid biomarker of dietary fat when using MSI-NACE-MS, which has utility for accurate assessment of fish/fish oil and full-fat dairy intake in women as compared to self-report tools (*Chapter IV*). Finally, comprehensive profiling of polar/ionic metabolites and non-polar/ionic lipids using MSI-(NA)CE-MS was applied for the discovery of novel metabolic

signatures for differentiation of chronic limb-threatening ischemia from earlier stages of intermittent claudication, shedding light on the role of circulating fatty acids as possible prognostic biomarkers of PAD progression in older persons at high risk for limb amputations and poor long-term survivorship (*Chapter V*).

1.4.1 High throughput analysis of nonesterified fatty acids in serum

Fatty acids are clinically relevant metabolites and essential nutrients due to their myriad roles relevant to human health. GC remains the gold standard method for fatty acids determination as it offers excellent resolution, but it suffers from low sample throughput due to complicated sample workup, pre-column chemical derivatization procedures, and long elution times. *Chapter II* introduces MSI-NACE-MS as a high throughput method for reliable determination of serum NEFAs following a simple MTBE extraction procedure. This multiplexed separation method enables the analysis of seven serum extracts simultaneously within a single run (< 4 min/sample), including a QC that monitors for long-term system drift and allows for robust batch correction. Rigorous method optimization was performed to enhance sensitivity with sub-micromolar detection limits with good linearity and intermediate precision. Also, a cross-platform comparison of MSI-CE-MS relative to GC-MS demonstrated good mutual agreement for NEFA determination with minimal bias. Overall, MSI-NACE-MS offers a rapid and low-cost platform for quantitative determination of 20 (or more) fatty acids in serum extracts on a single instrument (> 200 samples/day) as required for large-scale epidemiological or clinical studies.

1.4.2 Rapid screening of perfluoroalkyl substances exposure in serum

There is growing concern on the harmful impact of PFASs exposures to human health due to their ubiquitous prevalence as synthetic surfactants in commercial products worldwide. These persistent

organic pollutants can function as endocrine disrupting compounds during critical stages of fetal development, which contribute to adverse birth outcomes and chronic disease risk later in life. In *Chapter III,* we further develop MSI-NACE-MS/MS as a high throughput yet targeted approach for biomonitoring of serum PFASs when using a simple MTBE extraction for sample cleanup and enrichment. Separation and ionization conditions were optimized to quantify low nanomolar concentration levels of perfluorooctanoic acid (PFOA) and perfluorooctanesulfonic acid (PFOS) in serum extracts when using multiple reaction monitoring with a triple quadrupole mass analyzer under negative ion mode conditions. Separation of these two PFASs was achieved with excellent sample throughput (< 3 min/sample) when using a serial injection of seven samples in a single run in MSI-NACE-MS/MS with low nanomolar detection limits. This method also demonstrated good technical precision over 3 consecutive days for reliable detection and quantification of PFOA and PFOS in maternal serum samples collected prior to 2009 marking the beginning of the global PFAS production phase out. In contrast, lower PFAS exposures were measured in a subset of maternal serum samples collected after 2009. Overall, this work is a proof-of-concept for a novel approach for biomonitoring of maternal exposures to harmful PFASs as required for new advances in public health and risk assessment of chronic disease in children later in life.

1.4.3 Investigation of serum NEFAs as dietary biomarkers of fat intake in women

Fatty acids are dietary components that have long been implicated in human health and chronic disease risk. However, fatty acids analysis encompasses diverse fractions derived from blood, erythrocytes or adipose tissue. To date, serum phospholipid and total lipid fractions are often used in epidemiological and nutrition studies with a notable gap regarding the clinical utility of more convenient lipid pools as putative dietary biomarkers, such as serum NEFA. In *Chapter IV*, we applied our newly validated MSI-NACE-MS method for serum fatty acids determination to

examine the utility of certain circulating NEFAs as dietary biomarkers of fat intake in women. We aimed to identify certain serum NEFAs correlated to habitual intake of specific foods in an observational cohort of pregnant women and track their changes in women following high dose ω-3 fatty acids supplementation in a randomized placebo-controlled intervention trial. In the cross-sectional analysis, pregnant women were selected with contrasting eating patterns based on a diet quality index score from a food frequency questionnaire. Serum ω-3 fatty acids concentrations correlated significantly with total ω-3 daily intake, notably EPA as its NEFA, whereas non-esterified 15:0 and 14:0 had the strongest correlation to full-fat dairy intake. As for the intervention study, a 2.5-fold significant increase in serum ω-3 NEFA concentrations from baseline was measured within 28 days for women following ω-3 supplementation similar to independent erythrocyte phospholipid fatty acids measurements. For the first time, we demonstrate through a combination of observational and intervention studies that direct NEFA analysis offers a reliable approach for assessment of dietary fat without serum fractionation or hydrolysis. Our findings endorse the objective use of specific serum NEFAs for investigating the role of maternal nutrition or supplementation on birth outcomes.

1.4.4 Metabolomics study of peripheral artery disease

PAD is a form of atherosclerosis manifested in the lower extremities that leads to symptoms as painful walking, fatigue, muscle cramping and tissue loss that is classified as intermittent claudication (IC) at early stages of disease.[101] PAD is often undiagnosed despite a high prevalence of 10-20% in older persons. Furthermore, some patients progress to the severe, end-stage form known as chronic limb-threatening ischemia (CLTI) that is characterized by rest pain, non-healing ischemic ulcers, and gangrene.[151] A better understanding of CLTI progression can reduce delays in surgical interventions, reduce amputations, improve quality of life, and decrease mortality.

Chapter V describes the metabolic phenotype of PAD patients using a comprehensive metabolomics and lipidomics approach using MSI-(NA)CE-MS with full-scan data acquisition. A total of 60 patients were recruited and stratified based on the Rutherford classification into CLTI, IC, and non-PAD controls; also, patients with diabetes, history of cancer, deep vein thrombosis, and acute coronary syndrome were excluded from this study to reduce confounding. Compared to non-PAD controls (n = 20), PAD patients had lower serum concentrations of creatine, histidine, lysine, oxoproline, monomethylarginine, as well as higher circulating phenylacetylglutamine ($p <$ 0.05). Moreover, CLTI cases exhibited higher serum concentrations of carnitine, creatinine, cystine and trimethylamine-*N*-oxide along with lower circulating fatty acids relative to well-matched IC patients. Most serum metabolites associated with PAD progression were also correlated with the ankle-brachial index (ABI); a PAD-specific diagnostic tool that is usually not available at the primary care level. Importantly, the ratio of stearic acid to carnitine, and arginine to propionylcarnitine differentiated CLTI from IC with good accuracy (AUC = 0.87). These metabolic perturbations are related to muscle energy deficits, vascular remodeling, myopathic ischemia, inflammation and oxidative stress and provide novel biochemical insights into the pathophysiology of CLTI.

1.5 References

1. Gilstrap LG, Wang TJ. Biomarkers and cardiovascular risk assessment for primary prevention: An Update. *Clin Chem*. 2012;58:72-82.
2. Jain KK. *The handbook of biomarkers*. Springer New York; 2017.
3. Park SH, Goo JM, Jo C-H. Receiver operating characteristic (ROC) curve: practical review for radiologists. *Korean J Radiol*. 2004;5:11-18.
4. Abdelmagid SA, Clarke SE, Nielsen DE, et al. Comprehensive profiling of plasma fatty acid concentrations in young healthy canadian adults. Aspichueta P, ed. *PLOS ONE*. 2015;10:e0116195.

5. DiBattista A, McIntosh N, Lamoureux M, Al-Dirbashi OY, Chakraborty P, Britz-McKibbin P. Temporal signal pattern recognition in mass spectrometry: a method for rapid identification and accurate quantification of biomarkers for inborn errors of metabolism with quality assurance. *Anal Chem*. 2017;89:8112-8121.

6. Wild J, Shanmuganathan M, Hayashi M, Potter M, Britz-McKibbin P. Metabolomics for improved treatment monitoring of phenylketonuria: urinary biomarkers for non-invasive assessment of dietary adherence and nutritional deficiencies. *Analyst*. 2019;144:6595-6608.

7. FDA-NIH Biomarker Working Group. BEST (Biomarkers, EndpointS, and other Tools) Resource [Internet]. Silver Spring (MD): Food and Drug Administration (US); 2016-. Monitoring Biomarker. 2016 Dec 22. Available from: https://www.ncbi.nlm.nih.gov/books/NBK402282/Co-published by National Institutes of Health (US), Bethesda (MD)

8. Ledermann JA, El-Khouly F. PARP inhibitors in ovarian cancer: clinical evidence for informed treatment decisions. *Br J Cancer*. 2015;113:S10-S16.

9. Maruvada P, Lampe JW, Wishart DS, et al. Perspective: Dietary biomarkers of intake and exposure—exploration with omics approaches. *Adv Nutr*. 2020;11:200-215.

10. Wishart DS. Metabolomics: applications to food science and nutrition research. *Trends Food Sci Technol*. 2008;19:482-493.

11. Wishart DS. Metabolomics for investigating physiological and pathophysiological processes. *Physiol Rev*. 2019;99:1819-1875.

12. Wang Z, Klipfell E, Bennett BJ, et al. Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. *Nature*. 2011;472:57-63.

13. Wellington N, Shanmuganathan M, de Souza RJ, et al. Metabolic trajectories following contrasting Prudent and Western diets from food provisions: identifying robust biomarkers of short-term changes in habitual diet. *Nutrients* 2019;11:2407.

14. Ioannidis JPA. The challenge of reforming nutritional epidemiologic research. *JAMA*. 2018;320:969-970.

15. Prentice RL. Dietary assessment and opportunities to enhance nutritional epidemiology evidence. *Ann Intern Med*. 2020;172:354.

16. Chu SH, Huang M, Kelly RS, et al. Integration of metabolomic and other omics data in population-based study designs: an epidemiological perspective. *Metabolites*. 2019;9:117.

17. Dunn WB, Broadhurst DI, Atherton HJ, Goodacre R, Griffin JL. Systems level studies of mammalian metabolomes: the roles of mass spectrometry and nuclear magnetic resonance spectroscopy. *Chem Soc Rev*. 2011;40:387-426.

18. Psychogios N, Hau DD, Peng J, et al. The human serum metabolome. Flower D, ed. *PLoS ONE*. 2011;6:e16957.

19. Poste G. Biospecimens, biomarkers, and burgeoning data: the imperative for more rigorous research standards. *Trends Mol Med*. 2012;18:717-722.

20. Lu W, Su X, Klein MS, Lewis IA, Fiehn O, Rabinowitz JD. Metabolite measurement: pitfalls to avoid and practices to follow. *Annu Rev Biochem*. 2017;86:277-304.

21. Brenna JT, Plourde M, Stark KD, Jones PJ, Lin Y-H. Best practices for the design, laboratory analysis, and reporting of trials involving fatty acids. *Am J Clin Nutr*. 2018;108:211-227.

22. Arab L. Biomarkers of fat and fatty acid intake. *J Nutr*. 2003;133:925S-932S.

23. Beckonert O, Keun HC, Ebbels TMD, et al. Metabolic profiling, metabolomic and metabonomic procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts. *Nat Protoc*. 2007;2:2692-2703.

24. Yu Z, Kastenmüller G, He Y, et al. Differences between human plasma and serum metabolite profiles. Oresic M, ed. *PLoS ONE*. 2011;6:e21230.

25. Dragsted LO, Gao Q, Scalbert A, et al. Validation of biomarkers of food intake—critical assessment of candidate biomarkers. *Genes Nutr*. 2018;13:14.

26. Domazet SL, Grøntved A, Timmermann AG, Nielsen F, Jensen TK. Longitudinal associations of exposure to perfluoroalkylated substances in childhood and adolescence and indicators of adiposity and glucose metabolism 6 and 12 years later: The European Youth Heart Study. *Diabetes Care*. 2016;39:1745-1751.

27. Jensen TK, Andersen LB, Kyhl HB, Nielsen F, Christesen HT, Grandjean P. Association between perfluorinated compound exposure and miscarriage in danish pregnant women. Szecsi PB, ed. *PLoS ONE*. 2015;10:e0123496.

28. Ramautar R, Demirci A, Jong GJ de. Capillary electrophoresis in metabolomics. *TrAC Trends Anal Chem*. 2006;25:455-466.

29. The Human Serum Metabolome (HUSERMET) Consortium, Dunn WB, Broadhurst D, et al. Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. *Nat Protoc*. 2011;6:1060-1083.

30. Metherel AH, Armstrong JM, Patterson AC, Stark KD. Assessment of blood measures of n-3 polyunsaturated fatty acids with acute fish oil supplementation and washout in men and women. *Prostaglandins Leukot Essent Fatty Acids*. 2009;81:23-29.

31. Hodson L, Skeaff CM, Fielding BA. Fatty acid composition of adipose tissue and blood in humans and its use as a biomarker of dietary intake. *Prog Lipid Res*. 2008;47:348-380.

32. Dudzik D, Zorawski M, Skotnicki M, et al. GC–MS based gestational diabetes mellitus longitudinal study: identification of 2-and 3-hydroxybutyrate as potential prognostic biomarkers. *J Pharm Biomed Anal*. 2017;144:90-98.

33. Saoi M, Li A, McGlory C, et al. Metabolic perturbations from step reduction in older persons at risk for sarcopenia: plasma biomarkers of abrupt changes in physical activity. *Metabolites*. 2019;9:134.

34. Macedo AN, Mathiaparanam S, Brick L, et al. The sweat metabolome of screen-positive cystic fibrosis infants: revealing mechanisms beyond impaired chloride transport. *ACS Cent Sci*. 2017;3:904-913.

35. Godzien J, Alonso-Herranz V, Barbas C, Armitage EG. Controlling the quality of metabolomics data: new strategies to get the best out of the QC sample. *Metabolomics*. 2015;11:518-528.

36. Wehrens R, Hageman JosA, van Eeuwijk F, et al. Improved batch correction in untargeted MS-based metabolomics. *Metabolomics*. 2016;12:88.

37. Brunius C, Shi L, Landberg R. Large-scale untargeted LC-MS metabolomics data correction using between-batch feature alignment and cluster-based within-batch signal intensity drift correction. *Metabolomics*. 2016;12:173.

38. Wishart DS. Emerging applications of metabolomics in drug discovery and precision medicine. *Nat Rev Drug Discov*. 2016;15:473-484.

39. Cajka T, Fiehn O. Toward merging untargeted and targeted methods in mass spectrometry-based metabolomics and lipidomics. *Anal Chem*. 2016;88:524-545.

40. Kristoff CJ, Bwanali L, Veltri LM, et al. Challenging bioanalyses with capillary electrophoresis. *Anal Chem*. 2020;92:49-66.

41. DiBattista A, Rampersaud D, Lee H, Kim M, Britz-McKibbin P. High throughput screening method for systematic surveillance of drugs of abuse by multisegment injection–capillary electrophoresis–mass spectrometry. *Anal Chem*. 2017;89:11853-11861.

42. de Macedo AN, Irfan Yasin Jiwa M, Macri J, Belostotsky V, Hill S, Britz-McKibbin P. Strong anion determination in biological fluids by capillary electrophoresis for clinical diagnostics. *Anal Chem*. 2013;85:11112-11120.

43. Macedo AN de, Teo K, Mente A, et al. A robust method for iodine status determination in epidemiological studies by capillary electrophoresis. *Anal Chem*. 2014;86:10010-10015.

44. Harada S, Hirayama A, Chan Q, et al. Reliability of plasma polar metabolite concentrations in a large-scale cohort study using capillary electrophoresis-mass spectrometry. Motta A, ed. *PLOS ONE*. 2018;13:e0191230.

45. Kuehnbaum NL, Britz-McKibbin P. New advances in separation science for metabolomics: resolving chemical diversity in a post-genomic era. *Chem Rev*. 2013;113:2437-2468.

46. Buko A. Capillary electrophoresis mass spectrometry based metabolomics. *J Appl Bioanal*. 2017;3:5-20.

47. Jorgenson JW, Lukacs KD. Capillary zone electrophoresis. *Science*; 1983;222:266.

48. Olivares JA, Nguyen NT, Yonker CR, Smith RD. On-line mass spectrometric detection for capillary zone electrophoresis. *Anal Chem*. 1987;59:1230-1232.

49. Shanmuganathan M, Macklai S, Barrenas Cárdenas C, et al. High-throughput and comprehensive drug surveillance using multisegment injection-capillary electrophoresis-mass spectrometry. *J Vis Exp*. 2019;146:58986.

50. Saoi M, Percival M, Nemr C, Li A, Gibala M, Britz-McKibbin P. Characterization of the human skeletal muscle metabolome for elucidating the mechanisms of bicarbonate ingestion on strenuous Interval exercise. *Anal Chem*. 2019;91:4709-4718.

51. Petersson MA, Hulthe G, Fogelqvist E. New sheathless interface for coupling capillary electrophoresis to electrospray mass spectrometry evaluated by the analysis of fatty acids and prostaglandins. *J Chromatogr A*. 1999;854:141-154.

52. Moini M. Simplifying CE–MS Operation. 2. Interfacing low-flow separation techniques to mass spectrometry using a porous tip. *Anal Chem*. 2007;79:4241-4246.

53. Wojcik R, Dada OO, Sadilek M, Dovichi NJ. Simplified capillary electrophoresis nanospray sheath-flow interface for high efficiency and sensitive peptide analysis: Capillary electrophoresis electrospray interface. *Rapid Commun Mass Spectrom*. 2010;24:2554-2560.

54. Gulersonmez MC, Lock S, Hankemeier T, Ramautar R. Sheathless capillary electrophoresis-mass spectrometry for anionic metabolic profiling: Liquid Phase Separations. *Electrophresis*. 2016;37:1007-1014.

55. Jiang Y, He M-Y, Zhang W-J, et al. Recent advances of capillary electrophoresis-mass spectrometry instrumentation and methodology. *Chin Chem Lett*. 2017;28:1640-1652.

56. Kuehnbaum NL, Kormendi A, Britz-McKibbin P. Multisegment injection-capillary electrophoresis-mass spectrometry: a high-throughput platform for metabolomics with high data fidelity. *Anal Chem*. 2013;85:10664-10669.

57. Cheng J, Wang L, Liu W, Chen DDY. Quantitative nonaqueous capillary electrophoresis–mass spectrometry method for determining active ingredients in plant extracts. *Anal Chem*. 2017;89:1411-1415.

58. Matysik FM. Special aspects of detection methodology in nonaqueous capillary electrophoresis. *Electrophoresis*. 2002;23:400-407.

59. Yamamoto M, Ly R, Gill B, Zhu Y, Moran-Mirabal J, Britz-McKibbin P. Robust and high-throughput method for anionic metabolite profiling: preventing polyimide aminolysis and capillary breakages under alkaline conditions in capillary electrophoresis-mass spectrometry. *Anal Chem*. 2016;88:10710-10719.

60. Mokaddem M, Gareil P, Belgaied J-E, Varenne A. A new insight into suction and dilution effects in capillary electrophoresis coupled to mass spectrometryvia an electrospray ionization interface. Part I-Suction effect. *Electrophoresis*. 2008;29:1957-1964.

61. Azab S, Ly R, Britz-McKibbin P. Robust method for high-throughput screening of fatty acids by multisegment injection-nonaqueous capillary electrophoresis–mass spectrometry with stringent quality control. *Anal Chem*. 2019;9:2329-2336.

62. Chong J, Soufan O, Li C, et al. MetaboAnalyst 4.0: towards more transparent and integrative metabolomics analysis. *Nucleic Acids Res*. 2018;46:W486-W494.

63. Courant F, Antignac J-P, Dervilly-Pinel G, Le Bizec B. Basics of mass spectrometry based metabolomics. *Proteomics*. 2014;14:2369-2388.

64. Mahieu NG, Genenbacher JL, Patti GJ. A roadmap for the XCMS family of software solutions in metabolomics. *Curr Opin Chem Biol*. 2016;30:87-93.

65. Mahieu NG, Patti GJ. Systems-level annotation of a metabolomics data set reduces 25 000 features to fewer than 1000 unique metabolites. *Anal Chem*. 2017;89:10397-10406.

66. Yao C-H, Liu G-Y, Yang K, Gross RW, Patti GJ. Inaccurate quantitation of palmitate in metabolomics and isotope tracer studies due to plastics. *Metabolomics*. 2016;12:143.

67. Katajamaa M, Miettinen J, Oresic M. MZmine: toolbox for processing and visualization of mass spectrometry based molecular profile data. *Bioinformatics*. 2006;22:634-636.

68. Smith CA, Want EJ, O'Maille G, Abagyan R, Siuzdak G. XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Anal Chem*. 2006;78:779-787.

69. Kamleh MA, Ebbels TMD, Spagou K, Masson P, Want EJ. Optimizing the use of quality control samples for signal drift correction in large-scale urine metabolic profiling studies. *Anal Chem*. 2012;84:2670-2677.

70. Hrydziuszko O, Viant MR. Missing values in mass spectrometry based metabolomics: an undervalued step in the data processing pipeline. *Metabolomics*. 2012;8:161-174.

71. Armstrong LE. Hydration assessment techniques. *Nutr Rev*. 2005;63:S40-S54.

72. Chetwynd AJ, Abdul-Sada A, Holt SG, Hill EM. Use of a pre-analysis osmolality normalisation method to correct for variable urine concentrations and for improved metabolomic analyses. *J Chromatogr A*. 2016;1431:103-110.

73. Dieterle F, Ross A, Schlotterbeck G, Senn H. Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. Application in ^{1}H NMR metabonomics. *Anal Chem*. 2006;78:4281-4290.

74. Van den Berg RA, Hoefsloot HC, Westerhuis JA, Smilde AK, Van der Werf MJ. Centering, scaling, and transformations: improving the biological information content of metabolomics data. *BMC Genomics*. 2006;7:142.

75. Yazici B, Yolacan S. A comparison of various tests of normality. *J Stat Comput Simul*. 2007;77:175-183.

76. Goodacre R, Broadhurst D, Smilde AK, et al. Proposed minimum reporting standards for data analysis in metabolomics. *Metabolomics*. 2007;3:231-241.

77. Lazar AG, Romanciuc F, Socaciu MA, Socaciu C. Bioinformatics tools for metabolomic data processing and analysis using untargeted liquid chromatography coupled with mass spectrometry. *Bull Univ Agric Sci Vet Med Cluj-Napoca Anim Sci Biotechnol*. 2015;72:103-115.

78. Saccenti E, Hoefsloot HCJ, Smilde AK, Westerhuis JA, Hendriks MMWB. Reflections on univariate and multivariate analysis of metabolomics data. *Metabolomics*. 2014;10:361-374.

79. Glickman ME, Rao SR, Schultz MR. False discovery rate control is a recommended alternative to Bonferroni-type adjustments in health studies. *J Clin Epidemiol*. 2014;67:850-857.

80. Vinaixa M, Samino S, Saez I, Duran J, Guinovart JJ, Yanes O. A Guideline to Univariate Statistical Analysis for LC/MS-Based Untargeted Metabolomics-Derived Data. *Metabolites*. 2012;2:775-795.

81. Alonso A, Marsal S, Julia A. Analytical Methods in Untargeted Metabolomics: State of the Art in 2015. *Front Bioeng Biotechnol*. 2015;3:20.

82. Gromski PS, Muhamadali H, Ellis DI, et al. A tutorial review: metabolomics and partial least squares-discriminant analysis – a marriage of convenience or a shotgun wedding. *Anal Chim Acta*. 2015;879:10-23.

83. Szymańska E, Saccenti E, Smilde AK, Westerhuis JA. Double-check: validation of diagnostic statistics for PLS-DA models in metabolomics studies. *Metabolomics*. 2012;8:3-16.

84. Sumner LW, Amberg A, Barrett D, et al. Proposed minimum reporting standards for chemical analysis: Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). *Metabolomics*. 2007;3:211-221.

85. Dunn WB, Erban A, Weber RJM, et al. Mass appeal: metabolite identification in mass spectrometry-focused untargeted metabolomics. *Metabolomics*. 2013;9:44-66.

86. D'Agostino LA, Lam KP, Lee R, Britz-McKibbin P. Comprehensive plasma thiol redox status determination for metabolomics. *J Proteome Res*. 2011;10:592-603.

87. Koulman A, Lane GA, Harrison SJ, Volmer DA. From differentiating metabolites to biomarkers. *Anal Bioanal Chem*. 2009;394:663-670.

88. Goodsaid FM, Frueh FW, Mattes W. Strategic paths for biomarker qualification. *Toxicology*. 2008;245:219-223.

89. Xia J, Wishart DS. MetPA: a web-based metabolomics tool for pathway analysis and visualization. *Bioinformatics*. 2010;26:2342-2344.

90. Baumgartner C, Osl M, Netzer M, Baumgartner D. Bioinformatic-driven search for metabolic biomarkers in disease. *J Clin Bioinforma*. 2011;1:2.

91. Johnson CH, Ivanisevic J, Siuzdak G. Metabolomics: beyond biomarkers and towards mechanisms. *Nat Rev Mol Cell Biol*. 2016;17:451-459.

92. Li S, Park Y, Duraisingham S, et al. Predicting network activity from high throughput metabolomics. Ouzounis CA, ed. *PLoS Comput Biol*. 2013;9:e1003123.

93. Xia J, Broadhurst DI, Wilson M, Wishart DS. Translational biomarker discovery in clinical metabolomics: an introductory tutorial. *Metabolomics*. 2013;9:280-299.

94. Feng K, Hong H, Tang K, Wang J. Decision making with machine learning and ROC curves. *SSRN Electron J*. Published online 2019. https://dx.doi.org/10.2139/ssrn.3382962

95. Obuchowski NA, Lieber ML, Wians FH. ROC curves in clinical chemistry: uses, misuses, and possible solutions. *Clin Chem*. 2004;50:1118-1125.

96. Florkowski CM. Sensitivity, specificity, receiver-operating characteristic (ROC) curves and likelihood ratios: communicating the performance of diagnostic tests. *Clin Biochem Rev*. 2008;1:S83-S87.

97. Cook NR. Use and Misuse of the receiver operating characteristic curve in risk prediction. *Circulation*. 2007;115:928-935.

98. Wilson PWF, D'Agostino RB, Levy D, Belanger AM, Silbershatz H, Kannel WB. Prediction of coronary heart disease using risk factor categories. *Circulation*. 1998;97:1837-1847.

99. Hyötyläinen T. Analytical methodologies utilized in the search for chronic disease biomarkers. *Bioanalysis*. 2010;2:919-923.

100. Cooke JP, Wilson AM. Biomarkers of peripheral arterial disease. *J Am Coll Cardiol*. 2010;55:2017-2023.

101. Hirsch AT, Criqui MH, Treat-Jacobson D, et al. Peripheral arterial disease detection, awareness, and treatment in primary care. *JAMA*. 2001;286:1317-1324.

102. Walker CM, Bunch FT, Cavros NG, Dippel EJ. Multidisciplinary approach to the diagnosis and management of patients with peripheral arterial disease. *Clin Interv Aging*. 2015;10:1147-1153.

103. Wenk MR. The emerging field of lipidomics. *Nat Rev Drug Discov*. 2005;4:594-610.

104. Spener F, Lagarde M, Géloën A, Record M. Editorial: what is lipidomics? *European Journal of Lipid Science and Technology*. 2003;105:481-560.

105. Li M, Zhou Z, Nie H, Bai Y, Liu H. Recent advances of chromatography and mass spectrometry in lipidomics. *Anal Bioanal Chem*. 2011;399:243-249.

106. Fahy E, Subramaniam S, Brown A, et al. A comprehensive classification system for lipids. Eur. *J. Lipid Sci. Technol*. 2005;107:337-364.

107. Cajka T, Fiehn O. Comprehensive analysis of lipids in biological systems by liquid chromatography-mass spectrometry. *TrAC Trends Anal Chem*. 2014;61:192-206.

108. Folch J, Lees A, Meath JA, Lebaron FN.Preparation of lipids extracts from brain tissue. *JBC*. 1951;191:833-841.

109. Bligh EG, Dyer WJ. A rapid method of total lipid extraction and purification. *Canadian Journal of Biochemistry and Physiology*. 1959;37:911-917

110. Matyash V, Liebisch G, Kurzchalia TV, Shevchenko A, Schwudke D. Lipid extraction by methyl- *tert* -butyl ether for high-throughput lipidomics. *J Lipid Res*. 2008;49:1137-1146.

111. Löfgren L, Ståhlman M, Forsberg G-B, Saarinen S, Nilsson R, Hansson GI. The BUME method: a novel automated chloroform-free 96-well total lipid extraction method for blood plasma. *J Lipid Res*. 2012;53:1690-1700.

112. Pizarro C, Arenzana-Rámila I, Pérez-del-Notario N, Pérez-Matute P, González-Sáiz J-M. Plasma lipidomic profiling method based on ultrasound extraction and liquid chromatography mass spectrometry. *Anal Chem*. 2013;85:12085-12092.

113. Ivanisevic J, Zhu Z-J, Plate L, et al. Toward 'omic scale metabolite profiling: a dual separation–mass spectrometry approach for coverage of lipid and central carbon metabolism. *Anal Chem*. 2013;85:6876-6884.

114. Jurowski K, Kochan K, Walczak J, Barańska M, Piekoszewski W, Buszewski B. Comprehensive review of trends and analytical strategies applied for biological samples preparation and storage in modern medical lipidomics: State of the art. *TrAC Trends Anal Chem*. 2017;86:276-289.

115. Bicalho B, David F, Rumplel K, Kindt E, Sandra P. Creating a fatty acid methyl ester database for lipid profiling in a single drop of human blood using high resolution capillary gas chromatography and mass spectrometry. *J Chromatogr A*. 2008;1211:120-128.

116. Christinat N, Morin-Rivron D, Masoodi M. High-throughput quantitative lipidomics analysis of nonesterified fatty acids in human plasma. *J Proteome Res*. 2016;15:2228-2235.

117. Dingess KA, Valentine CJ, Ollberding NJ, et al. Branched-chain fatty acid composition of human milk and the impact of maternal diet: the Global Exploration of Human Milk (GEHM) Study. *Am J Clin Nutr*. 2017;105:177-184.

118. Baylin A, Campos H. The use of fatty acid biomarkers to reflect dietary intake: *Curr Opin Lipidol*. 2006;17:22-27.

119. Albani V, Celis-Morales C, Marsaux CFM, et al. Exploring the association of dairy product intake with the fatty acids C15:0 and C17:0 measured from dried blood spots in a multipopulation cohort: Findings from the Food4Me study. *Mol Nutr Food Res*. 2016;60:834-845.

120. Weir JM, Wong G, Barlow CK, et al. Plasma lipid profiling in a large population-based cohort. *J Lipid Res*. 2013;54:2898-2908.

121. Stark K. Analytical implications of routine clinical testing for omega-3 fatty acid biomarkers. *Lipid Technol*. 2008;20:177-179.

122. Stark KD, Van Elswyk ME, Higgins MR, Weatherford CA, Salem N. Global survey of the omega-3 fatty acids, docosahexaenoic acid and eicosapentaenoic acid in the blood stream of healthy adults. *Prog Lipid Res*. 2016;63:132-152.

123. Harcombe Z, Baker JS, Cooper SM, et al. Evidence from randomised controlled trials did not support the introduction of dietary fat guidelines in 1977 and 1983: a systematic review and meta-analysis. *Open Heart*. 2015;2:e000196.

124. Astrup A, Magkos F, Bier DM, et al. Saturated Fats and Health: A reassessment and proposal for food-based recommendations: JACC State-of -the-Art Review, *Journal of the American College of Cardiology*. 2020; in press, doi: https://doi.org/10.1016/j.jacc.2020.05.077.

125. Rizos EC, Ntzani EE, Bika E, Kostapanos MS, Elisaf MS. Association between omega-3 fatty acid supplementation and risk of major cardiovascular disease events: a systematic review and meta-analysis. *JAMA*. 2012;308:1024.

126. Miller M, Ballantyne CM, Bays HE, et al. Effects of icosapent ethyl (eicosapentaenoic acid ethyl ester) on atherogenic lipid/lipoprotein, apolipoprotein, and inflammatory parameters in patients with elevated high-sensitivity C-Reactive Protein (from the ANCHOR Study). *Am J Cardiol*. 2019;124:696-701.

127. Ballantyne CM, Bays HE, Kastelein JJ, et al. Efficacy and safety of eicosapentaenoic acid ethyl ester (AMR101) therapy in statin-treated patients with persistent high triglycerides (from the ANCHOR Study). *Am J Cardiol*. 2012;110:984-992.

128. Bisgaard H, Stokholm J, Chawes BL, et al. Fish oil–derived fatty acids in pregnancy and wheeze and asthma in offspring. *N Engl J Med*. 2016;375:2530-2539.

129. Rago D, Rasmussen MA, Lee-Sarwar KA, et al. Fish-oil supplementation in pregnancy, child metabolomics and asthma risk. *EBioMedicine*. 2019;46:399-410.

130. Siri-Tarino PW, Sun Q, Hu FB, Krauss RM. Meta-analysis of prospective cohort studies evaluating the association of saturated fat with cardiovascular disease. *Am J Clin Nutr*. 2010;91:535-546.

131. Dehghan M, Mente A, Zhang X, et al. Associations of fats and carbohydrate intake with cardiovascular disease and mortality in 18 countries from five continents (PURE): a prospective cohort study. *The Lancet*. 2017;390:2050-2062.

132. Matti M, Wu Jason H.Y., Imamura Fumiaki, et al. Biomarkers of dietary omega-6 fatty acids and incident cardiovascular disease and mortality. *Circulation*. 2019;139:2422-2436.

133. Sanders Thomas A.B. Omega-6 fatty acids and cardiovascular disease. *Circulation*. 2019;139:2437-2439.

134. Würtz P, Havulinna AS, Soininen P, et al. Metabolite profiling and cardiovascular event risk: a prospective study of 3 population-based cohorts. *Circulation*. 2015;131:774-785.

135. Forouhi NG, Koulman A, Sharp SJ, et al. Differences in the prospective association between individual plasma phospholipid saturated fatty acids and incident type 2 diabetes: the EPIC-InterAct case-cohort study. *Lancet Diabetes Endocrinol*. 2014;2:810-818.

136. Khaw K-T, Friesen MD, Riboli E, Luben R, Wareham N. Plasma phospholipid fatty acid concentration and incident coronary heart disease in men and women: The EPIC-Norfolk Prospective Study. Katan MB, ed. *PLoS Med*. 2012;9:e1001255.

137. Wagenknecht LE, Mayer EJ, Rewers M, et al. The insulin resistance atherosclerosis study (IRAS). *Ann Epidemiol*. 1995;5:464-472.

138. Kurotani K, Sato M, Ejima Y, et al. High levels of stearic acid, palmitoleic acid, and dihomo-γ-linolenic acid and low levels of linoleic acid in serum cholesterol ester are associated with high insulin resistance. *Nutr Res*. 2012;32:669-675.e3.

139. Allalou A, Nalla A, Prentice KJ, et al. A predictive metabolic signature for the transition from gestational diabetes mellitus to type 2 diabetes. *Diabetes*. 2016;65:2529-2539.

140. Yi L, He J, Liang Y, Yuan D, Gao H, Zhou H. Simultaneously quantitative measurement of comprehensive profiles of esterified and non-esterified fatty acid in plasma of type 2 diabetic patients. *Chem Phys Lipids*. 2007;150:204-216.

141. Suhre K, Meisinger C, Döring A, et al. Metabolic Footprint of diabetes: a multiplatform metabolomics study in an epidemiological setting. Breant B, ed. *PLoS ONE*. 2010;5:e13953.

142. Jenkins BJ, Seyssel K, Chiu S, et al. Odd chain fatty acids; new insights of the relationship between the gut microbiota, dietary intake, biosynthesis and glucose intolerance. *Sci Rep*. 2017;7:44845.

143. Oda E, Hatada K, Kimura J, Aizawa Y, Thanikachalam PV, Watanabe K. Relationships between serum nnsaturated fatty acids and coronary risk factors. *Int Heart J*. 2005;46:975-985.

144. Manfredi M, Conte E, Barberis E, et al. Integrated serum proteins and fatty acids analysis for putative biomarker discovery in inflammatory bowel disease. *J Proteomics*. 2019;195:138-149.

145. Perreault M, Zulyniak MA, Badoud F, et al. A distinct fatty acid profile underlies the reduced inflammatory state of metabolically healthy obese individuals. Eckel J, ed. *PLoS ONE*. 2014;9:e88539.

146. Volk BM, Kunces LJ, Freidenreich DJ, et al. Effects of step-wise increases in dietary carbohydrate on circulating saturated fatty acids and palmitoleic acid in adults with metabolic syndrome. Singh PK, ed. *PLoS ONE*. 2014;9:e113605.

147. Weitkunat K, Schumann S, Nickel D, et al. Odd-chain fatty acids as a biomarker for dietary fiber intake: a novel pathway for endogenous production from propionate. *Am J Clin Nutr*. 2017;105:1544-1551.

148. Mcmillen IC, Robinson JS. Developmental origins of the metabolic syndrome: prediction, plasticity, and programming. *Physiol Rev*. 2005;85:571-633.

149. Tamayo-Uria I, Maitre L, Thomsen C, et al. The early-life exposome: description and patterns in six European countries. *Environ Int*. 2019;123:189-200.

150. Harding J. The nutritional basis of the fetal origins of adult disease. *Int J Epidemiol*. 2001;30:15-23.

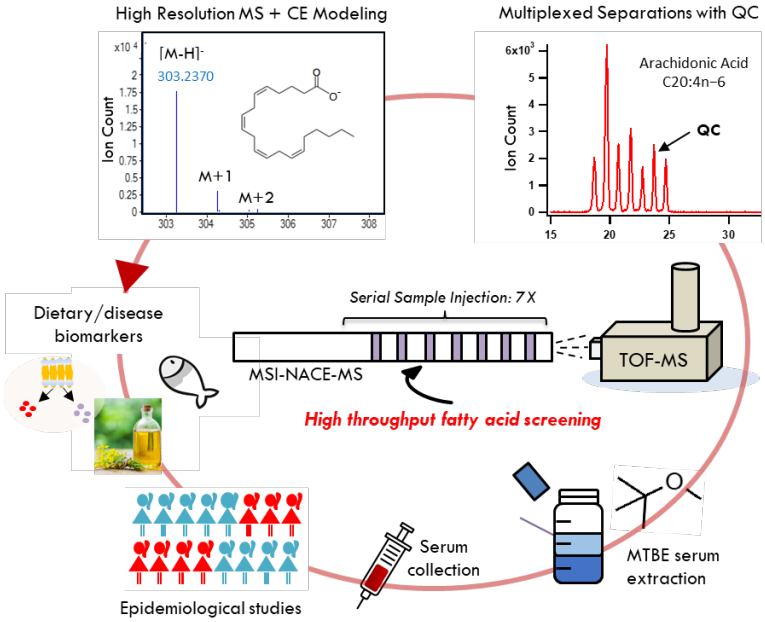
151. Varu VN, Hogg ME, Kibbe MR. Critical limb ischemia. *J Vasc Surg*. 2010;51:230-241.

Chapter II:

A Robust Method for High Throughput Screening of Fatty Acids by Multisegment Injection-Nonaqueous Capillary Electrophoresis-Mass Spectrometry with Stringent Quality Control

Thesis chapter is derived from a published peer-reviewed article:

S.M. Azab, R. Ly and P. Britz-McKibbin. Robust method for high-throughput screening of fatty acids by multisegment injection-nonaqueous capillary electrophoresis–mass spectrometry with stringent quality control. *Anal. Chem.* **2019**, 91, 2329-2336



S.M.A and R.L. are both co-first authors of this work. R.L. conducted initial experiments for the development of MSI-NACE-MS. S.M.A and R.L. further conducted the experiments for method optimization of MSI-NACE-MS. S.M.A. conducted the inter-method comparison for validation of MSI-NACE-MS, data processing and data analysis as well as optimization of sample processing protocol for lipidomic extraction. R.L. contributed to data analysis and conducted the interlaboratory method-comparison for total hydrolyzed fatty acids in NIST reference sample. S.M.A wrote the initial manuscript draft and R.L. contributed to further edits to the manuscript including edits to figures and tables. P.B.M. provided critical feedback for experiments and data analysis and edited manuscript for publication.

Chapter II: A Robust Method for High Throughput Screening of Fatty Acids by Multisegment Injection-Nonaqueous Capillary Electrophoresis-Mass Spectrometry with Stringent Quality Control

"Two are better than one… a three-fold rope is not readily broken."

2.1 Abstract

High throughput screening methods for fatty acid (FA) determination are urgently needed due to their critical biochemical roles in human health while serving as biomarkers of habitual diet and chronic disease risk assessment. Herein, we introduce multisegment injection-nonaqueous-capillary electrophoresis-mass spectrometry (MSI-NACE-MS) as a multiplexed separation platform for analysis of more than twenty non-esterified FAs in human serum or plasma. Optimization of experimental conditions was required to overcome major technical hurdles in MSI-NACE-MS prior to a rigorous method validation and inter-method comparison with gas chromatography-mass spectrometry (GC-MS). Following a simple methyl-*tert*-butyl ether extraction, seven serum extracts were analyzed directly by MSI-NACE-MS within a single run (< 4 min/sample) under negative ion mode detection that incorporates stringent measures for quality control, including batch correction adjustment. Overall, excellent technical variance (RSD = 10%) and good mutual agreement was demonstrated for twelve non-esterified FAs consistently measured in 50 serum samples analyzed independently by MSI-NACE-MS and GC-MS within the same laboratory (mean bias = 24%, n = 600). Also, total hydrolyzed plasma FAs using MSI-NACE-MS was compared to mean concentrations reported from a NIST standard reference material as an inter-laboratory method validation (mean bias = 15%, n = 20). Accurate prediction of ion migration behavior in CE also supports structural elucidation of FAs in conjunction with high resolution MS. For the first time, we demonstrate that MSI-NACE-MS offers a rapid yet robust platform for direct quantification of circulating FA using volume-restricted blood

specimens that expands metabolome coverage to encompass anionic classes of lipids as required for large-scale epidemiological studies.

2.2 Introduction

Fatty acids (FA) are fundamental building units of lipids that function as essential dietary macronutrients for energy production/storage, structural constituents of cell membranes, as well as bioactive molecules involved in the regulation of inflammation and immunity.[1] For instance, alterations in maternal dietary FA intake impact metabolic programming and neuroendocrine function during fetal development that can contribute to adverse health outcomes in children later in life.[2] However, public health policies to promote lower dietary saturated fat (*i.e.*, diet-heart hypothesis) have proven largely ineffective to reduce coronary heart disease risk[3-5] while contributing to a global obesity and type 2 diabetes epidemic.[6] Similarly, the putative health benefits of supplementing omega-3 FAs in the diet remain controversial with mixed clinical outcomes[7] despite the establishment of a billion dollar fish oil pill industry. As a result, there is an urgent need of reliable analytical methods for rapid yet accurate FA determination that can serve as biomarkers of habitual diet[8, 9] or predictive biomarkers of incident type 2 diabetes risk.[10,11] This is also critical for validating well-designed nutritional interventions and longitudinal/observational studies for chronic disease prevention and population health.

Nuclear magnetic resonance (NMR) and gas chromatography (GC) coupled to flame ionization detection (FID) and mass spectrometry (MS) are widely used methods for the resolution and quantification of FAs and their stereoisomers in complex biological samples.[12] However, there is considerable variation in protocols used to measure FAs in the literature in terms of biospecimen collected and pre-analytical sample workup conditions that complicates data comparisons. For example, free/unbound FAs, non-esterified (protein-bound) FAs or total (hydrolyzed) FAs from

specific lipid fractions (*e.g.*, phospholipid) can be measured depending on study objectives and biospecimen availability, but they provide different biochemical information regarding FA metabolism.[13] Indeed, FA determination has been reported from adipose tissue, erythrocytes/platelets, and blood samples, including plasma, serum or dried (whole) blood spots.[8] GC-MS is considered the gold standard for total serum FA analysis due to its high separation efficiency for resolving geometrical and, in some cases, positional isomers, following pre-column chemical derivatization to generate volatile fatty acid methyl esters (FAMEs).[14,15] However, long analysis times and complicated sample pretreatment necessitate robotic liquid handling systems with multiple GC instruments operated in parallel to achieve adequate throughput (1,200 samples/mo).[16,17] In contrast, liquid chromatography-mass spectrometry (LC-MS) allows for shorter analysis times that is ideal for direct analysis of non-esterified FAs (NEFAs) without chemical derivatization or prolonged heating.[18,19] This is important to prevent lipid hydrolysis and oxidation of labile polyunsaturated FAs ensuring their quantitative recovery.[20] As a result, new strategies are still needed to enable high throughput screening of FAs that is also cost-effective when analyzing bio-banked specimens in support of large-scale epidemiological studies.

Capillary electrophoresis (CE) offers an orthogonal separation platform for FA analysis that is applicable to volume-restricted samples when using detergents or organic solvents in the background electrolyte (BGE) in conjunction with direct or indirect UV absorbance, capacitively-coupled contactless conductivity or fluorescence detection.[21] Yet, there have been few reported studies for FA determination when using nonaqueous-CE when coupled to electrospray ionization-MS (NACE-MS)[22,23] with most methods applied to better solubilize pharmaceuticals, peptides, or natural products.[24-27] Recently, Lee *et al*.[23] reported the use of a dicationic ion pair reagent added to the sheath liquid to form a dynamic complex with saturated FAs during ion desorption to

improve their ionization responses under positive ion mode conditions. However, this method was not extensively validated and longer chain FAs (> 16:0) were prone to deleterious band broadening due to insufficient acetonitrile content in the BGE with modest enhancements to concentration sensitivity.[23] Although NACE-MS reports have long claimed theoretical advantages in term of separation performance,[28,29] it has not been implemented in metabolomic/lipidomic studies to date.[30] This is likely due to the lack of robust NACE-MS methods amenable to routine application underlying several major technical obstacles. For instance, NACE-MS is prone to acetonitrile-induced polyimide swelling[31] and siphoning effects[32] when using organic solvents of low viscosity that contribute to frequent capillary fractures and current drops, respectively.[33] In this context, careful optimization of separation and ionization conditions is vital for developing standardized NACE-MS methods that can be successfully replicated in different laboratories. Herein, we introduce multisegment injection (MSI)-NACE-MS as a high throughput screening method for FAs that expands metabolome coverage beyond hydrophilic/polar metabolites when using conventional aqueous BGE conditions.[34] Importantly, multiplexed separations enable the design of customized serial injection formats for analysis of seven independent samples within a single run that encodes mass spectral information based on temporal signal pattern recognition.[35-37] In this work, an extensive validation and an inter-laboratory method comparison were conducted relative to GC-MS to demonstrate accurate quantification of NEFAs or total FAs in blood specimens following a simple methyl-*tert*-butyl ether (MTBE) extraction and reconstitution step. Additional benefits of this method include accurate prediction of ion electrophoretic mobility to facilitate FA identification that is complementary to high resolution MS. Batch-correction adjustment to data is also demonstrated when using a reference serum as QC that is included within each run to correct for long-term signal drift in ESI-MS.

2.3 Experimental Section

2.3.1 CE-MS Instrumentation

An Agilent 6230 time-of-flight (TOF) mass spectrometer with a coaxial sheath liquid electrospray (ESI) ionization source equipped with an Agilent G7100A CE unit was used for all experiments (Agilent Technologies Inc., Mississauga, ON, Canada). An Agilent 1260 Infinity Isocratic pump and a 1260 Infinity degasser were applied to deliver an 80:20 MeOH:water with 0.5% *v* NH_4OH at a flow rate of 10 μL/min using a CE-MS coaxial sheath liquid interface kit. For real-time mass correction, reference ions purine and hexakis(2,2,3,3-tetrafluoropropoxy)phosphazine (HP-921) were spiked into the sheath liquid at 0.02% *v/v* to provide constant mass signals at *m/z* 119.0363 and *m/z* 1033.9881, which were also utilized for monitoring for potential ion suppression/enhancement effects during separation. The nebulizer spray was set off during serial sample injection but then subsequently turned on at a low pressure of 4 psi (27.6 kPa) following voltage application with the source temperature at 300 °C and drying gas delivered at 4 L/min. The instrument was operated in 2 GHz extended dynamic range with negative mode detection. V_{cap} was set at 3500 V while fragmentor was 120 V, the skimmer was 65 V and the Octopole RF was 750 V. Separations were performed on bare fused-silica capillaries with 50 μm internal diameter, 360 μm outer diameter and 95 cm total length (Polymicro Technologies Inc., AZ, USA). A capillary window maker (MicroSolv, Leland, NC, USA) was used to remove about 7 mm of the polyimide coating on both ends of the capillary.[37] The applied voltage was set to 30 kV at 25 °C for CE separations together with an pressure application of 20 mbar (2 kPa). The BGE was 35 mM ammonium acetate in 70% *v* acetonitrile, 15% *v* MeOH, 10% H_2O, and 5% *v* isopropanol with an apparent pH of 9.5 adjusted by addition of 12% *v* of ammonium hydroxide. Serum or plasma extracts were injected hydrodynamically at 50 mbar (5 kPa) alternating between 5 s for each

sample plug and 40 s for the BGE spacer plug for a total of seven discrete samples analyzed within one run. Prior to first use, capillaries were conditioned by flushing for 15 min at 950 mbar (95 kPa) sequentially with MeOH, 0.1 M sodium hydroxide, deionized water, 1 M formic acid, deionized water then BGE for 30 min. Between runs, the capillary was flushed with BGE for 10 min at 950 mbar (95 kPa). BGE and sheath liquid were degassed before use. Further description of chemicals/reagents, GC-MS methodology, sample workup protocols for NEFAs and total (hydrolyzed) FAs from serum and plasma, as well as method validation, data processing and statistical methods are described in the experimental of the Supporting Information.

2.4 Results and Discussion

2.4.1 MSI-NACE-MS Method Development and Optimization

Due to the importance of developing a fully compatible non-aqueous BGE for routine use in NACE-MS, initial experiments were focused on separation optimization based on several factors when using a coaxial sheath liquid interface, including electrolyte type, ionic strength, BGE solvent/sheath liquid composition, nebulizer gas and cone voltage. Overall, ammonium acetate was selected as the preferred volatile electrolyte due to its better solubility in acetonitrile (ACN)-rich solvents as compared to ammonium bicarbonate in order to ensure complete dissolution of very-long chain FAs (> 20:0).[38] Improved peak shape and sensitivity was realized, irrespective of carbon chain-length (3:0-22:0), in higher ACN content solutions (> 60% *v*). Nevertheless, a minimal fraction of deionized water (10% *v*) was still necessary to solubilize ammonium acetate at sufficiently high concentrations (> 10 mM) for stable current generation in NACE. Additionally, a series of miscible solvent mixtures with different viscosity and polarity than ACN were also investigated to further optimize FA separations, including methanol (MeOH) and isopropanol

(iPrOH). For instance, ACN content > 80% *v* drastically shortened apparent migration times with a concomitant loss in resolution and peak capacity, whereas increasing MeOH content improved overall separation performance due to higher solution viscosity. The addition of 5% *v* of iPrOH as an organic modifier to the non-aqueous BGE was also found to enhance peak shapes notably for long-chain FAs. As a result, a compatible non-aqueous BGE for solubilizing and resolving complex FA mixtures by NACE-MS was comprised of 70% *v* ACN, 15% *v* MeOH, 10% *v* H_2O, and 5% *v* iPrOH in ammonium acetate.

In all cases, the polyimide outer coating was removed from the terminal ends (≈ 7 mm length) of the fused-silica capillary to prevent deleterious swelling and irreversible aminolysis which causes frequent capillary fractures with prolonged contact with organic solvents[32] and alkaline BGE systems (pH > 9.0) with reactive ammonia.[33] This also has the benefit of reducing sample carry-over when using serial injections without compromising capillary integrity.[37] In our study, the apparent pH (pH_{app} = 8.0-11.0) and ionic strength of ammonium acetate (10-75 mM) were varied to further optimize non-aqueous BGE conditions for resolution of anionic FAs from neutral compounds co-migrating with the electrosomotic flow (EOF), including butylated hydroxytoluene (BHT) that was used as an antioxidant in all FA calibrant solutions. A pH_{app} of 9.5 was found to improve resolution of the slowest migrating very-long chain FA, lignoceric acid (24:0) from the EOF in order to minimize ion suppression/enhancement effects as shown in **Figure S2.1**. However, use of a higher pH_{app} (> 10) resulted in excessive band broadening without further resolution of FAs when using MSI-NACE-MS. Additionally, a lower ionic strength non-aqueous BGE (35 mM ammonium acetate) was used to shorten total analysis times due to a faster EOF that also minimizes Joule heating effects. Pressure-assisted NACE during electrophoretic separation was also performed using an optimal pressure setting of 20 mbar since it allowed for resolution of

very long-chain fatty acids from the EOF without deleterious band broadening and long analysis times at lower pressure settings (< 15 mbar) as depicted in **Figure S2.2**. A major challenge in implementing serial hydrodynamic sample injections in MSI-NACE-MS was the siphoning caused by the nebulizer gas for stable electrospray formation. This siphoning effect is most pronounced in NACE-MS as compared to aqueous BGE systems due to the lower viscosity of ACN-rich solvents, which contributes to frequent current drops due to aspiration of air plugs within the capillary during sample vial switching. In our case, the gas nebulizer was turned off during capillary flushing and sample injection in MSI-NACE-MS resulting in reliable analysis of FAs without incidental currents drops as spray formation is initiated only following electrophoretic separation. One unique feature of the sheath liquid in NACE-MS is the ability to independently optimize solute ionization conditions as distinct from separation conditions. For example, the addition of 0.5% *v* NH_4OH as a base modifier in the sheath liquid composed of 80% MeOH *v* provided maximum ion responses in terms of signal-to-noise ratio (SNR) for most FAs detected as their intact deprotonated molecular ions $[M-H]^-$ under negative ion mode as shown in **Figure S2.3**. Lastly, an optimal sprayer voltage of -3.5 kV was critical for maintaining spray stability without corona discharge[39] prevalent at higher voltage settings (> -4 kV).

2.4.2 Multiplexed Separations of Fatty Acids in Serum Extracts by MSI-NACE-MS

A major benefit of MSI-NACE-MS is that seven distinct samples are analyzed within a single run to enable high throughput lipid profiling (< 4 min/sample) as illustrated in **Figure 2.1A**. Unlike shotgun or direct infusion (DI)-ESI-MS, this retains the benefits of a high efficiency separation including better resolution of isobaric ions in complex sample mixtures, reduced ion suppression for FA quantification, as well as higher quality MS and MS/MS spectra for improved identification

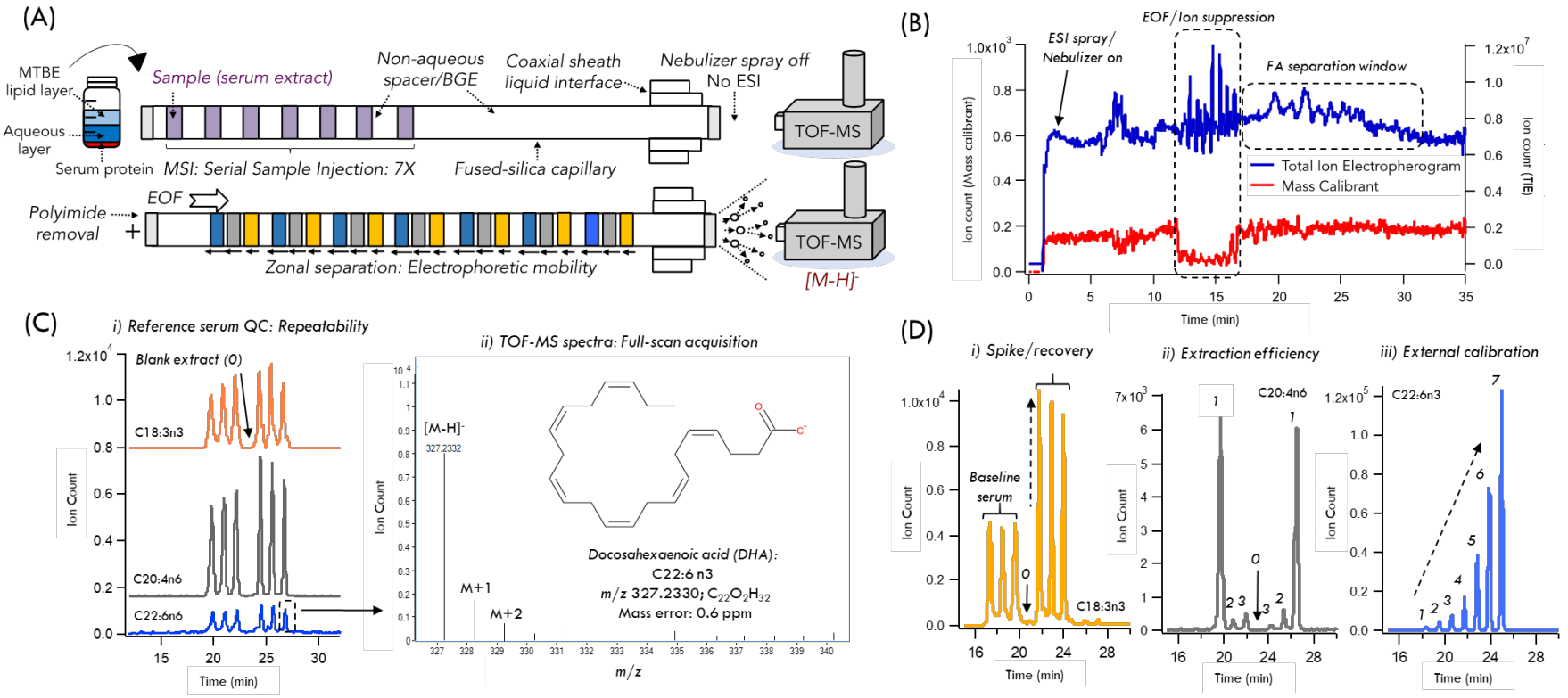


Figure 2.1. (A) Multiplexed separation of free fatty acids using MSI-NACE-MS based on serial injection of seven or more discrete sample segments and their zonal electrophoretic separation following MTBE serum extraction with full-scan data acquisition under negative ion mode detection. Careful optimization of non-aqueous electrolyte and sheath liquid conditions, including nebulizer gas operation and removal of polyimide segments at capillary ends are critical for robust performance. (B) An overlay of the total ion electropherogram (TIE) and mass calibrant (included in sheath liquid) depicts resolution of free fatty acids from serum extracts without ionization suppression from other neutral/zwitter-ionic lipids that co-migrate with the EOF following delayed spray formation with nebulizer set on following serial sample injection sequence. (C) An overlay of extracted ion electropherograms for representative free polyunsaturated fatty acids from pooled serum sample as QC (six replicate injections), including blank extract, as well as high resolution MS for determination of the most likely molecular formula for the deprotonated molecular ion ($M\text{-}H^-$) with low mass error (< 1 ppm). (D) Customized serial injection configurations used in MSI-NACE-MS for free fatty acid quantification, including spike/recovery study for determination of method accuracy (in triplicate), repeated MTBE serum extracts to evaluate extraction efficiency (in duplicate) along with seven-point external calibration curve over a 200-fold concentration range (1-200 μM).

of unknown lipids. In our case, FAs are analyzed directly under negative ion mode conditions without pre-column chemical derivatization or dynamic complexation with reagents added to the sheath liquid[24] that can contaminate the ion source. Also, zonal separation of FAs from different sample segments occurs within a non-aqueous BGE system based on their characteristic (*i.e.*, steady-state) electrophoretic mobility (μ_{ep}) in free solution whereas any overlapping ions can be further resolved by the TOF-MS. Matyash *et al*.[40] first introduced a simple protocol using methyl-*tert*-butylether (MTBE) for yielding cleaner lipid extracts from biological samples with similar or better recoveries than classic Folch extraction procedures using chloroform/MeOH.[41] Due to the lower density of MTBE relative to water, lipid extracts are readily collected from the upper ether phase, whereas a protein pellet is sedimented at the bottom aqueous layer following centrifugation. This strategy avoids contamination of the lipid ether layer with better technical variance.[42] We first examined the efficiency of the MTBE extraction procedure from an aliquot of human serum (50 μL) for NEFA analysis using repeat/fresh ether aliquots after vigorous mixing, centrifugation and biphasic separation by collecting consecutively the first, second and third ether fractions that were analyzed separately. As shown in **Table S2.3**, 14 representative serum NEFAs from the first MTBE fraction demonstrated excellent overall recovery of 92% (ranging from 82 to 96%), whereas a much lower mean recovery of about 5.6% and 2.6% was measured in the second and third ether fractions, respectively. Consequently, a single MTBE aliquot allowed for near exhaustive serum extraction thus allowing for reliable NEFA quantification while using a simple sample processing protocol.

Figure 2.1B demonstrates that anionic FAs from a serum extract were optimally resolved after the EOF where ion suppression is most evident, when comparing traces for the total ion electropherogram (TIE) and a mass calibrant in the sheath liquid. The same trace overlay also

highlights the delayed onset of spray formation since the gas nebulizer is turned on only after voltage application following the serial sample injection to avoid current drops due to siphoning. **Figure 2.1C** depicts a series of extracted ion electropherograms (EIEs) for representative ω-3 and ω-6 polyunsaturated NEFAs from six replicate injections of a standard serum extract, as well as a solvent blank extract (0) to confirm the lack of sample carry-over between sample injections. Full-scan data acquisition using TOF-MS allows for determination of the most likely molecular formula for FAs with low mass error (< 2 ppm) based on accurate mass and isotopic pattern for the deprotonated molecular ion $[M-H]^-$. **Figure 2.1D** shows that various serial injection configurations in MSI-NACE-MS can be designed to encode mass spectral information temporally within a separation depending on experimental design. For example, recovery studies for FA calibrants spiked into serum (in triplicate) were used to evaluate method accuracy as shown for linolenic acid (18:3n-3), quantitative extraction from a single fraction was confirmed by analyzing three repeat MTBE extractions of serum in duplicate as depicted for arachidonic acid (20:4n-6), and a seven-point calibration curve was acquired within a single run for NEFA quantification as highlighted for docosahexaenoic acid (22:6n-3).

2.4.3 Modeling Electromigration Behavior to Support Fatty Acid Identification

The apparent μ_{ep} of a homologous series of saturated, monounsaturated and polyunsaturated FAs was next investigated to better understand selectivity in NACE since it represents an intrinsic solute property related to the effective charge density of an ion. Indeed, accurate prediction of ion electromigration behavior is useful to support the identification of unknown metabolites *in silico* that is complementary to high resolution MS/MS notably when reference spectra and commercial standards are unavailable.[43] Since all model FAs have a single anionic charge upon deprotonation of the carboxylic acid under alkaline conditions, changes in μ_{ep} reflect primarily structural

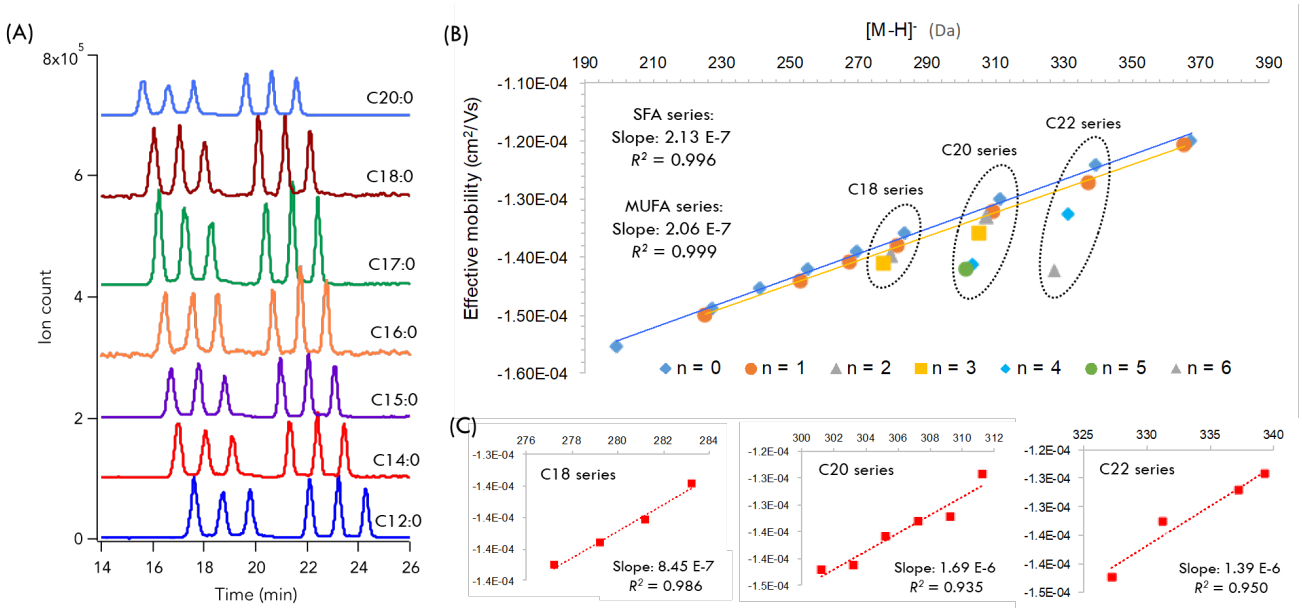


Figure 2.2. (A) Characteristic electromigration behavior of a homologous series of saturated fatty acids when using MSI-NACE-MS based on replicate injections of an equimolar calibrant mixture (6X) together with a solvent blank (1X) as the fourth sample segment. (B) Modeling of the negative electrophoretic mobility of different classes of FAs when using a non-aqueous background electrolyte based on their molecular weight/deprotonated molecular ion $[M\text{-}H]^-$ using a least-squares linear regression. Good linearity of the model is demonstrated for a homologous series of saturated (SFA), monounsaturated (MUFA), as well as (C) a series of polyunsaturated FAs among 18:0, 20:0 and 22:0 with up to five degrees of unsaturation.

differences related to carbon chain length and degrees of unsaturation. **Figure 2.2A** highlights the migration order for a series of saturated FAs in a calibrant mixture, where shorter chain FAs migrate with a larger negative mobility (or longer migration time) counter to the EOF. **Figure 2.2B** confirms that there was an excellent linear relationship ($R^2 = 0.996$) between carbon chain length in terms of molecular weight and mean apparent μ_{ep} among saturated FAs (from 12:0 to 24:0) using a least-squares regression model. Similarly, a homologous series of monounsaturated FAs (from 14:1 to 24:1) was also well described by a linear regression model ($R^2 = 0.999$), including three separate series of polyunsaturated FAs corresponding to 18:0, 20:0 and 22:0 ($R^2 > 0.94$) that vary from one to up to six degrees of unsaturation as shown in **Figure 2.2C.** Overall, greater FA unsaturation resulted in a progressively larger negative μ_{ep} due to their more compact size as compared to linear/saturated FA analogs with equivalent carbon chain length. However, certain

positional (18:1, ω-3/6) and geometric (18:1, *cis*/*trans*) isomers examined by MSI-NACE-MS were not resolved under these conditions as they co-migrate with similar mobilities in free solution. Additional selectivity for resolving FA stereoisomers by NACE thus requires differential partitioning using non-ionic surfactants in the BGE, but this is not readily compatible with ESI-MS.[44] Nevertheless, MSI-NACE-MS offers adequate selectivity for resolution of complex mixtures of FAs while minimizing isobaric interferences and supporting their identification given the limited diagnostic information typically acquired with MS/MS under negative ion mode conditions.[45]

2.4.4 Method Validation and Figures of Merit of MSI-NACE-MS

Validation of the optimized MSI-NACE-MS method was next performed for 18 representative FAs (from 10:0 to 24:0) based on several figures of merit. Calibration curves were generated based on triplicate analysis of seven calibrant solutions ranging in concentration from 1.0 to 200 μM, each containing 50 μM stearic acid-d_{35} as a single stable-isotope internal standard for data normalization. Overall, there was good linearity over a 200-fold linear dynamic range with correlation coefficients (R^2) ranging from 0.992 to 0.997 as highlighted in **Figure S2.4**. The median LOD (*SNR* = 3) and LOQ (*SNR* = 10) for native FAs were about 0.70 μM and 2.4 μM, respectively. There were however two notable exceptions, including short/medium-chain FAs (< 10:0) and two saturated FAs (16:0, 18:0). In the former case, as ionization efficiency in ESI-MS is highly solute-dependent, concentration sensitivity was lower for the least hydrophobic medium-chain FA, such as 10:0. Similarly, more water-soluble medium/short-chain NEFAs (3:0-9:0) had higher detection limits and they were also not readily detected in serum due to their poor extraction efficiency in MTBE. Additionally, higher LOD and LOQ was also measured for 16:0 and 18:0 since they are ubiquitous plasticizer additives in the environment.[46] In fact, both saturated FAs are

present at low concentrations levels in the background likely due to the unavoidable use of plastic liners for delivery of the sheath liquid that is recirculated using an isocratic LC pump. Despite the higher LOD for palmitate (16:0) and stearate (18:0), both saturated FAs are highly abundant in serum nonetheless, and thus can still be reliably measured after background correction when implementing stringent cleaning protocols to all plastic and glassware used in the method. Overall, the low micromolar to sub-micromolar detection limits reported in this work is comparable to outcomes when using a dication ion pair reagent in the sheath liquid in CE-MS for detection of saturated FAs under positive ion mode.[23] Also, method detection/quantification limits for FAs are superior to GC-EI-MS methods applied in this work (**Table S2.4**), and similar to GC-FID that is used in large-scale epidemiological studies.[17] However, a major advantage of MSI-NACE-MS is the simplicity of the protocol as it enables rapid yet comprehensive profiling of FAs derived from volume-restricted samples (< 3 μL), such as dried blood spot punches or infant sweat specimens.[35,36] Method reproducibility was evaluated by determining the intraday (n = 18) and interday (n = 78) precision for FA calibrant standards at 50 μM when using MSI-NACE-MS. Overall, the median RSD was 7.3% and 12.1% based on RPA integration of measured FA responses for the intraday and interday precision, respectively, whereas RMTs for all FAs ranged from 0.2-2.0% for 78 repeated injections over three consecutive days. Additionally, method accuracy was assessed based on spike-recovery experiments in standard human serum extracts for five representative FAs performed in triplicate at two different concentration levels. Recoveries for the spiked FAs ranged from 73–103% with a mean recovery of 93% as summarized in **Table S2.5**. These results again support the valid use of a single MTBE extraction fraction for quantitative recovery of FAs from serum samples. **Table S2.6** summarizes other major figures of merit of MSI-NACE-MS for determination of FAs based on their characteristic accurate mass

(*m/z*) and migration behavior (μ_{ep} or RMT), including method reproducibility, linearity/linear dynamic range, LOD/LOQ, and separation efficiency.

2.4.5 Inter-method Comparison for Accurate NEFA or Total Fatty Acid Quantification.

An extensive inter-method comparison of MSI-NACE-MS relative to GC-MS was next performed within the same laboratory/analyst for quantitative analysis of serum NEFAs. In order to minimize hydrolysis of intact serum lipids for reliable NEFA determination, optimization of acid-catalyzed transesterification for FAMEs using GC-MS was performed under "mild" reaction conditions using 0.2% *v/v* HCl in methanol at 45 °C within 1 h.[47] This takes advantage of the higher reactivity of NEFAs as compared to other classes of lipids in serum/plasma, including triglycerides, phospholipids, and cholesteryl esters that otherwise require longer reaction times to ensure complete derivatization.[47] **Figure S2.5** confirms that low background hydrolysis of glyceryl tridecanoate (with < 8% C13:0 formation) when using GC-MS under mild reaction conditions (45 °C for 30 min) as compared to more vigorous reaction conditions (80 °C for 4 h). In addition, the overall yield of FAMEs from other endogenous NEFAs was not significantly affected due to the shorter incubation times and milder reaction conditions to minimize lipid hydrolysis. Furthermore, a comparison of these two reactions conditions were also performed on standard serum samples when using GC-MS as shown in **Figure S2.6**. As expected, there was about a 7 to 12-fold higher concentration for total hydrolyzed FAs (free FA, protein-bound FA and FA esterified lipids) as compared to their corresponding NEFAs in standard human serum.

An inter-method comparison of commonly measured NEFAs from a cohort of maternal serum samples was next examined as a way to rigorously validate MSI-NACE-MS relative to the GC-MS protocol. Serum aliquots were independently processed in randomized order while using a standard human serum as a QC analyzed intermittently after every batch of seven samples/runs

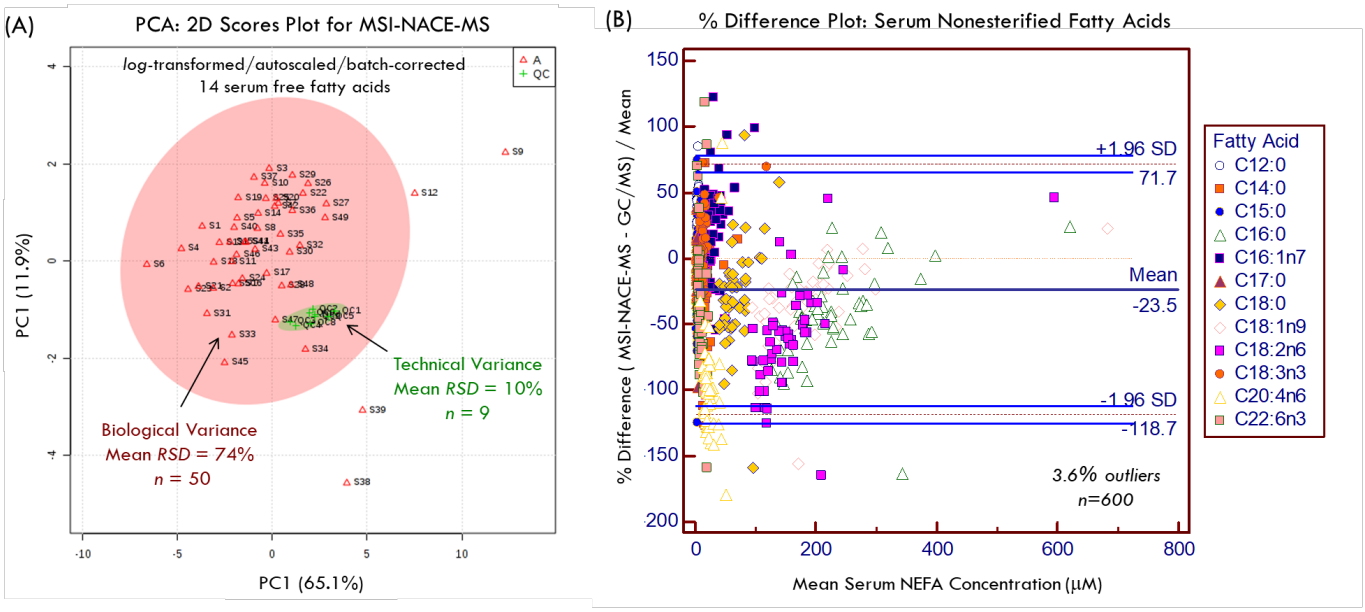


Figure 2.3. (A) A PCA 2D scores plot to summarize total biological variance of 14 free fatty acids measured in fasting maternal serum extracts (n=50) as compared to technical variance from QCs (standard human serum, n = 9) when using MSI-NACE-MS following a batch correction algorithm. (B) A Bland-Altman % difference plot showing an inter-method comparison for quantification of 14 serum NEFAs consistently measured in 50 serum extracts when using GC-EI-MS and MSI-NACE-MS (n = 600), which highlights a modest extent negative bias (24%) with few outliers (3.6%) exceeding agreement limits (95% CI). GC-MS used a standard extraction procedure in hexane following NEFA analysis as their FAMEs derivatives, whereas MSI-NACE-MS utilized an MTBE extraction protocol for direct analysis of NEFAs without chemical derivatization.

when using GC-MS. In contrast, a QC was measured within every run when using a seven-sample serial injection format[35,36] in MSI-NACE-MS (*i.e.*, 6 individual serum extracts + 1 QC) for direct analysis of NEFAs without chemical derivatization. QC samples were randomized within the injection sequence in order to assess technical variance, which also allows for long-term signal drift adjustment that is critical when analyzing large batches of samples by ESI-MS. **Figure 2.3A** depicts a 2D scores plot from a principle component analysis (PCA) of serum extracts (n = 50) analyzed by MSI-NACE-MS as compared to reference QCs (n = 9) following a batch correction algorithm,[48] which highlights excellent technical variance (mean RSD = 10%) as compared to the much greater biological variance (mean RSD = 74%) for fourteen NEFAs measured consistently in all serum extracts. In fact, the technical variance for NEFA determination by GC-MS was slightly higher than MSI-CE-MS with a mean RSD of 21% (n = 12) as compared to 17% (n = 9)

before QC batch correction adjustment, respectively. Moreover, control charts for deuterated recovery standards added to all serum samples and QCs (n = 57 or 62) prior to extraction highlights robust analytical performance for both instrumental platforms as reflected by a random distribution about the mean with few outliers exceeding warning or action limits as shown in **Figure S2.7**. Also, **Figure 2.3(B)** depicts a Bland-Altman %difference plot, which demonstrates good mutual agreement when comparing serum NEFA concentrations measured independently by MSI-NACE-MS and GC-MS protocols. Overall, there is evidence of a modest extent of negative bias of about 25% for NACE-MSI-MS as compared to GC-MS, whereas all data is normally distributed with 579 out of 600 measurements falling within agreement limits ($p < 0.05$ with 3.5% error for 21 outliers). This bias was likely attributed to the limited extent of background hydrolysis of serum lipids (< 10%) even when using mild reaction conditions for GC-MS notably for abundant serum FAs, such as oleic acid (18:1n-9), linoleic acid (18:2n-6), palmitic (16:0), and arachidonic acid (20:4n-6). Indeed, these FAs have among the highest fold-change increase when analyzing their total hydrolyzed lipid pool as compared to serum NEFA (protein-bound) concentration levels (**Figure S2.5**). Lastly, as a way to further validate the accuracy of MSI-NACE-MS, an inter-laboratory method comparison was also performed based on analysis of total hydrolyzed FAs from a reference plasma sample (SRM 1950) with results compared with mean concentrations acquired at accredited laboratories using different protocols based on GC-MS/GC-FID (NIST) and GC-MS (CDC). In this case, a Passing-Bablok regression and Bland-Altman %difference plots demonstrated good mutual agreement with a slope of 1.32 and mean bias of 15% when comparing twenty total FAs from plasma extracts that were measured consistently across both instrumental platforms as shown in **Figure S2.8**. A modest degree of negative bias was noted for MSI-NACE-MS as compared to GC-MS/GC-FID methods for total hydrolyzed FAs that was comparable to

technical precision especially considering that different extraction protocols were implemented. Although GC methods offer better selectivity based on resolution of certain FA geometric/regional isomers in hydrolyzed plasma extracts, this did not significantly impact accuracy given their low relative abundances in the diet,[49] such as *cis*- and *trans*-vaccenic acid (18:1n-7); these FAs are minor positional and geometric isomers (< 10% of total) as compared to the predominant isomer from the diet, namely oleic acid (18:1n-9). Future work will examine strategies to improve FA isomer resolution while also enhancing ionization efficiency for lower detection limits. Comprehensive profiling of intact lipids also offers an intriguing approach to expand metabolomic coverage when using MSI-NACE-MS notably for the analysis of volume-limited or mass-restricted samples ranging from bio-banked biofluids and single cells.

2.5 Conclusion

For the first time, we demonstrate high throughput determination of FAs from blood specimens when using MSI-NACE-MS following rigorous method optimization and validation to ensure robust performance. Several major technical obstacles were addressed during method development that have long hindered progress in NACE-MS for routine lipid profiling applications, including capillary fractures due to polyimide swelling, current drops from nebulizer siphoning and corona discharges when using non-aqueous electrolyte systems with negative ion mode detection. The development of a simple yet efficient MTBE extraction protocol for serum/plasma FAs without chemical derivatization was achieved when using multiplexed separations comprising seven discrete samples serially injected within a single run. This not only greatly improves sample throughput without added infrastructure costs, but also ensures more stringent QC since every run contains a reference sample for monitoring technical variance, which also allows for long-term batch correction. Moreover, accurate modeling of ion migration behavior in NACE offers a

complementary tool to support unknown FA identification in conjunction with high resolution MS or MS/MS. Overall, method sensitivity is comparable to conventional GC methods, however certain minor positional or geometric FA isomers in serum extracts were unresolved. Nevertheless, excellent technical precision and accuracy were demonstrated following extensive inter-method and inter-laboratory validation with good mutual agreement in serum NEFA and total plasma FA concentrations measured by MSI-NACE-MS as compared to GC-MS methods. This work opens the door for new advances in comprehensive FA profiling in support of large-scale epidemiological studies and population health, such as objective biomarkers of habitual diet and prognostic indicators of metabolic syndrome.

2.6 Acknowledgements

P.B.M. acknowledges funding from the Natural Sciences and Engineering Research Council of Canada, the Canada Foundation for Innovation, and Genome Canada. The authors would like to thank Dr. Kirk Green and Dr. Fan Fei at MRCMS for their support with GC/MS training. Further thanks are extended to Dr. Koon Teo, Dr. Sonia Anand and the FAMILY cohort co-investigators for access to anonymized maternal serum specimens used in the inter-method comparison.

2.7 References

1. Galli C, Calder PC. Effects of fat and fatty acid intake on inflammatory and immune responses: a critical review. *Ann Nutr Metab Basel*. 2009;55:123-139.

2. Kabaran S, Besler HT. Do fatty acids affect fetal programming? *J Health Popul Nutr*. 2015;33:14.

3. Dehghan M, Mente A, Zhang X, et al. Associations of fats and carbohydrate intake with cardiovascular disease and mortality in 18 countries from five continents (PURE): a prospective cohort study. *The Lancet*. 2017;390:2050-2062.

4. DiNicolantonio JJ, Lucan SC, O'Keefe JH. The Evidence for saturated fat and for sugar related to coronary heart disease. *Prog Cardiovasc Dis*. 2016;58:464-472.

5. Willett WC, Leibel RL. Dietary fat is not a major determinant of body fat. *Am J Med*. 2002;113:47-59.

6. Basu S, Yoffe P, Hills N, Lustig RH. The relationship of sugar to population-level diabetes prevalence: an Econometric analysis of repeated cross-sectional data. *PLoS ONE*. 2013;8:e57873.

7. Abdelhamid AS, Brown TJ, Brainard JS, et al. Omega-3 fatty acids for the primary and secondary prevention of cardiovascular disease. *Cochrane Database Syst Rev*. 2018;(7):CD003177.

8. Hodson L, Skeaff CM, Fielding BA. Fatty acid composition of adipose tissue and blood in humans and its use as a biomarker of dietary intake. *Prog Lipid Res*. 2008;47:348-380.

9. Weitkunat K, Schumann S, Nickel D, et al. Odd-chain fatty acids as a biomarker for dietary fiber intake: a novel pathway for endogenous production from propionate. *Am J Clin Nutr*. 2017;105:1544–51.

10. Allalou A, Nalla A, Prentice KJ, et al. A predictive metabolic signature for the transition from gestational diabetes mellitus to type 2 diabetes. *Diabetes*. 2016;65:2529-2539.

11. Forouhi NG, Koulman A, Sharp SJ, et al. Differences in the prospective association between individual plasma phospholipid saturated fatty acids and incident type 2 diabetes: the EPIC-InterAct case-cohort study. *Lancet Diabetes Endocrinol*. 2014;2:810-818.

12. Wu Z, Zhang Q, Li N, Pu Y, Wang B, Zhang T. Comparison of critical methods developed for fatty acid analysis: A review. *J Sep Sci*. 2017;40:288-298.

13. Karpe F, Dickmann JR, Frayn KN. Fatty acids, obesity, and insulin resistance: time for a reevaluation. *Diabetes*. 2011;60:2441-2449.

14. Bicalho B, David F, Rumplel K, Kindt E, Sandra P. Creating a fatty acid methyl ester database for lipid profiling in a single drop of human blood using high resolution capillary gas chromatography and mass spectrometry. *J Chromatogr A*. 2008;1211(1-2):120-128.

15. Sánchez-Ávila N, Mata-Granados JM, Ruiz-Jiménez J, Luque de Castro MD. Fast, sensitive and highly discriminant gas chromatography–mass spectrometry method for profiling analysis of fatty acids in serum. *J Chromatogr A*. 2009;1216:6864-6872.

16. Lin YH, Salem N, Wells EM, et al. Automated high-throughput fatty acid analysis of umbilical cord serum and application to an epidemiological study. *Lipids*. 2012;47:527-539.

17. Wang L, Summerhill K, Rodriguez-Canas C, et al. Development and validation of a robust automated analysis of plasma phospholipid fatty acids for metabolic phenotyping of large epidemiological studies. *Genome Med*. 2013;5:39.

18. Hellmuth C, Weber M, Koletzko B, Peissner W. Nonesterified fatty acid determination for functional lipidomics: comprehensive ultrahigh performance liquid chromatography–tandem

mass spectrometry quantitation, qualification, and parameter prediction. *Anal Chem*. 2012;84:1483-1490.

19. Christinat N, Morin-Rivron D, Masoodi M. High-throughput quantitative lipidomics analysis of nonesterified fatty acids in human plasma. *J Proteome Res*. 2016;15:2228-2235.

20. Jurowski K, Kochan K, Walczak J, Barańska M, Piekoszewski W, Buszewski B. Comprehensive review of trends and analytical strategies applied for biological samples preparation and storage in modern medical lipidomics: State of the art. *TrAC Trends Anal Chem*. 2017;86:276-289.

21. De Oliveira M,Porto B,Faria I, et al. 20 Years of Fatty Acid Analysis by Capillary Electrophoresis. *Molecules*. 2014;19:14094-14113.

22. Petersson MA, Hulthe G, Fogelqvist E. New sheathless interface for coupling capillary electrophoresis to electrospray mass spectrometry evaluated by the analysis of fatty acids and prostaglandins. *J Chromatogr A*. 1999;854:141-154.

23. Lee J-H, Kim S-J, Lee S, Rhee J-K, Lee SY, Na Y-C. Saturated fatty acid determination method using paired ion electrospray ionization mass spectrometry coupled with capillary electrophoresis. *Anal Chim Acta*. 2017;984:223-231.

24. Cheng J, Wang L, Liu W, Chen DDY. Quantitative nonaqueous capillary electrophoresis–mass spectrometry method for determining active ingredients in plant extracts. *Anal Chem*. 2017;89:1411-1415.

25. Bonvin G, Schappler J, Rudaz S. Non-aqueous capillary electrophoresis for the analysis of acidic compounds using negative electrospray ionization mass spectrometry. *J Chromatogr A*. 2014;1323:163-173.

26. Klampfl CW, Himmelsbach M. Nonaqueous capillary electrophoresis mass spectrometry. In: Schmitt-Kopplin P, ed. *Capillary Electrophoresis: Methods and Protocols*. Methods in Molecular Biology. Springer; 2016:111-130.

27. Scriba GKE. Nonaqueous capillary electrophoresis–mass spectrometry. *J Chromatogr A*. 2007;1159:28-41.

28. Porras SP, Kenndler E. Are the asserted advantages of organic solvents in capillary electrophoresis real? A critical discussion. *Electrophoresis*. 2005;26:3203-3220.

29. Kenndler E. A critical overview of non-aqueous capillary electrophoresis. Part I: Mobility and separation selectivity. *J Chromatogr A*. 2014;1335:16-30.

30. Zhang W, Hankemeier T, Ramautar R. Next-generation capillary electrophoresis–mass spectrometry approaches in metabolomics. *Curr Opin Biotechnol*. 2017;43:1-7.

31. Baeuml F, Welsch T. Improvement of the long-term stability of polyimide-coated fused-silica capillaries used in capillary electrophoresis and capillary electrochromatography. *J Chromatogr A*. 2002;961:35-44.

32. Mokaddem M, Gareil P, Belgaied J-E, Varenne A. A new insight into suction and dilution effects in capillary electrophoresis coupled to mass spectrometryvia an electrospray ionization interface. Part I-Suction effect. *Electrophoresis*. 2008;29:1957-1964.

33. Yamamoto M, Ly R, Gill B, Zhu Y, Moran-Mirabal J, Britz-McKibbin P. Robust and high-throughput method for anionic metabolite profiling: preventing polyimide aminolysis and capillary breakages under alkaline conditions in capillary electrophoresis-mass spectrometry. *Anal Chem*. 2016;88:10710-10719.

34. Kuehnbaum NL, Britz-McKibbin P. New advances in separation science for metabolomics: resolving chemical diversity in a post-genomic era. *Chem Rev*. 2013;113:2437-2468.

35. Macedo AN, Mathiaparanam S, Brick L, et al. The sweat metabolome of screen-positive cystic fibrosis infants: revealing mechanisms beyond impaired chloride transport. *ACS Cent Sci*. 2017;3:904-913.

36. DiBattista A, McIntosh N, Lamoureux M, Al-Dirbashi OY, Chakraborty P, Britz-McKibbin P. Temporal signal pattern recognition in mass spectrometry: a method for rapid identification and accurate quantification of biomarkers for inborn errors of metabolism with quality assurance. *Anal Chem*. 2017;89:8112-8121.

37. DiBattista A, Rampersaud D, Lee H, Kim M, Britz-McKibbin P. High throughput screening method for systematic surveillance of drugs of abuse by multisegment injection–capillary electrophoresis–mass spectrometry. *Anal Chem*. 2017;89:11853-11861.

38. Sassa T, Kihara A. Metabolism of very long-chain fatty acids: genes and pathophysiology. *Biomol Ther*. 2014;22:83-92.

39. McClory PJ, Håkansson K. Corona discharge suppression in negative ion mode nanoelectrospray ionization via trifluoroethanol addition. *Anal Chem*. 2017;89:10188-10193.

40. Matyash V, Liebisch G, Kurzchalia TV, Shevchenko A, Schwudke D. Lipid extraction by methyl- *tert* -butyl ether for high-throughput lipidomics. *J Lipid Res*. 2008;49:1137-1146.

41. Jafari N, Ahmed R, Gloyd M, Bloomfield J, Britz-McKibbin P, Melacini G. Allosteric sensing of fatty acid binding by NMR: application to human serum albumin. *J Med Chem*. 2016;59:7457-7465.

42. Godzien J, Ciborowski M, Armitage EG, et al. A single in-vial dual extraction strategy for the simultaneous lipidomics and proteomics analysis of HDL and LDL fractions. *J Proteome Res*. 2016;15:1762-1775.

43. Lee R, Ptolemy AS, Niewczas L, Britz-McKibbin P. Integrative metabolomics for characterizing unknown low-abundance metabolites by capillary electrophoresis-mass spectrometry with computer simulations. *Anal Chem*. 2007;79:403-415.

44. de Castro Barra PM, Castro R de JC, de Oliveira PL, Aued-Pimentel S, da Silva SA, de Oliveira MAL. An alternative method for rapid quantitative analysis of majority cis–trans fatty acids by CZE. *Food Res Int*. 2013;52:33-41.

45. Bollinger JG, Rohan G, Sadilek M, Gelb MH. LC/ESI-MS/MS detection of FAs by charge reversal derivatization with more than four orders of magnitude improvement in sensitivity. *J Lipid Res*. 2013;54:3523-3530.

46. Yao C-H, Liu G-Y, Yang K, Gross RW, Patti GJ. Inaccurate quantitation of palmitate in metabolomics and isotope tracer studies due to plastics. *Metabolomics*. 2016;12:143.

47. Ichihara K, Fukubayashi Y. Preparation of fatty acid methyl esters for gas-liquid chromatography. *J Lipid Res*. 2010;51:635-640.

48. Wehrens R, Hageman JosA, van Eeuwijk F, et al. Improved batch correction in untargeted MS-based metabolomics. *Metabolomics*. 2016;12:88.

49. de Souza RJ, Mente A, Maroleanu A, et al. Intake of saturated and trans unsaturated fatty acids and risk of all cause mortality, cardiovascular disease, and type 2 diabetes: systematic review and meta-analysis of observational studies. *BMJ*. 2015;351.

2.8 Supporting Information

2.8.1 Chemicals and Reagents

Ultra LC-MS grade methanol (Caledon Inc., Georgetown, ON, Canada) and ultra LC-MS grade acetonitrile (Honeywell Inc., Muskegon, MI, USA) were used to prepare sheath liquid and background electrolyte (BGE), respectively. Ammonium acetate, ammonium hydroxide, butylated hydroxytoluene (BHT), methyl-*tert*-butyl ether (MTBE), stearic acid-d_{35}, myristic acid-d_{27}, glyceryl tridecaonate, and other FA standards, chemicals, solvents, as well as standard human serum (S7023) and normal fasting human plasma from NIST (SRM 1950) were purchased from Sigma-Aldrich Inc. (St. Louis, MO, USA).

2.8.2 GC-MS Instrumentation

Comparative analysis of serum FA for validation purposes were also performed by GC-MS using an Agilent 6890 gas chromatograph equipped with a Supelco SP-2380 column (30m x 0.25mm) coupled to an Agilent 5973 single quadrupole mass spectrometer with electron impact ionization (EI-MS) operating in selected ion monitoring (SIM) mode. FA were analyzed by GC/MS as their FAME derivatives, where quantification was based on the relative ion response ratio of their most abundant fragment ion relative to that of pyrene-d_{10} used as internal standard. In most cases, characteristic fragment ions for saturated/monoenoic and polyenoic FAs were monitored at *m/z* 74 (*i.e.,* McLafferty rearrangement product ion) and *m/z* 79 (*i.e.,* cyclohexadienyl ion), respectively unless otherwise stated.[1] The temperature program used for optimal resolution of FAMEs was set initially at 60 °C for 1 min followed by 12 °C/min temperature gradient for 10 min to 180 °C similar to elution conditions described elsewhere.[2] The temperature was then increased for 10 min to 210 °C at a rate of 3 °C/min to resolve the most abundant C18 species and isomers. The temperature was further increased for the next 2 min with a ramp of 30 °C/min until reaching 270 °C, at which point the temperature was held for 7 min corresponding to a total run time of 30 min (**Table S2.1** of the Supporting Information). Samples were injected in 1.0 μL volumes using a splitless injector held at 250 °C. Helium was used as a carrier gas flowing at 1.0 mL/min and the transfer line was held at 270 °C. SIM was used for targeted quantification of 12 representative FAs in serum extracts for inter-method comparison studies, which began with *m/z* 55, 74, 77 for the first 13 min followed by selected ions *m/z* 67 and *m/z* 79 for the next 10 min and a final segment at *m/z* 212 for the last 7 min for detection of the internal standard with dwell time set at 100 msec (**Table S2.2** of the Supporting Information).

2.8.3 Sample Workup Procedure for NEFAs by GC-MS

Aliquots of the same 50 serum samples from the FAMILY cohort,[3] as well as standard human serum (Sigma) that were used as QCs, were prepared randomly for an inter-method comparison for NEFA determination when using GC-MS relative to MSI-NACE-MS. Extractions were carried out in glass GC vials that were pre-rinsed with dichloromethane and all pipette tips used during this procedure were pre-rinsed with methanol. To 50 µL of serum, 10 µL of 1 mM stearic acid-d_{35} (recovery standard), 25 µL of 8% *v/v* methanolic HCl reagent, 100 µL toluene and 815 µL of methanol with 0.05% *m/v* BTH were added and thoroughly mixed by vortexing. Next, vials were tightly sealed and incubated at 45 °C for 30 min under "mild" reaction conditions for NEFA determination while preventing serum lipid hydrolysis,[3] which was shortened from 1 h to 30 min based on optimization experiments. Phase separation was then induced by adding 100 µL of deionized water and 200 µL of hexane. Samples were then vortexed and centrifuged at 3000 *g* to sediment serum protein prior to equilibration for 5 min at 4 °C. A fixed volume (45 µL) was collected of the upper hexane layer into a new GC vial insert with addition of 5 µL of 10 mg/L pyrene-d_{10} that was used as an internal standard. The serum extract was thoroughly mixed followed by splitless injection and analysis by GC-MS, where a QC specimen was analyzed intermittently every 8 runs to evaluate technical precision. For absolute quantification of FAMEs, a stock solution of FA standards were diluted from 5.0 to 500 µM in methanol and external calibration curves were constructed using seven different concentration levels, where their integrated peak areas were normalized to the internal standard. Calibrants were derivatized under similar conditions as serum samples, but 765 µL of 0.05% *w/v* BHT in methanol and 50 µL of deionized water needed to match the serum matrix for equivalent rates of trans-esterification since higher FAMEs recoveries were consistently observed for samples prepared in methanol as compared to

water or serum. Further experiments were also conducted to test "mild" reaction conditions used for FAME generation from NEFAs without significant hydrolysis of major serum lipids. In this case, glyceryl tridecanoate (Sigma-Aldrich) was used as an intact yet non-native lipid standard in serum with reactions performed under "harsh" reaction conditions as required for total lipid hydrolysis using 10 µL of concentrated sulfuric acid at 80 °C for 4 h.[4] GC-MS analysis of 13:0 as its FAMEs derivative (*m/z* 74) was then compared under mild (NEFAs) and harsh (total hydrolyzed FAs) reaction conditions since triglycerides are among the most susceptible lipid class prone to hydrolysis.[4]

2.8.4 Sample Workup Protocol for Nonesterified Fatty Acids by MSI-NACE-ME

Anonymized fasting serum samples from pregnant women (n = 50) in the FAMILY cohort,[3] standard human serum (Sigma), and NIST (SRM 1950) plasma samples were prepared using a slightly modified extraction protocol using MTBE originally developed by Matyash *et al.*[5] Extractions were carried out in glass GC vials that were pre-rinsed with dichloromethane and all pipette tips used during this procedure were pre-rinsed with methanol. First, 100 µL of 0.01% *v/v* BHT in methanol was mixed with a 50 µL aliquot of serum. In this case, BHT was used as a lipophilic antioxidant to prevent degradation of labile polyunsaturated FAs during sample processing/storage, as well as a neutral EOF marker for CE separations for determination of apparent electrophoretic mobilities (μ_{ep}) for FAs under non-aqueous buffer conditions. Next, 250 µL of 50 µM myristic acid-d_{27} (recovery standard) in MTBE and 12.5 µL of 1 M HCl were added to the mixture followed by vigorous shaking for 30 min at room temperature. Phase separation was then induced by addition of 100 µL of deionized water. Samples were then centrifuged at 3000 *g* at +4 °C for 30 min to sediment protein at bottom of vial followed by a biphasic water and ether (top) layer. A fixed volume (200 µL) was collected from the upper MTBE layer into a new vial

then dried under a gentle stream of nitrogen gas at room temperature. Serum or plasma extracts were then stored at -80 °C and before analysis reconstituted in 25 µL of acetonitrile/isopropanol/water (70:20:10) with 10 mM ammonium acetate and 50 µM stearic acid-d_{35} as internal standard.

2.8.5 Sample Workup Protocol for Total Hydrolyzed fatty acids by MSI-NACE-MS

The following protocol was used using NIST human plasma (SRM 1950), which is a standard reference material for method validation and inter-laboratory comparisons in support of growing metabolomic and lipidomic initiatives.[6] The hydrolysis reaction for analysis of total FAs is based on a modified protocol in which MTBE and toluene was used as a substitute over the original methanol and hexane protocol to avoid producing FAMEs and problems with solvent immiscibility.[7] First, a hydrolysis reaction for the analysis of total FAs was conducted by the addition of 25 µL of MTBE and 25 µL of concentrated sulfuric acid to 50 µL of serum. Next, 50 µL of toluene is added followed by 500 µL of water and vortexed vigorously for 15 min, which was followed by incubation of sample in an oven at 80 °C for 4 h. Then, MTBE extraction was carried out to recover total hydrolyzed FAs for subsequent analysis by MSI-NACE-MS. All hydrolysis reactions and extractions were carried out in triplicate using glass GC vials that were pre-rinsed with (dichloromethane) DCM and all pipette tips used during this procedure were pre-rinsed with MeOH. After hydrolyzed plasma sample was transferred into the vial, 100 µL of MeOH was added and the mixture was shaken for 30 min. Next, 250 µL of MTBE and 12.5 µL of 1 M HCl was added to the mixture. This is then shaken for an additional 30 min. Next, 87.5 µL of distilled, deionized water is added which induces separation of the MTBE layer. The solutions were then centrifuged for 30 min to sediment protein followed by transferring a fixed 200 µL volume from the top MTBE layer into a new vial. Fractions were stored dried at -80 °C after

solvent evaporation under a gentle stream of nitrogen for 10 min at room temperature.

2.8.6 Calibration and Method Validation of MSI-NACE-MS

10 mM stock solutions of FA calibrants were prepared in MTBE with addition of 0.1% *w/v* BHT to prevent oxidation during storage. A serial dilution of calibrant solutions from 1.0 to 200 μM was prepared in triplicate when constructing seven-point calibration curves for FAs using least-squares linear regression. All integrated peak areas for the deprotonated molecular ion [M-H$^-$] of FAs were normalized using 50 μM stearic acid-d_{35} as an stable-isotope internal standard unless otherwise noted. FA concentrations corresponding to limits of detection (LOD) and limits of quantification (LOQ) were calculated based on a serial dilution of calibrant solutions equivalent to a signal-to-noise ratio (*SNR*) of 3 and 10, respectively. Blank (methanol) extracts were also prepared intermittently to confirm lack of sample carry-over effects and background interferences. Reproducibility was evaluated based on intraday (n = 18) and interday (n = 54) precision based on three independent MSI-NACE-MS runs, at the beginning, middle and end of day, each with six replicate injections of a 50 μM FA calibrant mixture and a blank (methanol) over three consecutive days. Method accuracy was assessed by representative FA standards spiked at two different concentration levels into standard human serum (Sigma S7023) prior to MTBE extraction and the % recovery was calculated based on the percentage difference between spiked and original (baseline) concentration of human serum divided by the spiked (known) concentration. In addition to evaluating the accuracy of MSI-CE-MS for NEFAs based on spike-recovery experiments, an inter-method comparison was also conducted on representative serum samples (n = 50) when using GC-MS by the same analyst in the same laboratory, whereas an inter-laboratory method comparison was performed on total FAs from reference plasma samples (SRM 1950). In this case, total plasma FA concentrations derived MSI-CE-MS were compared to reference ranges reported

in a certificate of analysis based on weighted means of results from NIST (GC-FID, GC-MS) and the Centers for Disease Control and Prevention (CDC, GC-MS).

2.8.7 Data Processing and Statistical Analysis

MSI-NACE-MS data was analyzed using Agilent Mass Hunter Workstation Software (Qualitative Analysis, version B.06.00, Agilent Technologies, 2012). Molecular features were extracted in profile mode using a 10 ppm mass window, and FAs were annotated based on their characteristic accurate mass for their deprotonated molecular ion (m/z, $[M\text{-}H]^-$) and relative migration time (RMT) or apparent electrophoretic mobility (μ_{ep}). Extracted ion electropherograms (EIEs) were integrated after smoothing using a quadratic/cubic Savitzky-Golay function (15 points) and peak areas, migration times and *SNR* were transferred to Excel (Microsoft Office, Redmond, WA, USA) for calculation of relative integrated peak area (RPA), RMT, μ_{ep} as well as LOD and LOQ respectively. GC-MS data was analyzed by GC MSD ChemStation Software, version D.03.00, Agilent Technologies Inc. All electropherograms and chromatograms were depicted using Igor Pro 5.0 software (Wavemetric Inc., Lake Oswego, OR, USA). Least-squares linear regression analysis for external calibration curves, figures of merit calculations, FA migration prediction model and control charts were performed using Excel. Principle component analysis (PCA) based on 2D scores plots were used for data visualization (*i.e.*, data trends/outlier detection) when comparing the technical variance (normal serum from Sigma as QC) and overall biological variance (individual serum samples; between-subject) using MetaboAnalyst 4.0,[8] where data was normalized using a generalized *log*-transformation and autoscaling. Also, a QC derived batch-correction algorithm was performed to correct for instrumental drift while improving long-term precision for FA determination using an algorithm available in the R Project for statistical computing.[9] Bland-Altman % difference plots and Passing-Bablok regression used in the inter-

method and inter-laboratory comparisons[10] for NEFA (serum) or total hydrolyzed FA (plasma) determination were performed in MedCalc version 12.5 (MedCalc Software, Ostend, Belgium).

2.8.8 Supporting References

1. Thurnhofer S, Vetter W. A Gas chromatography/electron ionization–mass spectrometry–selected ion monitoring method for determining the fatty acid pattern in food after formation of fatty acid methyl esters. *J Agric Food Chem*. 2005;53:8896-8903.

2. Jafari N, Ahmed R, Gloyd M, Bloomfield J, Britz-McKibbin P, Melacini G. Allosteric sensing of fatty acid binding by NMR: application to human serum albumin. *J Med Chem*. 2016;59:7457-7465.

3. Morrison KM, Atkinson SA, Yusuf S, et al. The Family Atherosclerosis Monitoring In earLY life (FAMILY) study. *Am Heart J*. 2009;158:533-539.

4. Ichihara K, Fukubayashi Y. Preparation of fatty acid methyl esters for gas-liquid chromatography. *J Lipid Res*. 2010;51:635-640.

5. Matyash V, Liebisch G, Kurzchalia TV, Shevchenko A, Schwudke D. Lipid extraction by methyl- *tert* -butyl ether for high-throughput lipidomics. *J Lipid Res*. 2008;49:1137-1146.

6. Phinney KW, Ballihaut G, Bedner M, et al. Development of a standard reference material for metabolomics research. *Anal Chem*. 2013;85:11732-11738.

7. Watts JL, Browse J. Genetic dissection of polyunsaturated fatty acid synthesis in Caenorhabditis elegans. *Proc Natl Acad Sci U S A*. 2002;99:5854-5859.

8. Chong J, Soufan O, Li C, et al. MetaboAnalyst 4.0: towards more transparent and integrative metabolomics analysis. *Nucleic Acids Res*. 2018;46:W486-W494.

9. Wehrens R, Hageman JosA, van Eeuwijk F, et al. Improved batch correction in untargeted MS-based metabolomics. *Metabolomics*. 2016;12:88.

10. de Macedo AN, Macri J, Hudecki PL, Saoi M, McQueen MJ, Britz-McKibbin P. Validation of a capillary electrophoresis assay for monitoring iodine nutrition in populations for prevention of iodine deficiency: An Interlaboratory Method Comparison. *J Appl Lab Med*. 2017;1:649-660.

Table S2.1. Temperature gradient program used for serum NEFA analysis by GC-MS.

Ramp (°C/min)	Temperature (°C)	Hold (min)	Time (min)
0.00	60	1.00	0.00
12.00	180	0.00	12.00
3.00	210	0.00	22.00
30.00	270	7.00	30.00

Table S2.2. EI-MS parameters for SIM optimized for serum NEFA analysis using GC-MS.

Group	Selected Ion (s)	Target	Time (min)	Dwell Time (ms)
1	55, 74, 77	MUFA, SAT. C18*d35*	8	100
2	67, 79	PUFA	13	100
3	212	Pyrene-*d10*	23	100

Table S2.3. Optimization of serum extraction protocol for NEFA analysis for quantitative recovery of 14 FAs when analyzing successive MTBE fractions of standard human serum with an overall average recovery of 92% using the first extract fraction.

Fatty Acid	Molecular Formula	*m/z*:RMT	Relative % Recovery[a]		
			1st Fraction[b]	2nd Fraction[b]	3rd Fraction[b]
Lauric acid (12:0)	$C_{12}H_{24}O_2$	199.1704:1.085	91.3 ± 1.6	4.7 ± 1.5	5.0 ± 3.1
Myristic acid (14:0)	$C_{14}H_{28}O_2$	227.2017:1.054	93.0 ± 0.42	5.5 ± 2.7	3.4 ± 2.3
Palmitic acid (16:0)	$C_{16}H_{32}O_2$	255.233:1.028	81.5 ± 1.5	13.2 ± 3.4	5.3 ± 2.0
Heptadecanoic acid (17:0)	$C_{17}H_{34}O_2$	269.2486:1.017	91.7 ± 8.1	7.02 ± 0.13	1.3 ± 1.8
Stearic acid (18:0)	$C_{18}H_{36}O_2$	283.2643:1.005	84.3 ± 4.8	10.5 ± 5.4	5.23 ± 0.61
Palmitoleic acid (16:1n-7) [c]	$C_{16}H_{30}O_2$	253.2173:1.035	94.3 ± 1.5	4.23 ± 0.01	1.49 ± 0.18
Heptadecenoic acid (17:1n-7)	$C_{17}H_{32}O_2$	267.233:1.024	93.1 ± 1.0	5.06 ± 0.44	1.84 ± 0.58
Oleic acid (18:1n-9) [c]	$C_{18}H_{34}O_2$	281.2486:1.012	95.11 ± 0.41	3.66 ± 0.27	1.24 ± 0.14
Eicosenoic acid (20:1n-9)	$C_{20}H_{38}O_2$	309.2799:0.990	90.85 ± 0.94	6.25 ± 0.15	2.9 ± 0.79
Linoleic acid (18:2n-6) [c]	$C_{18}H_{32}O_2$	279.233:1.018	95.97 ± 0.32	3.03 ± 0.15	0.98 ± 0.17
Eicosadienoic acid (20:2n-6)	$C_{20}H_{36}O_2$	307.2642:0.995	93.9 ± 1.8	4.3 ± 1.8	1.78 ± 0.03
Linolenic acid (18:3n-3) [c]	$C_{18}H_{30}O_2$	277.2173:1.021	93.1 ± 4.9	2.57 ± 0.09	4.4 ± 5.0
Arachidonic acid (20:4n-6)	$C_{20}H_{32}O_2$	303.233:1.025	95.4 ± 0.41	3.66 ± 0.27	1.23 ± 0.14
Docosahexaenoic acid (22:6n-3)	$C_{22}H_{32}O_2$	327.233:1.029	94.4 ± 4.4	4.5 ± 3.8	1.12 ± 0.56

[a] *% Recovery was calculated based on relative peak areas (RPA) of each fraction for each FA assuming 100% recovery from all three extraction fractions collectively. RPA was calculated based on integrated peak area normalized to the integrated peak area of the deuterated internal standard, stearic acid-d35.*

[b] *50 µL of standard human serum (S7023) was repeatedly extracted with 3 X 250 µL portions of MTBE, where three consecutive ether aliquots of 200 µL were withdrawn after addition of 100 µL of deionized water followed by centrifugation for phase separation. Each ether extract fraction was analyzed by MSI-NACE-MS following an evaporation and reconstitution step with deuterated internal standard into a final volume of 25 µL.*

[c] *For serum NEFA with positional or geometric isomers, total FA extraction fractions are reported since they are unresolved by MSI-NACE-MS.*

Table S2.4. Method sensitivity and linearity based on FA calibration curves using the GC-EI-MS[a] method used in this study.

Fatty Acid	Molecular Formula	RT (min)	LOD (μM)	LOQ (μM)	Range (μM)	Linearity (R^2)
Lauric acid (12:0)	$C_{12}H_{24}O_2$	8.32	5.27	6.03	5-500	0.992
Myristic acid (14:0)	$C_{14}H_{28}O_2$	9.69	3.62	4.17	5-500	0.993
Pentadyclic acid (15:0)	$C_{15}H_{30}O_2$	10.33	8.36	8.76	5-500	0.990
Palmitic acid (16:0)	$C_{16}H_{32}O_2$	10.94	0.42	1.39	5-500	0.997
Heptadecanoic acid (17:0)	$C_{17}H_{34}O_2$	11.54	0.84	2.79	5-500	0.998
Stearic acid (18:0)	$C_{18}H_{36}O_2$	12.18	0.51	1.69	5-500	0.996
Palmitoleic acid (16:1)	$C_{16}H_{30}O_2$	11.3	8.33	27.78	5-500	0.99
Elaidic acid (18:1, *trans*)	$C_{18}H_{34}O_2$	12.45	7.89	26.32	5-500	0.995
Oleic acid (18:1, *cis*)	$C_{18}H_{34}O_2$	12.56	7.50	25.00	5-500	0.994
Linoleic acid (18:2n-6)	$C_{18}H_{32}O_2$	13.21	1.90	6.33	5-500	0.992
Linolenic acid (18:3n-3)	$C_{18}H_{30}O_2$	14.06	0.95	3.17	5-500	0.993
Arachidonic acid (20:4n-6)	$C_{20}H_{32}O_2$	16.02	1.20	4.00	5-500	0.996
Docosahexaenoic acid (22:6n-3)	$C_{22}H_{32}O_2$	20.22	1.40	4.50	5-500	0.997

[a] *Analysis of FA standards for GC-EI-MS method validation was performed on a Supelco SP-2380 column (30m x 0.25mm) operating in selected ion monitoring (SIM) mode. Characteristic fragment ions for saturated, monoenoic and polyenoic FAs were monitored at m/z 74, 55 and 79, respectively. The temperature program used for resolution of FAMEs is described in Table S1.*

Table S2.5. Spike and recovery studies for 5 saturated/polyunsaturated FA standards at two concentration levels in human serum.

	Fatty Acid	Molecular Formula	*m/z*:RMT	Concentration in normal serum (μM)	Level 1 spike (μM)	% Recovery[a]	Level 2 spike (μM)	% Recovery[a]
1	Myristic acid (14:0)	$C_{14}H_{28}O_2$	227.2017:1.054	14.10 ± 0.07	+25	90 ± 7	+50	73 ± 5
2	Pentadyclic acid (15:0)	$C_{15}H_{30}O_2$	241.2173:1.041	3.21 ± 0.02	+8	106 ± 1	+16	80 ± 3
3	Linolenic acid (18:3n-3)	$C_{18}H_{30}O_2$	277.2173:1.021	17.95 ± 0.43	+30	97 ± 2	+60	95 ± 12
4	Docosahexaenoic acid (22:6n-3)	$C_{22}H_{32}O_2$	327.233:1.029	6.56 ± 0.04	+10	99 ± 11	+20	80 ± 5
5	Eicosapentaenoic acid (20:5n-3)	$C_{20}H_{30}O_2$	301.2173:1.027	3.36 ± 0.09	+10	80 ± 4	+20	***Average Recovery = 93 %***

[a] *% recovery for NEFAs was calculated based on the percentage difference between spiked and original concentrations (in normal human serum from Sigma) divided by the spiked concentration with triplicate measurements performed (n=3).*

Table S2.6. Method validation and summary of figures of merit for high throughput screening of fatty acids by MSI-NACE-MS.

						Precision[b] (%RSD)					Linearity[d]			
Fatty Acid	**Molecular Formula**	**Mobility[a] (cm^2/Vs)**	***m/z* [M-H]⁻**	**Mass Error (ppm)**	**RMT**	**RMT (n=78)**	**RPA Interday (n=78)**	**RPA Intraday (n=18)**	**LOD[c] (µM)**	**LOQ[c] (µM)**	**Range (µM)**	***R²***	**Peak width (min)**	**Number of Theoretical Plates[e]**
Capric acid (10:0)	$C_{10}H_{20}O_2$	-0.0001636	171.1390	2.6	1.123	1.06	23.3	7.62	2.5	8.4	1-200	0.994	0.71	23,702
Lauric acid (12:0)	$C_{12}H_{24}O_2$	-0.0001554	199.1704	2.0	1.085	0.73	10.7	9.32	0.70	2.4	1-200	0.995	0.78	26,294
Myristic acid (14:0)	$C_{14}H_{28}O_2$	-0.0001489	227.2017	3.5	1.054	0.49	11.8	14.4	0.90	3.0	1-200	0.991	0.59	27,970
Pentadyclic acid (15:0)	$C_{15}H_{30}O_2$	-0.0001454	241.2173	2.1	1.041	0.38	11.7	7.55	0.80	2.7	1-200	0.995	0.64	28,670
Palmitic acid (16:0)	$C_{16}H_{32}O_2$	-0.0001421	255.2330	1.2	1.028	0.28	13.9	4.59	6.9	23	5-200	0.991	0.67	42,832
Heptadecanoic acid (17:0)	$C_{17}H_{34}O_2$	-0.0001391	269.2486	2.2	1.017	0.25	7.00	3.97	0.30	1.1	1-200	0.997	0.63	25,131
Stearic acid (18:0)	$C_{18}H_{36}O_2$	-0.0001359	283.2643	4.6	1.005	0.23	14.4	8.56	4.2	14	5-200	0.992	0.64	32,198
Arachidic acid (20:0)	$C_{20}H_{40}O_2$	-0.0001302	311.2956	3.8	0.983	0.26	29.3	29.3	0.40	1.3	1-200	0.995	0.71	22,388
Behenic acid (22:0)	$C_{22}H_{44}O_2$	-0.0001243	339.3269	0.59	0.960	0.39	12.3	6.15	0.7	2.4	1-200	0.991	0.63	25,610
Lignoceric acid (24:0)	$C_{24}H_{48}O_2$	-0.0001200	367.3582	0.54	0.945	0.65	16.0	14.2	1.4	4.5	1-200	0.974	0.48	26,523
Palmitoleic acid (16:1n-7)	$C_{16}H_{30}O_2$	-0.0001442	253.2173	2.4	1.035	0.36	13.8	10.73	0.70	2.3	1-200	0.993	0.66	24,550
Oleic acid (18:1n-9)	$C_{18}H_{34}O_2$	-0.0001382	281.2486	1.1	1.012	0.22	13.7	7.03	0.60	2.2	1-200	0.995	0.73	32,705
Erucic acid (22:1n-9)	$C_{22}H_{42}O_2$	-0.0001272	337.3112	0.59	0.966	0.28	8.66	5.8	0.90	3.1	1-200	0.995	0.54	32,699
Nervonic acid (24:1n-9)	$C_{24}H_{46}O_2$	-0.0001208	365.3425	5.2	0.946	0.48	10.2	9.66	0.10	0.4	1-200	0.996	0.68	28,290
Linoleic acid (18:2n-6)	$C_{18}H_{32}O_2$	-0.0001396	279.2330	0.36	1.018	0.26	8.41	5.35	0.30	1.1	1-200	0.996	0.69	24,350
Linolenic acid (18:3n-3)	$C_{18}H_{30}O_2$	-0.0001411	277.2173	1.4	1.021	0.26	7.93	3.34	0.20	0.6	1-200	0.996	0.68	24,628
Arachidonic acid (20:4n-6)	$C_{20}H_{32}O_2$	-0.0001413	303.2330	2.3	1.025	0.27	9.05	4.54	0.08	0.27	1-200	0.993	0.63	36,869
Docosahexaenoic acid (22:6n-3)	$C_{22}H_{32}O_2$	-0.0001423	327.2330	2.4	1.029	0.26	9.91	5.22	0.10	0.38	1-200	0.997	0.57	64,004
					Median RSD	**0.32%**	**12%**	**7.3%**						

[a] Apparent average negative electrophoretic mobility for FAs was measured using 35 mM ammonium acetate in ACN (70% v/v), MeOH (15% v/v), IPA (5% v/v), pH_{app} 9.5 under an applied voltage of 30 kV with 20 mbar pressure applied during separation within a 95 cm fused-silica capillary at 25 °C using BHT as a neutral EOF marker.

[b] Precision were assessed by analyzing six replicate injections of 50 µM FA calibrants along with a blank at the beginning, middle and end of day, over three consecutive days.

[c] LOD and LOQ were estimated at the lowest FA concentration that generated a SNR ≈ 3 and 10, respectively.

[d] External calibration curves were derived from triplicate analysis of seven calibrant solutions for FAs over a 200-fold concentration range normalized to stearic acid-d35.

[e] Average plate number (n=6) for separation of fatty acids by MSI-NACE-MS based on migration time (MT) and width at half peak height ($w_{0.5}$) method, where $N = 5.54 (MT/w_{0.5})^2$

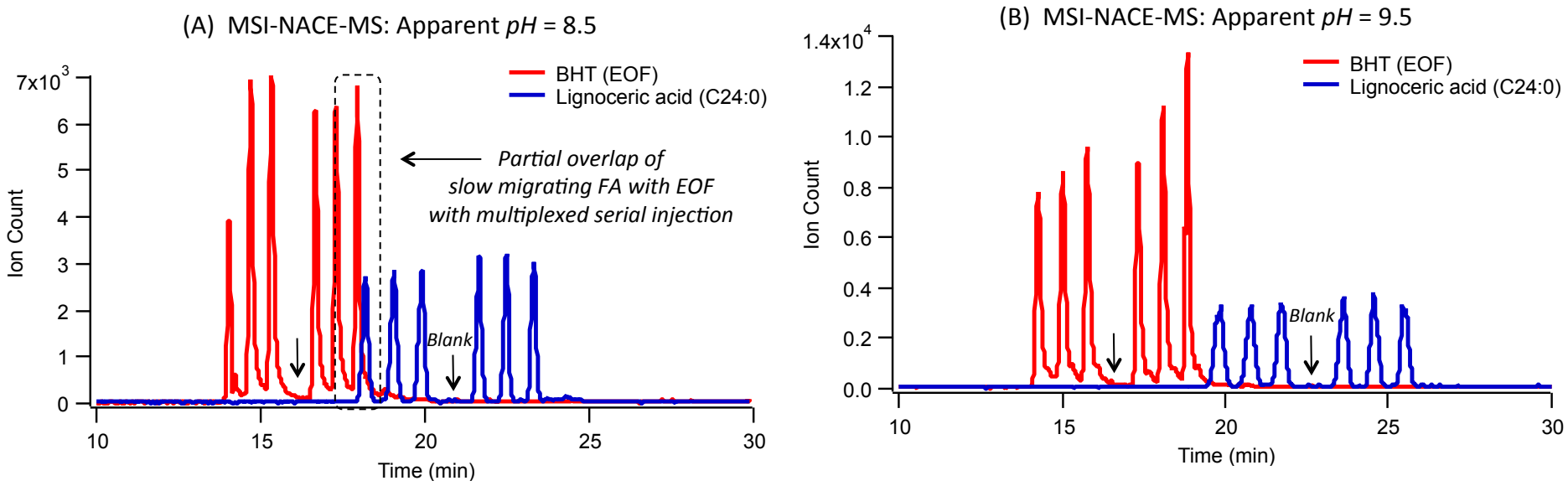


Figure S2.1. Extracted ion electropherograms (EIEs) illustrating the impact of apparent pH of non-aqueous background electrolyte on the resolution of the neutral EOF marker and antioxidant added to all FA calibrant solutions, butylated hydroxytoluene (BHT), from lignoceric acid (24:0) that is a very-long chain FA with the lowest negative electrophoretic mobility. Serial injection of a six calibrant samples at 50 µM together with a blank (sample#4) using MSI-NACE-MS highlighting (A) partial overlap of 24:0 in the late migrating zone of sample#7 with BHT/EOF from sample#1 at pH_{app} = 8.5 that is prone to ion suppression effects, as compared to improved baseline resolution at pH_{app} = 9.5. MSI-NACE-MS conditions: fused-silica capillary: 95 cm total length and 50 µm, id; BGE: 75 mM ammonium acetate in ACN (70% *v/v*), MeOH (15% *v/v*), IPA (5% *v/v*), adjusted to desired pH_{app} with ammonium hydroxide; applied voltage of 30 kV with 20 mbar applied pressure during separation; temperature, 25°C; serial hydrodynamic injection at 40 mbar alternating 5 s for sample and 40 s for BGE spacer. TOF-MS conditions: Full-scan data acquisition under negative ion mode detection; capillary voltage 3500 V; nebulizer gas N_2 at 0 psi during flushing/injection and then set at 4 psi following electrophoretic separation; dry gas 4 L/min at 300 °C; sheath liquid, MeOH/H_2O (80:20 *v/v*) with 0.5% NH_4OH at 1.0 mL/min.

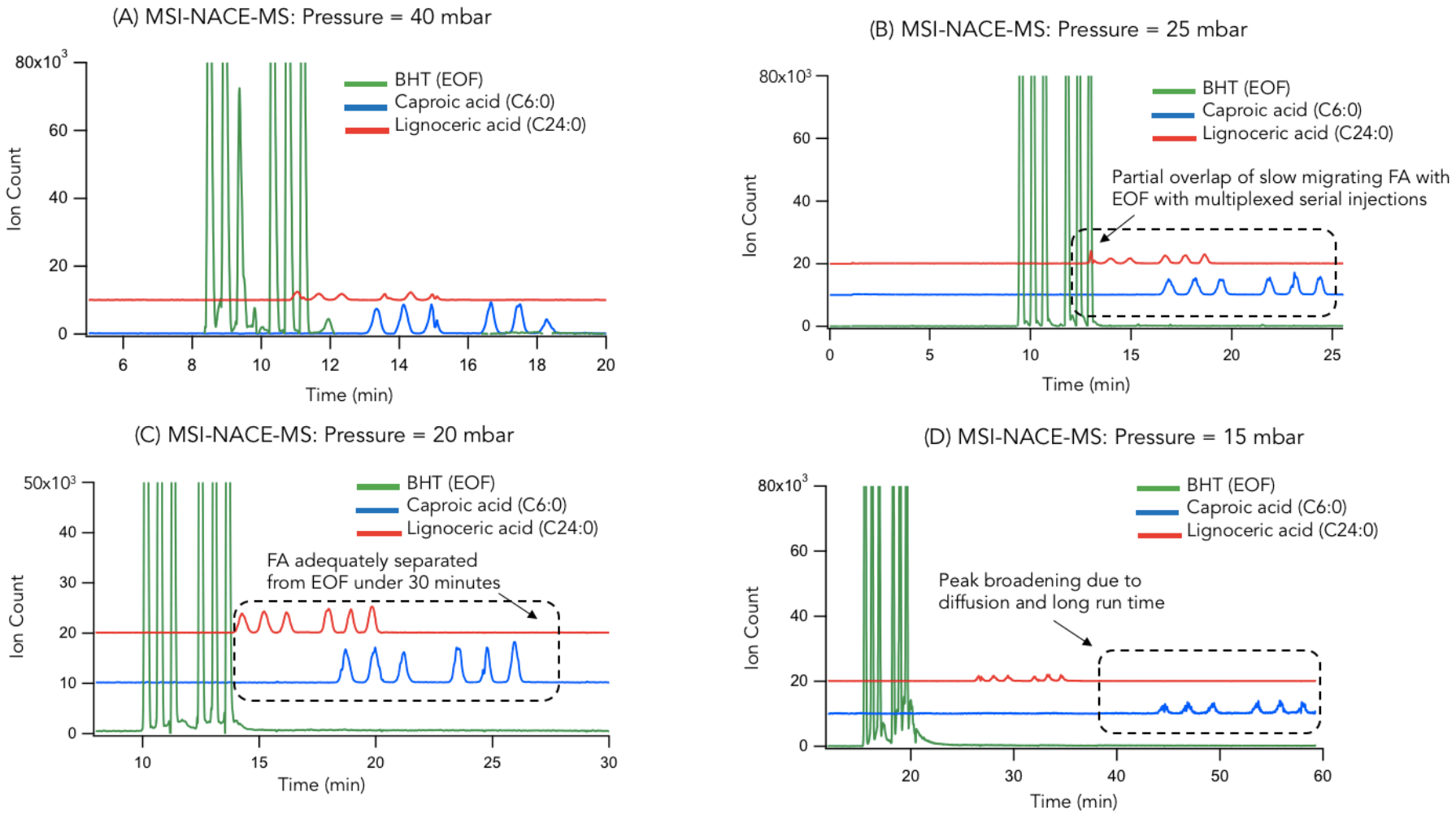


Figure S2.2. Extracted ion electropherograms (EIEs) illustrating the impact of pressure assistance on the resolution of the neutral EOF marker, butylated hydroxytoluene (BHT), from lignoceric acid (24:0) that is a very-long chain FA with the lowest negative electrophoretic mobility and on the migration of caproic acid (6:0) that is a short-chain FA determining total run time. Serial injection of a six calibrant samples at 50 µM together with a blank (sample#4) using MSI-NACE-MS highlighting (A, B) partial overlap of 24:0 in the late migrating zone of sample#1 with BHT/EOF from sample#7 at pressures = 40 or 35 mbar that is prone to ion suppression effects, as compared to (C) improved baseline resolution at pressure = 20 mbar and total run time under 30 min for seven samples. (D) Further decrease of pressure = 15 mbar drastically increases run times to 60 min and results in deleterious band broadening. MSI-NACE-MS conditions: fused-silica capillary: 95 cm total length and 50 µm, id; BGE: 35 mM ammonium acetate in ACN (70% *v/v*), MeOH (15% *v/v*), IPA (5% *v/v*), adjusted to pH_{app} = 9.5; applied voltage of 30 kV with desired applied pressure during separation; temperature, 25°C; serial hydrodynamic injection at 40 mbar alternating 5 s for sample and 40 s for BGE spacer.

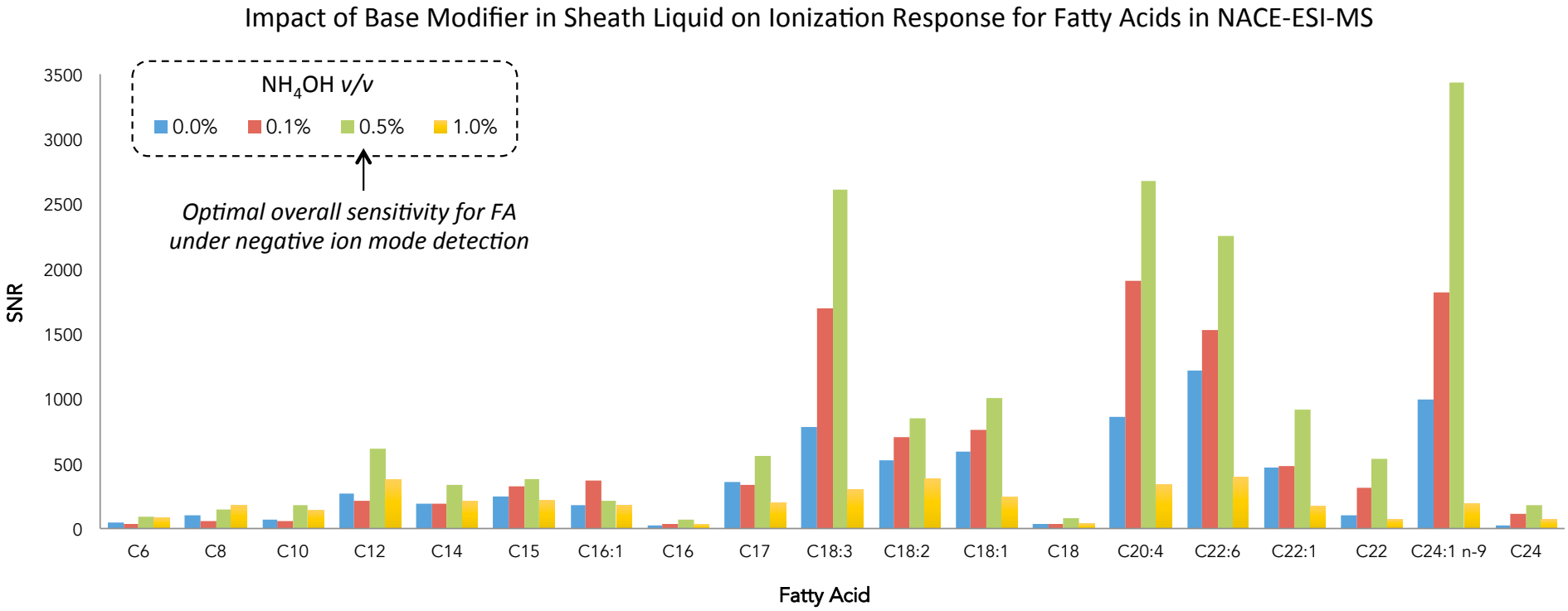


Figure S2.3. Bar graphs comparing the effect of NH_4OH as a base modifier in the sheath liquid to improve ionization efficiency of fatty acids (FAs) when using MSI-NACE-MS under negative ion mode detection. A maximum ionization response based on measured signal-to-noise ratio (SNR) for majority of FAs was achieved with 0.5% *v/v* NH_4OH when using a coaxial sheath liquid interface in CE-MS to enhance deprotonation of FAs during spray formation. However above 1% *v/v* NH_4OH, the average SNR generally decreased for most FAs likely due to ion suppression effects at higher concentrations as compared to no base modifier added to the sheath liquid (80:20 *v/v* MeOH:H_2O).

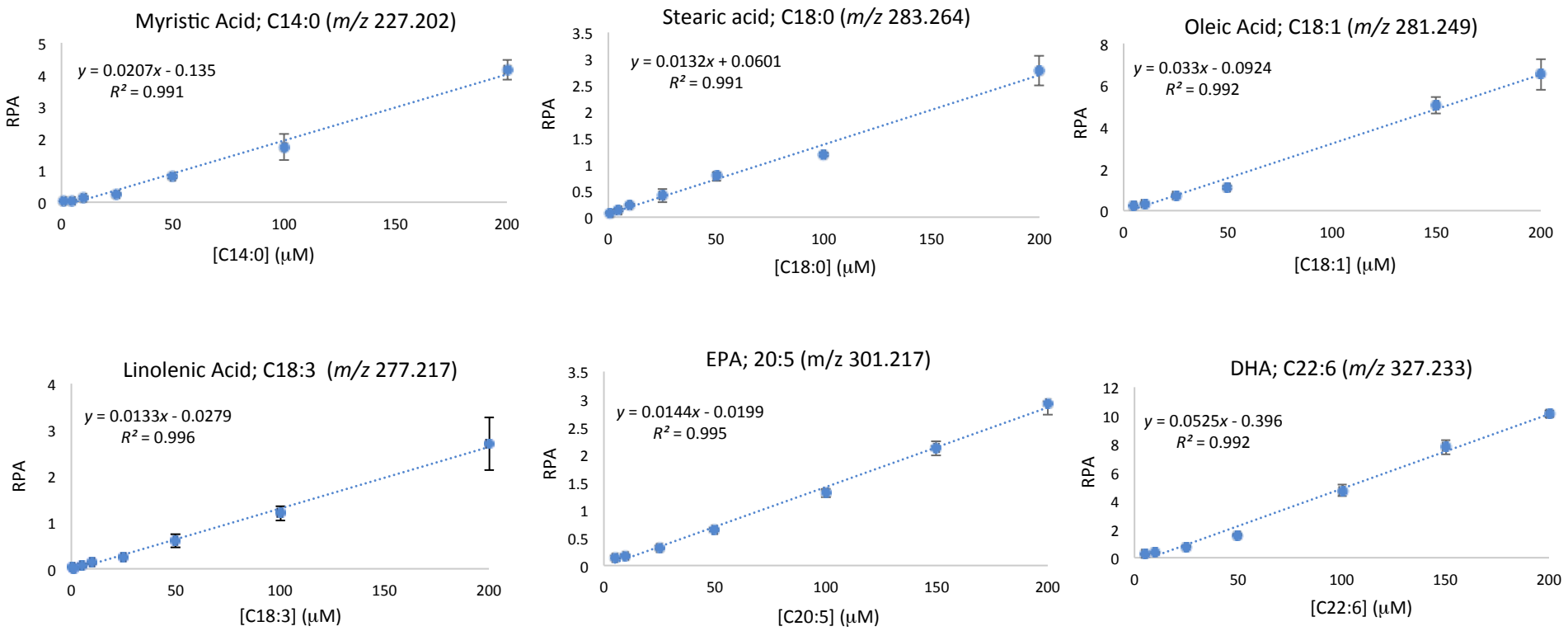


Figure S2.4. Representative calibration curves for saturated, monounsaturated and polyunsaturated FAs when using MSI-NACE-MS under negative ion mode detection, where their ion responses are normalized to a single deuterated internal standard, stearic acid-*d35*. Overall, good linearity ($R^2 > 0.99$) was achieved over a 200-fold dynamic range when using a least-squares linear regression model with low micromolar to sub-micromolar detection limits. Importantly, a single MTBE extract enables quantitative recovery of serum NEFAs that is amenable to automation without complicated sample handling or pre-column chemical derivatization.

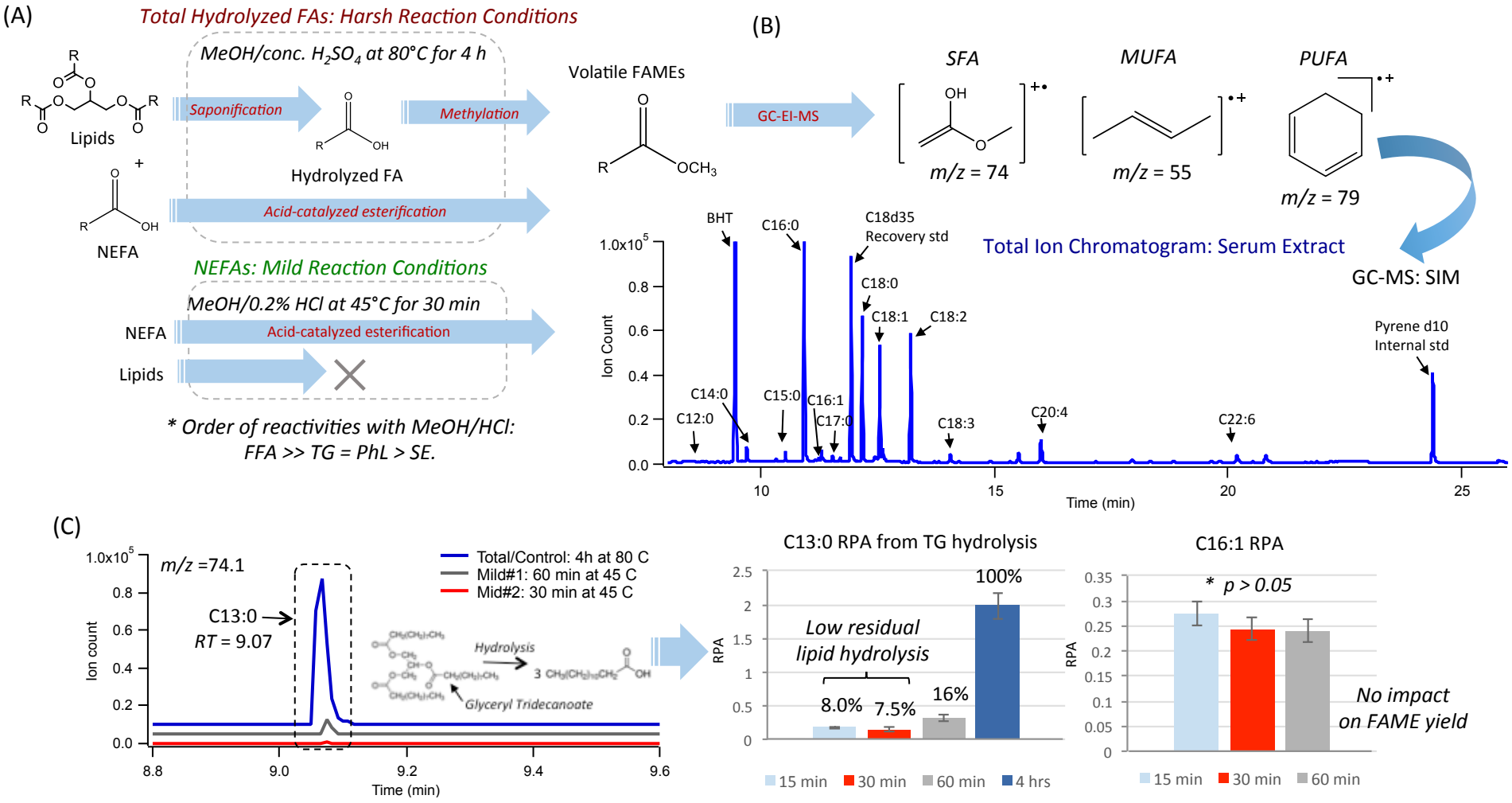


Figure S2.5. (A) Overview of sample workup protocol developed for analysis of fatty acid methyl esters (FAMEs) from serum extracts under mild (for NEFA) and harsh (for total hydrolyzed FAs) reaction conditions. (B) Characteristic fragment ions for saturated, monounsaturated and polyunsaturated FAs were measured by GC-EI-MS with using selective ion monitoring (SIM) as reflected in the total ion chromatogram for a representative serum extract. (C) Reduction of temperature (from 80 to 45 °C) and incubation time (from 4 h to 30 min) was able to ensure quantitative yield of FAMEs for protein-bound/free FAs under mild conditions (*e.g.*,16:1) with a low extent of lipid hydrolysis (*e.g.*, spiked glyceryl tridecanoate in serum) as reflected by low residual amount of 13:0 formation (< 8% of total) as compared to harsh reaction conditions. This method takes advantages of the different reactivities of FAs, such as protein-bound or free fatty acids (FFA) as compared to FAs incorporated into various lipid classes, including phospholipids (PhL), triglycerides (TG) and steroid esters (SE), such as cholesterol esters.

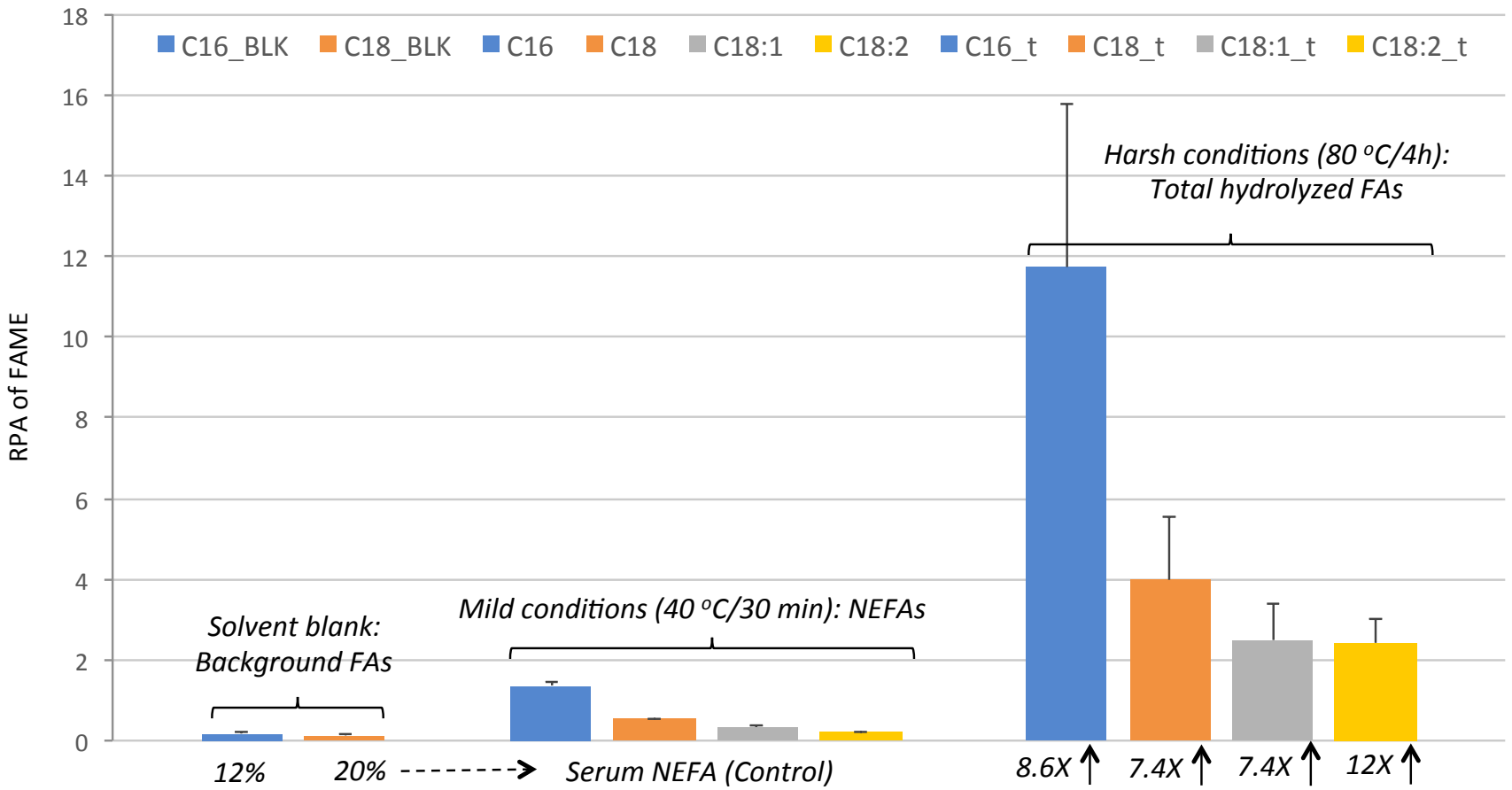


Figure S2.6. Bar graph showing extent of FA derivatization into FAMEs under mild reaction conditions (NEFAs) as compared to harsh reaction conditions (total hydrolyzed FAs) when using GC-EI-MS based on triplicate analysis of standard human serum samples along with a solvent blank mainly attributed to 16:0 and 18:0 from background due to plastic tubing and liners.

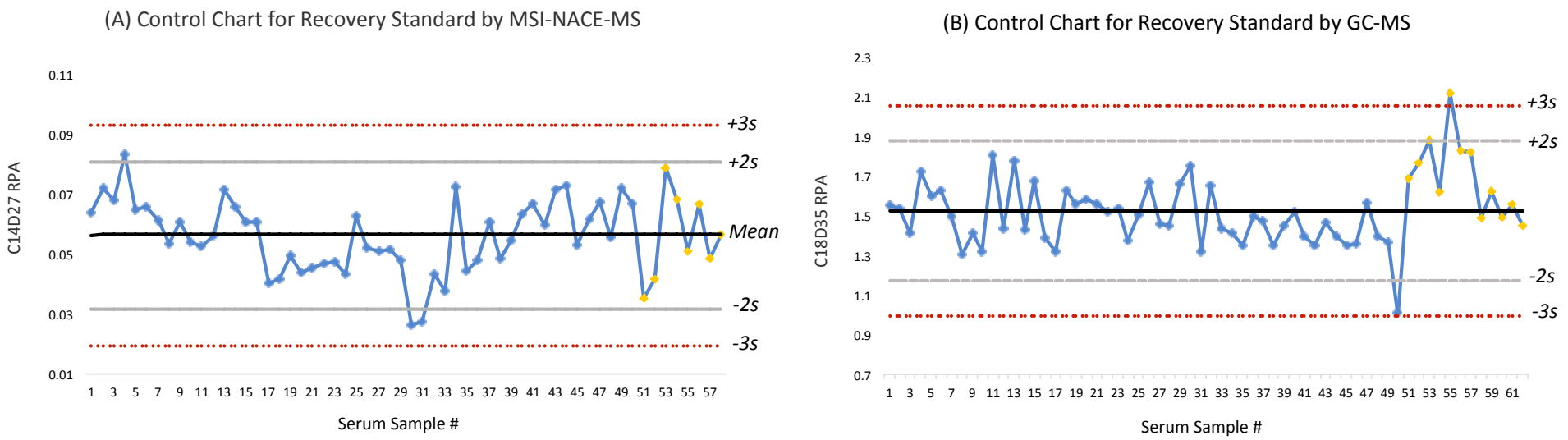


Figure S2.7. Control charts for deuterated recovery standards added to all serum samples (n = 50) and QCs (n = 7-12) depicting the long-term robustness of MSI-NACE-MS and GC-EI-MS methods for reliable serum NEFA determination. Both methods display good overall technical precision when measuring recovery standards added to serum samples prior to extraction at the same concentration as reflected by a random distribution about a mean with few outliers exceeding both warning (± 2s) or action (± 3s) limits.

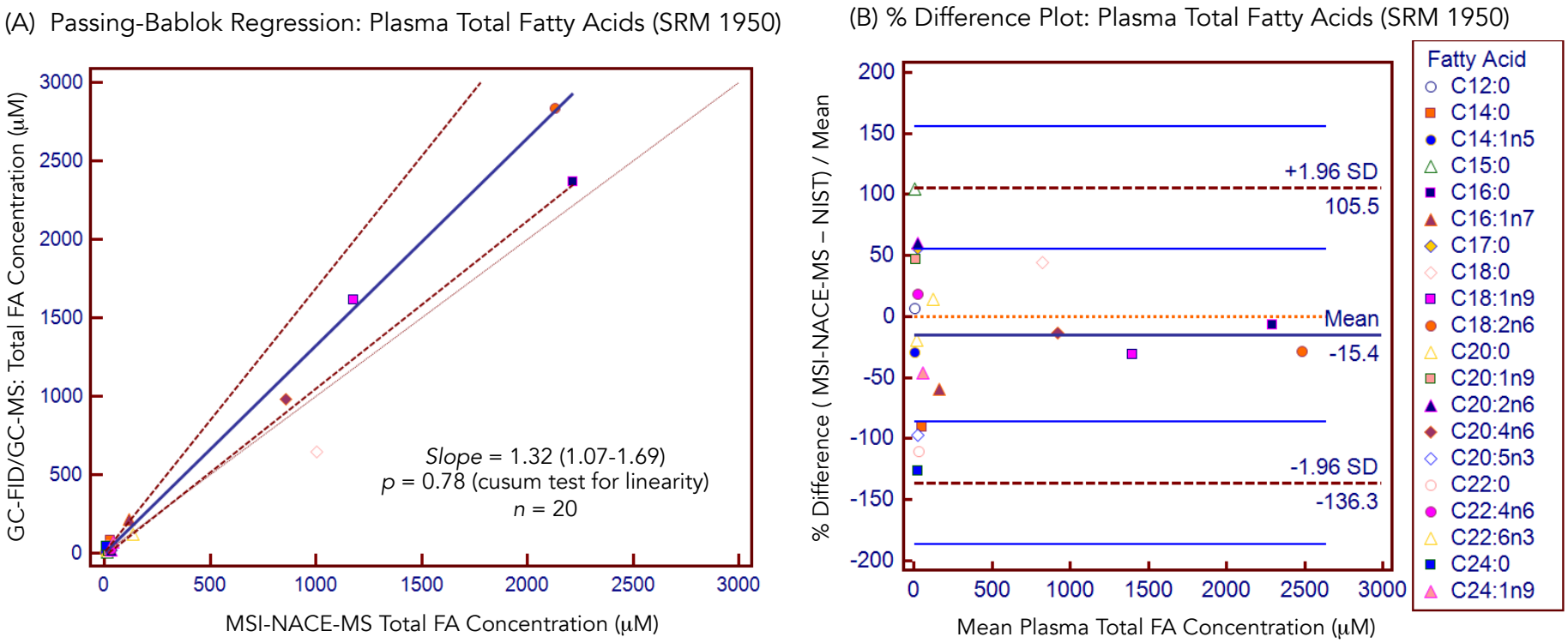


Figure S2.8. An inter-laboratory method comparison for accurate quantification of 20 total hydrolyzed FAs measured in a certified reference plasma sample from NIST (SRM 1950). (A) A Passing-Bablok regression plot highlights good correlation with a modest positive bias (linear slope of 1.32) for total serum FAs measured independently by MSI-CE-MS as compared to mean results from GC-FID and GC-MS methods (NIST; CDC) using different extraction protocols. (B) Similarly, a Bland-Altman % difference plot also shows good mutual agreement between results with a mean bias of 15% with few outliers exceeding agreement limits (95% CI).

Chapter III:

Rapid Biomonitoring of Perfluoroalkyl Substance Exposures in Serum by Multisegment Injection-Nonaqueous Capillary Electrophoresis-Tandem Mass Spectrometry

Thesis chapter is derived from a published peer-reviewed article:

S.M. Azab, R. Hum and P. Britz-McKibbin. Rapid biomonitoring of perfluoroalkyl substance exposures in serum by multisegment injection-nonaqueous capillary electrophoresis-tandem mass spectrometry. *Anal. Sci. Adv.* **2020**, 1: 1-10. https://doi.org/10.1002/ansa.202000053.

S.M.A. and R.H. conducted the experiments for development, optimization, and validation of the MSI-NACE-MS/MS method as well as data processing and data analysis. S.M.A. conducted serum sample processing and analysis for method application. S.M.A. wrote the initial manuscript draft. P.B.M. provided critical feedback for experiments and data analysis and edited manuscript for publication.

Chapter III: Rapid Biomonitoring of Perfluoroalkyl Substance Exposures in Serum by Multisegment Injection-Nonaqueous Capillary Electrophoresis-Tandem Mass Spectrometry

"Test all things; hold fast to that which is good"

3.1 Abstract

Perfluoroalkyl substances (PFASs) are a major contaminant class due to their ubiquitous prevalence, persistence and putative endocrine disrupting activity that may contribute to chronic disease risk notably with exposures early in life. Herein, multisegment injection-nonaqueous capillary electrophoresis-tandem mass spectrometry (MSI-NACE-MS/MS) is introduced as a high throughput approach for PFAS screening in serum samples following a simple methyl-*tert*-butyl ether (MTBE) liquid extraction. Separation and ionization conditions were optimized to quantify low nanomolar concentration levels of perfluorooctanoic acid (PFOA) and perfluorooctanesulfonic acid (PFOS) from serum extracts when using multiple reaction monitoring under negative ion mode conditions. Multiplexed separations of PFOA and PFOS were achieved with excellent throughput (< 3 min/sample), adequate concentration sensitivity (LOD ~ 20 nM, *SNR* = 3) and good technical precision over three consecutive days of analysis (mean CV = 9.1%, n = 84). Accurate quantification of PFASs was demonstrated in maternal serum samples (n = 16) when using MSI-CE-MS/MS following pre-column sample enrichment with median concentrations of 3.46 nM (0.7 to 9.0 nM) and 3.29 nM (1.5 to 6.6 nM) for PFOA and PFOS, respectively. This was lower than average PFAS exposures assessed for pregnant women who had serum collected prior to 2009 due to subsequent phase out of their production. Overall, this method offers a convenient approach for large-scale biomonitoring of environmental exposures to legacy PFASs and their emerging replacements that is relevant to maternal health and chronic disease risk in children.

3.2 Introduction

Perfluoroalkyl substances (PFASs) are synthetic compounds first developed in the 1940s due to their unique thermal and chemical properties that render them useful in a myriad of consumer products and industrial applications.[1,2] These ubiquitous chemicals are both water and oil-repellant, with heat-resistant and surface-active properties comprising over 4,000 different perfluoroalkyl acid analogs.[3] PFASs have been incorporated in food packaging, electronics, stain-resistant textiles, non-stick coatings, as well as firefighting foams.[3-5] Yet, PFASs are a major contaminant class in environmental toxicology given their persistence and tendency to bioaccumulate with widespread human exposures from the ingestion of contaminated food and drinking water,[6,7] as well as household cookware, treated carpets and waterproofed clothing.[8] Biomonitoring studies have largely focused on analyzing the two most abundant contaminants from this chemical class, namely perfluorooctanoic acid (PFOA) and perfluorooctanesulfonic acid (PFOS), which are classified as persistent organic pollutants (POPs) under the Stockholm Convention. With a global phase-out of PFOA and PFOS production with exceptions granted to certain products (*e.g.*, medical devices), other alternatives have replaced these legacy contaminants, including short-chain (C4-C6) PFAS substitutes, and chlorinated polyfluoroalkyl ether sulfonic acids (Cl-PFESAs);[2,9] however, the long-term impacts of these emerging replacements for regulated PFASs on human and wildlife health remain poorly understood.[10] A study by the U.S. Center for Disease Control and Prevention (CDC) reported declining serum concentrations of PFOS in the population, constant levels of PFOA and even increased concentrations of certain other non-regulated PFASs (*e.g.*, perfluorononanoic acid).[11] Long-range atmospheric and/or oceanic transport to polar regions with subsequent biotic uptake may explain these observations due to unregulated manufacturing and emission of PFASs occurring in developing countries.[12,13] For these reasons, there is urgent

need for continued surveillance of PFASs as they are not effectively removed from wastewater treatment plants, which can also generate unexpected chlorinated by-products.[14]

Human exposure studies have shown that PFASs are predominately stored within the liver and in the circulation bound to plasma proteins with an estimated half-life ranging from 4 to 8 years for PFOS and PFOA.[3,15] As a result, blood is a specimen of choice for biomonitoring of PFAS exposures in the population.[16] Also, human breast milk represents a complementary specimen for assessment of PFAS exposures during infancy,[17] as well as urine which is correlated with PFAS in serum.[18] PFASs may function as potential endocrine disrupting chemicals due to their ability to compete with thyroxine binding to human thyroid hormone transport protein transthyretin[19] along with antagonist effects on receptors for testosterone[20] and progesterone.[21] Nevertheless, safe levels of PFAS exposures and their exact mechanisms of toxicity remain unclear[22] with health advisory limits of 0.07 μg/L PFOS and PFOA in drinking/ground water.[23] Similar to other environmental pollutants, children have a higher burden of PFASs as compared to adults,[24] attributed to transfer *in utero* from maternal exposures, feeding of breastmilk and/or commercial baby formula,[25] and ingestion from active mouthing behaviour that is exacerbated by indoor dust exposure.[26] Several studies have reported an inverse relation of PFOA and PFOS exposures to birth weight, as well as a direct relation to dyslipidemia and impaired gestational glucose homeostasis later in life.[4,27,28] In a study evaluating the association of maternal POPs serum levels with the risk of childhood obesity, PFOA and PFOS have been associated with increased BMI and risk for overweight/obesity in early childhood.[29] On the other hand, serum PFAS concentrations in children did not show a similar association, suggesting that exposure to environmental stressors *in-utero* may represent a critical period of susceptibility during gestation.[29,30] Further research is needed to assess the long-term health impacts of early life

exposures to PFASs and mixtures of other endocrine disrupting chemicals[31] using high throughput methods that enable low-cost and large-scale biomonitoring studies in environmental epidemiology.[32]

Analytical methods for PFAS determination require exquisite sensitivity and selectivity without complicated sample workup or background/matrix interferences.[33] Liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS) with multiple reaction monitoring (MRM) is the gold standard for analysis of nanomolar levels of PFAS in complex biological and environmental samples.[16,34,35] Standardized protocols have been developed for reliable PFAS analysis by LC-MS/MS following pre-column sample enrichment and off-line clean-up based on ion-pair extraction or anion-exchange solid-phase extraction (SPE).[34,36,37] Recent advances for PFAS screening by LC-MS/MS include on-line SPE with column switching to automate sample processing and reduce total analysis times (~ 28 min run/sample).[38] Capillary electrophoresis (CE) offers an alternative microseparation platform, but is far less commonly used for routine analysis of PFASs due to poor separation performance in aqueous buffer systems and/or inadequate concentration sensitivity and selectivity when using UV detection.[39-41] Recently, CE-MS using a novel nanospray interface was reported for anionic micropollutants analysis in drinking water samples, however deleterious peak broadening was noted for the surface-active contaminants, PFOA and PFOS.[42] Herein, we introduce a new strategy for high throughput screening of PFOA and PFOS in serum extracts when using multisegment injection-nonaqueous capillary electrophoresis-tandem mass spectrometry (MSI-NACE-MS/MS) following a simple methyl-*tert*-butyl ether (MTBE) extraction protocol.[43,44] Extensive method optimization and validation was performed to demonstrate reliable serum PFOA and PFOS analyses by MSI-NACE-MS/MS with

stringent quality control (QC), which was applied to assess PFAS exposures in women during pregnancy.

3.3 Experimental Section

3.3.1 Chemicals and reagents

Ultra LC-MS grade methanol (Caledon Inc., Georgetown, ON, Canada) and ultra LC-MS grade acetonitrile (Honeywell Inc., Muskegon, MI, USA) were used to prepare sheath liquid and background electrolyte (BGE), respectively. Ammonium acetate, ammonium hydroxide, butylated hydroxytoluene (BHT), methyl-*tert*-butyl ether (MTBE), PFOA and PFOS standards, chemicals, solvents, as well as standard human serum (S7023) used for initial optimization, were purchased from Sigma-Aldrich Inc. (St. Louis, MO, USA). The stable isotope sodium salt of perfluorooctanesulfonate ($^{13}C_8$-PFOS) was obtained from Cambridge Isotope Laboratories, Inc (Tewksbury, MA, USA).

3.3.2 CE-MS instrumentation

An Agilent 6470 triple quadrupole (QQQ) mass spectrometer with a coaxial sheath liquid electrospray ionization (ESI) source coupled to an Agilent 7100 CE unit was used for all experiments (Agilent Technologies Inc., Mississauga, ON, Canada). An Agilent 1260 infinity isocratic pump and degasser were used to deliver a sheath liquid mixture of methanol:water (80:20 *vol*) with 0.5% *vol* NH_4OH at a flow rate of 10 µL/min using a CE-MS coaxial sheath liquid interface kit. The nebulizing gas (N_2) was set off during serial sample injection and then turned on at a pressure of 10 psi during separation following voltage application.[43, 44] The source capillary voltage (Vcap) was set at 3500 V with the drying gas at a flow rate of 4 L/min and the source

temperature at 300 °C in negative ion mode. For all PFAS standards, an MRM scan was performed including quantifier and qualifier ion transitions generated under optimal fragmentation voltages as summarized in **Table 3.1**. Separations were performed on bare fused-silica capillaries with 50 µm internal diameter, 360 µm outer diameter and 90 cm total length (Polymicro Technologies Inc., AZ, USA). A capillary window maker (MicroSolv, Leland, NC, USA) was used to remove about 7 mm of the polyimide coating on both ends of the capillary to prevent polymer swelling when in contact with organic solvents.[45] The applied voltage was set to 30 kV at 25 °C for CE separations together with a pressure application of 20 mbar (2 kPa) for the first 8 min followed by a 2 mbar per minute gradient pressure increase. The nonaqueous background electrolyte (BGE) was composed of 35 mM of ammonium acetate in acetonitrile (70% *vol*), methanol (15% *vol*) and isopropanol (5% *vol*) with an apparent pH of 9.5 adjusted by addition of 12% *vol* of concentrated ammonium hydroxide. Samples were injected hydrodynamically at 50 mbar (5 kPa) alternating between 10 s for each sample plug and 40 s for the nonaqueous BGE spacer plug for a total of seven discrete samples analyzed within a single run. Prior to first use, capillaries were conditioned by flushing for 5 min at 950 mbar (95 kPa) sequentially with methanol, 0.1 M sodium hydroxide, deionized water, 1.0 M formic acid, deionized water then nonaqueous BGE for 15 min. Between runs, the capillary was flushed with nonaqueous BGE for 5 min at 950 mbar (95 kPa). Both nonaqueous BGE and sheath liquid solutions were degassed before use.

3.3.3 Calibration and method validation of MSI-NACE-MS/MS

10 mM stock solutions of PFOA and PFOS calibrants were prepared in MTBE with addition of 0.1% *w/v* BHT as a neutral electroosmotic flow (EOF) marker. A serial dilution of calibrant solutions from 0.025 to 1.0 µM was prepared in triplicate when constructing seven-point calibration curves for PFASs using least-squares linear regression. All integrated peak areas for

the quantifier ions were normalized using 1.0 μM $^{13}C_8$-PFOS as a stable-isotope internal standard. PFAS concentrations corresponding to limits of detection (LOD) and limits of quantification (LOQ) were calculated based on a serial dilution of calibrant solutions equivalent to a signal-to-noise ratio (*SNR*) of 3 and 10, respectively. Blank extracts were also prepared intermittently to confirm lack of sample carry-over effects and background interferences. Reproducibility was evaluated via intra-day ($n = 28$) and inter-day ($n = 84$) precision studies based on three independent MSI-NACE-MS runs, at the beginning, middle and end of day, each with seven replicate injections of a 1.0 μM PFAS calibrant mixture over three consecutive days. Method accuracy was assessed by spike-recovery studies using PFOA and PFOS standards at three different concentration levels (100, 500, 1000 nM) into standard human serum (Sigma S7023) prior to MTBE extraction. PFAS recovery was calculated based on the percentage difference between spiked and original (baseline) concentration of human serum divided by the spiked (known) concentration.

3.3.4 Study birth cohorts

SouTh Asian birth cohoRT (START) study is a prospective birth cohort study involving 1006 predominantly South Asian pregnant women recruited from Brampton and Mississauga, Ontario between 2011 and 2015.[46] Fasting blood samples were collected in the second trimester, and serum was fractionated within 2 h from collection according to standard protocols and stored at -80 °C. A subset of residual serum samples from START birth cohort ($n = 16$) were analyzed for PFOA and PFOS in this work. The Family Atherosclerosis Monitoring In earLY life (FAMILY) study is a prospective birth cohort study involving 839 predominantly white European pregnant women recruited from the greater Hamilton area between 2002 and 2009.[47] Fasting blood samples were collected in the second trimester, and serum was fractionated within 2 h from collection according to standard protocols and stored at -80 °C. A pooled serum sample from FAMILY was used as a

reference sample for PFAS analysis during method validation and was also used as a QC and introduced in every run to monitor technical precision when using MSI-NACE-MS/MS. Ethical approval and informed consent from all study participants were obtained.[46,47]

3.3.5 Sample workup procedure for serum extracts

Human serum samples were prepared using a slightly modified extraction protocol using MTBE originally developed by Matyash *et al*.[44, 48] First, 100 µL of 0.01% *vol* BHT in methanol was mixed with a 200 µL aliquot of serum. Next, 500 µL of 0.01 µM $^{13}C_8$-PFOS in MTBE and 25 µL of 1.0 M HCl were added to the mixture followed by vigorous shaking for 30 min at room temperature. Phase separation was then induced by addition of 200 µL of deionized water. Samples were then centrifuged at 3000 *g* at 4 °C for 30 min to sediment protein at the bottom of the vial followed by a biphasic water and ether (top) layer. A fixed volume (400 µL) was collected from the upper MTBE layer into a new vial, then dried under a gentle stream of nitrogen gas at room temperature. Serum extracts were then stored at -80 °C and reconstituted in 5.0 µL of acetonitrile/isopropanol/water (70:20:10 *vol*) with 10 mM ammonium acetate prior to analysis. This extraction procedure results in an overall 40-fold enrichment of PFAS from serum with good quantitative recovery. Standard human serum samples (100 μL) were spiked with 100, 500 and 1000 nM PFAS in triplicates and extracted following the same protocol.

3.3.6 Data processing and statistical analysis

MSI-NACE-MS data was analyzed using Agilent Mass Hunter Workstation Software (Qualitative Analysis, version B.06.00, Agilent Technologies Inc., 2012). Molecular features were extracted in profile mode using a 10 ppm mass window for all transitions. Extracted ion electropherograms (EIEs) were integrated after smoothing using a quadratic/cubic Savitzky-Golay function (15

points) and peak areas, migration times and *SNR* were transferred to Excel (Microsoft Office, Redmond, WA, USA) for calculation of relative integrated peak area (RPA), relative migration time (RMT), LOD and LOQ. Least-squares linear regression analysis for external calibration curves and figures of merit calculations were performed using Excel. MedCalc version 12.5.0 (MedCalc Software, Ostend, Belgium) was used for generation of boxplots and control charts, and all extracted ion electropherograms were depicted using Igor Pro 5.0 software (Wavemetric Inc., Lake Oswego, OR, USA).

3.4 Results and Discussion

3.4.1 Method optimization for PFAS analysis by MSI-NACE-MS/MS

Optimization of separation, ionization and fragmentation conditions for PFASs was first explored when using MSI-NACE-MS/MS coupled to a coaxial sheath liquid interface (**Figure 3.1A**), which was recently validated for rapid quantification of serum fatty acids and accurate assessment of dietary fat intake.[43,44] In this case, multiplexed separation was performed using a serial hydrodynamic injection program involving the introduction of alternating sample and BGE segments within a bare fused-silica capillary to increase sample throughput without complicated column switching or hardware modifications.[49-51] However, perfluorinated anionic surfactants possess strong adsorption properties that contribute to band broadening and/or poor resolution in CE when using aqueous BGE conditions with inadequate organic modifier content.[42] As a result, a nonaqueous BGE system was used for separation of PFASs in this work, which was composed of 70% *vol* acetonitrile, 15% *vol* methanol, and 5% *vol* isopropanol; however, residual water (10% *vol*) was still needed for solubilization of ammonium acetate as the electrolyte (apparent pH of

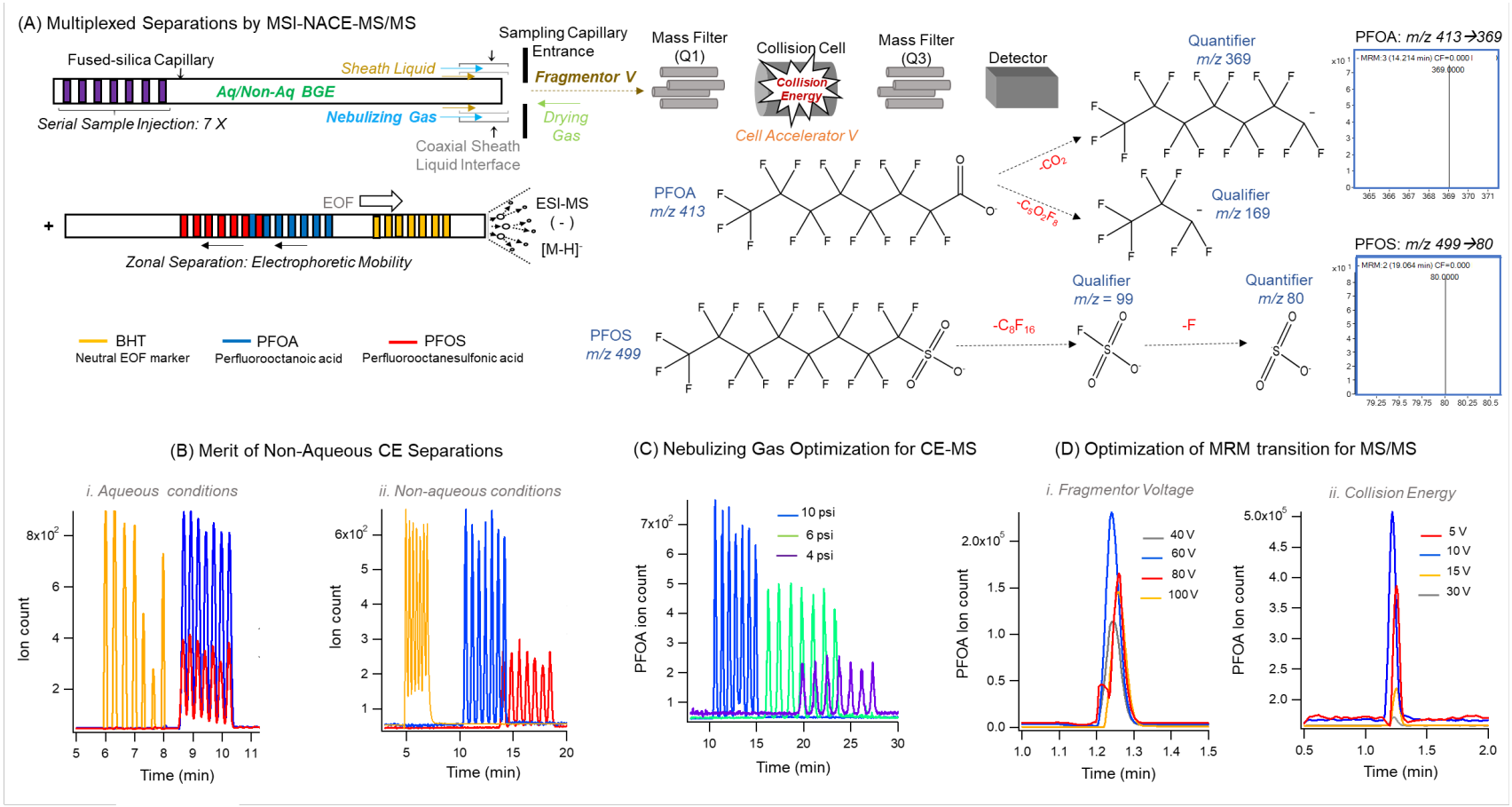


Figure 3.1. (A) Multiplexed separations of surface-active PFOA and PFOS when using MSI-NACE-MS based on serial injection of seven discrete samples followed by their zonal electrophoretic separation with MRM-scan data acquisition of quantifier (*m/z* 413→369 for PFOA *m/z* 499→80 for PFOS) and qualifier ions (*m/z* 413→169 for PFOA *m/z* 499→99 for PFOS) under negative ion mode detection. (B) Comparison of PFOA and PFOS separation resolution under aqueous (left) and non-aqueous (right) background electrolyte conditions. (C) Optimization of nebulizer gas pressure for the coaxial sheath liquid interface. (D) Optimization of MRM transition parameters including fragmentor voltage (V) and collision energy for maximal precursor and product ion response, respectively.

9.5) to generate a stable current during electrophoretic separation. **Figure 3.1B** demonstrates a clear benefit of the non-aqueous BGE system as required for resolution of PFOA from PFOS unlike equivalent aqueous buffer conditions (pH 9.5) without organic modifiers that resulted in their co-migration after the EOF, where BHT serves as a neutral marker. In fact, baseline resolution of all seven independently introduced sample plugs in the same run was achieved within 20 min (~ 3 min/sample) when using the non-aqueous BGE system with PFOS migrating with a larger apparent negative electrophoretic mobility (*i.e.*, longer migration time) as compared to PFOA. Improved solubilization and specific solvent-solute interactions have long been attributed to the unique selectivity in NACE[52] with PFOS being bulkier, less volatile and more hydrophobic than PFOA.[23] We recently demonstrated that a homologous series of fatty acids are accurately modeled in NACE based on their characteristic mobilities reflecting differences in carbon chain length and degree of unsaturation supporting their unambiguous identification that is complementary to high resolution MS.[43]

Other experimental variables also impacted PFAS separation performance, such as drying gas flow rate, nebulizer gas pressure, as well as a hydrodynamic pressure applied during electromigration (*i.e.*, pressure-assisted NACE). Nebulizer gas pressure is required in the CE-MS interface to stabilize spray formation, and higher nebulizer gas velocities contributed to faster migration times due to a greater siphoning effect,[53] which in turn gave rise to sharper peaks and a 3-fold higher *SNR* as depicted for PFOA in **Figure 3.1C**. Nevertheless, the nebulizer gas was required to be shut off during serial sample introduction in MSI-NACE-MS/MS when using low viscosity nonaqueous BGE solutions that can give rise to suctioning of air within the capillary inlet resulting in a current drop upon voltage application.[43] In contrast, increasing drying gas flow rate from 4 L/min to 8 L/min resulted in longer migration times exceeding 20 min for both PFOA and

Table 3.1. Optimized parameters for MSI-NACE-MS/MS determination of PFOS and PFOA with $^{13}C_8$-PFOS used as recovery standard in serum extracts.

Compound	MRM transition	Fragmentor Voltage (V)	Collision energy (V)	Cell accelerator voltage (V)	Dwell time (msec)
PFOS Quantifier	499→80	60	60	5	100
PFOS Qualifier	499→99	60	60	5	200
PFOA Quantifier	413→369	60	10	7	100
PFOA Qualifier	413→169	60	10	7	100
$^{13}C_8$-PFOS Quantifier	507→80	60	60	5	100
$^{13}C_8$-PFOS Qualifier	507→99	60	60	5	200

PFOS with greater band dispersion, whereas reducing it to 2L/min shortened the separation window with insufficient resolution from the EOF. Next, a pressure gradient of 2 mbar/min starting at 8 min was also introduced to further sharpen peaks for improved resolution of PFASs while also reducing total analysis times. Lastly, selective yet sensitive analysis of PFASs was achieved by MSI-NACE-MS/MS when performing collisional-induced dissociation of precursor ions to generate characteristic product ions, including quantifier and qualifier ions that were detected via MRM as shown in **Figure 3.1D**. **Table 3.1** summarizes optimal voltage settings for MRM transitions for quantifier and qualifier ions associated with PFOS, PFOA and $^{13}C_8$-PFOS (as internal/recovery standard), including fragmentor voltage, cell accelerator voltage, and collision energy. Noteworthy, a conventional electrospray ion source was optimal for MSI-NACE-MS/MS since it allowed for uninterrupted analyses without incidental capillary failures and/or current instabilities upon voltage application. In contrast, a dual Jet Stream electrospray source that uses a high flow of heated sheath gas together with a nozzle voltage increases the susceptibility to

corona discharge under negative ion mode conditions resulting in frequent capillary fractures.[43,45] Also, the Jet Stream electrospray source precludes the ability of the sheath gas to be turned off completely during sample introduction that is critical for stable runs when using non-aqueous BGE systems. As a result, systematic method optimization of several experimental variables in MSI-NACE-MS/MS was critical to enable high efficiency separation of PFASs with excellent selectivity, sensitivity, and robustness.

3.4.2 Method validation for PFAS determination from serum extracts

Validation of the optimized MSI-NACE-MS/MS method was next performed for reliable quantification of PFOA and PFOS from serum extracts based on several figures of merit as summarized in **Table 3.2**. External calibration curves were generated based on triplicate analysis of seven calibrant solutions that were acquired in a single run from 25 to 1000 nM, each containing 1.0 μM $^{13}C_8$-PFOS as a single stable-isotope internal/recovery standard for peak area normalization to correct for variations in injection volume between samples as shown in **Figure 3.2A**. As expected, there is co-migration of $^{13}C_8$-PFOS and PFOS in all seven independent introduced sample plugs in MSI-NACE-MS/MS, which facilitates alignment and unambiguous confirmation of sample positions in cases when PFOS is not detected. Good linearity was found for PFOA and PFOS over a 40-fold linear dynamic range with correlation coefficients (R^2) of 0.997 and 0.995, respectively as highlighted in **Figure 3.2A**. The LOD (*SNR* ~ 3) and LOQ (*SNR* ~ 10) for PFOA were 20 nM and 102 nM, and for PFOS were 25 nM and 117 nM, respectively. As expected, these detection limits are considerably lower than previous CE methods coupled to UV detection (LOD ~ 3300 nM),[39] as well as NACE methods with on-line sample preconcentration using large-volume sample stacking or field-amplified sample injections (LOD ~ 30-280 nM).41 Yet, LC-MS/MS methods with pre-column SPE reported significantly lower LODs for PFASs ranging from 0.02 to

Table 3.2. Method validation and figures of merit for rapid screening of PFOS and PFOA from serum extracts using MSI-NACE-MS/MS

PFAS	Molecular Formula	RMT[a]	LOD[b] (nM)	LOQ[b] (nM)	Range[c] (nM)	Linearity[c] (R^2)	Mean % Recovery[d]	Intraday RMT[e] (CV)	Intraday RPA[e] (CV)	Interday RMT[e] (CV)	Interday RPA[e] (CV)
PFOA	$C_8HF_{15}O_2$	0.76	20	103	25-1000	0.997	97 ± 12	1.11	4.58	1.43	9.56
PFOS	$C_8HF_{17}O_3S$	1.00	25	117	25-1000	0.995	109 ± 9	1.55	4.75	1.78	8.59

[a] Relative migration time (RMT) calculated by normalization to $^{13}C_8$-PFOS, which was also used for PFAS quantification based on relative peak area.
[b] LOD and LOQ were estimated at the lowest PFAS calibrant concentration that generated a SNR ≈ 3 and 10, respectively.
[c] Calibration curves from triplicate analysis of 7 calibrant solutions for PFASs over a 40-fold concentration range normalized to $^{13}C_8$-PFOS.
[d] Average % recovery for PFASs was calculated based on the percentage difference between spiked and original concentrations (in normal human serum) divided by the spiked concentration with triplicate measurements performed (n=3) at 3 concentration levels (100, 500 and 1000 nM).
[e] Precision was assessed by analyzing 6 replicate injections of 1.0 µM PFAS calibrants along with a blank at the beginning, middle and end of day, over 3 consecutive days analyzed using a single capillary. Method reproducibility determined by intraday (n = 28) and interday (n = 84) precision for both quantifier ions for PFOA and PFOS when using MSI-NACE-MS.

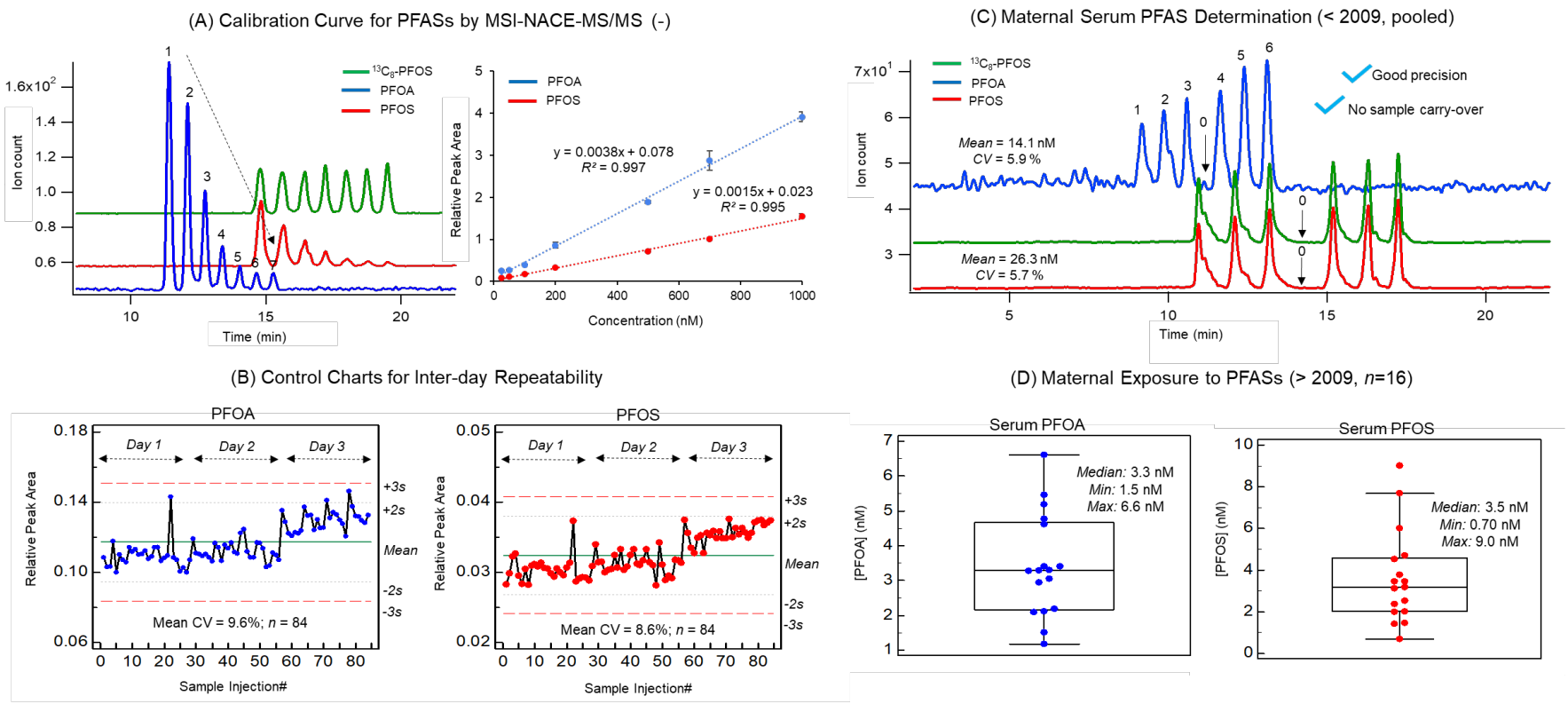


Figure 3.2. (A) Seven-point calibration curve for PFAS quantification by MSI-NACE-MS/MS over a 40-fold concentration range (25 to 1000 nM) based on their relative peak areas (RPA) normalized to $^{13}C_8$-PFOS (1000 nM) as a single stable-isotope internal standard. (B) Control charts depicting long-term method precision over 3 days of analysis for a calibrant mixture (n=84) of PFOA and PFOS when using MSI-NACE-MS/MS with mean CV < 10% and no outlier data exceeding action limits (± 3s). (C) Method application on pooled maternal serum samples collected before 2009 from FAMILY birth cohort highlighting higher PFAS exposures with precise measurements of six replicate serum extracts with no background signal in blank (0) extract. (D) Boxplots showing the distribution of PFOA and PFOS serum concentrations in a subset of maternal serum samples (n=16) from START birth cohort collected after 2009.

0.20 nM due to the much larger sample volumes injected on-column (~ 100-fold) as compared to CE (~ 10 nL).[34,38] For these reasons, off-line sample enrichment was critical to enable the quantification of low nanomolar levels of PFOSs from serum extracts when using MSI-CE-MS/MS. In this case, a simple MTBE extraction protocol using a 200 μL serum aliquot was able to preconcentrate PFASs by a factor of 40-fold when reconstituted into 5.0 μL, which effectively reduced LOD and LOQ to about 0.5 nM and 2.5 nM, respectively. This compares well to a recent CE-MS method using a sheathless interface for analysis of trace levels of environmental contaminants with LOD of 0.1 and 0.8 nM reported for PFOA and PFOS, respectively.[42] Importantly, spike-recovery studies confirmed good method accuracy in standard human serum, which was performed in triplicate for PFOA and PFOS at three different concentration levels (100, 500 and 1000 nM). Overall, measured recoveries for PFASs in serum ranged from 86–123% with a mean recovery of 103% as summarized in **Table 3.2**. These results support the valid use of a single MTBE extraction fraction to ensure accurate quantification of PFASs that are predominately bound to human serum albumin in circulation, where PFOS has greater binding affinity than PFOA.[54] Lastly, method precision was evaluated by assessment of both intraday (n = 28) and interday (n = 84) reproducibility for analysis of a standard mixture of PFASs by MSI-NACE-MS/MS. Overall, the mean CV for the quantifier ions was 4.7 % and 9.1 % for intra-day and inter-day precision, respectively when responses were normalized to $^{13}C_8$-PFOS, whereas the mean CV for RMTs of PFOS and PFOA were < 2.0% for 84 repeated injections performed over three consecutive days (**Table 3.2**). **Figure 3.2B** depicts control charts for analysis of PFOA and PFOS highlighting acceptable long-term technical precision over 3 days of intermittent analysis (CV < 10%) with few outliers exceeding warning (± 2s), and no data exceeding action (± 3 s) limits. As

a result, high throughput screening of serum PFASs is feasible with acceptable accuracy, linearity and precision when using MSI-NACE-MS/MS following a simple ether extraction procedure.

3.4.3 Assessment of PFAS exposures in maternal serum

Following method optimization and validation, MSI-NACE-MS/MS was next applied to evaluate PFOS exposures in a sub-set of second trimester pregnant women (n = 16) from START. Additionally, a pooled maternal serum sample from FAMILY was used as a QC sample, as well as a reference comparator to evaluate potential changes in mean PFAS exposures prior to regulations to restrict PFAS production in 2009. **Figure 3.2C** shows representative extracted ion electropherograms for PFOA and PFOS detected from replicate serial injection of serum extracts for pooled QC (n = 6) together with a blank extract (n = 1) when using MSI-NACE-MS/MS, which highlights good technical precision (CV < 6%) with no sample carryover effects/background contamination. Overall, pregnant women from FAMILY (serum collected prior to 2009) were found to have a higher average exposure of PFASs as compared to START (serum collected after 2009) with mean serum concentrations of PFOS and PFOA of 26 nM and 14 nM, respectively. In contrast, PFAS exposures assessed from individual pregnant women in START were variable yet much lower in concentration notably for PFOS, with median concentrations of 3.5 nM (range of 0.70 to 9.0 nM) and 3.3 nM (range of 1.5 to 6.6 nM) for PFOS and PFOA, respectively as depicted in box-whisker plots in **Figure 3.2D**. These concentrations are consistent with exposures measured in a large cohort of Swedish women early in pregnancy with median concentrations of 10.8 and 3.9 nM for PFOS and PFOA respectively, which also reported a higher risk for preeclampsia in women with greater PFAS exposures after adjustment for confounders.[55] Also, the lower PFAS exposures in START as compared to FAMILY is in agreement with longitudinal studies in the Fernald Community cohort that demonstrated decreasing serum concentration trajectories

occurring from 2000-2008, especially for PFOS, yet with a corresponding increase in certain unregulated PFAS analogs, such as perflourononanoic acid.[56] Future studies will investigate temporal changes in PFAS exposures in larger numbers of ethnically diverse birth cohorts, including assessing their impacts on childhood health outcomes. Also, the application of nanospray/sheathless interfaces in CE-MS[42,57] is needed to further lower detection limits under negative ion mode conditions for measurement of legacy PFASs and an expanding array of PFAS replacements relevant to contemporary exposures.

3.5 Conclusion

PFASs pose an on-going challenge to environmental epidemiology due to their widespread pervasiveness and bioaccumulation that ensures a continued exposure risk given the rise of unregulated PFAS substitutes. For the first time, we demonstrate a rapid method for PFAS determination from serum when using MSI-NACE-MS/MS that is optimal for large-scale biomonitoring applications. This approach allows for reliable PFOS and PFOA quantification using a simple ether extraction protocol following rigorous method optimization and validation to ensure adequate robustness, accuracy, and precision. Our work overcomes previous technical challenges of CE/CE-MS methods using aqueous background electrolyte systems related to poor solubilization, deleterious band broadening and/or inadequate sensitivity for PFAS determination with low nanomolar detection limits. Multiplexed separations comprising seven or more discrete samples serially injected within a single run offers higher throughput (< 3 min/sample) than conventional chromatographic separations with greater data fidelity since a reference sample and/or blank extract can be incorporated in each run for improved quality control and batch correction. Preliminary studies confirm that maternal exposures to PFOA and PFOS have largely decreased since 2009, with considerable between-subject variability in PFAS serum

concentrations reflecting different exposure mechanisms in pregnant women. We anticipate that this method may allow for comprehensive PFAS surveillance when using full data acquisition with high resolution MS for new advances in maternal health and the developmental origins of health and disease.

3.6 Acknowledgements

P.B.M. acknowledges funding from the Natural Sciences and Engineering Research Council of Canada, and Genome Canada. S.M.A. acknowledges funding from the Egyptian Ministry of Higher Education. Further thanks are extended to Dr. Sonia Anand, Dr. Koon Teo, Dr. Stephanie Atkinson and other coinvestigators from START and FAMILY cohorts for access to residual maternal serum used in the method application. We also acknowledge Marcus Kim and John Sausen from Agilent Technologies Inc. for helpful discussions, including loan of the triple quadrupole mass spectrometer.

Conflict of Interest

The authors declare no conflict of interest.

Data Availability

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

3.7 References

1. Houde M, Martin JW, Letcher RJ, Solomon KR, Muir DCG. Biological monitoring of polyfluoroalkyl substances: a review. *Environ Sci Technol*. 2006;40:3463-3473.

2. Wang Y, Shi Y, Vestergren R, Zhou Z, Liang Y, Cai Y. Using hair, nail and urine samples for human exposure assessment of legacy and emerging per- and polyfluoroalkyl substances. *Sci Total Environ*. 2018;636:383-391.

3. Lindstrom AB, Strynar MJ, Libelo EL. Polyfluorinated compounds: past, present, and future. *Environ Sci Technol*. 2011;45:7954-7961.

4. Wang Y, Zhang L, Teng Y, et al. Association of serum levels of perfluoroalkyl substances with gestational diabetes mellitus and postpartum blood glucose. *J Environ Sci*. 2018;69:5-11.

5. Dobraca D, Israel L, McNeel S, et al. Biomonitoring in California firefighters: metals and perfluorinated chemicals. *J Occup Environ Med*. 2015;57:88-97.

6. Trudel D, Horowitz L, Wormuth M, Scheringer M, Cousins IT, Hungerbühler K. Estimating consumer exposure to PFOS and PFOA. *Risk Anal*. 2008;28:251-269.

7. Trier X, Granby K, Christensen JH. Polyfluorinated surfactants (PFS) in paper and board coatings for food packaging. *Environ Sci Pollut Res*. 2011;18:1108-1120.

8. Domazet SL, Grøntved A, Timmermann AG, Nielsen F, Jensen TK. Longitudinal associations of exposure to perfluoroalkylated substances in childhood and adolescence and indicators of adiposity and glucose metabolism 6 and 12 years later: the European youth heart study. *Diabetes Care*. 2016;39:1745-1751.

9. Jian J-M, Chen D, Han F-J, et al. A short review on human exposure to and tissue distribution of per- and polyfluoroalkyl substances (PFASs). *Sci Total Environ*. 2018;636:1058-1069.

10. Liu Y, Ruan T, Lin Y, et al. Chlorinated polyfluoroalkyl ether sulfonic acids in marine organisms from Bohai Sea, China: occurrence, temporal variations, and trophic transfer behavior. *Environ Sci Technol*. 2017;51:4407-4414.

11. Kato K, Wong L-Y, Jia LT, Kuklenyik Z, Calafat AM. Trends in exposure to polyfluoroalkyl chemicals in the U.S. population: 1999−2008 †. *Environ Sci Technol*. 2011;45:8037-8045.

12. Zhao Z, Xie Z, Möller A, et al. Distribution and long-range transport of polyfluoroalkyl substances in the Arctic, Atlantic Ocean and Antarctic coast. *Environ Pollut*. 2012;170:71-77.

13. Yeung LWY, Dassuncao C, Mabury S, Sunderland EM, Zhang X, Lohmann R. Vertical profiles, sources, and transport of PFASs in the Arctic Ocean. *Environ Sci Technol*. 2017;51:6735-6744.

14. Wang Y, Yu N, Zhu X, et al. Suspect and nontarget screening of per- and polyfluoroalkyl substances in wastewater from a fluorochemical manufacturing park. *Environ Sci Technol*. 2018;52:11007-11016.

15. Olsen GW, Burris JM, Ehresman DJ, et al. Half-life of serum elimination of perfluorooctanesulfonate, perfluorohexanesulfonate, and perfluorooctanoate in retired fluorochemical production workers. *Environ Health Perspect*. 2007;115:1298-1305.

16. Haug LS, Thomsen C, Becher G. A sensitive method for determination of a broad range of perfluorinated compounds in serum suitable for large-scale human biomonitoring. *J Chromatogr A*. 2009;1216:385-393.

17. Kärrman A, Ericson I, van Bavel B, et al. Exposure of perfluorinated chemicals through lactation: levels of matched human milk and serum and a temporal trend, 1996–2004, in Sweden. *Environ Health Perspect*. 2007;115:226-230.

18. Worley RR, Moore SM, Tierney BC, et al. Per- and polyfluoroalkyl substances in human serum and urine samples from a residentially exposed community. *Environ Int*. 2017;106:135-143.

19. Kar S, Sepúlveda MS, Roy K, Leszczynski J. Endocrine-disrupting activity of per- and polyfluoroalkyl substances: Exploring combined approaches of ligand and structure based modeling. *Chemosphere*. 2017;184:514-523.

20. Di Nisio A, Sabovic I, Valente U, et al. Endocrine disruption of androgenic activity by perfluoroalkyl substances: clinical and experimental evidence. *J Clin Endocrinol Metab*. 2019;104:1259-1271.

21. Di Nisio A, Rocca MS, Sabovic I, et al. Perfluorooctanoic acid alters progesterone activity in human endometrial cells and induces reproductive alterations in young women. *Chemosphere*. 2020;242:125208.

22. Kotthoff M, Bücking M. Four chemical trends will shape the next decade's directions in perfluoroalkyl and polyfluoroalkyl substances research. *Front Chem*. 2018;6:103.

23. Viberg H, Eriksson P. Perfluorooctane sulfonate (PFOS) and perfluorooctanoic acid (PFOA). In: *Reproductive and Developmental Toxicology*. Elsevier; 2011:623-635.

24. Koponen J, Winkens K, Airaksinen R, et al. Longitudinal trends of per- and polyfluoroalkyl substances in children's serum. *Environ Int*. 2018;121:591-599.

25. Llorca M, Farré M, Picó Y, Teijón ML, Álvarez JG, Barceló D. Infant exposure of perfluorinated compounds: levels in breast milk and commercial baby food. *Environ Int*. 2010;36:584-592.

26. Winkens K, Giovanoulis G, Koponen J, et al. Perfluoroalkyl acids and their precursors in floor dust of children's bedrooms – Implications for indoor exposure. *Environ Int*. 2018;119:493-502.

27. Uppal JS, Zheng Q, Le XC. Maternal exposure to specific perfluoroalkyl substances is associated with increasing blood glucose in pregnant women. *J Environ Sci*. 2018;69:1-2.

28. Sunderland EM, Hu XC, Dassuncao C, Tokranov AK, Wagner CC, Allen JG. A review of the pathways of human exposure to poly- and perfluoroalkyl substances (PFASs) and present understanding of health effects. *J Expo Sci Environ Epidemiol*. 2019;29:131-147.

29. Karlsen M, Grandjean P, Weihe P, Steuerwald U, Oulhote Y, Valvi D. Early-life exposures to persistent organic pollutants in relation to overweight in preschool children. *Reprod Toxicol*. 2017;68:145-153.

30. Wang Y, Han W, Wang C, et al. Efficiency of maternal-fetal transfer of perfluoroalkyl and polyfluoroalkyl substances. *Environ Sci Pollut Res*. 2019;26:2691-2698.

31. Nadal A, Quesada I, Tudurí E, Nogueiras R, Alonso-Magdalena P. Endocrine-disrupting chemicals and the regulation of energy balance. *Nat Rev Endocrinol*. 2017;13:536-546.

32. Heindel JJ, Skalla LA, Joubert BR, Dilworth CH, Gray KA. Review of developmental origins of health and disease publications in environmental epidemiology. *Reprod Toxicol*. 2017;68:34-48.

33. Barceló D, Ruan T. Challenges and perspectives on the analysis of traditional perfluoroalkyl substances and emerging alternatives. *TrAC Trends Anal Chem*. 2019;121:115605.

34. Nakayama SF, Isobe T, Iwai-Shimada M, et al. Poly- and perfluoroalkyl substances in maternal serum: method development and application in pilot study of the Japan environment and children's study. *J Chromatogr A*. 2020.

35. Trojanowicz M, Koc M. Recent developments in methods for analysis of perfluorinated persistent pollutants. *Microchim Acta*. 2013;180:957-971.

36. Onghena M, Moliner-Martinez Y, Picó Y, Campíns-Falcó P, Barceló D. Analysis of 18 perfluorinated compounds in river waters: comparison of high performance liquid chromatography–tandem mass spectrometry, ultra-high-performance liquid chromatography–tandem mass spectrometry and capillary liquid chromatography–mass spectrometry. *J Chromatogr A*. 2012;1244:88-97.

37. Keller JM, Calafat AM, Kato K, et al. Determination of perfluorinated alkyl acid concentrations in human serum and milk standard reference materials. *Anal Bioanal Chem*. 2010;397:439-451.

38. Poothong S, Lundanes E, Thomsen C, Haug LS. High throughput online solid phase extraction-ultra high performance liquid chromatography-tandem mass spectrometry method for polyfluoroalkyl phosphate esters, perfluoroalkyl phosphonates, and other perfluoroalkyl substances in human serum, plasma, and whole blood. *Anal Chim Acta*. 2017;957:10-19.

39. Wójcik L, Szostek B, Maruszak W, Trojanowicz M. Separation of perfluorocarboxylic acids using capillary electrophoresis with UV detection. *Electrophoresis*. 2005;26:1080-1088.

40. Wójcik L, Korczak K, Szostek B, Trojanowicz M. Separation and determination of perfluorinated carboxylic acids using capillary zone electrophoresis with indirect photometric detection. *J Chromatogr A*. 2006;1128:290-297.

41. Knob R, Maier V, Petr J, Ranc V, Ševčík J. On-line preconcentration of perfluorooctanoic acid and perfluorooctanesulfonic acid by nonaqueous capillary electrophoresis. *Electrophoresis*. 2012;33:2159-2166.

42. Höcker O, Bader T, Schmidt TC, Schulz W, Neusüß C. Enrichment-free analysis of anionic micropollutants in the sub-ppb range in drinking water by capillary electrophoresis-high resolution mass spectrometry. *Anal Bioanal Chem*. 2020; doi:10.1007/s00216-020-02525-8

43. Azab S, Ly R, Britz-McKibbin P. Robust method for high-throughput screening of fatty acids by multisegment injection-nonaqueous capillary electrophoresis–mass spectrometry with stringent quality control. *Anal Chem*. 2019;91:2329-2336.

44. Azab SM, de Souza RJ, Teo KK, Anand SS, Williams NC, Holzschuher J, McGlory C, Phillips SM; Britz-McKibbin P. Serum non-esterified fatty acids have utility as dietary biomarkers of fat intake from fish, fish oil and dairy in women. *J. Lipid Res*. 2020, *61*, 933–944.

45. Yamamoto M, Ly R, Gill B, Zhu Y, Moran-Mirabal J, Britz-McKibbin P. Robust and high-throughput method for anionic metabolite profiling: preventing polyimide aminolysis and capillary breakages under alkaline conditions in capillary electrophoresis-mass spectrometry. *Anal Chem*. 2016;88:10710-10719.

46. Anand SS, Gupta M, Teo KK, et al. Causes and consequences of gestational diabetes in South Asians living in Canada: results from a prospective cohort study. *CMAJ Open*. 2017;5:E604-E611.

47. Morrison KM, Atkinson SA, Yusuf S, et al. The Family Atherosclerosis Monitoring In earLY life (FAMILY) study. *Am Heart J*. 2009;158:533-539.

48. Matyash V, Liebisch G, Kurzchalia TV, Shevchenko A, Schwudke D. Lipid extraction by methyl- *tert* -butyl ether for high-throughput lipidomics. *J Lipid Res*. 2008;49:1137-1146.

49. Kuehnbaum NL, Kormendi A, Britz-McKibbin P. Multisegment injection-capillary electrophoresis-mass spectrometry: a high-throughput platform for metabolomics with high data fidelity. *Anal Chem*. 2013;85:10664-10669.

50. DiBattista A, Rampersaud D, Lee H, Kim M, Britz-McKibbin P. High throughput screening method for systematic surveillance of drugs of abuse by multisegment injection–capillary electrophoresis–mass spectrometry. *Anal Chem*. 2017;89:11853-11861.

51. Shanmuganathan M, Macklai S, Barrenas Cárdenas C, et al. High-throughput and comprehensive drug surveillance using multisegment injection-capillary electrophoresis mass spectrometry. *J Vis Exp*. 2019;146:58986.

52. Kenndler E. A critical overview of non-aqueous capillary electrophoresis. Part I: Mobility and separation selectivity. *J Chromatogr A*. 2014;1335:16-30.

53. Huikko K, Kotiaho T, Kostiainen R. Effects of nebulizing and drying gas flow on capillary electrophoresis/mass spectrometry. *Rapid Commun Mass Spectrom*. 2002;16:1562-1568.

54. Chi Q, Li Z, Huang J, Ma J, Wang X. Interactions of perfluorooctanoic acid and perfluorooctanesulfonic acid with serum albumins by native mass spectrometry, fluorescence and molecular docking. *Chemosphere*. 2018;198:442-449.

55. Wikström S, Lindh CH, Shu H, Bornehag C-G. Early pregnancy serum levels of perfluoroalkyl substances and risk of preeclampsia in Swedish women. *Sci Rep*. 2019;9:9179.

56. Blake BE, Pinney SM, Hines EP, Fenton SE, Ferguson KK. Associations between longitudinal serum perfluoroalkyl substance (PFAS) levels and measures of thyroid hormone, kidney function, and body mass index in the Fernald Community Cohort. *Environ Pollut*. 2018;242:894-904.

57. Lin L, Liu X, Zhang F, Chi L, Amster IJ, Leach FE 3rd, Xia Q, Linhardt RJ. Analysis of heparin oligosaccharides by capillary electrophoresis-negative-ion electrospray ionization mass spectrometry. *Anal Bioanal Chem.* 2017;409:411-420.

Chapter IV:

Serum Non-Esterified Fatty Acids have Utility as Dietary Biomarkers of Fat Intake from Fish, Fish Oil and Dairy in Women

Thesis chapter is derived from a published peer-reviewed article:

S.M. Azab, R.J. de Souza, K.K. Teo, S.S. Anand, N.C. Williams, J. Holzschuher, C. McGlory, S.M. Philips and P. Britz-McKibbin. Serum non-esterified fatty acids have utility as dietary biomarkers of fat intake from fish, fish oil and dairy in women. *J. Lipid Res*. **2020**, 61: 933–944.

P.B.M., S.M.A., and S.S.A. designed the research; S.M.A conducted the analysis and data processing including NEFA and total FA analysis; C.M.G and S.M.P. devised the original design and provided the samples for the intervention study; S.M.A. and J.H. processed the samples for the intervention study; R.J.dS., and N.C.W. devised the diet index and nutrient calculations for the observational study. K.K.T. is the lead investigator of the FAMILY birth cohort. S.M.A., R.J.dS. and P.B.M. analyzed the data; S.M.A. drafted the manuscript that was edited and guided by R.J.dS. and P.B.M. All authors read and approved the final manuscript.

Chapter IV: Serum Non-Esterified Fatty Acids have Utility as Dietary Biomarkers of Fat Intake from Fish, Fish Oil and Dairy in Women

"All we know is what we're told?"

4.1 Abstract

Nutritional studies rely on various biological specimens for fatty acid (FA) determination, yet there remains a knowledge gap regarding the relationship of serum non-esterified FA (NEFA) to other circulating lipid pools. We performed two studies to assess the utility of specific serum NEFA as biomarkers of dietary fat intake in women when using a high throughput method (< 4 min/sample) based on multisegment injection-non-aqueous-capillary electrophoresis–mass spectrometry (MSI-NACE-MS). We first aimed to identify circulating NEFA correlated to habitual intake of specific foods among pregnant women with contrasting dietary patterns (n = 50). Acute changes in serum NEFA trajectories were also measured in young women (n = 18) following high-dose (5 g/day) fish oil supplementation or isoenergetic sunflower oil placebo over 56 days. In the cross-sectional study, serum omega-3 (ω-3) FA correlated with self-reported total ω-3 daily intake, notably eicosapentaenoic acid (EPA) as its NEFA (r = 0.46; p = 0.001), whereas pentadecanoic acid was strongly associated with full-fat dairy intake (r = 0.43; p = 0.002); these outcomes were consistent with FA measurements from total serum hydrolysates. In the intervention study, a 2.5-fold increase in serum ω-3 NEFA from baseline was achieved within 28 days following fish oil supplementation that was most pronounced for EPA (p = 0.0004). Circulating EPA was also correlated to its erythrocyte phospholipid fraction (r = 0.66; $p = 4.6 \times 10^{-10}$) unlike docosahexaenoic acid, and was more sensitive to detect dietary non-adherence. MSI-NACE-MS offers a rapid approach for serum NEFA quantification as required for accurate biomonitoring of dietary fat intake in support of maternal health.

4.2 Introduction

Accurate assessment of dietary fat intake remains a methodological challenge reflecting decades of conflicting evidence regarding the benefits of a low-fat diet for public health.[1] Validated semi-quantitative food frequency questionnaires (FFQ) are widely used dietary assessment tools in large-scale observational studies as they can reliably differentiate habitual dietary patterns, as well as estimate micro- and macronutrient intake in a cost-effective manner.[2] However, FFQ are prone to recall bias, errors in estimation of true portion sizes, as well as selective reporting;[3] this problem is exacerbated when assessing habitual fat intake due to the large variation of fatty acid (FA) species in the diet, and the tendency for underreporting fat consumption.[4] Errors associated with participant self-reporting have been recognized as one of the greatest obstacles in nutritional epidemiology, limiting our ability to capture food exposures in contemporary societies.[5] Comprehensive metabolite profiling (*i.e.,* metabolomics) offers a strategy to objectively measure complex dietary patterns, including the discovery of new biomarkers of recent food intake.[6] An optimal dietary biomarker is readily measurable in a minimally invasive human biofluid (urine or blood), specific to a single food group (selective), responsive to changes in the amount of food consumed (dose-response) over a desired time frame (time-response), and is ideally not generated *in vivo* nor extensively biotransformed (exogenous).[7] For example, proline betaine is a reliable dietary biomarker of recent citrus intake (< 24 h) in plasma and urine samples that has been validated in several independent observational and intervention studies.[3,8-10] Such biomarkers generally do not exist for most FA since they are synthesized *de novo* from carbohydrates and other FA precursors; however, there are some exceptions, such as certain polyunsaturated FA (PUFA), odd-chain FA (OCFA), and *trans*-FA (TFA) since they are primarily derived from specific food sources and thus largely not synthesized *in-vivo*.[11]

There is an evolving consensus regarding optimal nutritional guidelines for dietary fat with greater emphasis placed on assessing the specific type of fat within complex dietary patterns as opposed to measuring total fat intake and single nutrients.[12] For example, many jurisdictions have now restricted industrial TFA exposures from processed foods due to their deleterious effects on cardiovascular health as compared to non-restricted populations.[13] Recent findings from the 21-country Prospective Urban and Rural Epidemiological (PURE) study reported that total fat is correlated with lower total mortality, and increased saturated fat or total dairy consumption with lower risk for cardiovascular events;[14, 15] however, these studies have relied on estimating intakes of FA from FFQ, and gathered health information from different countries with the potential for residual confounding. As a result, high throughput methods for objective measurements of circulating FA concentrations are thus needed to provide a more standardized yet accurate approach for assessment of habitual fat consumption. This is important given conflicting data regarding the putative health benefits of dietary intake of essential omega-3 (ω-3) PUFA,[16] such as clinical trials involving supplementation of eicosapentaenoic acid (EPA) and docosahexaenoic acid (DHA) from fish oil (FO) or a prescription based EPA analog for primary or secondary prevention of cardiovascular disease events.[17,18] Various biospecimen types have been used for FA determination and each reflects a different time interval associated with dietary fat intake, including adipose tissue (1-2 years), erythrocyte membrane (2-3 months), serum phospholipids (PL) or cholesterylesters (CE) (past few days), and triglycerides (TG) (past few hours) fractions.[4,11,19] While adipose tissue may be useful for long-term assessment of dietary fat intake, such biopsies are invasive and cannot be routinely collected. Also, erythrocytes isolated from whole blood are prone to hemolysis during processing and long-term storage[20, 21] and they are not often available in most bio-bank repositories unlike serum or plasma. As of 2016, it was reported

that 90% of studies analyzing ω-3 PUFA used erythrocytes membrane PL, serum PL, or total (hydrolyzed) plasma lipids.[22] Alternatively, fasting serum non-esterified FA (NEFA) are a more accessible protein-bound lipid pool, which are released into circulation by the hydrolyzing action of lipases on TG from adipose tissue.[4] Nevertheless, there have been few reports to date examining the utility of NEFA as biomarkers of dietary intake in nutritional studies as compared to total FA from serum extracts, and other blood fractions or adipose tissue samples.[19, 23] For instance, gas chromatography (GC) methods allow for high efficiency separation of FA and their isomers,[24] but are limited by long analysis times and pre-column chemical derivatization procedures that contribute to bias due to hydrolysis of esterified FA from other lipid classes, impeding reliable serum NEFA determination.[25]

Herein, we performed two studies to assess the utility of serum NEFA as convenient dietary biomarkers of fat intake in women when using multisegment injection-non-aqueous-capillary electrophoresis–mass spectrometry (MSI-NACE-MS). This multiplexed separation method offers higher sample throughput (< 4 min/sample) and stringent quality control (QC) for direct analysis of NEFA from serum extracts without fractionation, hydrolysis and chemical derivatization unlike conventional GC methods.[26] For the first time, we assess a cross-section of pregnant women with contrasting diet quality patterns[27] when using MSI-NACE-MS to identify specific serum NEFA that serve as biomarkers of habitual intake of fish/seafood, full-fat dairy products, and fiber as compared to FA from total serum hydrolysates. Also, time-dependent changes in serum EPA and DHA in young women participating in a placebo-controlled, repeated-measures trial of high-dose FO supplementation[28] were also examined as their circulating NEFA relative to independent erythrocyte membrane PL measurements.

4.3 Experimental Section

4.3.1 Serum NEFA biomarkers of dietary fat intake in pregnant women from FAMILY

The Family Atherosclerosis Monitoring In earLY life (FAMILY) study is a prospective birth cohort study involving 839 predominantly white European pregnant women recruited from the greater Hamilton area between 2002 and 2009.[27] Fasting blood samples were collected in the second trimester, and serum was fractionated within 2 h from collection according to standard protocols and stored at -80 °C. Comprehensive clinical and dietary data from all participants were also collected. Ethical approval and informed consent from all study participants were obtained. In this study, we used purposive sampling, after exclusion of smokers and women with gestational diabetes mellitus, to generate a subset of women (n = 50) from 226 eligible participants, half of whom consumed a healthy diet and half of whom consumed a poor quality diet as assessed by a diet quality index (DQI) score[29] with a mean age of 32 years (range of 17 to 43 years) and pre-pregnancy BMI of 27 kg/m^2 (range of 18 to 50 kg/m^2). Briefly, a semi-quantitative FFQ developed for the Study of Health and Risk in Ethnic Groups (SHARE) study[30] was used to assess maternal dietary intake on one occasion after recruitment at about mid-pregnancy by asking the participants to answer the questions in the context of the usual eating habits during the last one year period. The complete FFQ was analyzed by using a database linked to the Canadian Nutrient File. Nutrient composition was calculated as previously described,[31] excluding records where the FFQ was 50% incomplete, or with implausible dietary intakes (< 500 or > 4500 kcal/d). Use of supplements was also assessed as part of the FFQ, which included a separate supplemental questionnaire for cod liver or halibut oil supplement usage. The DQI score used to classify the nutritional status of pregnant women into healthy and unhealthy eating categories was based on reported daily servings of foods from 36 harmonized food groups as described previously.[29] This aggregate score reflects

differences in the overall nutritional quality of foods consumed, which was based on the sum of the daily number of servings of healthy and nutrient-rich foods (*e.g.,* fermented dairy, fish and seafood, vegetables, fruits, whole grains, nuts and seeds) minus the daily number of servings of unhealthy and processed foods (*e.g.,* processed meats, refined grains, fries, snacks, sweets and sweet drinks). A positive DQI score signifies consumption of greater amounts of healthy than unhealthy foods and vice versa for a negative DQI score. In this study, 25 participants were selected from the top 10th percentile to form the "good diet" group, and the bottom 10th percentile of the cohort comprised the "poor diet" group as summarized in **Figure S4.1**.

4.3.2 High dose ω-3 PUFA supplementation in women and serum NEFA trajectories

Serum NEFA were analyzed using fasting serum samples from a clinical intervention trial investigating the effect of ω-3 PUFA supplementation from FO on attenuating skeletal muscle atrophy following leg immobilization.[28] The trial was registered at the U.S. National Library of Medicine (https://clinicaltrials.gov/) as NCT03059836. Briefly, a cohort of healthy young women with a mean age of 22 years (range of 19 to 31 years) and BMI of 24 kg/m^2 (range of 18 to 26 kg/m^2), were recruited from the greater Hamilton area to participate in a randomized, double-blinded, placebo-controlled intervention study. Participants received either the active treatment arm of a high-dose ω-3 PUFA from FO (3.0 g EPA, and 2.0 g DHA daily, n = 9), or a control based on an isoenergetic and volume equivalent sunflower oil daily (n = 9). Serum samples in the resting fasted state was collected from participants at baseline, and at 28, 42 and 56 days following initiation of the intervention, which were then stored at -80 °C. Participants taking FO supplements within 6 months of the study were excluded, 2 participants had missing residual serum samples from the original study, and 1 participant was reported to be non-compliant to leg immobilization intervention, and possibly dietary treatment.[28] Comprehensive analysis of serum NEFA in this

work was compared to FA concentrations reported for the erythrocyte membrane PL fraction measured using a validated protocol based on GC-FID. [28]

4.3.3 Validated method protocol for serum NEFA and total FA analysis by MSI-NACE-MS

NEFA from fasting serum samples collected from young women in both studies, along with standard human serum (Sigma S7023) used as a quality control (QC) specimen, were analyzed using a previously validated protocol for rapid serum NEFA determination based on MSI-NACE-MS that included an extensive inter-method comparison to GC-MS.[26] In all cases, frozen serum aliquots were thawed once slowly on ice prior to analysis. Briefly, protein-bound circulating FA were extracted from serum in acidified (3.7 % *vol* of 1.0 M HCl) methyl-*tert*-butyl ether (MTBE):methanol (5:1, *vol*) containing butylated hydroxytoluene (BHT, 0.01% *vol*) as an antioxidant additive during sample processing, and a deuterated analog of myristic acid, 14:0-d27 as a recovery standard. Following vigorous shaking, phase separation was then induced by addition of deionized water, and samples were then centrifuged to sediment protein at bottom of vial (at 3000 *g* at 4 °C for 30 min) followed by a biphasic water and ether (top) layer. A fixed volume (200 µL) was collected from the upper MTBE layer into a new vial then dried under a gentle stream of nitrogen gas at room temperature. Dried serum extracts were then stored at -80 °C and at time of analysis reconstituted in 25 µL of acetonitrile/isopropanol/water (70:20:10) with 10 mM ammonium acetate and 50 µM deuterated stearic acid, 18:0-d35 as an internal standard. For analysis of total (hydrolyzed) serum FA, acid-catalyzed hydrolysis was performed using 2.5 M sulfuric acid and 0.01% *vol* BHT in toluene followed by incubation at 80 °C for 1 h.[26] MTBE extraction was then carried out to recover total serum FA similar to protocol outlined for serum NEFA. An Agilent 6230 time-of-flight (TOF) mass spectrometer with a coaxial sheath liquid electrospray (ESI) ionization source equipped with an Agilent G7100A CE unit was used for all

experiments (Agilent Technologies Inc., Mississauga, ON, Canada). An Agilent 1260 Infinity isocratic pump was used to deliver an 80% *vol* methanol with 0.5% *vol* ammonium hydroxide at a flow rate of 10 μL/min using a CE-MS coaxial sheath liquid interface kit. Separations were performed on bare fused-silica capillaries with 50 μm internal diameter, 360 μm outer diameter and 95 cm total length (Polymicro Technologies Inc., AZ, USA). The applied voltage was set to 30 kV at 25 °C for CE separations together with an isocratic pressure of 20 mbar (2 kPa). The background electrolyte (BGE) was 35 mM ammonium acetate in 70% *vol* acetonitrile, 15% *vol* MeOH, 5% *vol* isopropanol with an apparent pH of 9.5 adjusted by addition of 12% *vol* of ammonium hydroxide. Serum extracts were injected hydrodynamically at 50 mbar (5 kPa) alternating between 5 s for each sample plug and 40 s for the BGE spacer plug for a total of 7 discrete samples analyzed within 30 min for a single run.[26,32] Repeat QC samples introduced in a randomized position for each MSI-NACE-MS run were analyzed for NEFA (n = 8) and total FA analysis (n = 8) for assessment of technical precision of the method. All FA extracts were analyzed directly by MSI-NACE-MS without chemical derivatization when using negative ion mode detection at 3500 V with full-scan data acquisition (m/z 50-1700), which allows for comprehensive screening of 24 FA consistently measured in human serum extracts.

4.3.4 Data processing, statistical analyses and data availability

MSI-NACE-MS data was analyzed using Agilent Mass Hunter Workstation software (Qualitative Analysis, version B.06.00, Agilent Technologies, 2012). Molecular features were extracted in profile mode within a 10 ppm mass window, and serum NEFA were annotated based on their characteristic accurate mass (m/z) corresponding to their intact deprotonated molecular ion $[M\text{-}H]^-$, and relative migration time (RMT) reflecting the electrophoretic mobility for anionic FA. Extracted ion electropherograms (EIE) were integrated after smoothing using a quadratic/cubic

Savitzky-Golay function (15 points), and integrated peak areas and apparent migration times were normalized to stearic acid-d35 for determination of relative peak area (RPA) and RMT. Least-squares linear regression analysis for external calibration curves and control charts were performed using Excel (Microsoft Office, Redmond, WA, USA). Principal component analysis (PCA) was used for data visualization (*i.e.,* data trends/outlier detection) when comparing the technical variance of QC samples as compared to the overall biological variance of serum NEFA concentrations between-subjects when using MetaboAnalyst 4.0.[33] For multivariate analysis for data visualization and univariate analysis comparing maternal diet sub-groups, data was normalized using a generalized *log* transformation and autoscaled with FDR correction applied for multiple hypothesis testing. Also, a QC based batch-correction algorithm was performed to correct for long-term signal drift in ESI-MS for robust serum NEFA determination using an algorithm available in the *R* Project for statistical computing.[34] Normality tests, Pearson and Spearman rank correlations, student's t-test, and nonparametric statistical analysis (Mann-Whitney U test) were performed using the Statistical Package for the Social Science (SPSS, version 18), whereas MedCalc version 12.5.0 (MedCalc Software, Ostend, Belgium) was used for generation of boxplots. To assess the validity of serum NEFA and total FA against reported dietary intakes of fish, full-fat dairy, and total fiber, as well as the diet quality index score, we used a Spearman's rank correlation coefficient (r) for non-transformed data. A correlation coefficient of $r = 0.1$ to 0.3 was considered a small effect, $r = 0.3$ to 0.5 a moderate effect, and $r > 0.5$ a large effect. In order to minimize error by accounting for EPA and DHA sources from both the diet and supplementation, a total daily ω-3 PUFA servings score (total ω-3) was devised, calculated from FFQ as the sum of EPA and DHA from self-reported dietary intake (g/day), as well as self-reported supplement use (g/day). Multiple regression models were constructed for *log*-transformed

measured serum NEFA concentrations to account for potential confounding variables (*i.e.*, BMI, cholesterol and HDL) that were different ($p < 0.05$) between the two dietary sub-groups. Circulating NEFA are reported in terms of absolute molar concentrations (μM) as a standardized way to enable data comparisons independent of analytical platform, sample workup protocols, and range of FA measured. A 2-way between and within mixed-model ANOVA (treatment; time) was used for assessing the impact of high-dose FO supplementation to alter circulating NEFA concentrations in healthy/non-pregnant women as compared to a placebo control. A Pearson correlation coefficient for non-transformed data was used to test the association between serum NEFA and FA from erythrocytes PL fraction. All processed serum FA data and de-identified clinical/dietary information for participants is available as an excel file in the supporting information section (NEFA-Serum-JLR.xlsx). This data includes original and batch-corrected RPA for NEFA measured by MSI-NACE-MS, as well as their serum concentrations from both observational and intervention studies, including matching total serum hydrolysate and independent erythrocyte PL fraction measurements, respectively.

4.4 Results

4.4.1 High throughput serum NEFA determination by MSI-NACE-MS

Overall, 24 serum FA (ranging from 9:0 to 24:1) were reliably measured as their NEFA and/or total hydrolyzed FA by MSI-NACE-MS from serum ether extracts (> 95%) with acceptable technical precision (CV < 15%) when using standard serum for QC as summarized in **Table S4.1**. Serum FA were analyzed after normalization of their ion responses to a single deuterated internal standard (18:0-d35) added to all samples, and most circulating FA were quantified in terms of their absolute concentration (μM) using an external calibration curve. Each run consisted of a serial

injection of 6 randomized serum samples together with a QC as shown in **Figure 4.1A** for representative serum NEFA annotated by their accurate mass and relative migration time (*m*/*z*:RMT). Anionic FA are resolved based on differences in their electrophoretic mobility (*i.e.,* carbon-chain length, degrees of unsaturation) that migrate after the electroosmotic flow (EOF) away from major neutral/zwitter-ionic lipids (*i.e.,* triacylglycerides, phospholipids, cholesterol) when using an alkaline non-aqueous buffer system, and detected as their intact molecular ion [M-H^-] under negative ion mode using a coaxial sheath liquid interface;[26] however, geometric isomers for certain FA are not baseline resolved. **Figure 4.1B** highlights that larger biological variance was evident for fasting serum NEFA (mean CV = 62%) as compared to corresponding total FA from serum lipid hydrolysates (mean CV = 32%) in pregnant women (n = 50). Also, good technical precision was confirmed based on repeat analysis of QC samples (mean CV ≈ 12%, n = 8), whereas a control chart for a recovery standard (14:0-d27) shows reliable long-term performance with few samples (≈ 2.6%, n = 114) exceeding warning limits (± 2s). Previous method validation studies demonstrated good mutual agreement for serum FA determination when using MSI-NACE-MS as compared to GC-MS,[26] which is optimal for higher throughput NEFA screening (< 4 min/sample) without pre-column chemical derivatization, lipid fractionation and/or hydrolysis artifacts. Only about 5-6% of total PUFA are protein-bound NEFA (*e.g.,* DHA, EPA and arachidonic acid) unlike other FA that are not extensively esterified into blood lipids, such as lauric acid (12:0) and myristic acid (14:0) as summarized in **Table S4.1**.

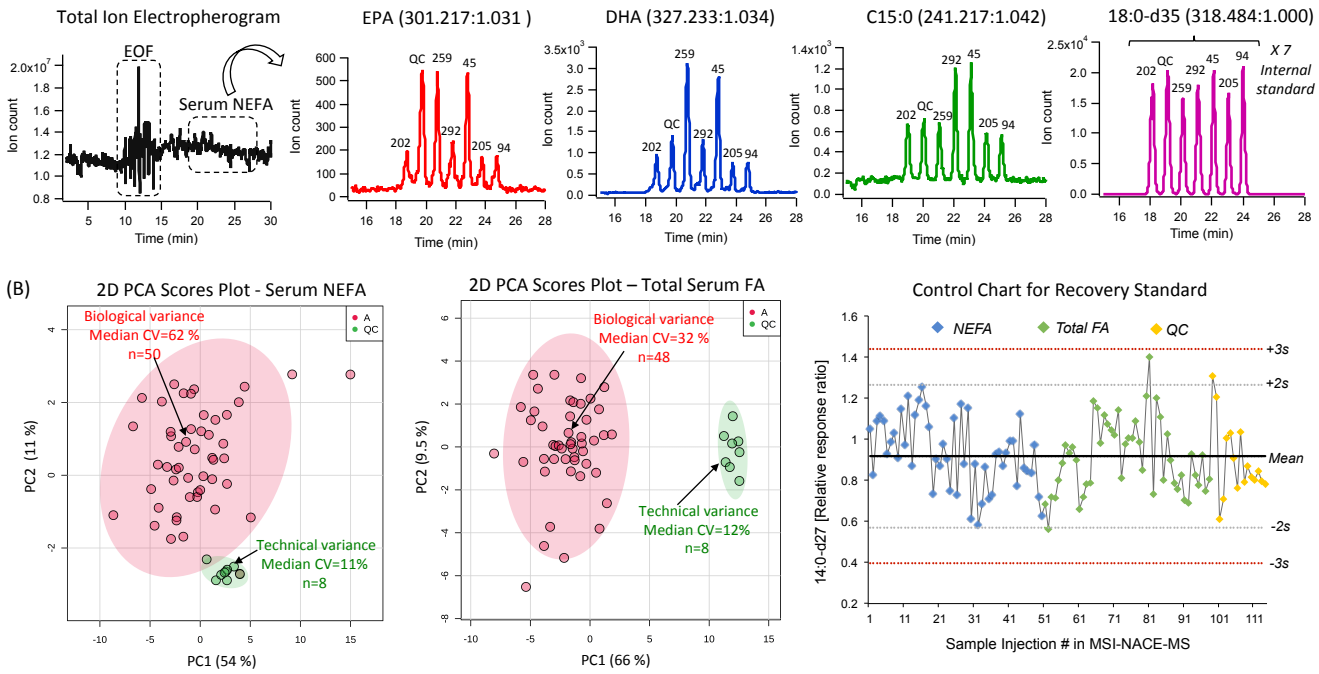


Figure 4.1. (A) Multiplexed separations of FA from serum extracts for assessment of dietary fat intake in pregnant women using MSI-NACE-MS under negative ion mode detection. This method relies on a serial injection of six randomized samples and a QC within each run to enhance sample throughput, where traces depict a total electropherogram, and a series of extracted ion electropherograms for representative serum NEFA annotated by their characteristic accurate mass and relative migration time (*m/z*:RMT), including a deuterated internal standard for data normalization. (B) Unsupervised multivariate data analysis using PCA depicts the biological variance from 24 FA as their NEFA or total hydrolysates, as compared to the technical variance from repeat QC, including a control chart for a recovery standard added to all processed serum samples.

4.4.2 Serum ω-3 PUFA status reflects differences in diet quality and habitual fish intake

Anthropometric and clinical data from second trimester pregnant women classified by their contrasting diets from FFQ are summarized in **Table 4.1**. Overall, 50 women were selected from eligible FAMILY participants (**Figure S4.1**) reflecting healthy (median DQI score = 12.0 ± 1.9) and non-healthy (median DQI score = - 9.1 ± 3.0) maternal eating patterns, respectively. As expected, pre-pregnancy BMI was lower in the healthy eating diet sub-group, but there were no differences in age, and weight gain during pregnancy, as well as fasting glucose concentrations, hemoglobulin glycation, serum triglycerides, and LDL; however, total and HDL cholesterol were

Table 4.1. Anthropometric and clinical characteristics of a cross-section of second-trimester pregnant women with contrasting dietary patterns from the FAMILY study.

Parameter	Poor diet group (*n*=25)	Good diet group (*n*=25)	*p*-value
Age (years)	31.3 ± 5.2	33.7 ± 5.7	0.14
Pre-pregnancy BMI (kg/m^2)*	29.3 ± 7.4	24.6 ± 5.3	0.011
Gestational weight gain (kg)	14.8 ± 6.9	14.1 ± 4.4	0.69
Total cholesterol (mmol/L)	6.0 ± 1.2	6.9 ± 1.0	0.015
LDL (mmol/L)	3.1 ± 1.0	3.5 ± 1.3	0.32
HDL (mmol/L)	1.8 ± 0.4	2.2 ± 0.5	0.006
Triglycerides* (mmol/L)	2.2 ± 0.8	2.3 ± 0.9	0.76
Fasting blood glucose (mmol/L)	4.4 ± 0.4	4.4 ± 0.5	0.69
HbA1c*	0.05 ± 0.01	0.05 ± 0.003	0.65
Diet quality index score* (range)	- 9.1 ± 3.0 (-16.5 to - 6.2)	12.0 ± 1.9 (10.8 to 17.4)	< 0.0001
Fish/seafood (#servings/day)*	0.1 ± 0.1	0.3 ± 0.2	< 0.0001
Total [EPA+ DHA] (g/day)*	0.07 ± 0.06	0.39 ± 0.50	< 0.0001
Total fiber (g/day)	18.3 ± 8.3	34.2 ± 7.6	< 0.0001
Full-fat dairy (#servings/day)	1.2 ± 1.1	1.6 ± 1.4	0.36

Presented data are mean and error as ± 1 s. Statistical comparisons assuming equal (t-test) or unequal variance (Welch's t test) or non-parametric Mann-Whitney test were performed as appropriate. Results were considered significant when $p < 0.05$. Full range of diet quality index score shown.*

modestly lower in the poor diet quality maternal group ($p < 0.02$). Importantly, total fiber intake, and daily fish/seafood servings were significantly higher in the healthy eating diet group ($p < 0.0001$) unlike full-fat dairy intake since it was not used a variable in the DQI score for participant selection. As expected, pregnant women consuming a healthy diet had consistently higher circulating concentrations of ω-3 PUFA, namely DHA and EPA in terms of their serum NEFA and total FA as compared to the poor diet quality maternal sub-group (**Figure 4.2A and 4.2B**). Moreover, moderate correlations ($r = 0.3$ to 0.5; $p < 0.05$) were measured between serum EPA, DHA, and their sum [EPA+DHA], relative to the DQI score, as well as total ω-3 PUFA from FFQ based on daily average intakes of EPA and DHA estimated during pregnancy from both dietary sources and FO supplement use as highlighted in **Table 4.2**.

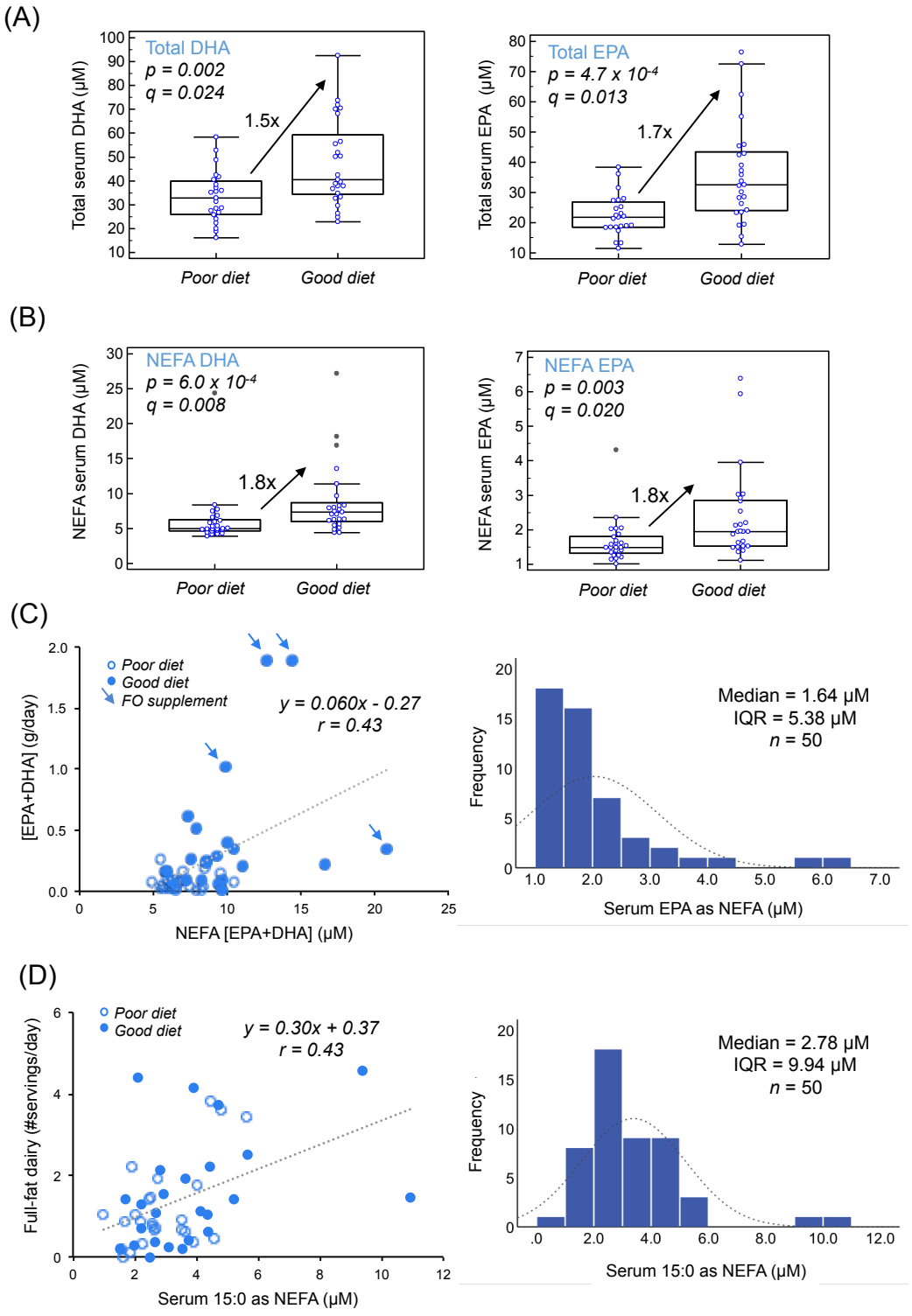


Figure 4.2. Boxplots and scatter plots for ω-3 PUFA in serum as (A) total hydrolyzed FA and (B) NEFA that are different among pregnant women with contrasting diets based on a DQI score using univariate Mann-Whitney test ($p < 0.05$) and after FDR adjustments ($q < 0.05$). (C) Scatter plot (left) showing the correlation of serum [EPA+DHA] measured as their NEFA by MSI-NACE-MS with self-reported total daily intake of ω-3 PUFA from FFQ; and histogram (right) showing the concentration distribution of EPA as its NEFA. (D) Scatter plot (left) showing the correlation of 15:0 as its NEFA with daily servings of full-fat dairy from FFQ, and histogram (right) showing the concentration distribution of 15:0 as its NEFA.

Table 4.2. Spearman rank correlation coefficients between serum ω-3 PUFA concentrations measured as their NEFA or total (hydrolyzed) FA fraction as compared to the diet quality index score, fish/seafood daily servings, and total ω-3 PUFA intake in pregnant women (n = 50) with contrasting diets from the FAMILY study.

	Diet quality index score[a]		**Fish/seafood (#servings/day)**		**Total ω-3 (g/day)[b]**	
Serum fatty acids	*r*	*p-value*	*r*	*p-value*	*r*	*p-value*
NEFA						
EPA (20:5*n*–3)	0.29*	0.043	0.45**	0.0010	0.46**	0.0010
DHA (22:6*n*–3)	0.38**	0.0060	0.36*	0.011	0.40**	0.0040
[EPA+DHA]	0.36**	0.010	0.40**	0.0040	0.43**	0.0020
Total FA						
EPA (20:5*n*–3)	0.37**	0.0090	0.47**	0.0010	0.50**	0.0003
DHA (22:6*n*–3)	0.28	0.052	0.29*	0.045	0.33*	0.024
[EPA+DHA]	0.40**	0.004	0.43**	0.0020	0.46**	0.0010

*Correlation is significant at the * 0.05 level (2-tailed) ** 0.01 level (2-tailed)*
[a] An aggregate score reflecting differences in nutritional quality of foods consumed based on the sum of the daily number of servings of healthy/nutrient-rich foods (e.g., fermented dairy, fish/seafood, vegetables, fruits, whole grains, nuts/seeds) minus the daily number of servings of unhealthy/processed foods (e.g., processed meats, refined grains, fries, snacks/sweets).
[b] Daily average servings of [EPA + DHA] intake estimated from FFQ, including diet (fish/seafood) and FO supplement use.

In fact, the strongest correlation was serum EPA as its NEFA or total FA with self-reported total ω-3 PUFA with r = 0.46 and 0.50 ($p < 0.001$), respectively. As for the correlation of circulating DHA with total ω-3 PUFA intake, it was found to be higher for NEFA (r = 0.40; p = 0.0040) as compared to total serum FA pool (r = 0.33; p = 0.024). Only 4 of 50 women were reported to be taking FO/ ω-3 supplements during pregnancy, and these women had the highest circulating NEFA concentrations for EPA. **Tables S4.2 and S4.3** summarize results from the linear regression model based on measured EPA and DHA concentrations as a function of the DQI score, and total ω-3 PUFA with adjustments for covariates between both diet groups, namely BMI, total cholesterol

and HDL. Overall, correlations remained significant ($p < 0.05$) after adjustments for BMI and total cholesterol, as well as HDL in most cases. The scatter plot in **Figure 4.2C** illustrates the positive correlation ($r = 0.43$, $p = 0.0020$) between measured concentrations for [EPA+DHA] as compared to self-reported ω-3 PUFA intake (g/day), which also includes a histogram for circulating EPA as its NEFA which had a median serum concentration of 1.64 μM. A correlation analysis (**Table S4.4**) between serum NEFA and total FA also demonstrated a much stronger association for EPA ($r = 0.57$; $p = 2.0 \times 10^{-5}$) as compared to DHA ($r = 0.29$; $p = 0.049$) highlighting the unique attribute of this low abundance circulating ω-3 PUFA.

4.4.3 Serum odd-chain/saturated FA reflect full-fat dairy and total fiber intake

Certain SFA as their serum NEFA and/or total FA hydrolysates were correlated with self-reported intake of full-fat dairy (**Table 4.3**). For instance, serum pentadecanoic acid (15:0) as its NEFA had the strongest association to full-fat dairy intake ($r = 0.43$; $p = 0.0020$), whereas heptadecanoic acid (17:0) was not significant ($r = 0.21$; $p = 0.15$). However, 17:0 from total hydrolyzed serum had a weak correlation to full-fat dairy ($r = 0.29$; $p = 0.043$). In this case, total 17:0 was also associated with daily fiber servings ($r = 0.29$; $p = 0.050$), including both soluble ($r = 0.38$; $p = 0.008$), and insoluble fiber ($r = 0.31$; $p = 0.034$) fractions. Also, serum myristic acid (14:0) showed a similar outcome as 15:0 as its NEFA ($r = 0.30$; $p = 0.034$) and total FA ($r = 0.35$; $p = 0.016$) albeit with a more moderate correlation to full-fat dairy intake. The scatter plot in **Figure 4.2D** highlights the positive correlation of fasting serum 15:0 as its NEFA to self-reported daily intake of dairy products (#servings/day). NEFA 14:0 and notably 15:0 are selective biomarkers of dairy fat since they were not correlated to either low-fat or fermented dairy intake. Also, circulating NEFA 15:0 status did not differentiate dietary sub-groups of pregnant women from FAMILY ($p = 0.36$) as

Table 4.3. Spearman rank correlation coefficients between serum FA measured as their NEFA or total (hydrolyzed) FA fraction as compared to full-fat dairy daily servings, and total fiber in pregnant women (n = 50) with contrasting diets from the FAMILY study.

	Full-fat Dairy (#servings/day)[a]		**Total Fiber (g/day)[b]**	
Serum Fatty Acids	*r*	*p-value*	*r*	*p-value*
NEFA				
Myristic acid (14:0)	0.30*	0.034	0.040	0.76
Pentadecanoic acid (15:0)	0.43**	0.0020	0.12	0.39
Heptadecanoic acid (17:0)	0.21	0.15	-0.26	0.86
Total				
Myristic acid (14:0)	0.35*	0.016	0.11	0.48
Pentadecanoic acid (15:0)	0.33*	0.023	0.22	0.14
Heptadecanoic acid (17:0)	0.29*	0.043	0.29*	0.050

*Correlation is significant at the * 0.05 level (2-tailed) ** 0.01 level (2-tailed)*
[a] Daily average servings of full-fat dairy products estimated from FFQ, including intake of cream of any kind, whole milk, milk, cottage/ricotta cheese, cream cheese, sour/whipping cream, full-fat cheese.
[b] Daily average intake of total fiber from various dietary sources estimated from FFQ.

they had similar consumption patterns for full-fat dairy (**Table 4.1**) with a median serum concentration of 2.78 μM for 15:0 as its NEFA.

4.4.4 Dietary intervention study in women: FO supplementation and serum NEFA trajectories

In this study, serum NEFA were analyzed in fasting serum samples from 18 young women collected at 4 time points over a 56 day intervention period, including baseline. For the active treatment arm, there was a mean 2.5-fold increase in serum NEFA concentrations for [EPA+DHA] from baseline after 28 days following high-dose FO supplementation as compared to the placebo group; however, there was no further increase in serum concentrations of EPA, DHA, or [EPA+DHA] at later sampling times (42 and 56 days) when using a 2-way mixed model ANOVA

(p = 0.012) as shown in **Figure 4.3A**. As expected, there was no changes in the temporal concentrations for other serum NEFA measured by MSI-NACE-MS either within-subjects, or between treatment arms at all time points (**Table 4.4**). This includes circulating linoleic acid and oleic acid despite being major constituents of sunflower oil consumed by the placebo group (**Figure 4.3B**); these two major FA in circulation are highly abundant in numerous other food sources in the diet. Overall, changes in serum EPA as its NEFA was found to be more sensitive to FO supplementation as compared to DHA or [EPA+DHA], which also was able to readily detect a non-adherent participant (S10) to FO supplementation who was also previously reported not to be compliant with leg immobilization protocols.[28] This is likely due to the lower circulating concentrations of EPA, and the higher dosage of EPA (≈ 3 g or 50% higher than DHA) used in FO supplement relative to DHA. Independent measurements available for hydrolyzed FA from erythrocyte membrane PL fraction showed a mean fold-change in [EPA+DHA] concentration of 2.6 from baseline, which was consistent with serum NEFA measurements. Further exploration of the underlying relationship between these two distinctive blood lipid pools demonstrated that there was a strong correlation only for EPA (r = 0.66; $p = 4.6 \times 10^{-10}$) at all time points (n = 69) when comparing concentrations from NEFA (protein-bound) and erythrocytes (membrane-bound) PL fractions in matching blood samples unlike DHA ($r = 0.22$; $p = 0.074$) as depicted in **Figure 4.3C**. A moderate correlation for DHA was only evident when comparing baseline and control cases (r = 0.35; p = 0.015; n = 44) after excluding data from the ω-3 PUFA treatment arm post-supplementation. No associations were found for other FA analyzed from these two blood fractions when using validated MSI-NACE-MS and GC-FID methods.

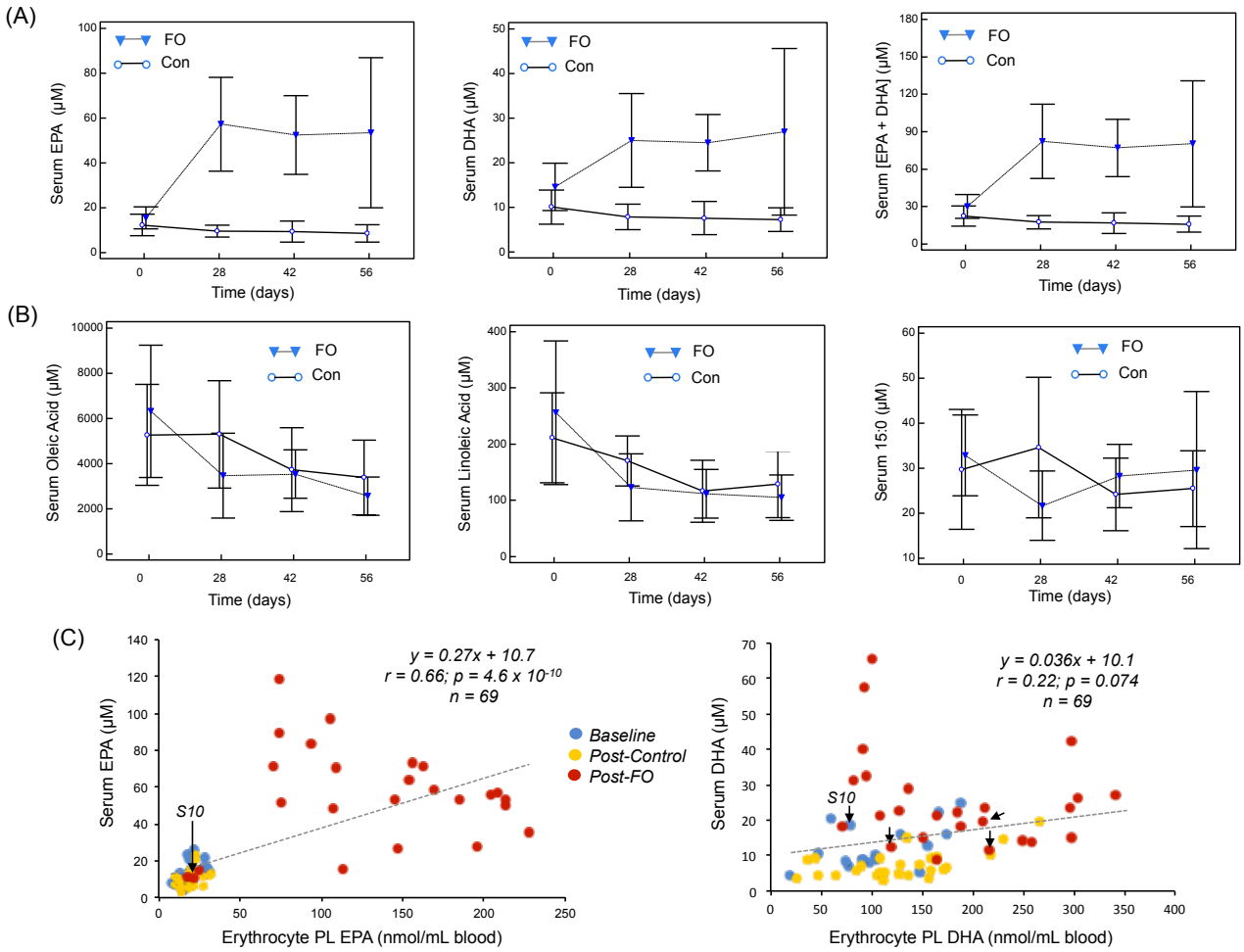


Figure 4.3. Graphs depicting dynamic changes in serum NEFA concentrations in the intervention study in young women for the high dose fish oil supplementation group (FO) ($n = 9$) as compared to sunflower oil (Con) ($n = 9$) based on means and error bars (± 2s), including (A) responsive EPA, DHA, and [EPA+DHA] in contrast to (B) non-responsive serum NEFA controls, including oleic acid, linoleic acid and pentadecanoic acid. The former two FA are major constituents in sunflower oil used that did not change in the control group. (C) Scatter plots showing a strong linear correlation ($r = 0.66$) between EPA concentrations as its serum NEFA as compared to corresponding erythrocyte phospholipid (PL) fraction at all time points ($n = 69$), unlike the much weaker correlation ($r = 0.22$) for DHA concentrations. Serum EPA was also more sensitive to detect dietary non-adherence to FO supplementation as indicated by arrows for one participant (S10) as compared to DHA as its NEFA.

Table 4.4. Fasting serum NEFA identified using a 2-way mixed-model ANOVA with repeated time points for within-subjects effects between high dose FO supplementation and control (sunflower oil) groups.

	Test of within-subjects effects				**Test of between-subjects effects**			
Serum Fatty Acids	*F*	*p-value*[a]	*Effect size*[b]	*Study power*	*F*	*p-value*[a]	*Effect size*[b]	*Study power*
EPA	8.36	0.0020*	0.39	0.99	22.2	0.0004*	0.63	0.99
[EPA+DHA]	5.59	0.0030*	0.30	0.92	19.0	0.0010*	0.59	0.98
DHA	3.54	0.023*	0.21	0.74	13.1	0.0030*	0.50	0.92
Myristic acid	1.93	0.14	0.13	0.46	1.5	0.24	0.10	0.21
Pentadecanoic acid	1.28	0.30	0.090	0.31	0.0	0.95	0.00	0.05
Heptadecanoic acid	0.28	0.84	0.020	0.10	0.26	0.62	0.02	0.080
Linoleic acid	0.34	0.80	0.030	0.11	1.5	0.24	0.11	0.21
Oleic acid	0.50	0.68	0.040	0.13	1.9	0.19	0.13	0.25

[a] *Mixed ANOVA model significant at the 0.001 level for EPA and DHA only for within and between subjects effects, where data sphericity was assumed/satisfied using Mauchly's test of sphericity.*
[b] *Based on partial Eta squared*

4.5 Discussion

For the first time, we report that fasting serum NEFA have promising utility for biomonitoring of dietary fat intake and FO supplementation in support of maternal health. Various blood fractions have been used for FA determination in nutritional studies[35,36] ranging from circulating lipid pools involved in transport (*e.g.,* NEFA, serum phospholipid fraction), cellular function (*e.g.,* erythrocyte or platelet membrane), to long-term storage (*e.g.,* adipose tissue triglycerides);[37] however, reports on serum NEFA as biomarkers of dietary fat have been sparse likely due to technical challenges in limiting background lipid hydrolysis even under mild reaction conditions for preparation of fatty acid methyl esters prior to GC analysis.[38] GC methods offer excellent selectivity for resolution of some FA geometric/positional isomers, but require longer analysis

times (> 20 min) even when using optimal column and elution conditions for comprehensive FA determination.[39] In this case, rapid serum/fasting NEFA screening can be achieved by enzymatic based colorimetric assays yet these less selective methods are prone to bias with discordant results as compared to LC-MS methods.[40] Alternatively, separation-free direct infusion-high resolution MS[41] or multiplexed chemical isotope labeling with LC-MS[42] offer greater sample throughput, but these approaches are better suited for analysis of total hydrolyzed FA from serum/plasma after sample processing. In this work, fasting serum NEFA were directly analyzed using a multiplexed separation platform based on MSI-NACE-MS,[26] which offers a higher throughput approach for assessment of complex dietary patterns associated with a health-promoting Prudent diet.[9] Equivalent or better correlations to self-reported intake of fish/seafood, full-fat dairy, as well as FO supplementation was achieved for certain serum NEFA as compared to their corresponding total FA from serum hydrolysates or erythrocyte membrane PL fractions. Indeed, there are conflicting reports on the exact relationship of fasting serum/plasma NEFA with corresponding FA from adipose tissue,[19,43,44] where circulating NEFA may serve as a virtual surrogate for tissue biopsies, and thus more convenient dietary biomarkers of habitual and acute changes in the intake of seafood and dairy products.[45]

Overall, our results from the cross-sectional study are consistent with a subset of participants from the EPIC study, where EPA and DHA had moderate correlations to self-reported fish intake from FFQ based on either total hydrolysates from plasma PL fraction (r = 0.33 and 0.29, respectively) or erythrocyte-membrane PL fractions (r = 0.29 and 0.40, respectively).[46] In our work, correlations for fasting serum EPA and DHA as their NEFA ranged from 0.36 to 0.46 indicating that it provides an analogous assessment of habitual fat intake without lipid fractionation and hydrolysis as required for FA determination from serum phospholipids.[36] For the high-dose

ω-3 PUFA supplementation study, our results demonstrated a relatively fast equilibration time for serum [EPA+DHA] as their NEFA as reflected by a mean 2.5 fold-change from baseline within 28 days, which is consistent with acute changes in circulating blood lipid pools in previous FO intervention studies;[37,47,48] this was a selective treatment effect as no other changes were measured in other serum NEFA from either active treatment or sunflower oil placebo arms. We hypothesize that this fast equilibration may be a consequence of the high ω-3 PUFA dosage regime used in this study (5 g/day) that is greater than the average intake of fish for Canadians (0.1-0.7 seafood meals/week) or used in commercial supplements (≈ 1 g/day).[22,49] Furthermore, the strong correlation between erythrocyte membrane PL and serum EPA as its NEFA over the duration of the study, but not DHA suggests that EPA is more responsive to changes in dietary patterns as compared to DHA, which is a consistent finding in both our observational and intervention studies involving young women. This is also in agreement with reports on other plasma fractions, where EPA responds to high-dose FO supplementation and cessation within 1 week[50] despite appreciable retroconversion of EPA to DHA.[51] In our work, serum EPA concentrations (median = 1.6 μM, range from 1.0 to 6.3 μM) were lower ($p = 0.00010$) in pregnant women than DHA (median = 6.1 μM, range from 3.8 to 27 μM) as their NEFA, which was more striking as compared to differences in their total serum hydrolysate ($p = 0.0010$) concentrations (**Figure 4.1**). Also, EPA as its NEFA was much more sensitive to detect self-reported FO supplement use among 4 pregnant women, as well as acute changes in non-pregnant women following high-dose ω-3 PUFA supplementation as compared to DHA or [EPA+DHA] (**Figure 4.3A**; **Tables 4.2**, **4.4**) while also revealing likely dietary non-adherence for a participant. Furthermore, serum EPA as its NEFA was strongly correlated ($r = 0.66$, $p = 4.6 \times 10^{-10}$) to independently measured erythrocyte PL membrane concentrations unlike DHA (**Figure 4.3C**). Consequently, we propose fasting serum EPA as its

NEFA as a robust and sensitive dietary biomarker that correlates well to long-term/habitual fish intake, as well as acute changes following FO supplementation. This is important given expanding interests in high dose ω-3 PUFA (either EPA+DHA or EPA only) for prevention of muscle atrophy,[28] reduction of asthma and persistent wheezing[52] that promotes also lean mass and bone growth in childhood,[53] as well as reducing atherosclerotic cardiovascular disease risk in patients with hypertriglyceridemia.[54] However, ω-3 PUFA can have quite distinctive lipid composition impacting their bioavailability while also varying up to 10-fold in natural abundance when comparing oily fish (*e.g.*, mackerel, salmon, sardines) to other commonly consumed lean fish (*e.g.*, haddock, canned tuna, cod) and other seafood sources (*e.g.*, algae, krill, prawns).[55] Interestingly, the poor diet quality sub-group of pregnant women had an estimated total ω-3 PUFA of only 71 mg/day from self-reported FFQ as compared to 217 mg/day for the healthy eating sub-group, which is still below 300 mg/day DHA recommended by the International Society for the Study of Fatty Acids and Lipids Working Group.[49] This information is valuable for prenatal screening of circulatory ω-3 PUFA status and biomonitoring of individual responses to dietary modifications or supplementation regimes for optimal maternal nutrition.

Next, OCFA are of special interest due to their role as promising food-specific biomarkers of full-fat dairy intake, which have also been reported to be inversely associated with type 2 diabetes risk.[56,57] Observational and intervention studies have reported that 15:0 and 17:0 are dietary biomarkers reflecting milk fat intake as measured from adipose tissue TG, serum PL, serum CE, total serum lipids, as well as dried blood spots;[56-60] however, there have been sparse reports from the analysis of serum OCFA as their NEFA.[61] In fact, serum 17:0 does not correlate with 15:0 since 17:0 can also be endogenously synthesized via α-oxidation, as well as generated via propionate via the action of gut microbiota on fermentable fiber;[56,57] for these reasons, 17:0 may

serve as a putative biomarker of dietary fiber intake. Alternatively, adipose tissue 14:0 has been proposed as a biomarker for long-term intake of dairy fat.[60] Our results confirmed that 14:0 (r = 0.30, p = 0.034), and especially 15:0 (r = 0.43, p = 0.0020) as their NEFA were robust dietary biomarkers of full-fat dairy intake in pregnant women (**Figure 4.2D**; **Table 4.4**), but not skim/low fat or fermented milk. Our results for 17:0 were inconsistent when comparing NEFA and total FA pools, with only the latter showing a weak association with self-reported intakes of full-fat dairy and total fiber, including soluble and insoluble fiber. Further studies that incorporate microbiome analyses are needed to better elucidate the utility of OCFA as biomarkers of fiber intake since it is a major source of biological variance. Nevertheless, our work validates the use of fasting serum NEFA as a convenient circulating lipid pool reflecting dietary intake of oily fish and full-fat dairy without invasive adipose tissue biopsies.

Strengths of our study include use of a rapid method based on MSI-NACE-MS for quantitative serum NEFA measurements with stringent quality control that was applied to two independent cohorts of women involving a validated FFQ, and a placebo-controlled, high-dose ω-3 PUFA clinical trial. Serum NEFA have been rarely investigated as dietary biomarkers in nutritional studies largely due to technical challenges when using low throughput GC protocols that are susceptible to background lipid hydrolysis and oxidation artifacts during sample processing. This work also compared analyses between NEFA and total serum FA pools for assessment of habitual fat consumption patterns, which represents a viable alternative to erythrocyte membrane PL hydrolysates when biomonitoring responses to ω-3 PUFA supplementation, including confirming dietary adherence. Our study has some limitations, including the modest sample size of each cohort involving a single biological sex, and the lack of self-reported diet records in the ω-3 PUFA clinical trial. Also, since extreme diet scores from

pregnant women were selected to maximize the effect size, this might have introduced a selection bias. For the intervention study, sampling shorter time points (< 28 days) is also needed to better assess the minimum time frame required for ω-3 PUFA equilibration that is likely both dose and sex-dependent. In conclusion, our study introduces a rapid yet inexpensive approach for quantification of serum NEFA that avoids serum lipid fractionation, hydrolysis and/or pre-column chemical derivatization procedures. This approach largely provides equivalent and in some cases superior results as compared to total FA hydrolyzed from serum, as well as erythrocyte PL fraction notably in the case of assessment of circulating EPA and 15:0 as optimal NEFA biomarkers of habitual intake of oily fish and full-fat dairy, respectively. MSI-NACE-MS is anticipated to enable large-scale blood-based testing of serum NEFA with greater sample throughput, lower costs, and better quality control than standard GC protocols, which is needed for more accurate assessment of dietary fat intake as compared to self-reported diet records and semi-quantitative FFQ.

The authors declare no conflict of interest.

Data availability statement

All processed serum FA data and deidentified clinical/dietary information for participants is available as an excel file in the supporting information section (NEFA-Serum-JLR.xlsx). This data includes original and batch-corrected RPA for NEFAs measured by MSI-NACE-MS, as well as their serum concentrations from both observational and intervention studies, including matching total serum hydrolysate and independent erythrocyte PL fraction measurements, respectively.

4.6 Acknowledgments

P.B.M. acknowledges funding from the Natural Sciences and Engineering Research Council of Canada, and Genome Canada. S.M.A acknowledges funding from the Egyptian Ministry of Higher

Education. S.S.A. acknowledges funding from Canadian Institutes for Health Research Team Grant. S.S.A. holds a Canada Research Chair in Ethnic Diversity and Cardiovascular Disease and Heart and Stroke, Michael DeGroote Chair in Population Health Research. S.M.P. acknowledges funding from the Natural Sciences and Engineering Research Council of Canada. We also acknowledge the contribution of all the FAMILY research team members notably Dr. Stephanie Atkinson for discussions on maternal diet assessment.

4.7 References

1. Liu AG, Ford NA, Hu FB, Zelman KM, Mozaffarian D, Kris-Etherton PM. A healthy approach to dietary fats: understanding the science and taking action to reduce consumer confusion. *Nutr J*. 2017;16:53.

2. Willett WC, Hu FB. Not the time to abandon the food frequency questionnaire: point. *Cancer Epidemiol Prev Biomark*. 2006;15:1757-1758.

3. Brennan L, Hu FB. Metabolomics-based dietary biomarkers in nutritional epidemiology-current status and future opportunities. *Mol Nutr Food Res*. 2019;63:1701064.

4. Arab L. Biomarkers of fat and fatty acid intake. *J Nutr*. 2003;133:925S-932S.

5. Ioannidis JPA. The challenge of reforming nutritional epidemiologic research. *JAMA*. 2018;320:969-970.

6. Wishart DS. Metabolomics: applications to food science and nutrition research. *Trends Food Sci Technol*. 2008;19:482-493.

7. Dragsted LO, Gao Q, Scalbert A, et al. Validation of biomarkers of food intake—critical assessment of candidate biomarkers. *Genes Nutr*. 2018;13:14.

8. Brennan L. Moving toward objective biomarkers of dietary intake. *J Nutr*. 2018;148:821-822.

9. Wellington N, Shanmuganathan M, de Souza RJ, et al. Metabolic trajectories following contrasting Prudent and Western diets from food provisions: identifying robust biomarkers of short-term changes in habitual diet. *Nutrients* 2019:11:2407

10. Gibbons H, Michielsen CJR, Rundle M, McNulty BA, Frost G, Nugent AP, Walton J, Flynn A, Gibney MJ, Brennan L. Demonstration of the utility of biomarkers for dietary intake assessment; proline betaine as an example. *Mol Nutr Food Res*. 2017;61:1700037.

11. Baylin A, Campos H. The use of fatty acid biomarkers to reflect dietary intake: *Curr Opin Lipidol*. 2006;17:22-27.

12. Forouhi NG, Krauss RM, Taubes G, Willett W. Dietary fat and cardiometabolic health: evidence, controversies, and consensus for guidance. *BMJ*. 2018;361:k2139.

13. Brandt EJ, Myerson R, Perraillon MC, Polonsky TS. Hospital admissions for myocardial infarction and stroke before and after the trans-fatty acid restrictions in New York. *JAMA Cardiol*. 2017;2:627-634.

14. Dehghan M, Mente A, Zhang X, et al. Associations of fats and carbohydrate intake with cardiovascular disease and mortality in 18 countries from five continents (PURE): a prospective cohort study. *The Lancet*. 2017;390:2050-2062.

15. Dehghan M, Mente A, Rangarajan S, et al. Association of dairy intake with cardiovascular disease and mortality in 21 countries from five continents (PURE): a prospective cohort study. *The Lancet*. 2018;392:2288-2297.

16. Andersen LF, Solvoll K, Drevon CA. Very-long-chain n–3 fatty acids as biomarkers for intake of fish and n–3 fatty acid concentrates. *Am. J. Clin. Nutr*. 1996;64:305–311.

17. Abdelhamid AS, Brown TJ, Brainard JS, et al. Omega-3 fatty acids for the primary and secondary prevention of cardiovascular disease. *Cochrane Database Syst Rev*. 2018;11: CD003177.

18. Miller M, Ballantyne CM, Bays HE, Granowitz C, Doyle RT, Juliano RA, Philip S. Effects of icosapent ethyl (eicosapentaenoic acid ethyl ester) on atherogenic lipid/lipoprotein, apolipoprotein, and inflammatory parameters in patients with elevated high-sensitivity C-reactive protein (from the ANCHOR study). *Am J Cardiol*. 2019;124:696-701.

19. Andersen LF, Solvoll K, Johansson LRK, Salminen I, Aro A, Drevon CA. Evaluation of a food frequency questionnaire with weighed records, fatty acids, and alpha-tocopherol in adipose tissue and serum. *Am. J. Epidemiol*. 1999;150:75–87.

20. Brenna JT, Plourde M, Stark KD, Jones PJ, Lin Y-H. Best practices for the design, laboratory analysis, and reporting of trials involving fatty acids. *Am J Clin Nutr*. 2018;108:211-227.

21. Stark K. Analytical implications of routine clinical testing for omega-3 fatty acid biomarkers. *Lipid Technol*. 2008;20:177-179.

22. Stark KD, Van Elswyk ME, Higgins MR, Weatherford CA, Salem N. Global survey of the omega-3 fatty acids, docosahexaenoic acid and eicosapentaenoic acid in the blood stream of healthy adults. *Prog Lipid Res*. 2016;63:132-152.

23. Hodson L, Skeaff CM, Fielding BA. Fatty acid composition of adipose tissue and blood in humans and its use as a biomarker of dietary intake. *Prog Lipid Res*. 2008;47:348-380.

24. Han LD, Xia JF, Liang QL, Wang Y, Wang YM, Hu P, Li P, Luo GA. Plasma esterified and non-esterified fatty acids metabolic profiling using gas chromatography-mass spectrometry and its application in the study of diabetic mellitus and diabetic nephropathy. *Anal. Chim. Acta*. 2011;689:85–91.

25. Hellmuth C, Weber M, Koletzko B, Peissner W. Nonesterified fatty acid determination for functional lipidomics: comprehensive ultrahigh performance liquid chromatography-tandem mass spectrometry quantitation, qualification, and parameter prediction. *Anal. Chem*. 2012;84:1483–1490.

26. Azab S, Ly R, Britz-McKibbin P. Robust method for high throughput screening of fatty acids by multisegment injection-nonaqueous capillary electrophoresis–mass spectrometry with stringent quality control. *Anal Chem*. 2019;91:2329-2336.

27. Morrison KM, Atkinson SA, Yusuf S, Burgeois J, McDonald S, McQueen MJ, Persadie R, Hunter B, Pogue J, Teo K. The Family Atherosclerosis Monitoring In earLY life (FAMILY) study. *Am Heart J*. 2009;158:533-539.

28. McGlory C, Gorissen SHM, Kamal M, Bahniwal R, Hector AJ, Baker SK, Chabowski A, Philips SM. Omega-3 fatty acid supplementation attenuates skeletal muscle disuse atrophy during two weeks of unilateral leg immobilization in healthy young women. *FASEB J*. 2019;33(3):4586-4597.

29. de Souza RJ, Zulyniak MA, Desai D, et al. Harmonization of food-frequency questionnaires and dietary pattern analysis in 4 ethnically diverse birth cohorts. *J Nutr*. 2016;146:2343-2350.

30. Anand SS, Yusuf S, Vuksan V, et al. Differences in risk factors, atherosclerosis, and cardiovascular disease between ethnic groups in Canada: the Study of Health Assessment and Risk in Ethnic groups (SHARE). *The Lancet*. 2000;356:6.

31. Merchant AT, Kelemen LE, de Koning L, Lonn E, Vuksan V, Jacobs R, Davis B, Teo KK, Yusuf S, Anand SS; SHARE and SHARE-AP investigators. Interrelation of saturated fat, trans fat, alcohol intake, and subclinical atherosclerosis. *Am. J. Clin. Nutr*. 2008;87:168–174

32. Kuehnbaum NL, Kormendi A, Britz-McKibbin P. Multisegment injection-capillary electrophoresis-mass spectrometry: a high throughput platform for metabolomics with high data fidelity. *Anal Chem*. 2013;85:10664-10669.

33. Chong J, Soufan O, Li C, Caraus L, Li S, Bourque G, Wishart D, Xia J. MetaboAnalyst 4.0: towards more transparent and integrative metabolomics analysis. *Nucleic Acids Res*. 2018;46:W486-W494.

34. Wehrens R, Hageman JosA, van Eeuwijk F, et al. Improved batch correction in untargeted MS-based metabolomics. *Metabolomics*. 2016;12:88.

35. Wolk A, Furuheim M, Vessby B. Fatty acid composition of adipose tissue and serum lipids are valid biological markers of dairy fat intake in men. *J Nutr*. 2001;131:828-833.

36. Furtado JD, Beqari J, Campos H. Comparison of the utility of total plasma fatty acids versus those in cholesteryl ester, phospholipid, and triglyceride as biomarkers of fatty acid intake. *Nutrients*. 2019;11:E2081.

37. Browning LM, Walker CG, Mander AP, West AL, Madden J, Gambell JM, Young S, Wang L, Jebb SA, Calder PC. Incorporation of eicosapentaenoic and docosahexaenoic acids into lipid pools when given as supplements providing doses equivalent to typical intakes of oily fish. *Am. J. Clin. Nutr*. 2012;96:748–758.

38. Ichihara K, Fukubayashi Y. Preparation of fatty acid methyl esters for gas-liquid chromatography. *J. Lipid Res*. 2010;51:635–640.

39. Ecker J, Scherer M, Schmitz G, Liebisch G. A rapid GC-MS method for quantification of positional and geometric isomers of fatty acid methyl esters. *J. Chromatogr. B*. 2012;897:98–104.

40. Song Y, Zhou L, Jensen MD. Errors in measuring plasma free fatty acid concentrations with a popular enzymatic colorimetric kit. *Clin. Biochem*. 2019;66:83–90.

41. Gallego SF, Hermansson M, Liebisch G, Hodson L, Ejsing CS. Total fatty acid analysis of human blood samples in one minute by high-resolution mass spectrometry. *Biomolecules*. 2018;9:E7.

42. Sun F, Choi AA, Wu R. Systematic analysis of fatty acids in human cells with a multiplexed isobaric tag (TMT)-based method. *J. Proteome Res*. 2018;17:1606–1614.

43. Walker CG, Browning LM, Stecher L, West AL, Madden J, Jebb SA, Calder PC. Fatty acid profile of plasma NEFA does not reflect adipose tissue fatty acid profile. *Br. J. Nutr*. 2015;114:756–762.

44. Baylin A, Kim MK, Donovan-Palmer A, Siles X, Dougherty L, Tocco P, Campos H. Fasting whole blood as a biomarker of essential fatty acid intake in epidemiologic studies: comparison with adipose tissue and plasma. *Am. J. Epidemiol*. 2005;162:373–381.

45. Hellmuth C, Demmelmair H, Schmitt I, Peissner W, Blüher M, Koletzko B. Association between plasma nonesterified fatty acids species and adipose tissue fatty acid composition. *PLoS One*. 2013;8:e74927.

46. Patel PS, Sharp SJ, Jansen E, Luben RN, Khaw KT, Wareham NJ, Forouhi NG. Fatty acids measured in plasma and erythrocyte-membrane phospholipids and derived by food-frequency questionnaire and the risk of new-onset type 2 diabetes: a pilot study in the European Prospective Investigation into Cancer and Nutrition (EPIC)–Norfolk cohort. *Am J Clin Nutr*. 2010;92:1214-1222.

47. Arterburn LM, Hall EB, Oken H. Distribution, interconversion, and dose response of n-3 fatty acids in humans. *Am. J. Clin. Nutr*. 2006;83(6):1467S–1476S.

48. Katan MB, Deslypere JP, van Birgelen AP, Penders M, Zegwaard M. Kinetics of the incorporation of dietary fatty acids into serum cholesteryl esters, erythrocyte membranes, and adipose tissue: an 18-month controlled study. *J. Lipid Res*. 1997;38:2012–2022.

49. Denomme J, Stark KD, Holub BJ. Directly quantitated dietary (n-3) fatty acid intakes of pregnant Canadian women are lower than current dietary recommendations. *J Nutr*. 2005;135:206-211.

50. Metherel AH, Armstrong JM, Patterson AC, Stark KD. Assessment of blood measures of n-3 polyunsaturated fatty acids with acute fish oil supplementation and washout in men and women. *Prostaglandins Leukot Essent Fatty Acids.* 2009;81:23–9.

51. Metherel AH, Irfan M, Klingel SL, Mutch DM, Bazinet RP. Compound-specific isotope analysis reveals no retroconversion of DHA to EPA but substantial conversion of EPA to DHA following supplementation: a randomized control trial. *Am J Clin Nutr*. 2019;110:823-831.

52. Bisgaard H, Stokholm J, Chawes BL, et al. Fish oil–derived fatty acids in pregnancy and wheeze and asthma in offspring. *N Engl J Med*. 2016;375:2530-2539.

53. Vinding RK, Stokholm J, Sevelsted A, Sejersen T, Chawes BL, Bønnelykke K, Thorsen J, Howe LD, Krakauer M, Bisgaard H. Effect of fish oil supplementation in pregnancy on bone, lean, and fat mass at six years: randomised clinical trial. *BMJ*. 2018;362:k3312.

54. Skulas-Ray AC, Wilson PWF, Harris WS, et al. Omega-3 fatty acids for the management of hypertriglyceridemia: a science advisory from the American Heart Association. *Circulation.* 2019; 140:e673–91.

55. Calder PC. Very long-chain n-3 fatty acids and human health: fact, fiction and the future. *Proc. Nutr. Soc*. 2018;77:52–72.

56. Jenkins BJ, Seyssel K, Chiu S, et al. Odd chain fatty acids; new insights of the relationship between the gut microbiota, dietary intake, biosynthesis and glucose intolerance. *Sci Rep*. 2017;7:44845.

57. Weitkunat K, Schumann S, Nickel D, Hornemann S, Petzke KJ, Schulze MB, Pfeiffer AFH, Klaus S. Odd-chain fatty acids as a biomarker for dietary fiber intake: a novel pathway for endogenous production from propionate. *Am J Clin Nutr*. 2017;105:1544–51.

58. Smedman AE, Gustafsson I-B, Berglund LG, Vessby BO. Pentadecanoic acid in serum as a marker for intake of milk fat: relations between intake of milk fat and metabolic risk factors. *Am J Clin Nutr*. 1999;69:22-29.

59. Brevik A, Veierød MB, Drevon CA, Andersen LF. Evaluation of the odd fatty acids 15:0 and 17:0 in serum and adipose tissue as markers of intake of milk and dairy fat. *Eur J Clin Nutr*. 2005;59:1417-1422.

60. Albani V, Celis-Morales C, Marsaux CFM, et al. Exploring the association of dairy product intake with the fatty acids C15:0 and C17:0 measured from dried blood spots in a multipopulation cohort: Findings from the Food4Me study. *Mol Nutr Food Res*. 2016;60:834-845.

61. Trimigno A, Münger L, Picone G, Freiburghaus C, Pimentel G, Vionnet N, Pralong F, Capozzi F, Badertscher R, Vergeres G. GC-MS based metabolomics and NMR spectroscopy investigation of food intake biomarkers for milk and cheese in serum of healthy humans. *Metabolites*. 2018;8:26.

4.8 Supplemental Information

Table S4.1. Serum fatty acids (NEFA and/or total FA) measured by MSI-NACE-MS from FAMILY cohort (*n*=50).

	Fatty Acid	Molecular Formula	*m/z*[a]	RMT[b]	Mass Error (ppm)	%RSD (*n* = 8)[c]	Ratio of NEFA/Total (%)[d]
1	Pelargonic acid (9:0)*	$C_9H_{18}O_2$	157.123	1.214	0.7	9.8	-
2	Capric acid (10:0)*	$C_{10}H_{20}O_2$	171.139	1.123	2.6	22.6	-
3	Lauric acid (12:0)	$C_{12}H_{24}O_2$	199.170	1.085	2.0	14.1	137 ± 80
4	Myristic acid (14:0)	$C_{14}H_{28}O_2$	227.202	1.054	3.5	8.6	85 ± 42
5	Pentadecanoic acid (15:0)	$C_{15}H_{30}O_2$	241.217	1.041	2.1	11.0	34 ± 18
6	Palmitic acid (16:0)	$C_{16}H_{32}O_2$	255.233	1.028	1.2	12.6	19 ± 9
7	Heptadecanoic acid (17:0)	$C_{17}H_{34}O_2$	269.249	1.017	2.2	13.2	28 ± 14
8	Stearic acid (18:0)	$C_{18}H_{36}O_2$	283.264	1.005	4.6	13.6	22 ± 13
9	Arachidic acid (20:0)*	$C_{20}H_{40}O_2$	311.296	0.983	3.8	13.6	-
10	Behenic acid (22:0)**	$C_{22}H_{44}O_2$	339.327	1.017	0.6	15.9	-
11	Myristelaidic acid (14:1)	$C_{14}H_{26}O_2$	225.186	1.061	1.8	13.1	42 ± 13
12	Palmitoleic acid (16:1)	$C_{16}H_{30}O_2$	253.217	1.035	2.4	8.5	85 ± 45
13	Heptadecenoic acid (17:1)	$C_{17}H_{32}O_2$	267.233	1.024	1.8	7.9	120 ± 50
14	Oleic acid (18:1)[x]	$C_{18}H_{34}O_2$	281.249	1.012	1.1	12.7	69 ± 33
15	Gondoic acid (20:1)	$C_{20}H_{38}O_2$	309.280	0.99	4.2	13.1	87 ± 50
16	Nervonic acid (24:1)**	$C_{24}H_{46}O_2$	365.342	0.946	5.2	15.5	-
18	Linoleic acid (18:2)[x]	$C_{18}H_{32}O_2$	279.233	1.018	0.4	7.7	18 ± 9
17	Eicosadienoic acid (20:2)	$C_{20}H_{36}O_2$	307.264	0.995	3.1	11.8	33 ± 30
19	Linolenic acid (18:3)[x]	$C_{18}H_{30}O_2$	277.217	1.021	1.4	8.1	71 ± 32
20	Dihomo-γ-linolenic acid (20:3*n*-6)	$C_{20}H_{34}O_2$	305.249	1.004	2.3	10.6	-
21	Arachidonic acid (20:4*n*-6)	$C_{20}H_{32}O_2$	303.233	1.025	2.3	5.1	5 ± 4
22	Eicosapentaenoic acid (20:5*n*-3)	$C_{20}H_{30}O_2$	301.217	1.027	5.0	9.6	5 ± 4
23	Adrenic acid (22:4*n*-6)**	$C_{22}H_{36}O_2$	331.264	0.991	4.5	8.5	-
24	Docosahexaenoic acid (22:6*n*-3)	$C_{22}H_{32}O_2$	327.233	1.029	2.4	9.0	6 ± 6

[a] *Accurate mass of fatty acid detected as its intact molecular ion [M-H⁻] under negative mode ionization.*
[b] *Migration time of fatty acids relative to the internal standard, C18d35 for data normalization.*
[c] *Relative standard deviation of repeated QC samples introduced in each run for assessment of technical precision.*
[d] *Relative abundance of NEFA as compared to total (hydrolyzed) FA based on a ratio of their relative peak areas.*
**Fatty acid only detected in serum NEFA analysis*
***Fatty acid only detected in serum for total fatty acid analysis*
[x] *Isomeric resolution of cis and trans geometrical isomers for fatty acids was not achieved.*

Table S4.2. Multiple linear regression model to predict levels of ω-3 fatty acids (dependent variable) as a function of the diet quality index score with adjustments for covariates, BMI, cholesterol, and HDL.

Independent variables	Coefficient	*p*-value	*r*
EPA-Total			
Diet index (DI)	0.018	0.0020	0.44
DI, BMI	0.017	0.0030	0.44
DI, BMI, cholesterol	0.013	0.029	0.52
DI, BMI, cholesterol, HDL	0.006	0.19	0.75
DHA-Total			
Diet index (DI)	0.015	0.0030	0.42
DI, BMI	0.015	0.0030	0.42
DI, BMI, cholesterol	0.009	0.061	0.57
DI, BMI, cholesterol, HDL	0.008	0.10	0.58
EPA-NEFA			
Diet index (DI)	0.012	0.017	0.34
DI, BMI	0.012	0.025	0.35
DI, BMI, cholesterol	0.013	0.019	0.37
DI, BMI, cholesterol, HDL	0.009	0.087	0.52
DHA-NEFA			
Diet index (DI)	0.013	0.017	0.34
DI, BMI	0.013	0.027	0.35
DI, BMI, cholesterol	0.014	0.024	0.36
DI, BMI, cholesterol, HDL	0.011	0.081	0.45

Table S4.3. Multiple linear regression model to predict levels of ω-3 PUFA (dependent variable) as a function of total ω-3 daily servings with adjustment for covariates, BMI, cholesterol, and HDL.

Independent variables	Coefficient	*p*-value	*r*
EPA-Total			
Tω-3	0.373	0.020	0.34
Tω-3, BMI	0.363	0.024	0.35
Tω-3, BMI, HDL	0.295	0.055	0.50
Tω-3, BMI, cholesterol, HDL	0.209	0.076	0.76
DHA-Total			
Tω-3	0.520	1.9×10^{-6}	0.53
Tω-3, BMI	0.518	2.7×10^{-6}	0.53
Tω-3, BMI, HDL	0.446	4.7×10^{-6}	0.68
Tω-3, BMI, cholesterol, HDL	0.434	8.5×10^{-6}	0.69
EPA-NEFA			
Tω-3	0.329	0.024	0.32
Tω-3, BMI	0.320	0.029	0.34
Tω-3, BMI, HDL	0.337	0.025	0.36
Tω-3, BMI, cholesterol, HDL	0.283	0.039	0.54
DHA-NEFA			
Tω-3	0.308	0.056	0.27
Tω-3, BMI	0.297	0.067	0.30
Tω-3, BMI, HDL	0.310	0.061	0.32
Tω-3, BMI, cholesterol, HDL	0.265	0.095	0.44

Table S4.4. Spearman correlation coefficients between circulating DHA and EPA (molar concentrations) measured as their NEFA and hydrolyzed (total) serum (n = 48) in pregnant women (n = 50) from the FAMILY study.

	Serum Total			
	DHA (22:6n–3)		EPA (20:5n–3)	
Serum NEFA	*r*	*p-value*	*r*	*p-value*
EPA (20:5*n*–3)	0.44**	0.0020	0.57**	0.000020**
DHA(22:6*n*–3)	0.29*	0.049	0.44**	0.0020**

*Correlation is significant at the * 0.05 level (2-tailed) ** 0.01 level (2-tailed)*

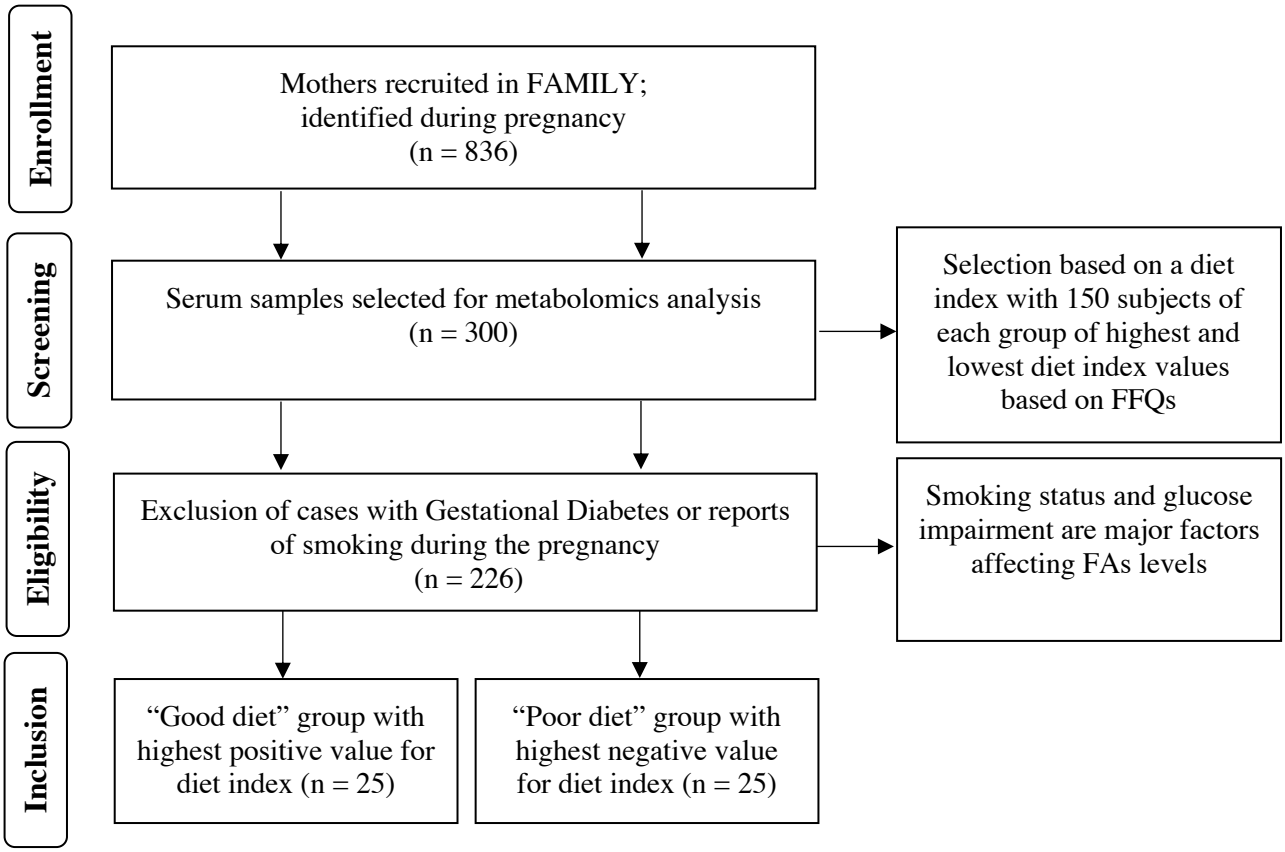


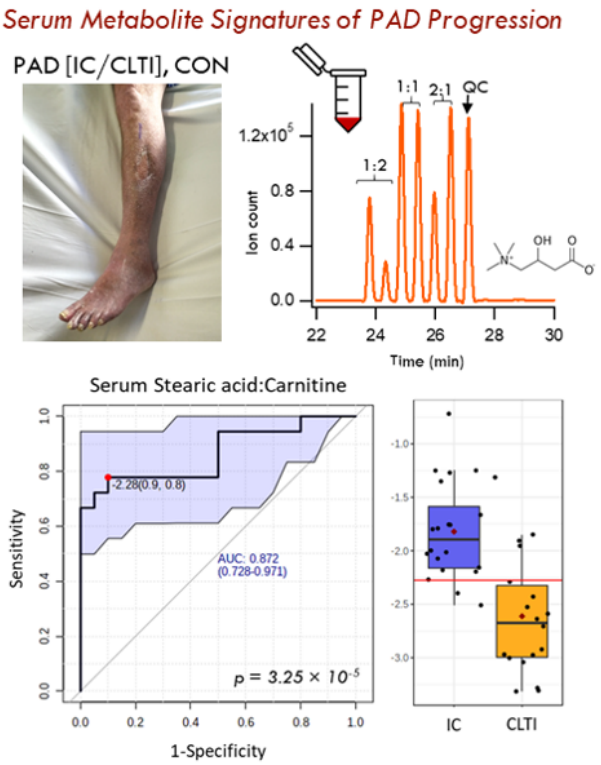
Figure S4.1. A consort flow diagram outlining selection criteria used in a cross-sectional study involving participants from the FAMILY birth cohort (n = 50) having contrasting eating patterns as measured by a diet quality index score. NEFA and total hydrolyzed fatty acids analysis was performed using MSI-NACE-MS on maternal serum samples collected during the second trimester of pregnancy.

Chapter V:

Serum Metabolic Signatures of Chronic Limb-Threatening Ischemia in Patients with Peripheral Artery Disease

Thesis chapter is derived from a published peer-reviewed article:

S.M. Azab, A. Zamzam, M. Syed, R. Abdin, M. Qadura and P. Britz-McKibbin. Serum metabolic signatures of chronic limb-threatening ischemia in patients with peripheral artery disease. *J. Clin. Med.* **2020**, 9: 1877.



M.Q. and P.B.M. designed the study. M.Q., M.H.S. and R.A. recruited the patients and collected blood samples. M.Q. was the principal investigator of the clinical recruitment protocol. S.M.A. conducted the metabolomics analysis including serum processing and data processing under the supervision of P.B.M. Statistical analysis was performed by S.M.A. and P.B.M., which was further evaluated by A.Z., M.Q. and S.M.A. drafted the manuscript that was edited by P.B.M. with critical feedback from M.Q. All authors read and approved the final manuscript.

Chapter V: Serum Metabolic Signatures of Chronic Limb-Threatening Ischemia in Patients with Peripheral Artery Disease

"It is far better to foresee even without certainty than not to foresee at all."

5.1 Abstract

Peripheral artery disease (PAD) is characterized by the atherosclerotic narrowing of lower limb vessels, leading to ischemic muscle pain in older persons. Some patients experience progression to advanced chronic limb-threatening ischemia (CLTI) with poor long-term survivorship. Herein, we performed serum metabolomics to reveal the mechanisms of PAD pathophysiology that may improve its diagnosis and prognosis to CLTI complementary to the ankle–brachial index (ABI) and clinical presentations. Non-targeted metabolite profiling of serum was performed by multisegment injection–capillary electrophoresis–mass spectrometry (MSI–CE–MS) from age and sex-matched, non-diabetic, PAD participants who were recruited and clinically stratified based on the Rutherford classification into CLTI (n = 18) and intermittent claudication (IC, n = 20). Compared to the non-PAD controls (n = 20), PAD patients had lower serum concentrations of creatine, histidine, lysine, oxoproline, monomethylarginine, as well as higher circulating phenylacetylglutamine ($p < 0.05$). Importantly, CLTI cases exhibited higher serum concentrations of carnitine, creatinine, cystine and trimethylamine-*N*-oxide along with lower circulating fatty acids relative to well matched IC patients. Most serum metabolites associated with PAD progression were also correlated with ABI ($r = \pm 0.24–0.59$, $p < 0.05$), whereas the ratio of stearic acid to carnitine, and arginine to propionylcarnitine differentiated CLTI from IC with good accuracy ($AUC = 0.87$, $p = 4.0 \times 10^{-5}$). This work provides new biochemical insights into PAD progression for the early detection and surveillance of high-risk patients who may require peripheral vascular intervention to prevent amputation and premature death.

5.2. Introduction

Peripheral artery disease (PAD) is a form of atherosclerosis that manifests in the lower extremities leading to a cascade of symptoms from insufficient blood flow, including impaired/painful walking, reduced functional capacity, ischemic myopathy, and recurrent skin lesions.[1] PAD is also associated with higher risk for cardiovascular events such as myocardial infarction, stroke, and vascular death.[2] Although most PAD patients are asymptomatic, patients with PAD usually present with intermittent claudication (IC), characterized with varying degrees of pain in leg muscles induced by walking. If left untreated, patients can progress to a severe end-stage form of PAD known as chronic limb-threatening ischemia (CLTI), which is characterized by rest pain, non-healing ischemic ulcers, and gangrene requiring limb amputation.[2] Disease progression in PAD is highly variable and unpredictable as some CLTI patients who undergo amputations do not exhibit any PAD symptoms 6 months before onset.[3] In fact, cardiovascular events are more prevalent among CLTI patients than coronary artery disease (CAD) indicating significant associated morbidity.[3] Although CLTI accounts for less than 5% of total PAD diagnoses, survivorship is poor with a 5-year mortality rate of about 50%.[4] As a result, there is an urgent need for understanding the mechanisms of PAD progression for the early detection of CLTI that also guides evidence-based treatment decisions.[5]

Despite a high estimated prevalence of 10–20% among older persons, PAD is often undiagnosed and/or untreated in the primary care setting with most physicians unaware of the diagnosis.[6] In practice, diagnosis is confirmed in symptomatic patients by an abnormal resting ankle–brachial index (ABI) below 0.90, which is determined by the ratio of the systolic blood pressure at the ankle of the affected leg to the upper arm, using a Doppler ultrasound blood flow detector.[1] Usually, ABI is a reliable diagnostic tool for PAD diagnosis that may also predict

atherosclerotic PAD mortality when performed in accredited laboratories with specialized equipment and training.[7] However, in 25% of diabetic patients, ABI lacks sensitivity to diagnose patients with PAD due to peripheral arterial stiffening from calcification.[8] Moreover, the benefit of the routine screening of PAD risk in asymptomatic patients using ABI remains inconclusive based on the recent findings from the US Preventive Services Task Force.[9] For these reasons, specific yet sensitive blood-based biomarkers are needed for PAD diagnosis and risk assessment that is applicable for routine testing in a clinical setting, including the surveillance of high-risk CLTI patients following revascularization interventions.

Metabolomics offers a systemic approach for the molecular phenotyping of complex biological processes underlying cardiovascular diseases (CVD) as required for new advances in precision medicine and drug development.[10] Comprehensive metabolite profiling using high-field nuclear magnetic resonance (NMR) and increasingly high-resolution mass spectrometry (MS) enables the discovery of clinically relevant biomarkers associated with atherosclerosis, that reflect the dynamic interactions between the host, gut microbiota, and dietary exposures, such as trimethylamine-*N*-oxide.[11] Growing evidence also demonstrates that elevated plasma branched-chain amino acid concentrations increase the risk for stroke that may be counteracted by Prudent diet modifications.[12] However, most metabolomic studies to date have focused on identifying aberrant metabolic pathways in CAD as compared to standard predictors.[13] In contrast, there have been few reports using metabolomics to understand the pathophysiology of PAD, which is prone to confounding since older patients often suffer from other comorbidities, including type 2 diabetes, chronic kidney disease and/or cardiovascular events.[14–17] Herein, we apply an untargeted serum metabolomics data workflow with stringent quality control (QC) on a clinically stratified non-diabetic patient cohort for the assessment of PAD progression that may allow for better

therapeutic monitoring of patients as compared to conventional ABI and clinical assessment tools.

5.3. Experimental Section

5.3.1. Study Cohort and Design

This study was approved by the research ethics board at St. Michael's Hospital-University of Toronto (REB #16-375), and informed consent was obtained from all participants in accordance with the Declaration of Helsinki principles. PAD diagnosis and classification into IC and CLTI were made according to specialists' clinical examination and arterial ultrasound.[18] Patients with CLTI or "Rutherford stage ≥ 4" referred to vascular surgery ambulatory clinics or emergency department at St. Michael's Hospital (Toronto, ON, Canada) from June 2017 to August 2017, were requested to participate in this study. Exclusion criteria included all patients on anticoagulants, chemotherapy or biological anti-inflammatory agents. Patients diagnosed with sepsis, type 2 diabetes, systematic inflammatory disease or with an active/history of any cancer or deep vein thrombosis were excluded as well. Moreover, patients with a 6-month history of acute coronary syndrome, heart failure, or uncontrolled arrhythmia, as defined by the American College of Cardiology, also failed to meet the inclusion criteria of this study.[19] A total of 128 consecutive ambulatory patients were initially recruited, however only 20 CLTI patients met the inclusion criteria and consented for this study. Upon acceptance to participate, CLTI patients were matched with non-diabetic IC cases or "Rutherford stage 1–3", and non-PAD controls in a ratio of 1:1:1 by age group and biological sex. To do so, during the months of January and February 2018, 20 IC patients and 20 non-PAD participants were recruited to match the CLTI cohort. PAD status was defined clinically as per the Rutherford classification, whereas non-PAD controls were defined as patients with cardiovascular risk factors alongside a normal arterial ultrasound of the lower limbs,

and palpable distal pulses without a significant clinical history of claudication. However, after matching our cohorts, two CLTI patients later withdrew their consent and were not included in the final study.

5.3.2. Baseline Patient Clinical Assessments

Medical history included details of any previous acute coronary syndrome, hyperlipidemia, arterial arrhythmia, arterial hypertension, renal disease, congestive heart failure, history of stroke or transient ischemic attack, history of cancer, diabetes, and smoking status. Hyperlipidemia was defined as total cholesterol > 5.2 mM or the use of anti-hyperlipidemic medication. Hypertension was defined as systolic blood pressure ≥ 130 mmHg or diastolic pressure ≥ 80 mm Hg, or the use of antihypertensive medication. Renal disease was defined as an estimated glomerular filtration rate of less than 60 mL/min/1.73 m^2 as per the Kidney Disease Outcomes Quality Initiative 2002 guidelines, since serum cystatin C and urinary creatinine clearance are not measured routinely for vascular patients. Diabetes mellitus was defined as glycosylated hemoglobin A1c ≥ 6.5% or the use of antidiabetic medication. Smoking status was recorded for each patient.

5.3.3. Chemicals and Reagents

All chemicals were purchased from Sigma Aldrich (St. Louis, MO, USA) unless otherwise stated. Stock solutions for internal standards and metabolite standards were prepared in deionized water from a Barnstead EASYpure® II LF system (Dubuque, IA, USA) for hydrophilic metabolites, and in methyl-*tert*-butyl ether (MTBE) for lipophilic fatty acids. Ultra-grade LC–MS solvents (acetonitrile, methanol, 2-isopropanol and water) purchased from Caledon Laboratories Ltd. (Georgetown, ON, Canada) were used to prepare sheath liquid solution for spray formation and aqueous or nonaqueous background electrolyte (BGE) for the capillary electrophoresis (CE)

separations. All stock solutions for chemical standards were stored at 4 ºC. Nanosep® 3k Omega™ ultrafiltration devices (Pall Life Sciences, Port Washington, NY, USA) were used for processing the diluted serum samples with a 3 kDa molecular weight cut-off filter for protein removal.

5.3.4. Serum Creatinine Measurement by Jaffé Method

Measurement of serum creatinine was performed at St. Michael's Hospital using a modified Jaffé colorimetric assay on a Beckman Coulter AU system/analyzer (Beckman Coulter, Inc., Brea, CA, USA). Briefly, creatinine standard reagent (OSR6678) was added to an aliquot of serum, where the serum creatinine reacted with picric acid under alkaline conditions to form a yellow-orange complex and the rate of change in absorbance at 520/800 nm was proportional to the serum creatinine concentration to minimize other reacting/absorbing interferences.[20] Serum creatinine measurements by the Jaffé colorimetric assay were compared with multisegment injection–capillary electrophoresis–mass spectrometry (MSI–CE–MS) using independent serum aliquots collected from the same participants.

5.3.5. Serum Sample Collection and Preparation

Fasting blood samples were collected and immediately centrifuged at 4 °C within 1 h after clotting at room temperature, where the serum was separated, aliquoted and stored frozen at –80 °C. Frozen serum was then thawed slowly on ice, vortexed for 30 s and aliquoted prior to ultrafiltration or liquid extraction. An aliquot of 50 µL of serum was diluted two-fold with ultra-grade LC–MS water containing 40 µM of two recovery standards, 3-chloro-*L*-tyrosine (Cl-Tyr) and 3-cyclohexylamino-1-propanesulfonic acid (CAPS). The diluted serum was vortexed for 30 s, transferred to a pre-rinsed ultrafiltration device, that was centrifuged at 14,000 × *g* for 10 min to separate the proteins from the serum filtrate used for the analysis of polar/hydrophilic metabolites.

Ultrafiltration devices were pre-rinsed with ultra-grade LC–MS water, centrifuged for 5 min at 14,000 × *g* and air dried for about 20 min prior to the first use to wash out background additives (*e.g.*, lactic acid). Thereafter, serum filtrates (15 μL) were diluted one or two-fold with ultra-grade LC–MS water containing three internal standards, 4-fluoro-*L*-phenylalanine (F-Phe), 2-napthalenesulfonic acid (NMS) and $^{13}C_6$-glucose. The final concentrations for all the internal and recovery standards were 10 μM with the exception of $^{13}C_6$-glucose (2 mM), and all diluted serum filtrate samples were analyzed using MSI–CE–MS with two aqueous BGE systems optimal for the separation of ionic/hydrophilic metabolites.[21,22]

A second aliquot of frozen serum was slowly thawed on ice and processed separately for lipid analysis following acid hydrolysis and liquid extraction as described previously.[23,24] In this work, the total (hydrolyzed) serum fatty acids were analyzed by multisegment injection–nonaqueous capillary electrophoresis–mass spectrometry (MSI–NACE–MS). Acid-catalyzed hydrolysis of esterified lipids was performed by the addition of 25 μL of serum, 25 μL of 2.5 M sulfuric acid and 25 μL of 0.01% vol butylated hydroxytouene (BHT) as an antioxidant additive in toluene followed by incubation at 80 °C for 1 h. Serum fatty acids were subsequently extracted using a slightly modified extraction protocol originally reported by Matyash et al.[25] using 500 μL of MTBE containing 50 μM of deuterated myristic acid (14:0-d27) as a recovery standard, with 12.5 μL of 1.0 M HCl for better extraction efficiency. Following vigorous shaking for 30 min at room temperature, phase separation was then induced by the addition of 100 μL of deionized water. Samples were then centrifuged at 3000× *g* at 4 °C to sediment the protein for 30 min resulting in phase separation into a water and ether (top) layer. A fixed volume (200 μL) was collected from the upper MTBE layer into a new vial then dried under a gentle stream of nitrogen gas at room temperature. Serum extracts were then stored dry at –80 °C and at the time of analysis reconstituted

in 25 μL of acetonitrile/isopropanol/water (70:20:10 vol) with 10 mM ammonium acetate and 50 μM deuterated stearic acid (18:0-d35) as an internal standard. All hydrolysis reactions and extractions were carried out using glass GC vials that were pre-rinsed with dichloromethane and all the pipette tips used during this procedure were pre-rinsed with methanol to minimize the background palmitic acid (16:0) and stearic acid (18:0) contamination.[23] An internal QC sample was prepared in-house by pooling together serum aliquots from all the study participants. Separate QC aliquots were processed using the same sample protocols described above, stored at –80 °C and thawed once prior to analysis by MSI–NACE–MS using a nonaqueous BGE system optimal for the separation of acidic lipids.

5.3.6. Hydrophilic Metabolome Profiling by MSI–CE–MS

MSI–CE–MS experiments were performed on an Agilent G7100A CE instrument (Agilent Technologies Inc., Mississauga, ON, Canada) equipped with a coaxial sheath liquid (Dual AJS ESI) Jetstream electrospray ionization source coupled to an Agilent 6550 quadrupole-time-of-flight (QTOF) system. An Agilent 1260 Infinity isocratic pump equipped with a 100:1 splitter and a 1260 Infinity degasser were used to deliver the sheath liquid at a rate of 10 μL/min. Separations were performed using uncoated fused-silica capillaries (Polymicro Technologies, Phoenix, AZ, USA) of a total length of 130 cm and an inner diameter of 50 μm. The BGE consisted of 1.0 M formic acid with 13% vol acetonitrile as the organic modifier (pH 1.8) under positive ion mode, and 35 mM ammonium acetate (pH 9.5) under negative ion mode for the comprehensive analysis of cationic and anionic serum metabolites, respectively.[26–28] Approximately 7 mm of the polyimide coating was removed from both the capillary inlet and the outlet using a capillary window maker (MicroSolv, Leland, NC, USA) to reduce the sample carry-over as well as the polymer swelling or degradation when in contact with organic solvent or buffer solutions containing ammonia.[29,30]

Samples were injected hydrodynamically at 50 mbar (5 kPa) for 5 s for each sample, interspaced with alternating BGE spacer plugs injected hydrodynamically for 40 s (for positive ion mode) and electrokinetically for 45 s at 30 kV (for negative ion mode) for a total of 7 discrete serum filtrates analyzed within a single run. Customized serial injection configurations using calibrant mixtures, filtrate blanks, and pooled QC samples were analyzed in MSI–CE–MS depending on the study design requirements. Prior to first use, each bare fused-silica capillary was conditioned by flushing for 15 min at 950 mbar (95 kPa), sequentially, with methanol, 1.0 M sodium hydroxide, water, and BGE. An applied voltage of +30 kV at 25 °C was used for all the CE separations under normal polarity, however a pressure gradient of 2 mbar/min was implemented for the faster elution of slow migrating anions under negative ion mode conditions.[29] Between runs, the capillary was flushed with BGE for 15 min at 950 mbar (95 kPa). The sheath liquid was comprised of 60% vol MeOH with 0.1 vol% concentrated formic acid for positive ion mode, and 80% vol MeOH for negative ion mode detection. Reference ions purine and hexakis(2,2,3,3-tetrafluoropropoxy)phosphazine (HP-921) were spiked into the sheath liquid at 0.02% vol, providing constant mass signals to enable the real-time mass calibration and to allow for the monitoring of potential ion suppression effects during separation. The instrument was operated in a GHz extended dynamic range. The Vcap and nozzle voltage were both set at 2000 V, while the fragmentor was 380 V, the skimmer was 65 V and the octopole RF was 750 V. The QTOF–MS system was operated with full-scan data acquisition over a mass range of m/z 50–1700 and an acquisition rate of 1 spectrum/s. At the beginning of each day, the QTOF–MS system was calibrated before analysis using an Agilent tune mixture to ensure residual mass ranges did not exceed 1 ppm. Additionally, daily cleaning of the CE electrode and ion source with 50% vol isopropanol was performed as preventative maintenance. A standard mixture run followed by a QC sample run with blank were analyzed at the start of each day to

equilibrate the CE–MS system and assess the system stability. Each serum filtrate sample from PAD and non-PAD participants in this study were analyzed in duplicate over two consecutive days by MSI–CE–MS under positive and negative ion modes for cationic/zwitter-ionic (e.g., amino acids) and anionic (e.g., organic acids) metabolites.

5.3.7. Lipophilic Metabolome Profiling by MSI–NACE–MS

Total serum (hydrolyzed) fatty acids were analyzed by MSI–NACE–MS under negative ion mode conditions as described elsewhere.[23,24] An Agilent 6230 time-of-flight (TOF) mass spectrometer with a coaxial sheath liquid electrospray (ESI) ionization source equipped with an Agilent G7100A CE unit was used for all experiments (Agilent Technologies Inc., Mississauga, ON, Canada). An Agilent 1260 Infinity Isocratic pump and a 1260 Infinity degasser were applied to deliver an 80:20 vol methanol:water with 0.5% vol ammonium hydroxide at a flow rate of 10 μL/min using a coaxial sheath liquid interface kit. For real-time mass correction, reference ions purine and hexakis(2,2,3,3-tetrafluoropropoxy)phosphazine (HP-921) were spiked into the sheath liquid at 0.02% vol to provide constant mass signals at *m/z* 119.0363 and *m/z* 1033.9881, which were also used to monitor for potential ion suppression/enhancement effects during separation. The nebulizer spray was set off during serial sample injection but then subsequently turned on at a low pressure of 4 psi (27.6 kPa), following voltage application with the source temperature at 300 °C and drying gas delivered at 4 L/min. The instrument was operated in a 2 GHz extended dynamic range with negative mode detection. Vcap was set at 3500 V while fragmentor was 120 V, the skimmer was 65 V and the Octopole RF was 750 V. Separations were performed on bare fused-silica capillaries with 50 μm internal diameter, 360 μm outer diameter and 95 cm total length (Polymicro Technologies Inc., Phoenix, AZ, USA), and a capillary window maker was used to remove about 7 mm of the polyimide coating on both ends of the capillary. The applied voltage was set to 30 kV

at 25 °C for the CE separations together with an applied pressure of 20 mbar (2 kPa) used during the CE separation. The nonaqueous BGE was 35 mM ammonium acetate in 70% vol acetonitrile, 15% vol methanol, 10% deionized water, and 5% vol isopropanol with an apparent pH of 9.5 adjusted by the addition of 12% vol of ammonium hydroxide. Serum extracts, pooled QC extracts, or fatty acid calibrants were injected hydrodynamically at 50 mbar (5 kPa) alternating between 5 s for each sample plug and 40 s for the BGE spacer plug for a total of seven discrete samples analyzed within one run. Prior to the first use, the capillaries were conditioned by flushing for 15 min at 950 mbar (95 kPa) sequentially with methanol, 0.1 M sodium hydroxide, deionized water, 1.0 M formic acid, deionized water then BGE for 30 min. Between runs, the capillary was flushed with BGE for 10 min at 950 mbar (95 kPa), and nonaqueous BGE and sheath liquid solutions were degassed before use.

5.3.8. Data Processing and Statistical Analyses

All the MSI–CE–MS and MSI–NACE–MS data were analyzed using Agilent Mass Hunter Workstation Software (Qualitative Analysis, version B.06.00, Agilent Technologies Inc., Mississauga, ON, Canada). Molecular features were extracted in centroid and profile modes using a 10 ppm mass window for hydrophilic and lipophilic metabolites. Polar metabolites and hydrolyzed fatty acids were annotated based on their characteristic accurate mass (*m/z*) for their protonated, $[M+H]^+$ or deprotonated, $[M-H]^-$ molecular ion together with their relative migration time (RMT), where apparent migration times were normalized to an internal standard migrating from the same sample position.[26] Extracted ion electropherograms (EIEs) were integrated after smoothing using a quadratic/cubic Savitzky Golay function (15 points) and the peak areas and migration times were transferred to Excel (Microsoft Office, Redmond, WA, USA) for the calculation of relative integrated peak areas (RPAs) and relative migration times (RMTs). Data

normalization to an internal standard (Cl-Tyr, CAPS, or 18:0-d35) improves the precision in CE by correcting for differences in sample volumes introduced in-capillary, as well as migration time drift due to the changes in electroosmotic flow between runs. Control charts for monitoring long-term method stability were derived from changes in measured responses (RPA) for recovery standards (F-Phe, NMS, or 14:0-d27) added to all the serum filtrates or extracts, including the QC samples. In most cases, serum metabolites and fatty acids from hydrolyzed lipids were identified with a high confidence (level 1)[31] using authentic standards with the exception of 8 metabolites (10% of total identified 85 metabolites; level 2) where standards were unavailable, as well as 7 unknown metabolites with only a putative molecular formula (level 4). Overall, 85 authentic metabolites were consistently detected (CV < 30%) in the majority of serum samples (>75%) when applying stringent quality control (QC) measures on 800 initially extracted features to filter redundant and spurious signals arising from contaminants, artifacts, isotopes, in-source fragments, adducts or dimers.

Least-squares linear regression analysis for external calibration curves and control charts were performed using Excel (Microsoft Office, Redmond, WA, USA). All the multivariate data analysis, including principal component analysis (PCA), hierarchical cluster analysis (HCA), correlation matrix analysis (CMA), partial least-square discriminant analysis (PLS-DA), as well as receiver operating characteristic (ROC) curves were processed using MetaboAnalyst 4.0 (wwww.metaboanalyst.ca)[32], where the data were normalized using a generalized *log*-transformation and autoscaling unless otherwise stated. PCA, HCA and CMA methods were used for data visualization (i.e., data trends, outlier detection), and comparing technical variance relative to the overall biological variance, whereas the PLS-DA was used for selecting significant serum metabolites associated with the PAD progression. Additionally, ROC curves were performed only

for the top-ranked ratiometric serum metabolites that discriminated between IC and CLTI based on the area under the curve (*AUC*).[32] Baseline characteristics of PAD participants, including the control (CON), IC, and CLTI sub-groups were compared using Fisher exact tests for categorical variables and an analysis of variance (ANOVA) for continuous variables in addition to an independent Student *t*-test for comparing the IC and CLTI sub-groups. A one-way ANOVA was used to identify the significant differences between the CON, IC, and CLTI groups for all the metabolites with a polynomial contrasts analysis to identify the linear trends associated with disease progression and Welch's F employed in the case of inequality of variance tested by Levene's homogeneity test. This was followed by planned contrasts with contrast 1 comparing CON to all the PAD cases [IC + CLTI], as well as contrast 2 comparing IC to CLTI (i.e., clinical PAD sub-groups), which was further confirmed by *post-hoc* analyses with Gabriel's and Games–Howell procedures. In addition, a two-tailed Student's *t*-test was employed on non-transformed and *log*-transformed data to compare the CLTI to IC separately in the subgroup analysis while applying Benjamini Hochberg false discovery rate (FDR) correction ($q < 0.05$) for multiple hypothesis testing. Pearson correlations were calculated to evaluate the associations between the ABI and the metabolites on non-transformed data for the majority of variables and *log*-transformed data for non-normally distributed variables which was further confirmed with partial correlations adjusting for the BMI and smoking status. Normality tests, a Shapiro–Wilk test ($p < 0.05$), Pearson and partial correlations, ANOVA, *t*-tests and nonparametric statistical analysis (Kruskal–Wallis and Mann–Whitney U test) were performed using the Statistical Package for the Social Science (IBM SPSS, version 18.0, Armonk, NY, USA). MedCalc version 12.5.0 (MedCalc Software, Ostend, Belgium) was used for boxplots, as well as Bland–Altman % difference plot and Passing–Bablok regression for the inter-laboratory method comparison of serum creatinine concentrations.

5.4. Results

5.4.1. Cohort Demographics and Clinical Characteristics

This study comprised a cohort of non-diabetic older persons ($n = 58$) with an mean age of 63 years, including PAD patients at different stages of disease progression as clinically defined by the six-stage Rutherford classification (stages 1–3: IC, and stage ≥ 4: CLTI), as well as non-PAD controls (CON). A summary of the study demographics and the other clinical characteristics for the participants is summarized in **Table 5.1**. There were no significant differences in age, sex, body mass index (BMI), glycated hemoglobin, as well as leukocyte and platelet counts between the three patient sub-groups. The IC and CLTI patients presented with a higher incidence of dyslipidemia (>80%), hypertension (>65%), coronary artery disease (>40%), statin/antiplatelet medication use (>80%), and notably smoking (~95%) than the controls. However, all patients with a 6-month history of acute coronary syndrome, heart failure, or uncontrolled arrhythmia were excluded in this study eliminating active coronary symptoms at the time of blood sampling. Importantly, when analyzing the differences between the two PAD subgroups, IC and CLTI, the patients were closely matched with no statistical differences ($p > 0.05$) in anthropometric properties, comorbidities, or medication use with the exception for ABI (<0.90; $p = 2.36 \times 10^{-9}$), which was lower in the CLTI (mean ABI = 0.38) as compared to the IC group (mean ABI = 0.57) and non-PAD controls (mean ABI = 1.08). Moreover, there were only two PAD patients diagnosed with renal insufficiency (1 CLTI; 1 IC), thus the vast majority of participants had normal kidney function at the time of recruitment.

Table 5.1. Baseline patient demographics and clinical characteristics for two peripheral artery disease (PAD) sub-groups (intermittent claudication (IC), chronic limb-threatening ischemia (CLTI)) and non-PAD controls (CON).

Parameter	CON (*n* = 20)	IC (*n* = 20)	CLTI (*n* = 18)	*p*-value
Rutherford stage	-	1-3 [2.75 ± 0.4]	≥ 4 [4.11 ± 0.3]	-
Walking distance (m)	> 1000	530	< 160	-
ABI	1.08 ± 0.09	0.57 ± 0.08	0.38 ± 0.07	3.06×10^{-33};2.4×10^{-9}
Age (years)	62.6 ± 6.6	61.0 ± 7.4	65.2 ± 5.6	0.151; 0.055
BMI (kg/m^2)	26.6 ± 2.5	24.3 ± 3.0	24.9 ± 3.6	0.061; 0.631
HbA1c (%)	5.75 ± 0.51	5.98 ± 0.50	5.58 ± 0.99	0.217; 0.124
Leukocytes	6.6 ± 2.2	7.8 ± 2.5	8.4 ± 3.4	0.156; 0.534
Platelets	251 ± 76	244 ± 64	209 ± 65	0.152; 0.118
Males (%)	50 (10/20)	55 (11/20)	72 (13/18)	0.401; 0.224
Smoking (%)	55 (11/20)	95 (19/20)	94 (17/18)	0.002; 0.730
Diabetes mellitus (%)	0	0	0	-
Hypertension (%)	40 (8/20)	65 (13/20)	72 (13/18)	0.099; 0.450
Hyperlipidaemia (%)	39 (7/20)	85 (17/20)	83 (15/18)	0.001; 0.616
Renal insufficiency (%)	0 (0/20)	5 (1/20)	6 (1/18)	0.76; 0.730
Coronary artery disease (%)	0 (0/20)	40 (8/20)	61 (11/18)	0.001; 0.165
Statin use (%)	30 (6/20)	80 (16/20)	100 (18/18)	<0.001; 0.066
Antiplatelet use (%)	50 (10/20)	100 (20/20)	100 (18/18)	<0.001; -

Data shown as the mean ± standard deviation for continuous variables and % (number of cases/total) for categorical variables. p-value represents the overall difference between the three groups where significant differences are observed for ($p < 0.05$) calculated Fisher exact tests for categorical variables and ANOVA for continuous variables; followed by p-value for PAD subgroup comparison between CLTI and IC calculated using independent samples Student's t-test or non-parametric Mann–Whitney U test (only for body mass index , BMI). Smoking status reflects numbers of past/current smokers.

5.4.2. The Serum Metabolome of PAD Patients

Nontargeted analysis of the serum metabolome was performed by MSI–CE–MS and MSI–NACE–MS under three configurations for polar/hydrophilic and non-polar/lipophilic metabolites, respectively. Each run comprised a serial injection of seven serum samples for the CON, IC and CLTI patient sub-groups that were randomly analyzed in duplicate (except for fatty acids that used a single analysis), together with a pooled serum sample as a QC, to monitor for the long-term

signal drift as shown in **Figure 5.1A**. Serum metabolites were authenticated based on their characteristic accurate mass and relative migration time under positive or negative ion mode (*m*/*z*:RMT:mode) when using multiplexed separations together with temporal signal pattern recognition.[33] For instance, only serum metabolites measured with adequate precision (CV < 30%, n = 6) with no signal detected in blank filtrate/extract[26] were initially selected after rejecting spurious signals, background artifacts, as well as redundant ions derived from the same metabolite (e.g., in-source fragments, isotopic signals, and adducts) that constitute a majority of signals (> 90%) in ESI–MS.[34] Additionally, the identity of most serum metabolites was confirmed by spiking with authentic standards (level 1) based on their co-migration (RMT < 1%) with low mass error (< 5 ppm), which was also used for evaluating method accuracy (*i.e.*, spike/recovery) (Figure 1B), and quantification using external calibration curves (**Figure 5.1C**). Otherwise, seven serum metabolites with unknown chemical structures were annotated based on their most likely molecular formula (level 4). Moreover, serum metabolites were reported only if they satisfied two additional inclusion criteria to reduce false discoveries in metabolomics, namely they were measured with adequate precision throughout the entire study (CV < 30% from QC runs) and detected with high frequency (> 75%) in all serum samples. This iterative process of data filtering and selection culminated in a final data matrix of 85 serum metabolites reliably measured in the majority of the samples in this cohort, including 42 hydrophilic cations, 18 hydrophilic anions and 25 lipophilic anions (**Table S5.1**). Overall, the coverage of the serum metabolome when using three configurations in MSI-(NA)CE-MS, included a diverse range of circulating metabolites ranging from amino acids, amines, organic acids, long-chain fatty acids, ketone bodies, hexoses, and osmolytes/uremic toxins.

As expected, the overall biological variance of the serum metabolome (median CV = 39%, n = 58) was considerably larger than the technical precision of the method based on the repeated analysis of a QC sample in each run (median CV = 14%, n = 17–22) as shown in the PCA 2D scores plots in **Figure 5.2A,B**. Moreover, control charts for the three recovery standards used in MSI–CE–MS (10 μM F-Phe and NMS) and MSI–NACE–MS (50 μM 14:0-d27) added to all serum samples, prior to ultrafiltration or MTBE extraction, further demonstrate the acceptable intermediate precision (mean CV < 15%) with few outliers (< 2%, n = 57 total runs) exceeding warning limits (± 2s) (**Figure 5.2C–E**). An inter-laboratory method comparison of the serum creatinine concentrations measured independently from 56 participants was also performed by MSI–CE–MS relative to the Jaffé colorimetric method, which was used for the estimation of the GFR for patients as an indicator of kidney function.[20] In this case, a Bland–Altman % difference plot confirms a good mutual agreement for serum creatinine determination by both methods with a mean bias of –5.6% (**Figure 5.2F**). In addition, there was a random distribution in the data with few outliers (4 of 56) exceeding the agreement limits (±2 s). Similarly, a Passing–Bablok regression analysis (**Figure 5.2G**) reveals no statistically significant difference from the line of equality (p = 0.93) with a modest positive slope of 1.17. Only two PAD patients (1 IC and 1 CLTI) out of the 58 study participants were diagnosed with renal insufficiency. Furthermore, no patients had mean serum creatinine concentrations exceeding a clinically relevant cutoff value (>150 μM) for older persons, that is indicative of renal failure[35] in the absence of urinary creatinine clearance for the calculation of the GFR.

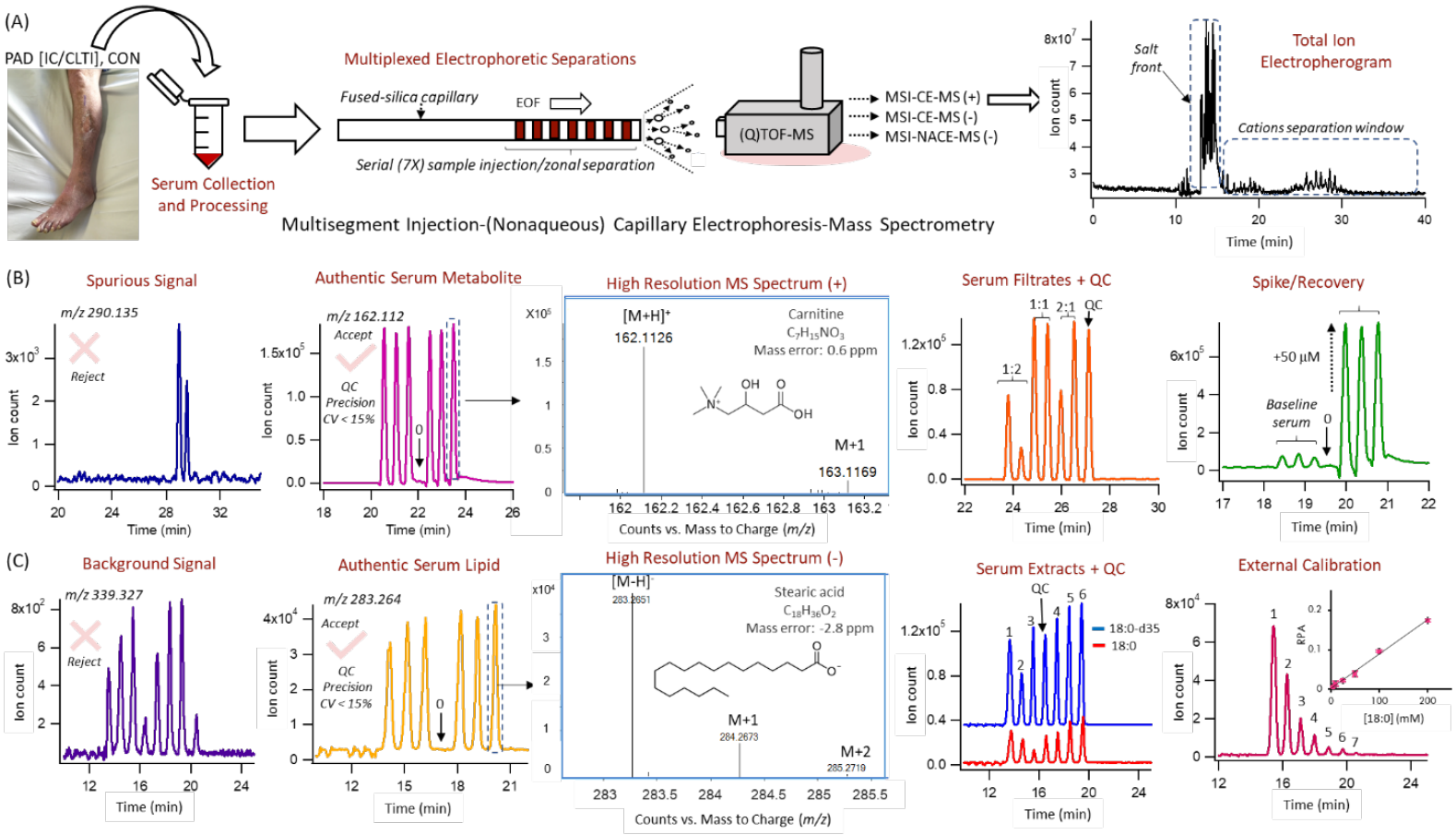


Figure 5.1. (**A**) Nontargeted metabolite profiling of serum samples from PAD patients using MSI–(NA)CE–MS under three different configurations, where the black trace depicts a total electropherogram using aqueous BGE conditions with positive mode detection. This multiplexed separation method relies on a serial injection of 6 serum samples and a quality control (QC) within each run to enhance sample throughput and data fidelity when using temporal signal pattern recognition. A rigorous data-filtering process allows for the reliable authentication of metabolites based on their accurate mass (*m/z*), which are measured in 6 replicate serum samples with acceptable precision (CV < 15%) with no background signal in blank (0) as shown for (**B**) carnitine and (**C**) stearic acid (18:0) under positive and negative ion mode detection, respectively. Various serial injection configurations are illustrated, such as a replicate injection of QC samples with a blank extract to filter out spurious and background signals in ESI–MS, the assessment of technical precision and potential sample carry-over for authentic serum metabolites, a randomized analysis of 6 serum samples from individual PAD patients along with a QC, a spike-recovery study to evaluate accuracy and confirm the identity based on co-migration, and a 7-point calibration curve for serum metabolite quantification.

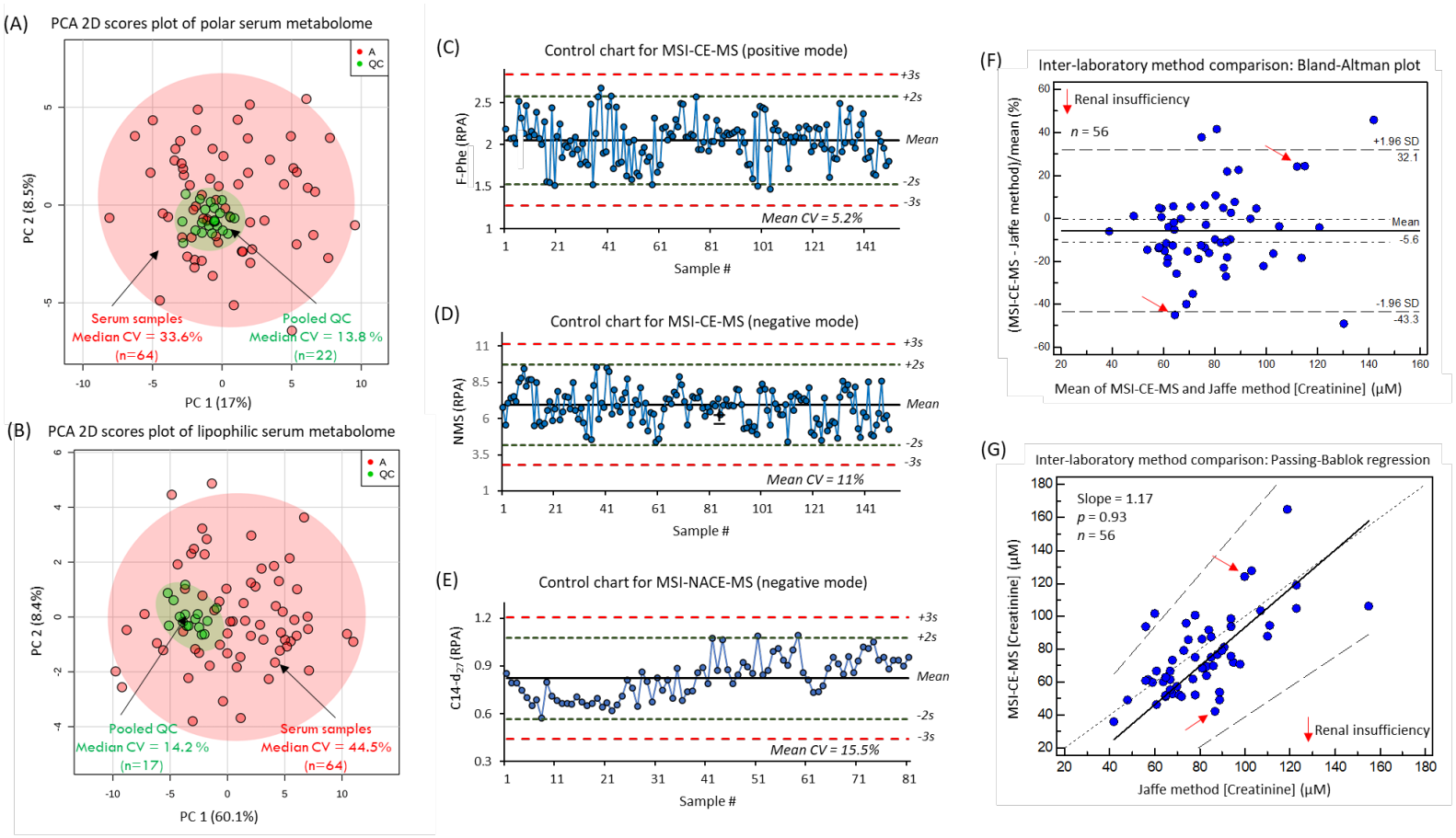


Figure 5.2. Two-dimensional scores plot from the principal component analysis (PCA) of the *glog*-transformed and autoscaled serum metabolome data used to compare the intersubject biological variance relative to the technical variance from the repeat analysis of pooled serum QC samples for (**A**) the hydrophilic serum metabolome and (**B**) the lipophilic serum metabolome. Control charts for (**C**) the recovery standard (F-Phe) measured under aqueous positive ion mode, for (**D**) the recovery standard 2-napthalenesulfonic acid (NMS) measured under aqueous negative ion mode, and for (**E**) the recovery standard 14-d27 measured under nonaqueous negative ion mode for all serum and QC samples demonstrate acceptable precision ($CV = 5.2–15.5\%$) with no outliers exceeding the warning limits (±3 s). (**F**) Bland–Altman percent difference plot for comparing the mutual agreement between the serum creatinine concentrations measured independently by MSI–CE–MS and Jaffé colorimetric methods at two different laboratories from 56 participants. Overall, the data are randomly distributed with a modest mean bias of –5.6% with four outliers outside agreement limits. (**G**) A Passing–Bablok regression analysis demonstrates no significant deviation (dotted lines; 95% confidence interval) from the line of equality ($p > 0.05$) with a slope of 1.17 (black line; regression line).

Table 5.2. Top-ranked serum metabolites showing significant changes reflecting disease progression when comparing non-PAD controls to PAD cases, as well as IC to CLTI patient sub-groups when using a one-way ANOVA with planned contrasts and their correlation to ABI.

Metabolite ID	*m/z*:RMT:mode	F-value	*p*-value overall	*p*-value linear[a]	Effect size[b]	*p*-value Contrast 1 PAD:CON[c]	FC PAD:CON[c]	*p*-value Contrast 2 CLTI:IC[d]	FC CLTI:IC[d]	*r* correlation to ABI[e]	*p*-value for *r*
Creatine HMDB000064	132.077:0.745:p	6.02*	0.006	0.002	0.422	0.008	0.65	0.097	0.75	0.44	0.001
Histidine HMDB000117	156.077:0.620:p	5.54	0.006	0.003	0.410	0.002	0.85	0.435	0.95	0.38	0.004
Phenylacetylglutamine HMDB0006344	263.104:0.899:n	5.43*	0.009	0.017	0.319	0.030	1.89	0.880	0.94	-0.30	0.020[f]
Lysine HMDB0000182	147.113:0.580:p	4.23	0.020	0.005	0.365	0.014	0.85	0.137	0.88	0.35	0.007[f]
Tyrosine HMDB0000158	182.080:0.9564:p	3.53	0.036	0.014	0.338	0.012	0.81	0.520	0.94	0.34	0.008[f]
Monomethylarginine HMDB0029416	189.134:0.606:p	3.19	0.049	0.022	0.332	0.005	0.74	0.378	0.93	0.32	0.014[f]
Oxo-proline HMDB0000267	128.035:1.137:n	2.89	0.054	0.028	0.316	0.021	0.69	0.638	0.86	0.34	0.013
Creatinine	Jaffé method	6.57	0.003	0.002	0.446	0.055	1.27	0.003	1.25	-0.31	0.020
Creatinine HMDB0000562	114.066:0.614:p	6.14	0.004	0.011	0.428	0.271	1.07	0.001	1.30	-0.30	0.035
Linoleic acid (18:2*n*-6) HMDB0000673	279.233:1.0189:l	4.96	0.010	0.007	0.390	0.101	0.78	0.009	0.95	0.24	0.066
Eicosadienoic acid (20:2) HMDB0005060	307.265:0.994:l	4.30	0.018	0.010	0.368	0.089	0.79	0.019	1	0.25	0.061
Nervonic acid (24:1) HMDB0002368	365.342:0.947:l	3.96	0.025	0.012	0.356	0.095	0.86	0.026	0.75	0.23	0.085
Phe/Tyr	-	3.67	0.032	0.026	0.343	0.22	1.10	0.018	1.15	-0.25	0.055
Behenic acid (22:0) HMDB0000944	339.327:0.969:l	3.49	0.038	0.026	0.336	0.187	0.91	0.024	0.75	0.22	0.105
Lignoceric acid (23:0) HMDB0002003	367.358:0.942:l	3.45	0.037	0.015	0.334	0.088	0.84	0.045	0.75	0.26	0.050
Cystine HMDB0000574	241.030:0.933:p	3.15*	0.050	0.028	0.377	0.305	1.07	0.019	1.29	-0.24	0.065

*[a] p-value for a linear trend when applying polynomial contrasts analysis; [b] Effect size calculated based on eta-squared; [c] p-value and mean FC for planned contrast 1 comparing PAD to CON; [d] p-value and mean FC for planned contrast 2 comparing CLTI to IC; [e] Pearson correlation (r) on normally-distributed non-transformed or log-transformed serum metabolites to the ABI, after adjusting for BMI and smoking. [f] not significant after adjusting for BMI and smoking. * Welch's F-test employed in case of inequality of variance tested by Levene's homogeneity test. Abbreviations correspond to RMT: relative migration time, p: positive aqueous mode, n: negative aqueous mode, l: negative nonaqueous mode, CON: non-PAD controls, PAD: peripheral artery disease [IC + CLTI], CLTI: chronic limb-threatening ischemia, IC: matched intermittent claudication, FC: fold-change, ABI: ankle–brachial index.*

5.4.3. Differentiating Serum Metabolites of PAD Progression

Figure 5.3A depicts a 2D heat map with HCA that summarizes the overall data structure involving 85 serum metabolites consistently measured in all 58 study participants (with only 0.7% missing values), the including CON, IC, and CLTI sub-groups. **Figure 5.3B** depicts a 2D scores plot when using PLS-DA for the differentiation of the serum metabolic phenotype in CLTI (n = 18) from the IC patients (n = 20), as well as CON (n = 20) based on *glog*-transformed and autoscaled data. Figure 3C summarizes the 10 top-ranked serum metabolites largely responsible for group separation along the first principal component (variable importance in projection, VIP > 1.5), including creatine, lysine (Lys), histidine (His), monomethylarginine (MMA), tyrosine (Tyr), phenylacetylglutamine (PAG), and several long-chain fatty acids from hydrolyzed lipids (18:2, 20:2, 23:0, 24:1). **Figure 5.3D** depicts a correlation matrix for the top-ranked serum metabolites associated with PAD progression, which highlights two major clusters of strongly co-linear serum metabolites not correlated to PAG, namely amino acids ($r > 0.60$ with Lys), and fatty acids ($r > 0.70$ with linoleic acid, 18:2*n*-6). Univariate statistical analysis was also performed to confirm the significance of the serum metabolites associated with PAD progression when using a one-way ANOVA with planned contrasts as summarized in **Table 5.2**. Overall, 14 serum metabolites were determined as significant ($p < 0.05$) when using a linear contrast analysis model across all three categories (i.e., CON–IC–CLTI), including 10 metabolites identified by the PLS-DA model, as well as four additional serum metabolites, namely oxo-proline (oxo-Pro), behenic acid (23:0), creatinine, and cystine. Moreover, the phenylalanine:tyrosine ratio (Phe/Tyr) was higher in CLTI as compared to IC, which is an indicator of inflammation in PAD.[16] Importantly, most serum metabolites exhibited a linear change in their concentrations (or RPAs) as a function of PAD status ($p < 0.05$). Overall, discrimination between the major patient sub-groups follows a linear trend

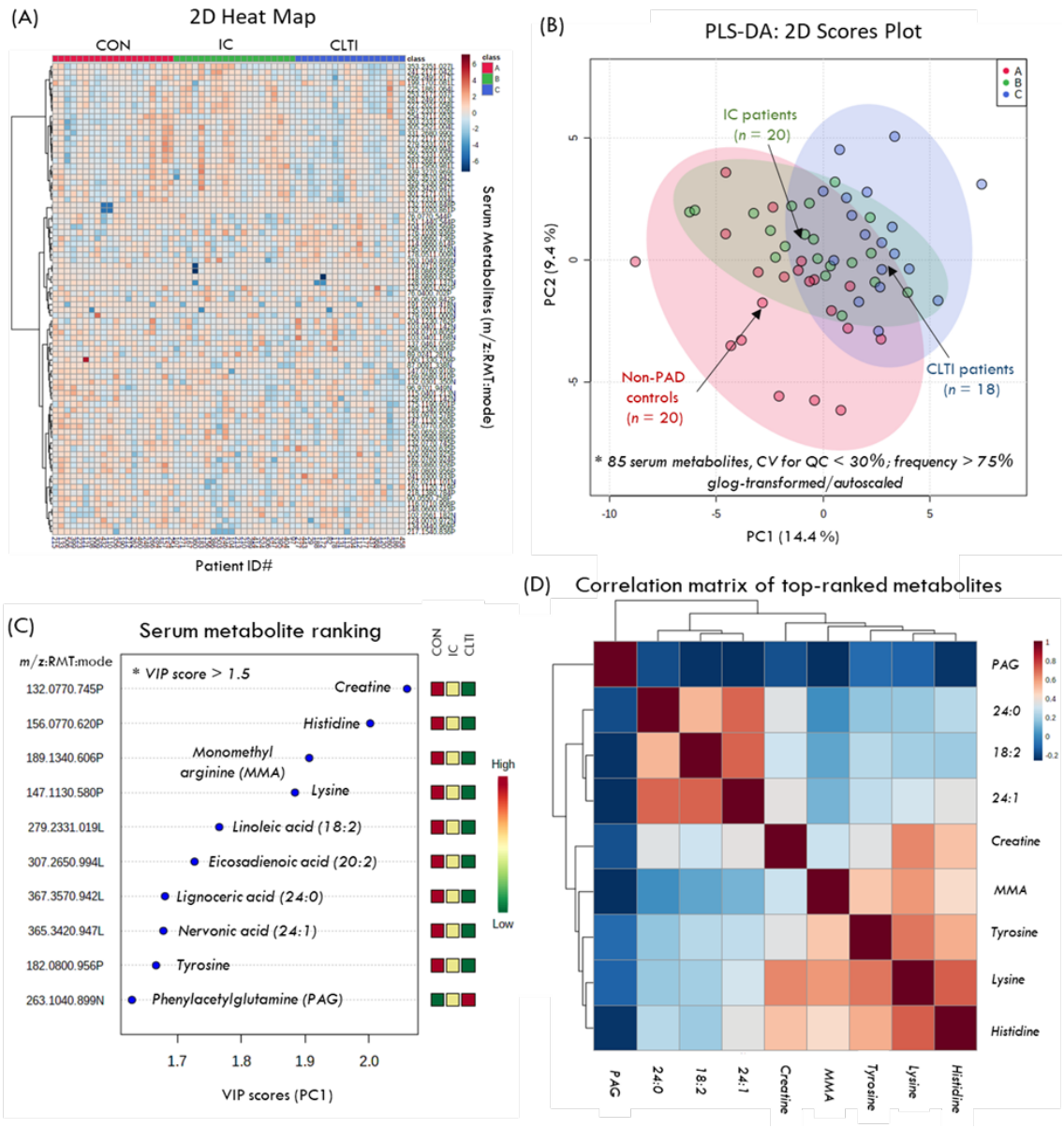


Figure 5.3. (**A**) Two-dimensional heat map of the serum metabolome of PAD patient sub-groups (IC, CLTI) and non-PAD controls (CON) that summarizes the overall data structure of this study. (**B**) Two-dimensional scores plot using a partial least-square discriminant analysis (PLS-DA) model to differentiate the metabolic phenotype of late-stage CLTI (n = 18) from early onset IC (n = 20) cases as compared to age and sex-matched CON (n = 20) based on 85 serum metabolites/lipids. (**C**) Ten top-ranked serum metabolites that differentiate PAD patients and non-PAD controls based on a variable importance in projection (VIP scores > 1.5). (**D**) Correlation matrix depicts two main clusters of circulating metabolites associated with PAD, including circulating amino acids/amines strongly correlated (r~0.70) to lysine (histidine, tyrosine, MMA: monomethylarginine, and creatine), as well as long-chain fatty acids (18:2, 20:2, 24:0, 24:1) unlike serum phenylacetylglutamine (PAG).

where IC clusters in the middle between the CON and the more severe CLTI cases; these results are analogous to trends depicted in the PLS-DA model (**Figure 5.3B**).

Further analysis was next performed to identify the specific between-group differences without inflating type I error by using an ANOVA with two discrete planned contrasts, namely CON–PAD (contrast 1) and CLTI–IC (contrast 2). The box–whisker plots in **Figure 5.4** show that circulating concentrations of creatine, His, oxo-Pro, Lys, Tyr and MMA were higher in non-PAD controls compared to PAD patients [IC + CLTI] unlike PAG (contrast 1, $p < 0.05$). Furthermore, serum creatinine, cystine, and Phe/Tyr were elevated in CLTI patients as compared to IC cases, whereas a series of circulating fatty acids (18:2, 22:0, 20:2, 24:0, 24:1) display the opposite trend. As expected, a similar outcome was found for serum creatinine independently measured by the Jaffé colorimetric method, confirming good mutual agreement with MSI–CE–MS results. An additional subgroup analysis using a two-tailed Student's *t*-test was used to better evaluate changes associated with PAD progression since IC and CLTI patients were closely matched in terms of age, sex, BMI, smoking, co-morbidities, and medication use (**Table 5.1**). **Table 5.3** summarizes 16 serum metabolites that were differentially expressed ($p < 0.05$) in the two PAD sub-groups, including 11 metabolites after a FDR adjustment ($q < 0.05$). In this case, serum creatinine, carnitine (C0), propionylcarnitine (C3), cystine, Phe/Tyr, and trimethylamine-*N*-oxide (TMAO) were elevated in CLTI as compared to IC cases, in contrast to several circulating fatty acids, including saturated/odd-chain fatty acids (15:0, 16:0, 17:0, 18:0). All putative serum biomarkers of PAD progression were also correlated with the ABI, notably stearic acid (18:0, $r = 0.51$, $p = 0.001$), as well as C0 and cystine ($r = –0.48$, $p = 0.002$).

Lastly, receiver operating characteristic (ROC) curve analysis was also performed on all the serum metabolites and their ratios to demonstrate the reliable discrimination of severe CLTI from lower risk IC patients. **Figure 5.5** shows two top-ranked ratiometric biomarkers in serum with an area under the curve or *AUC*~0.87 along with their 95% confidence intervals (0.73–0.98), namely

Table 5.3. Top-ranked serum metabolites comparing IC (n = 20) to CLTI (n = 18) using student's t-test and their correlation to ABI.

Metabolite ID	*m/z*:RMT:mode	*p*-value*	FDR *q*-value[a]	FC [b] (CLTI/IC)	*r* correlation to ABI [c]	*p*-value for *r*
Stearic acid; 18:0 HMDB0000827	283.264:1.005:l	0.001	0.014	0.72	0.51	0.001
Linoleic acid; 18:2*n*-6 HMDB0000673	279.233:1.019:l	0.003	0.028	0.68	0.39	0.016
Heptadecanoic acid; 17:0 HMDB0002259	269.249:1.030:l	0.003	0.029	0.72	0.43	0.007
Palmitic acid; 16:0 HMDB0000220	255.233:1.030:l	0.004	0.030	0.73	0.37	0.024
Creatinine HMDB0000562	114.066:0.614:p	0.004	0.031	1.30	-0.45	0.004
Carnitine HMDB0000062	162.112:0.719:p	0.005	0.031	1.28	-0.48	0.002
Oleic acid; 18:1*n*-9 HMDB0000207	281.249:1.013:l	0.005	0.031	0.71	-0.04	0.756
Heptadecenoic acid; 17:1*n*-9 HMDB0062437	267.233:1.026:l	0.008	0.043	0.73	-0.01	0.961
Propionylcarnitine HMDB0000824	218.138:0.784:p	0.008	0.043	1.37	0.09	0.507
Eicosadienoic acid; 20:2*n*-6 HMDB0005060	307.265:0.994:l	0.009	0.047	0.72	0.37	0.023
Pentadecanoic acid; 15:0 HMDB0000826	241.217:1.042:l	0.010	0.047	0.66	0.33	0.044
Cystine HMDB0000574	241.0299:0.933:p	0.014	0.061	1.29	-0.48	0.002
Arachidic acid; 20:0*n*-3 HMDB0002212	311.296:0.981:l	0.015	0.061	0.68	0.39	0.015
Trimethylamine-*N*-oxide HMDB0000925	76.077:0.544:p	0.019	0.080	1.60	-0.44	0.005
Nervonic acid; 24:1 HMDB0002368	365.342:0.947:l	0.024	0.091	0.75	0.29	0.083
Phe/Tyr ratio	-	0.022	0.103	1.19	-0.33	0.041

Two-tailed exact p-value on log-transformed serum metabolome data. [a] FDR correction for multiple hypothesis testing; [b] mean fold-change (FC) ratio when comparing relative ion response ratio for each metabolite as a ratio of CLTI/IC; [c] Pearson correlation on normally-distributed non-transformed or log-transformed data.

18:0/C0 and arginine (Arg)/C3. These two ratiometric biomarkers also exhibited a strong linear correlation with the ABI from PAD patients (r = 0.54 – 0.59, $p < 0.001$, n = 38). This is relevant for biomarker discovery in pilot studies to anchor aberrant metabolism to a validated physiological measure used for risk stratification of symptomatic PAD patients, which reflects increasing blockage of peripheral blood flow to the lower limbs.

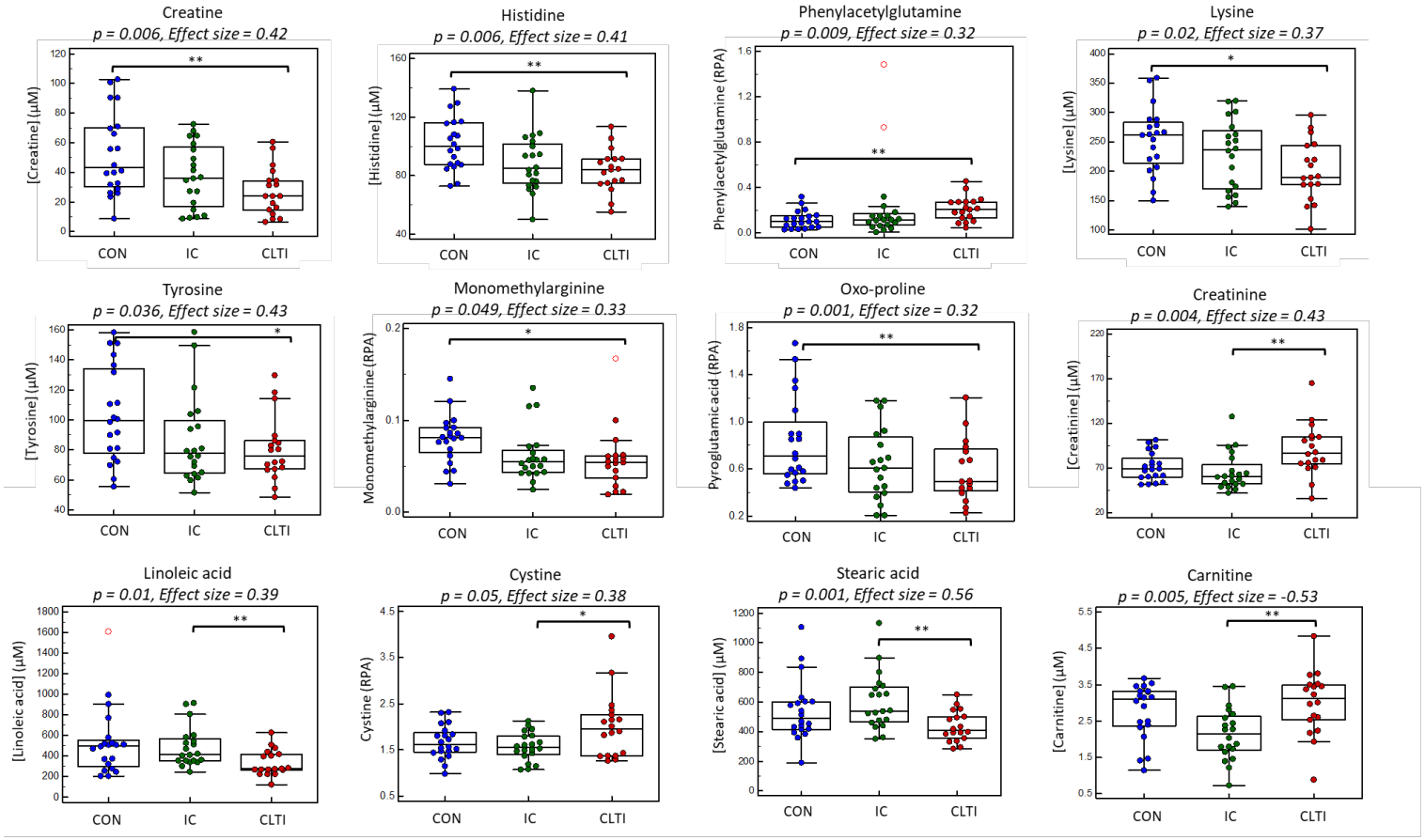


Figure 5.4. Box–whisker plots illustrating differences in the twelve top-ranked serum metabolites compared between the chronic limb-threatening ischemia (CLTI) patients ($n = 18$), matched intermittent claudication (IC) patients ($n = 20$) and the non-PAD controls ($n = 20$). A one-way ANOVA test was performed to compare the means and identify significant changes in circulating metabolite concentrations between the three groups, as summarized in Table 2, where a polynomial contrasts analysis depicts most metabolites having a significant linear trend, proportional with disease progression. Planned contrasts were conducted by comparing non-PAD controls to PAD cases (IC + CLTI) (contrast 1; long bracket) followed by comparing IC to CLTI (contrast 2; short bracket) reflecting disease status and PAD progression, respectively where test significance is denoted as * $p < 0.05$ and ** $p < 0.01$. Serum stearic acid and carnitine were different between CLTI and IC when using an unpaired Student's *t*-test after a FDR adjustment ($q < 0.05$). Serum metabolites responses in terms of absolute concentrations (mM), or relative peak area (RPA) if standards were not available

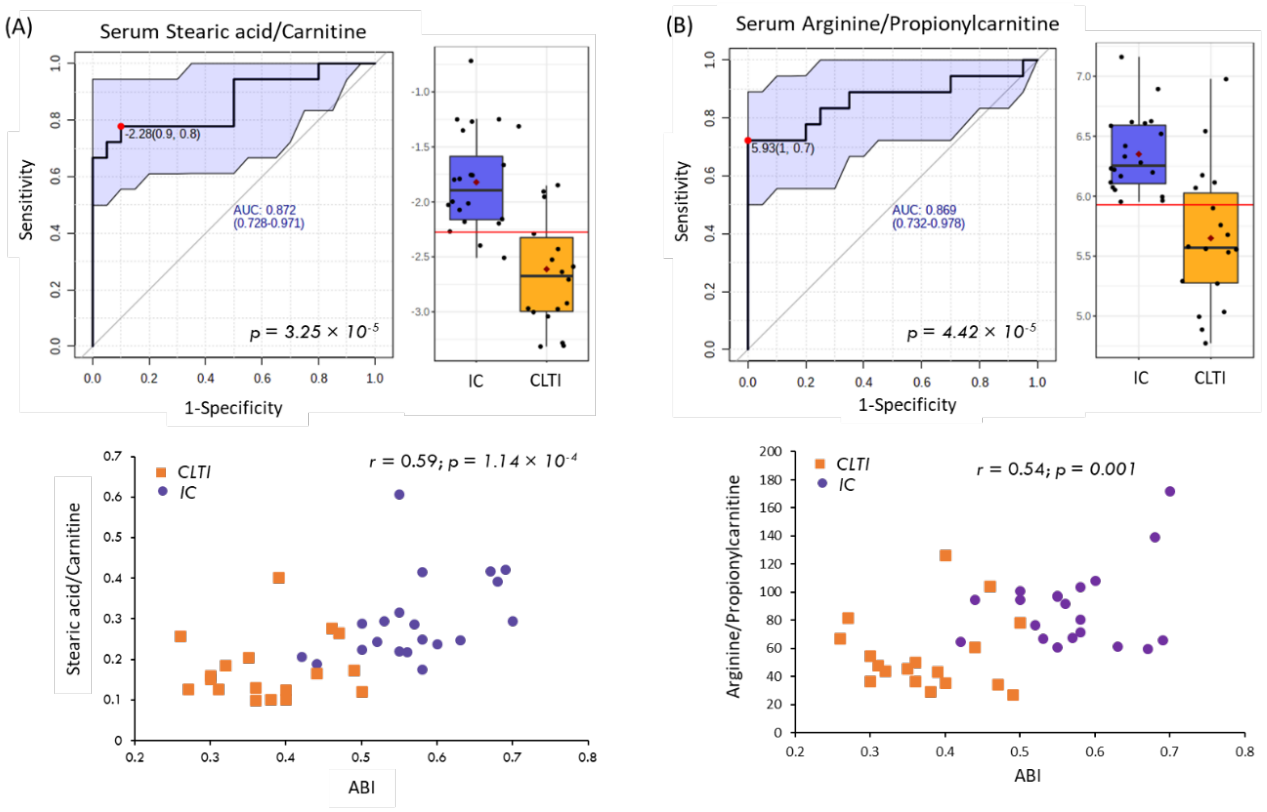


Figure 5.5. Upper panels show the receiver operating characteristic (ROC) curves and their corresponding box–whisker plots for the two top-ranked serum biomarker ratios used for discriminating chronic limb-threatening ischemia (CLTI, $n = 18$) from intermittent claudication (IC, $n = 20$) patients, including (**A**) stearic acid/carnitine (18:0/C0) and (**B**) arginine/propionylcarnitine (Arg/C3). Ratiometric ROC curves depict the area under the curve (*AUC*) and their 95% confidence intervals (blue shaded area). Lower panels depict the linear relationship of the serum biomarkers of PAD disease progression as a function of abnormal ankle–brachial index (ABI < 0.90) measurements with moderately strong Pearson correlation coefficients ($r > 0.50$; $p < 0.002$).

5.5. Discussion

The lack of PAD awareness among physicians continues to pose a major diagnostic challenge due to its variable clinical manifestations and unpredictable aggressive progression. An improved screening strategy for the early detection of PAD in asymptomatic patients is required, given the poor sensitivity of specialized Doppler methods for ABI assessment at early stages of atherosclerosis in peripheral tissue.[36] In this case, prognostic biomarkers will augment well established traditional risk factors (*e.g.*, smoking, diabetes, age, hyperlipidemia, renal dysfunction) while allowing for the reliable diagnosis of PAD, especially in high-risk patients with calcified

vascular tissue not suitable for the ABI. Regrettably, blood-based protein biomarkers (e.g., inflammatory cytokines, C-reactive protein) have yet to be clinically validated for the routine screening of PAD, predict disease progression, and/or monitor the treatment responses of patients.[6] In this pilot study, we identified a panel of serum metabolites associated with PAD progression, which is important given the poor survivorship of patients with CLTI following invasive surgical interventions, including revascularization procedures and limb loss from amputation.[2]

Untargeted metabolite profiling was performed on fasting serum samples collected from PAD patients, including well matched CLTI and IC cases, as well as non-PAD controls. Overall, 85 serum metabolites were consistently detected in most serum samples with good technical precision when using three different configurations in MSI–(NA)CE–MS. This multiplexed separation method takes advantage of unique data workflows and iterative data filtering processes to authenticate metabolites while also applying stringent QC to reduce false discoveries. We identified a panel of serum metabolites that differentiated CLTI from IC patients, as well as PAD from non-PAD controls. Serum creatine was found to be one of the most significant metabolites lower in PAD as compared to CON ($F = 6.0$, $p = 0.006$, effect size = 0.42), which also exhibited a linear change in concentration ($p = 0.002$) across the three patient sub-groups. Creatine is derived from dietary protein, as well as endogenously produced in the liver that is actively transported within muscle tissue, where it accumulates against a large concentration gradient. Creatine plays a key role in energy metabolism within skeletal muscle tissue by conversion into phosphocreatine via creatine kinase, which is an abundant phosphagen required for ATP regeneration during active muscle contractions.[37] As a result, lower serum creatine concentrations in CLTI likely reflects inadequate muscle energy storage and a loss of function with more advanced stages of PAD progression.[22]

Additionally, lower antioxidant capacity within ischemic muscle tissue likely plays a role in the 0.83-fold reduction of serum His concentrations in PAD as compared to CON ($F = 5.5$, $p = 0.006$, effect size = 0.41). The antioxidant and anti-inflammatory properties of His, whether in free, peptide or protein form (e.g., intramuscular carnosine), have been attributed to the capacity of the imidazole ring to scavenge hydroxyl radicals and singlet oxygen species.[38] Our findings suggest that lower circulating His may contribute to a greater susceptibility to oxidative stress in PAD in accordance with reports of histidine protection of cardiac and vascular tissue injury.[16,38] Similar to trends identified in metabolite trajectories for creatine and His, PAD patients also exhibited a decrease in serum Lys, MMA and oxo-Pro concentrations relative to non-PAD controls (Table 2, Figures 3 and 4). For example, the essential amino acid Lys acts as an exogenous inhibitor of plasmin-induced proteolysis that contributes to matrix remodeling, continued connective tissue degradation in the vascular wall and the formation of atherosclerotic lesions.[39] Since oxo-Pro is a degradation product of the major intracellular antioxidant glutathione,[22,40] the 0.63-fold decrease in serum oxo-Pro in PAD cases as compared to non-PAD controls may indicate increased glutathione depletion and the activation of the glutathione salvage pathway that is a hallmark of deleterious oxidative stress.[41,42] This is consistent with elevated serum cystine in CTLI as compared to IC ($p = 0.014$), which is a biomarker of systemic oxidative stress prevalent in PAD, and is associated with adverse clinical outcomes independent yet synergistic to inflammation.[43] An opposing trend was found for serum PAG with a 1.9-fold higher concentration in PAD as compared to CON ($F = 5.4$, $p = 0.009$, effect size = 0.32). This circulatory uremic toxin is an independent risk factor for cardiovascular disease, where elevated serum PAG may serve as a predictor of overall mortality in high-risk patients with chronic kidney disease,[44] but it has not been reported in PAD patients without renal dysfunction as was the case for most participants (> 96%) in our study.

Plasma metabolomic profiles of near-term death PAD patients was previously reported to be strongly correlated with lipid metabolism based on differentiating chemical signals from fatty acid acyl chain protons in lipoprotein species.[14] However, as this study used NMR, the exact identity of fatty acids associated with PAD progression was not fully elucidated. Our results confirmed aberrant circulating lipid metabolism in terms of several serum fatty acids (total hydrolyzed) that were highly co-linear (**Figure 5.3D**) and consistently reduced in PAD patients relative to CON participants (**Table 5.2**), notably when comparing CLTI to IC cases in PAD patient sub-groups (**Table 5.3**). The essential omega-6 fatty acid, linoleic acid (18:2*n*-6), derived largely from dietary vegetable oils and nuts, has been the subject of conflicting reports regarding its putative pro- or anti-inflammatory roles.[45] A recent meta-analysis of 30 cohort studies supports that higher blood or adipose tissue levels of 18:2*n*-6 are associated with lower, and not higher, risk of cardiovascular disease, in agreement with our results.[46] Besides their roles in cell membrane phospholipid structure and energy metabolism, polyunsaturated fatty acids also serve as precursors of bioactive mediators of inflammation that have also been implicated in PAD pathophysiology.[47] For example, a randomized clinical trial involving 7435 participants reported that a Mediterranean diet supplemented with either extra-virgin olive oil or nuts was associated with a lower risk of PAD incidence as compared to a low-fat counseled control group.[48] Similarly, odd-chain/saturated fatty acids, 15:0 and 17:0, have been associated with lower ischemic heart disease and type 2 diabetes incidence,[49] which are derived from dietary intake of full-fat dairy, as well as fiber and gut microflora activity.[24] Overall, serum 18:0 was the most significantly depleted in CLTI as compared to IC patients after a FDR adjustment ($q < 0.05$), that was also moderately correlated with the ABI ($r = 0.51$, $p = 0.01$). Mounting evidence supports that increasing circulating 18:0 lipids are associated with reduced blood pressure and improved heart function, which also promotes

mitochondrial fusion with greater beta-oxidation activity[50] in late-stage PAD patients with severe muscle ischemia. Further work is warranted to confirm the exact role of serum fatty acids that are depleted with PAD progression, which interestingly did not include omega-3 fatty acids, such as eicosapentaenoic acid (20:5*n*-3) and docosahexaenoic acid (22:6*n*-3) that are associated with dietary intake of oily fish.[24]

Due to potential confounding when comparing PAD patients to non-PAD controls, several serum metabolites were also differentially expressed (most satisfying a FDR adjustment, $q < 0.05$) in-well matched CI and CLTI patients (**Table 3**). A mean 1.3-fold increase in serum carnitine ($p = 0.005$) likely reflects severe myofiber degeneration occurring in CLTI as compared to a less acute muscle atrophy in IC due to its key role in mediating the transport of long-chain fatty acids within the mitochondria for beta-oxidation.[51] Moreover, TMAO generated from dietary phosphatidylcholine and carnitine via the action of gut microbiota has been widely reported as a pro-atherogenic metabolite in CAD, and circulating TMAO predicts higher risk of 5 year all-cause PAD mortality.[52,53] A higher degree of atherosclerosis associated with CLTI, as compared to IC, is consistent with elevated serum TMAO concentrations ($p = 0.019$) in the PAD sub-group analysis as reflected by their lower ABI measurements (**Table 5.1**). In addition, creatine deficiency in circulation indicates a state of a diminished energy supply among CLTI patients, possibly contributing to disease progression. Creatine undergoes spontaneous irreversible non-enzymatic degradation within skeletal muscle tissue into creatinine, which diffuses out of the muscle tissue into the circulation to be excreted by glomerular filtration through the kidneys. In our study, a mean 1.3-fold increase in serum creatinine was measured in CLTI as compared to IC patients ($p = 0.004$), which was replicated independently on two different instrumental platforms/laboratories when using the Jaffé colorimetric assay and MSI–CE–MS (**Figure 5.1G; Table 5.2**). This finding

highlights the consequence of ischemic muscle with advanced CLTI contributing to elevated serum creatinine, which was reported to predict mortality in PAD patients with renal failure independent of hypertension and type 2 diabetes.[54] An obvious confounder would be impaired renal function in CLTI, but our IC and CLTI patients were well balanced with only one participant from each group diagnosed with renal insufficiency. Furthermore, no other known differences exist among major PAD sub-groups, including comorbidities, smoking and medication use, with the exception of the ABI and clinical symptoms (i.e., leg pain and reduced mobility). Lastly, we observed a higher serum Phe/Tyr in CLTI relative to IC patients, consistent with recent reports of this as a ratiometric biomarker of inflammation in advanced stages of PAD,[16] as well as acute ischemic stroke that is also correlated with serum cytokines, such as interleukin-6.[55] Given the need for the improved risk stratification of PAD, other ratiometric biomarkers were found to be superior to Phe/Tyr ($AUC = 0.708$, $p = 0.0216$) as a diagnostic biomarker of CLTI in our study. For instance, the ratio of serum 18:0/C0 and Arg/C3 demonstrated better accuracy in differentiating CLTI from IC patients, with an $AUC \sim 0.87$ ($p = 4.0 \times 10^{-5}$) from the ROC curves (**Figure 5.5**). As expected, these ratiometric serum biomarkers were also strongly correlated ($r = 0.54–0.59$, $p < 0.001$) with the ABI and may thus prove useful for routine monitoring PAD progression and/or allow for the early detection of PAD in asymptomatic patients in a primary care setting. Since PAD is a complex clinical syndrome contributing to peripheral ischemia and claudication, **Figure 5.6** provides a systemic overview of the serum metabolites identified in this study that reflect its underlying pathophysiology, including elevated oxidative stress, inflammation, glutathione depletion, and perturbed energy homeostasis that contribute to severe

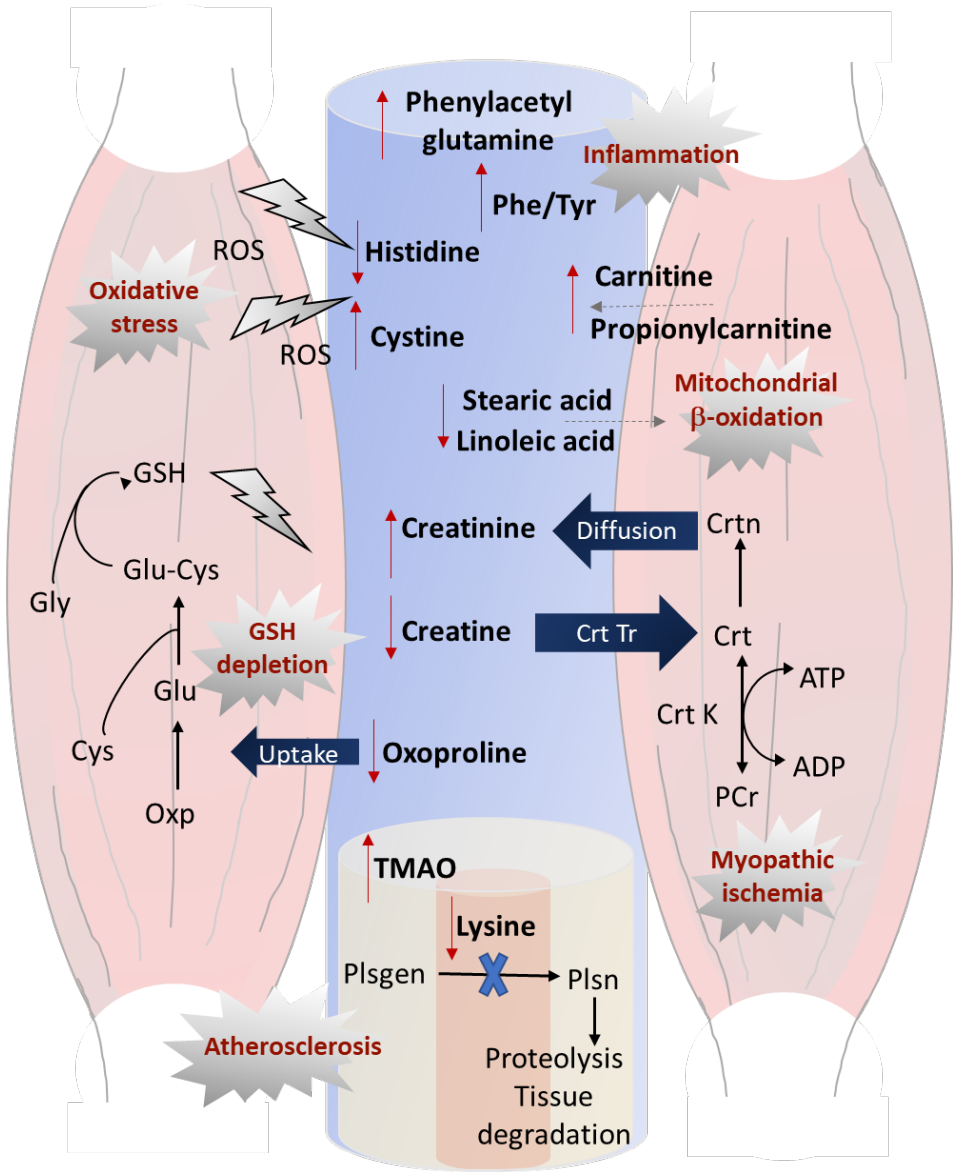


Figure 5.6. Schematic illustrating the systemic effects of an aberrant circulatory metabolism reflecting PAD progression in serum that also reflects the localized ischemia of skeletal muscle in lower limbs. Reduced serum lysine, an inhibitor of plasmin activation, in pathological states, leads to excessive proteolysis and vascular tissue degradation that is exacerbated by elevated TMAO, further promoting atherosclerosis in CLTI. Serum oxoproline, a key precursor used in the glutathione (GSH) salvage pathway, decreases in circulation as it is transported within the muscle to support intracellular glutathione recycling in response to elevated oxidative stress as reflected by lower histidine, and higher cystine in serum. Reduced creatine availability leads to lower intramuscular phosphocreatine, attributing to myopathic ischemia and perturbed energy homeostasis within the muscle that coincides with elevated creatinine concentrations in serum for CLTI. Impaired mitochondrial beta-oxidation is reflected by higher circulating carnitine and propionylcarnitine, that are not stored within ischemic muscle tissue in conjunction with lower circulating lipids/fatty acids, such as stearic acid and linoleic acid. Lastly, increased serum phenylacetylglutamine and Phe/Tyr reflect increased inflammation in CLTI, which are risk factors for all-causes mortality and cardiovascular disease. Abbreviations correspond to ROS: reactive oxygen species, ADP: adenosine diphosphate, ATP: adenosine triphosphate, Crt K: creatine kinase, Crt Tr: creatine transporter, Plsgen: plasminogen, Plsn: plasmin.

ischemic muscle dysfunction. This panel of serum metabolites offers a convenient approach for diagnostic testing complementary to the ABI, that may also predict CLTI onset, while guiding optimal dietary, pharmacological and/or surgical interventions to mitigate PAD progression relevant to improving long-term patient survivorship and their quality of life.

5.6. Conclusions

In summary, our work has revealed distinctive metabolic signatures associated with PAD progression in age/sex-matched non-diabetic patients, largely without renal failure or differences in prescribed medications (*e.g.,* statins, antiplatelets). We applied a high throughput metabolomics platform for the nontargeted analysis of circulating metabolites and hydrolyzed lipids from fasting serum samples, while applying a rigorous approach to data filtering to reduce the false discoveries with good technical precision. Certain serum metabolites have been reported in previous targeted and nontargeted metabolomic studies as indicative of PAD, using different instrumental platforms (NMR, GC–MS, LC–MS) that further corroborates our findings, including His, C0, TMAO, creatinine, Tyr and Phe/Tyr[14–17,52–54]; however, we did not observe differences in central energy metabolites (e.g., glutamate, 3-hydroxybuyrate, ketoglutarate) and metabolites from the urea cycle (e.g., ornithine, citrulline) that were also measured by MSI–CE–MS with stringent QC. In fact, discordant results were reported in two studies as related to the role of the urea cycle, likely reflecting potential confounding due to comorbidities in recruited PAD patients.[15,16] Importantly, we report for the first time that lower serum creatine, MMA, oxo-Pro, cystine, and several long-chain fatty acids are indicative of CLTI as compared to IC and/or non-PAD controls, with opposing trends for the uremic toxin, PAG. Overall, serum 18:0/C0 and Arg/C3 were among the most promising ratiometric biomarkers of PAD to accurately differentiate CLTI from IC, which also had strong positive correlations to the ABI. This panel of serum metabolites were related to

multiple pathological processes associated with the clinical syndrome of PAD, including muscle energy perturbations, vascular remodeling, atherosclerosis, myopathic ischemia, inflammation, and oxidative stress. Major limitations to our pilot study include the modest sample size of PAD patients recruited from a single center, which did not include patient diet records to assess the influence of habitual diet on PAD progression. Future work is needed to better validate the clinical utility of lead serum metabolites as potential diagnostic, prognostic and/or treatment biomarkers of PAD in prospective studies involving larger patient cohorts, including monitoring responses following vascularization surgery. Moreover, muscle metabolomic studies[37] may prove useful to better understand the deleterious impacts of chronic muscle ischemia in localized tissues, as related to the pathophysiology of PAD when compared to systemic changes of metabolism in serum. The specificity of PAD biomarkers also needs to be evaluated as compared to other related cardiometabolic diseases, including CAD, type 2 diabetes, chronic kidney disease, as well as sarcopenia in older persons.[22]

5.7 Acknowledgment

P.B.M. acknowledges funding from the Natural Sciences and Engineering Research Council of Canada, and Genome Canada. S.M.A. acknowledges funding from the Egyptian Ministry of Higher Education. M.Q. acknowledges funding from the Blair Foundation.

The authors declare no conflict of interest.

5.8 References

1. Hirsch AT, Criqui MH, Treat-Jacobson D, et al. Peripheral arterial disease detection, awareness, and treatment in primary care. *JAMA*. 2001;286:1317–1324.

2. Conte MS, Bradbury AW, Kolh P, White, JV, Dick F, FitridgeR, Mills JL, Ricco JB, Suresh KR, Murad MH, et al. Global vascular guidelines on the management of chronic limb-threatening ischemia. *J. Vasc. Surg*. 2019;69:3S–125S.

3. Dormandy J, Belcher G, Broos P, Eikelboom B, Laszlo G, Konrad P, Moggi L, Mueller U. Prospective study of 713 below–knee amputations for ischaemia and the effect of a prostacyclin analogue on healing. *Br. J. Surg*.1994;81:33–37.

4. Varu VN, Hogg ME, Kibbe MR. Critical limb ischemia. *J. Vasc. Surg*. 2010;51:230–241.

5. Almasri J, Adusumalli J, Asi N, Lakis S, Alsawas M, Prokop LJ, Bradbury A, Kolh P, Conte, MS, Murad MH. A systematic review and meta-analysis of revascularization outcomes of infrainguinal chronic limb-threatening ischemia. *J. Vasc. Surg*. 2019;69:126S–136S.

6. Cooke JP, Wilson AM. Biomarkers of peripheral arterial disease. *J Am. Coll. Cardiol*. 2010; 55:2017–2023.

7. Hyun S, Forbang NI, Allison MA, Denenberg JO, Criqui MH, Ix JH. Ankle-brachial index, toe-brachial index, and cardiovascular mortality in persons with and without diabetes mellitus. *J. Vasc. Surg*. 2014;60:390–395.

8. Clark N. American Diabetes Association. Peripheral arterial disease in people with diabetes. *Diabetes Care*. 2003;26:3333–3341.

9. Curry SJ. Screening for peripheral artery disease and cardiovascular disease risk assessment with the ankle-brachial index: US Preventive Services Task Force Recommendation Statement. *JAMA*. 2018;320:177–183.

10. McGarrah RW, Crown SB, Zhang GF, Shah SH, Newgard CB. Cardiovascular metabolomics. *Circ. Res*. 2018;122:1238–1258.

11. Wang Z, Klipfell E, Bennett BJ, et al. Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. *Nature*. 2011;472:57–63.

12. Ruiz-Canela M, Toledo E, Clish CB, et al. Plasma branched-chain amino acids and incident cardiovascular disease in the PREDIMED Trial. *Clin. Chem*. 2016;62:582–592.

13. Shah SH, Sun J-L, Stevens RD, et al. Baseline metabolomic profiles predict cardiovascular events in patients at risk for coronary artery disease. *Am. Heart J*. 2012;163:844–850.

14. Huang CC, McDermott MM, Liu K, Kuo CH, Wang SY, Tao H, Tseng YJ. Plasma metabolomic profiles predict near-term death among individuals with lower extremity peripheral arterial disease. *J. Vasc. Surg*. 2013;58:989–996.

15. Zagura M, Kals J, Kilk K, Serg M, Kampus P, Eha J, Soomets U, Zilmer M. Metabolomic signature of arterial stiffness in male patients with peripheral arterial disease. *Hypertens. Res*. 2015;38:840–846.

16. Ismaeel A, Franco ME, Lavado R, et al. Altered metabolomic profile in patients with peripheral artery disease. *J. Clin. Med*. 2019;8:1463.

17. Hernández-Aguilera A, Fernández-Arroyo S, Cabre N, Luciano-Mateo F, Baiges-Gaya G, Fibla M, Martín-Paredero V, Menendez JA, Camps J, Joven J. Plasma energy-balance metabolites discriminate asymptomatic patients with peripheral artery disease. *Mediat. Inflamm*. 2018;18:1–12.

18. Rutherford RB, Baker JD, Ernst C, Johnston KW, Porter JM, Ahn S, Jones DN. Recommended standards for reports dealing with lower extremity ischemia: Revised version. *J. Vasc. Surg*.1997; 26:517–538.

19. Al-Khatib SM, Stevenson WG, Ackerman MJ, et al. 2017 AHA/ACC/HRS Guideline for management of patients with ventricular arrhythmias and the prevention of sudden cardiac death: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines and the Heart Rhythm Society. *J. Am. Coll. Cardiol*. 2018;72:e91–e220.

20. Ou M, Song Y, Li S, Liu G, Jia J, Zhang M, Zhang H, Yu C. LC-MS/MS method for serum creatinine: Comparison with enzymatic method and Jaffé method. *PLoS ONE*. 2015;10: e0133912.

21. Wild J, Shanmuganathan M, Hayashi M, Potter M, Britz-McKibbin P. Metabolomics for improved treatment monitoring of phenylketonuria: Urinary biomarkers for non-invasive assessment of dietary adherence and nutritional deficiencies. *Analyst*. 2019;144:6595-6608.

22. Saoi M, Li A, McGlory C, Stokes T, von Allmen MT, Phillips SM, Britz-McKibbin P. Metabolic perturbations from step reduction in older persons at risk for sarcopenia: Plasma biomarkers of abrupt changes in physical activity. *Metabolites*. 2019;9:134.

23. Azab S, Ly R, Britz-McKibbin P. Robust method for high-throughput screening of fatty acids by multisegment injection-nonaqueous capillary electrophoresis–mass spectrometry with stringent quality control. *Anal. Chem*. 2019;91:2329–2336.

24. Azab SM, de Souza RJ, Teo KK, Anand SS, Williams NC, Holzschuher J, McGlory C, Phillips SM; Britz-McKibbin P. Serum non-esterified fatty acids have utility as dietary biomarkers of fat intake from fish, fish oil and dairy in women. *J. Lipid Res*. 2020;61:933–944.

25. Matyash V, Liebisch G, Kurzchalia TV, Shevchenko A, Schwudke D. Lipid extraction by methyl-*tert*-butyl ether for high-throughput lipidomics. *J. Lipid Res*. 2008;49:1137–1146.

26. Kuehnbaum NL, Kormendi A, Britz-McKibbin P. Multisegment injection-capillary electrophoresis-mass spectrometry: A high-throughput platform for metabolomics with high data fidelity. *Anal. Chem*. 2013;85:10664–10669.

27. Nori de Macedo A, Mathiaparanam S, Brick L, Keenan K, Gonska T, Pedder L, Hill S, Britz-McKibbin P. The sweat metabolome of screen-positive cystic fibrosis infants: Revealing mechanisms beyond impaired chloride transport. *ACS Cent. Sci*. 2017;3:904–913.

28. Wellington N, Shanmuganathan M, de Souza RJ, et al. Metabolic trajectories following contrasting Prudent and Western diets from food provisions: Robust biomarkers of short-term changes in habitual diet. *Nutrients* 2019;11:2407.

29. Yamamoto M, Ly R, Gill B, Zhu Y, Moran-Mirabal J, Britz-McKibbin P. Robust and high-throughput method for anionic metabolite profiling: Preventing polyimide aminolysis and capillary breakages under alkaline conditions in capillary electrophoresis-mass spectrometry. *Anal. Chem*. 2016;88:10710–10719.

30. DiBattista A, Rampersaud D, Lee H, Kim M, Britz-McKibbin P. High throughput screening method for systematic surveillance of drugs of abuse by multisegment injection–capillary electrophoresis–mass spectrometry. *Anal. Chem*. 2017;89:11853–11861.

31. Dunn WB, Erban A, Weber RJM, et al. Mass appeal: Metabolite identification in mass spectrometry-focused untargeted metabolomics. *Metabolomics*. 2013;9:44–66.

32. Chong J, Soufan O, Li C, Caraus I, Li S, Bourque G, Wishart DS, Xia J. MetaboAnalyst 4.0: Towards more transparent and integrative metabolomics analysis. *Nucleic Acids Res*. 2018; 46:W486–W494.

33. Yamamoto M, Pinto-Sanchez MI, Bercik P, Britz-McKibbin P. Metabolomics reveals elevated urinary excretion of collagen degradation and epithelial cell turnover products in irritable bowel syndrome patients. *Metabolomics*. 2019;15:82.

34. Mahieu NG, Patti GJ. Systems-level annotation of a metabolomics data set reduces 25,000 features to fewer than 1000 unique metabolites. *Anal Chem*. 2017;89:10397–10406.

35. Swedko PJ, Clark HD, Paramsothy K, Akbari A. Serum creatinine is an inadequate screening test for renal failure in elderly patients. *Arch. Intern. Med*. 2003;163:356–360.

36. Gilstrap LG, Wang TJ. Biomarkers and cardiovascular risk assessment for primary prevention: An update. *Clin. Chem*. 2012;58:72–82.

37. Saoi M, Percival M, Nemr C, Li A, Gibala M, Britz-McKibbin P. Characterization of the human skeletal muscle metabolome for elucidating the mechanisms of bicarbonate ingestion on strenuous interval exercise. *Anal. Chem*. 2019;91:4709–4718.

38. Wade AM, Tucker HN. Antioxidant characteristics of L-histidine. *J. Nutr. Biochem*. 1998;9: 308–315.

39. Tomé D, Bos C. Lysine requirement through the human life cycle. *J. Nutr*. 2007;137:1642S–1645S.

40. Castellano I, Merlino A. (Eds). Gamma-glutamyl transpeptidases: Structure and function. In *Gamma-Glutamyl Transpeptidases: Structure and Function*; Springer Briefs in Biochemistry and Molecular Biology; Springer: Basel, Switzerland, 2013; pp. 1–57. ISBN 978-3-0348-0682-4.

41. Steven S, Daiber A, Dopheide JF, Münzel T, Espinola-Klein C. Peripheral artery disease, redox signaling, oxidative stress—Basic and clinical aspects. *Redox Biol*. 2017;12:787–797.

42. Saoi M, Sasaki K, Sagawa H, Abe K, Kogiso T, Tokushige K, Hashimoto E, Ohashi Y, Britz-McKibbin P. High throughput screening of serum γ-glutamyl dipeptides for risk assessment of nonalcoholic steatohepatitis with impaired glutathione salvage pathway. *J. Proteome Res*. 2020;19:2689-2699.

43. Hajjari J, Tahhan AS, Alkhoder A, et al. Markers of oxidative stress and peripheral artery disease. *J. Am. Coll. Cardiol*. 2018;71:A2076.

44. Poesen R, Claes K, Evenepoel P, de Loor H, Augustijns P, Kuypers D, Meijers B. Microbiota-derived phenylacetylglutamine associates with overall mortality and cardiovascular disease in patients with CKD. *J. Am. Soc. Nephrol*. 2016:27:3479–3487.

45. Sanders Thomas AB. Omega-6 fatty acids and cardiovascular disease. *Circulation* 2019;139: 2437–2439.

46. Matti M, Wu JHY, Imamura F, et al. Biomarkers of dietary omega-6 fatty acids and incident cardiovascular disease and mortality. *Circulation*. 2019;139:2422–2436.

47. Grenon, SM, Hughes-Fulford M, Rapp J, Conte MS. Polyunsaturated fatty acids and peripheral artery disease. *Vasc. Med*. 2012;17:51–63.

48. Ruiz-Canela M, Estruch R, Corella D, Salas-Salvadó J, Martínez-González MA. Association of Mediterranean diet with peripheral artery disease: The PREDIMED Randomized Trial. *JAMA* 2014:311:415.

49. Weitkunat K, Schumann S, Nickel D, Hornemann S, Petzke KJ, Schulze MB, Pfeiffer AF, Klaus S. Odd-chain fatty acids as a biomarker for dietary fiber intake: A novel pathway for endogenous production from propionate. *Am. Clin. Nutr*. 2017;105:1544–1551.

50. Senyilmaz-Tiebe D, Pfaff DH, Virtue S, et al. Dietary stearic acid regulates mitochondria in vivo in humans. *Nat. Commun*. 2018;9:3129.

51. Weiss DJ, Casale GP, Koutakis P, Nella AA, Swanson SA, Zhu Z, Miserlis D, Johanning JM, Pipinos II. Oxidative damage and myofiber degeneration in the gastrocnemius of patients with peripheral arterial disease. *J. Transl. Med*. 2013;11:230.

52. Senthong V, Wang Z, Fan Y, Wu Y, Hazen SL, Tang WHW. Trimethylamine *N* -oxide and mortality risk in patients with peripheral artery disease. *J. Am. Heart Assoc*. 2016;5:e004237.

53. Roncal C, Martínez-Aguilar E, Orbe J, et al. Trimethylamine-n-oxide (TMAO) predicts cardiovascular mortality in peripheral artery disease. *Sci. Rep*. 2019;9:15580.

54. Mlekusch W, Exner M, Sabeti S, Amighi J, Schlager O, Wagner O, Minar E, Schillinger M, Serum creatinine predicts mortality in patients with peripheral artery disease: Influence of diabetes and hypertension. *Atherosclerosis*. 2004;175:361–367.

55. Ormstad H, Verkerk R, Sandvik L. Serum phenylalanine, tyrosine, and their ratio in acute ischemic stroke: On the trail of a biomarker? *J. Mol. Neurosci*. 2016;58:102–108.

5.9 Supplemental Information:

Serum Metabolome Data Matrix and Deidentified Patient Data.

An excel file (PAD-Metabolomics-Data.xls) containing serum metabolome data matrix measured for all PAD patients (IC, CLTI) and non-PAD controls for authenticated metabolites/lipids measured by MSI-(NA)CE-MS under three different configurations is provided, including quality controls. All serum metabolites are annotated by their accurate mass and relative migration time (*m/z*:RMT) and name (if identified), where responses reflect their ion response ratio normalized to an internal standard. This excel file also contains deidentified patient demographic and clinical information from this pilot study for full data transparency.

Table S5.1. Summary of 85 serum metabolites detected in PAD patients that are annotated based on their accurate mass (*m/z*), relative migration time (RMT), ionization mode (l = non-aqueous negative mode, p = aqueous positive mode, n = aqueous negative mode), metabolite ID, most likely molecular formula, confidence level for identification, and technical precision of repeated QCs analyzed in each run.

***m/z*:RMT:mode**	**Metabolite ID**	**Molecular Formula**	**Confidence level**	**% CV**
Myristic-d27 acid (14:0-d27)	254.371:1.053:l	$C_{14}HD_{22}O_2$	1	6.0
Lauric acid (12:0)	199.170:1.081:l	$C_{12}H_{24}O_2$	1	14.2
Myristelaidic acid (14:1)	225.186:1.064:l	$C_{14}H_{26}O_2$	2	16.1
Myristic acid (14:0)	227.202:1.058:l	$C_{14}H_{28}O_2$	1	16.1
Pentadyclic acid (15:0)	241.217:1.042:l	$C_{15}H_{30}O_2$	1	29.2
Palmitoleic acid (16:1n-7)	253.217:1.037:l	$C_{16}H_{30}O_2$	1	7.3
Palmitic acid (16:0)	255.233:1.030:l	$C_{16}H_{32}O_2$	1	11.6
Heptadecenoic acid (17:1n-7)	267.233:1.026:l	$C_{17}H_{32}O_2$	2	14.5
Heptadecanoic acid (17:0)	269.249:1.017:l	$C_{17}H_{34}O_2$	1	12.0
Linolenic acid (18:3*n*-3)	277.217:1.023:l	$C_{18}H_{30}O_2$	1	6.8
Linoleic acid (18:2*n*-6)	279.233:1.019:l	$C_{18}H_{32}O_2$	1	5.2
Oleic acid (18:1*n*-9)	281.249:1.013:l	$C_{18}H_{34}O_2$	1	5.4
Stearic acid (18:0)	283.2676:1.0051:l	$C_{18}H_{36}O_2$	1	14.9
Eicosapentaenoic acid (20:5*n*-3)	301.217:1.031:l	$C_{20}H_{30}O_2$	1	4.9
Arachidonic acid (20:4*n*-6)	303.233:1.028:l	$C_{20}H_{32}O_2$	1	6.5
Dihomo-linolenic acid (20:3*n*-6)	305.252:1.004:l	$C_{20}H_{34}O_2$	2	7.8
Eicosadienoic acid (20:2)	307.265:0.994:l	$C_{20}H_{36}O_2$	2	7.6
Arachidic acid (C20:0)	311.295:0.981:l	$C_{20}H_{40}O_2$	1	27.1
Docosahexaenoic acid (22:6*n*-3)	327.233:1.034:l	$C_{22}H_{32}O_2$	1	6.3
Docosapentaenoic acid (22:5*n*-6)	329.253:0.992:l	$C_{22}H_{34}O_2$	2	14.3
Adrenic acid (22:4*n*-6)	331.268:0.990:l	$C_{22}H_{36}O_2$	1	6.2
Behenic acid (22:0)	339.327:0.969:l	$C_{22}H_{44}O_2$	1	21.8
Unknown#1	353.235:1.027:l	$C_{20}H_{34}O_5$	3	23.4
Nervonic acid (24:1*n*-9)	365.342:0.947:l	$C_{24}H_{46}O_2$	1	20.9
Lignoceric acid (24:0)	367.357:0.942:l	$C_{24}H_{48}O_2$	1	26.5
Glycine	76.040:0.702:p	$C_2H_5NO_2$	1	27.5
Trimethylamine-*N*-oxide	76.077:0.544:p	C_3H_9NO	1	19.4
Alanine	90.055:0.758:p	$C_3H_7NO_2$	1	11.7
γ-Aminobutyric acid	104.071:0.805:p	$C_4H_9NO_2$	1	18.0
Dimethylglycine	104.071:0.926:p	$C_4H_9NO_2$	1	24.2
Choline	104.108:0.569:p	$C_5H_{14}NO$	1	14.6

Serine	106.050:0.842:p	$C_3H_7NO_3$	1	7.2
Creatinine	114.066:0.614:p	$C_4H_7N_3O$	1	7.8
Proline	116.071:0.908:p	$C_5H_9NO_2$	1	6.3
Guanidoacetate	118.086:0.835:p	$C_3H_7N_3O_2$	1	17.0
Betaine	118.086:0.956:p	$C_5H_{11}NO_2$	1	11.9
Threonine	120.065:0.885:p	$C_4H_9NO_3$	1	6.7
Unknown#2	129.066:0.736:p	$C_5H_8N_2O_2$	4	7.5
Hydroxyproline	132.065:1.022:p	$C_5H_9NO_3$	1	6.4
Creatine	132.077:0.745:p	$C_4H_9N_3O_2$	2	8.2
Isoleucine	132.102:0.848:p	$C_6H_{13}NO_2$	1	8.6
Leucine	132.102:0.861:p	$C_6H_{13}NO_2$	1	7.5
Ornithine	133.097:0.578:p	$C_5H_{12}N_2O_2$	1	9.3
Unknown#3	134.044:0.969:p	$C_4H_7NO_4$	4	14.2
Hypoxanthine	137.046:1.058:p	$C_5H_4N_4O$	1	13.9
Glutamine	147.076:0.910:p	$C_5H_{10}N_2O_3$	1	5.5
Lysine	147.113:0.580:p	$C_6H_{14}N_2O_2$	1	6.6
Glutamic acid	148.060:0.923:p	$C_5H_9NO_4$	1	5.6
Methionine	150.058:0.896:p	$C_5H_{11}NO_2S$	1	5.3
TMAO (dimer)	151.144:0.544:p	C_3H_9NO	1	16.8
Histidine	156.077:0.620:p	$C_6H_9N_3O_2$	1	7.8
Unknown#4	160.133:0.709:p	$C_8H_{17}NO_2$	4	15.3
α-Aminoadipate	162.076:0.924:p	$C_6H_{11}NO_4$	1	12.9
Carnitine	162.112:0.719:p	$C_7H_{15}NO_3$	1	7.5
Phenylalanine	166.086:0.926:p	$C_9H_{11}NO_2$	1	6.7
Unknown #5	169.058:0.910:p	$C_5H_{10}N_2O_3$	4	25.5
Methylhistidine	170.092:0.635:p	$C_7H_{11}N_3O_2$	1	5.4
Arginine	175.119:0.601:p	$C_6H_{14}N_4O_2$	1	7.3
Citrulline	176.103:0.936:p	$C_6H_{13}N_3O_3$	1	5.3
Tyrosine	182.080:0.9564:p	$C_9H_{11}NO_3$	1	2.8
Monomethylarginine	189.134:0.606:p	$C_7H_{16}N_4O_2$	1	18.9
Acetylcarnitine	204.123:0.762:p	$C_9H_{17}NO_4$	1	7.5
Tryptophan	205.097:0.925:p	$C_{11}H_{12}N_2O_2$	1	7.4
Unknown#6	217.154:0.836:p	$C_{15}H_{20}O$	4	28.8
Propionylcarnitine	218.138:0.784:p	$C_{10}H_{19}NO_4$	1	13.5
Cystine	241.030:0.933:p	$C_6H_{12}N_2O_4S_2$	1	5.8
Cysteinylglycine disulfide	298.052:0.806:p	$C_8H_{15}N_3O_5S_2$	2	6.6
Pyruvic acid	87.009:1.338:n	$C_3H_4O_3$	1	26.9
Lactic acid	89.024:1.281:n	$C_3H_6O_3$	1	16.6
Phosphoric acid	96.970:1.949:n	H_3O_4P	2	14.7
Dimethylglycine	102.056:1.182:n	$C_4H_9NO_2$	1	25.5

3-Hydroxybutyric acid	103.040:1.142:n	$C_4H_8O_3$	1	29.8
2-Hydroxybutyric acid	103.040:1.168:n	$C_4H_8O_3$	1	23.4
α-Ketoisovaleric acid	115.040:1.213:n	$C_5H_8O_3$	1	30.9
Taurine	124.007:0.972:n	$C_2H_7NO_3S$	1	15.2
Pyroglutamic acid	128.035:1.137:n	$C_5H_7NO_3$	1	31.4
3-Methyl-2-oxovaleric acid	129.055:1.142:n	$C_6H_{10}O_3$	1	22.3
Aspartic acid	132.030:1.350:n	$C_4H_7NO_4$	1	22.4
Unknown#7	135.031:1.110:n	$C_4H_8O_5$	4	19.1
Uric acid	167.021:1.101:n	$C_5H_4N_4O_3$	1	16.3
Hippuric acid	178.051:1.009:n	$C_9H_9NO_3$	1	13.8
Glucose	179.056:1.000:n	$C_6H_{12}O_6$	1	16.1
Citric acid	191.020:2.418:n	$C_6H_8O_7$	1	26.1
Gluconic acid	195.050:0.976:n	$C_6H_{12}O_7$	1	26.6
Phenylacetylglutamine	263.104:0.899:n	$C_{13}H_{16}N_2O_4$	1	15.5

Chapter VI:

Future Directions in Biomarkers, Metabolomics and Lipidomics for Clinical Medicine

Chapter VI: Future Directions in Biomarkers, Metabolomics and Lipidomics for Clinical Medicine

"Let us not lose heart in doing good, for in due time we will reap if we do not grow weary"

6.1 Overview of major thesis contributions

The research presented in this thesis has contributed two main directions in metabolomics and biomarker discovery: (1) development of novel analytical methods for high-throughput, simple and robust measurement of long-chain fatty acids and perfluoroalkyl substances (PFASs) in serum in support of large-scale epidemiological studies of metabolic disorders in birth cohorts as well as new advancements in clinical medicine and public health; and (2) application of these metabolomic and lipidomic analyses in nutritional science and cardiovascular disease for the purpose of biomarker discovery and dietary biomarker validation. Herein, various metabolomics studies were exemplified either through a targeted hypothesis-testing approach to assess maternal fat intake by analyzing serum non-esterified fatty acids (NEFAs) or an untargeted hypothesis-generating approach for characterization of the serum metabolome in peripheral artery disease (PAD).

Chapter I of this thesis provided a comprehensive review of the types and roles of biomarkers in clinical medicine and the role of metabolomics in biomarker discovery. It provides a thorough overview of the metabolomics experiment, major steps and challenges of the metabolomics workflow as well as the process of biomarker validation for translational research and concludes with a specific focus on fatty acids, their chemistry, biochemistry and clinical significance, as an important class of lipids and key molecules for human health. *Chapter II* introduced a novel assay based on multisegment injection - non-aqueous capillary electrophoresis – mass-spectrometry (MSI-NACE-MS) following rigorous method optimization and validation for fatty acids profiling in serum that is relevant for a multitude of clinical applications.[1] In this study,

a major limitation of the gold standard gas chromatographic (GC) methods, which require lengthy sample derivatization and complicated workup protocols was evaded, requiring only a simple single-step liquid extraction step using methyl-*tert*-butyl ether (MTBE) was needed.[2,3] Furthermore, multiplexed separations of seven sequential sample injections in a single run of 30 minutes offered a 7 x higher throughput in addition to superior quality management granted by the incorporation of a (Quality Control) QC sample in each run for monitoring instrumental drift and overall technical variance with subsequent batch correction as needed.[4,5,6] Accurate modeling of ion migration behavior of fatty acids in NACE was also achieved and served as a complementary physicochemical tool useful for identification of unknown fatty acids notably standards are unavailable. Importantly, longstanding technical challenges associated with NACE-MS that constitute major obstacles to an otherwise promising technique for lipidomics, were carefully studied and systematically overcome, including capillary swelling and incidental fractures, current instabilities from nebulizer suctioning effects, as well as frequent corona discharge under negative ion mode detection when using organic solvent solutions.

A rigorous validation of MSI-NACE-MS was further expanded in *Chapter III* using tandem mass spectrometry (MS) on a triple quadrupole (QQQ) system for targeted multiplexed analysis of nanomolar levels of PFASs in human serum with rigorous quality control. PFASs are a class of synthetic persistent organic pollutants and endocrine-disrupting chemicals linked to metabolic abnormalities in children from prenatal exposures and pose a global challenge to scientists, industry leaders and public health.[7,8,9] We demonstrated clear advantage of the use of non-aqueous background electrolyte systems over aqueous ones for capillary electrophoresis (CE) and CE-MS for superior solubilization, resolution and peak sharpness of perfluorooctanoic acid (PFOA) and perfluorooctanesulfonic acid (PFOS), the two major perflourinated contaminants in

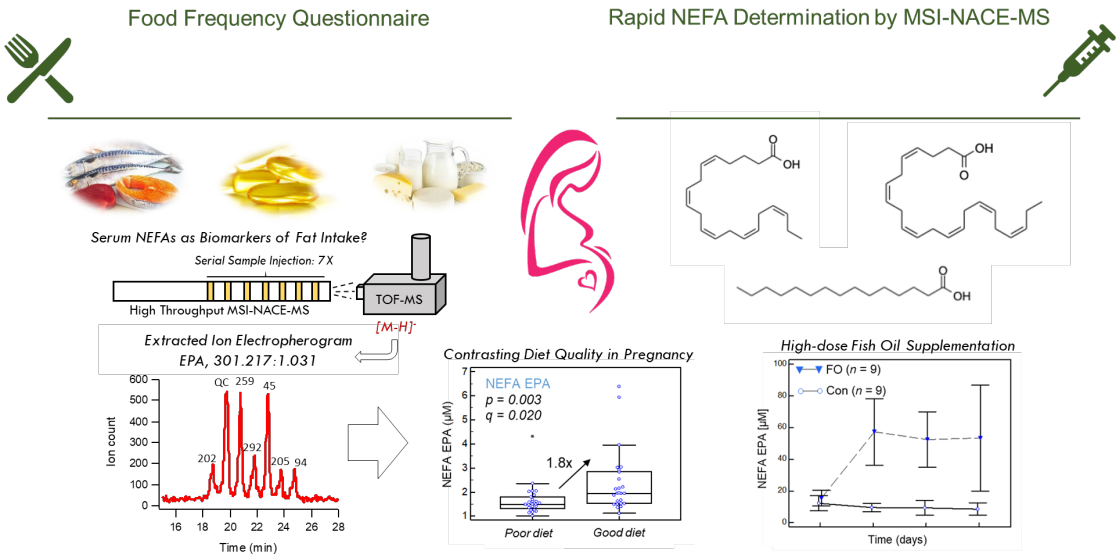
serum extracts. Similar to fatty acids, the sequential injection of seven discrete samples including a repeated QC sample within one run, offers higher throughput (< 3 min/sample) and adequate quality control for rapid biomonitoring of PFASs by MSI-NACE-MS/MS. Finally, the newly developed method was applied on a small subset of samples from birth cohorts collected after 2009 and compared to a pooled sample from before 2009 with a notable decrease in serum concentration in accordance with enforced PFASs regulation by the Stockholm Convention with considerable variation of individual serum concentrations of PFOA and PFOS in pregnant women.

Chapter IV of this thesis details a direct application in nutritional epidemiology of the rigorously validated lipidomic MSI-NACE-MS method developed in *Chapter II* for fatty acids analysis through an integration of two observational and intervention studies to assess the utility of circulating NEFA as biomarkers of fat intake in women and how this underreported lipid fraction correlated with other lipid pools.[10] From a methodological point of view, NEFA analysis is challenging as its chemical derivatization for GC without background hydrolysis is difficult and thus NEFA is an underreported fraction with 90% of studies focusing on phospholipid fraction, either from serum or erythrocytes, or total hydrolyzed serum lipids.[1,11,12] MSI-NACE-MS offers direct NEFA analysis with no need for derivatization, fractionation or hydrolysis and is thus optimal for large-scale epidemiological studies urgently needed for evidence-based dietary guidelines of fat intake in relation to disease. Furthermore, in search of candidate biomarkers of food intake (BFIs), ideally these should be readily measurable, specific to food and not endogenously produced by the body, as in the special case of ω-3 fatty acid and odd-chain fatty acids, which are the focus of this study. In the cross-sectional study of 50 pregnant women, eicosapentaenoic acid (EPA) and docosahexaenoic acid (DHA) in terms of their serum NEFA and similar to total hydrolyzed fraction, were significantly higher in women consuming a healthy diet

than women with a poor diet quality and were also correlated to self-reported fish/ω-3 intake. As for biomarkers of dairy fat intake, serum pentadecanoic aid (15:0) as its NEFA followed by myristic acid (14:0) had the strongest association and selectivity to self-reported full-fat dairy intake with no correlation found to either low-fat or fermented dairy intakes. In parallel, the randomized controlled study of serum NEFA following a high-dose fish oil supplementation as compared to a placebo group in 18 young women followed over a 56 day intervention period revealed a significant 2.5-fold increase from baseline in serum EPA and DHA as their NEFA within a month in the treatment arm with no further changes at later time points. These outcomes were comparable to independent erythrocyte phospholipid measurements with no other changes in other measured serum NEFAs. Interestingly, we have noted a strong correlation between erythrocyte PL and NEFA for EPA but not DHA, to support non-esterified EPA as a likely more responsive and sensitive biomarker of high-dose fish oil supplementation while also revealing non-adherence to fish oil supplementation for participants. These results are especially important in light of various interests focused on high-dose ω-3 fatty acids for the prevention and/or treatment of childhood asthma, as well as muscle atrophy, and cardiovascular disease in older persons. For the first time, we demonstrated that NEFA analysis using MSI-NACE-MS offers a rapid approach for accurate assessment of dietary/supplemental fat intake, in specific fat from fish, fish oil and dairy in women, consistent with other commonly used lipid pools, as needed to complement dietary self-report tools and food frequency questionnaires that are subjective and often biased and prone to error.

Chapter V described the non-targeted characterization of serum hydrophilic and lipophilic metabolome of age and sex-matched nondiabetic PAD patients. In addition to measuring total hydrolyzed serum fatty acids using MSI-NACE-MS as described in *Chapter II*, we have applied a

high throughput metabolomics platform for polar ionic compounds using MSI-CE-MS in both positive and negative ion modes. The incorporation of a QC sample in each MSI run, randomly allocated in the sample injection sequence, allowed for rigorous quality control measures of data with subsequent data filtering to exclude spurious signals/redundant features and reduce false discoveries. Overall, 85 authenticated metabolites were consistently detected in the majority of serum samples. We have revealed for the first time that lower serum creatine, monomethylarginine, oxoproline, cystine, and several fatty acids are indicative of chronic limb-threatening ischemia (CLTI) as compared to intermittent claudication (IC) and/or non-PAD controls with opposing trends for the uremic toxin, phenylacetylglutamine. Consistent with recent literature, we observed higher serum trimethylamine-*N*-oxide and phenylalanine/tyrosine ratio in CLTI relative to IC, as markers of atherosclerosis and inflammation, respectively, in advanced stages of PAD, as well as lower serum histidine levels in PAD patients versus non-PAD controls; a hallmark of reduced antioxidant capacity.[13,14] Noteworthy, this work outlined the discovery of promising ratiometric metabolites, namely serum stearic acid/carnitine and arginine/propionylcarnitine, that allow for differential discrimination of CLTI with good accuracy, not predictable by ABI measurements alone. These metabolites are related to multiple pathological processes associated with PAD, including skeletal muscle energy perturbations, vascular remodeling, atherosclerosis, myopathic ischemia, inflammation, and oxidative stress. In conclusion, a panel of serum metabolites that reflect underlying pathophysiology of PAD has been identified, that can putatively aid in early diagnosis or risk assessment of PAD and serve as a risk-stratifying tool for IC and CLTI for timely identification and therapeutic monitoring of high-risk patients to reduce delays in surgical interventions, amputations and mortality.



6.2 Further advancement of lipidomics analyses by MSI-NACE-MS

Successful non-traditional lipidomics analyses using MSI-NACE-MS were initiated in this thesis and show promise of future advancement for high throughput lipidomic studies, which largely rely on lower throughput LC-MS methods with reversed-phase or hydrophilic interaction chromatographic separations with continuous need for interlaboratory harmonization.[15] *Chapter II* introduced a rapid method based on MSI-NACE-MS for the reliable determination of fatty acids in human serum extracts that is optimal for large-scale analyses in nutritional epidemiological.[1] This method was rigorously validated and demonstrated acceptable linearity, precision, accuracy and sensitivity for accurate quantification of 24 fatty acids in addition to putative identification of other anionic lipids. However, isomeric resolution of geometrical or positional isomers was not achieved and improved resolution of low levels of branched-chain fatty acids and fatty acids isomers (cis/trans) is sought for in future studies through several directions. As with the advantage offered by micellar electrokinetic capillary chromatography (MEKC), use of a complexing agent that binds differently to geometric isomers could be expected to segregate co-migrating isomers. Unfortunately, this is less compatible with electrospray ionization-MS due to persistence and

memory effects of the used surfactants, nevertheless, use of volatile MS-compatible surfactants has shown some potential.[16] A different approach is one that supresses the EOF by employing a linear polyacrylamide (LPA) coated capillary which supresses electroosmosis; a major source of migration time variability in NACE.[17] Under reversed polarity, fatty acids would then migrate towards the capillary outlet held at the anode, which would theoretically increase separation window and peak capacity.

Moreover, certain low abundance eicosanoids, ideal for MSI-NACE-MS analysis in negative mode (*e.g.*, leukotrienes and prostaglandins), could benefit from improved concentration sensitivity in MSI-CE-MS when transitioning to sheathless/nanospray interfaces for CE-MS which can enhance ionization responses by up to two orders of magnitude.[18] In addition, use of a wider-bore inner diameter fused-silica capillary (100 μm) could be explored for increased sample loading. Similar approaches could also be applied to the method introduced in *Chapter III* for biomonitoring of serum PFASs by MSI-NACE-MS/MS to boost sensitivity and detection limits. Additionally, expanding coverage of the analytical assay by further optimizing multiple reaction monitoring (MRM) transitions seems a simple and quite promising step for quantitation of various other legacy PFASs that are not regulated, including perfluorononanoic acid (PFNA) and perfluorohexane sulfonic acid (PFHxS), as well as emerging fluorinated replacements (*e.g.*, chlorinated polyfluoroalkyl ether sulfonic acids). Importantly, method development for intact phospholipid analysis by MSI-NACE-MS under positive ion mode is currently undergoing (by Ritchie Ly supervised by Dr. Britz-Mckibbin) and is anticipated to offer expanded and extensive cationic lipid profiling of more than 80 different cationic lipid species to include phosphatidylcholines, sphingomyelins and phosphatidylethanolamines. This strategy relies on a rapid yet safe reaction for derivatization of zwitterionic phospholipids, that would otherwise co-

migrate close to EOF, by fixing a net permanent positive charge using 3-methyl-1-p-toyltriazene (MTT) to methylate phosphoric acid residues; optimal for positive ion mode detection. By means of the resulting increased mobility shift, more efficient separation and improved ESI signal responses are achieved while significantly increasing the lipidome coverage of MSI-NACE-MS.

6.3 NEFA analysis of maternal serum in the FAMILY birth cohort

In *Chapter III*, a subset of the maternal serum samples (n = 50) of the FAMILY birth cohort was used to investigate the utility of NEFA as objective BFIs, in specific, fish and full-fat dairy.[19] In this context, fatty acids profiling is undergoing for a larger sample size (n = 300) from FAMILY as part of the collaborative initiative for "Deciphering the metabolic signature of the metabolic syndrome in children".[20] This project first serves as a validation for pilot studies presented in *Chapter III* on the utility of serum non-esterified ω-3 fatty acids and 15:0 as biomarkers of maternal intake of fatty fish and full-fat dairy, respectively.[21] Furthermore, this work will provide reference ranges for absolute concentrations of circulating NEFA from a representative population of pregnant women in Canada, which is valuable for the research community to provide average baseline measures during pregnancy, that are not estimated from dietary assessment, in support of nutritional science as well as biomarker cutoffs as compared to normal distributions.[22] Absolute concentrations are also advantageous compared to weight percentages when pooling/investigating multiple studies to allow for direct integration of data from different study populations.[22] Additionally, an untargeted metabolomics approach will be conducted to precisely explore and expand MSI-NACE-MS coverage with previous putative identification of other classes of acidic metabolites in serum extracts e.g. bilirubin, pregnanediole-3-glucuronide, lysophosphatidic acid and some bile acids, for more comprehensive analyses as exemplified in **Figure 6.1**. Finally, this study will investigate the associations of the maternal metabolome/lipidome, which reflects

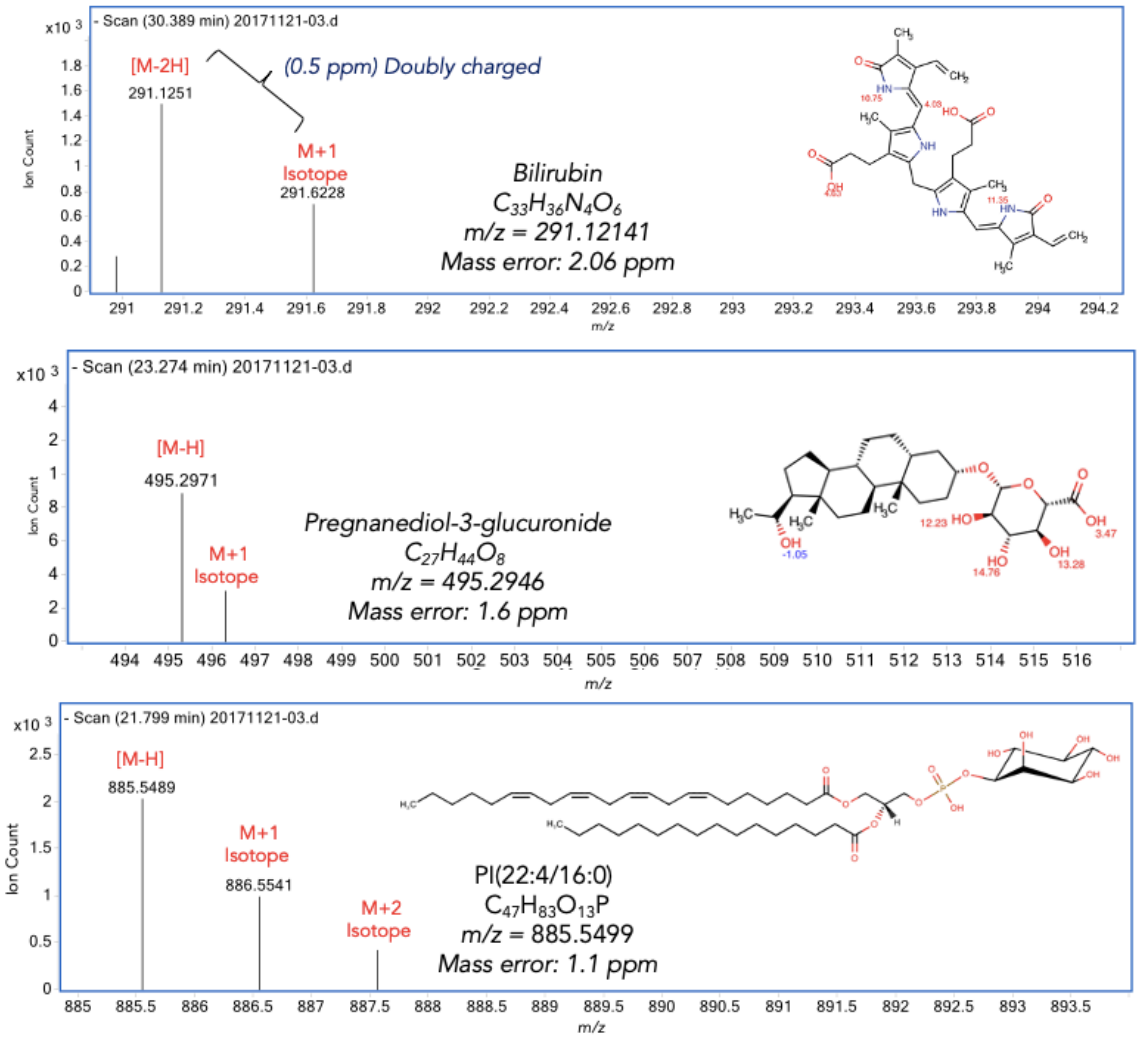


Figure 6.1 Representative mass spectra and tentative identification of various classes of acidic lipids in serum extracts from pregnant women when using nontargeted data analysis with MSI-NACE-MS. Tentative identification based on accurate mass since unambiguous elucidation of exact chemical structures for anionic lipids requires comparison to reference MS/MS spectra that are also confirmed with standards

maternal exposures and the prenatal milieu, with various disease and metabolic outcomes in the offspring later in life such as obesity, hyperlipidemia or asthma, as predisposed during fetal programming.

6.4 Validation of preliminary findings of PAD-specific metabolic differences in serum

Chapter V reported, for the first time, differential metabolic and ratiometric differences indicative of CLTI as compared to IC and/or non-PAD controls. Future studies to validate results of this work

are necessary to confirm the findings in larger and independent cohorts including a multi-center study design. In order to move the identified panel of serum metabolites from exploratory to advanced phases of biomarker discovery and translation to clinical settings, results must be first replicated on different samples drawn from the same population using receiver operating characteristic (ROC) curves to confirm the diagnostic performance ($AUC > 0.90$) of an optimal panel of serum metabolites or metabolite ratios.[23] Only when discovery and validation results are consistent, is the candidate biomarker(s) worth pursuing in advanced clinical testing on a much larger scale across multiple centres. Therefore, next phases include developing an ultra-fast assay for quantitative analysis of lead candidate serum biomarkers (ideally with stable-isotope internal standards) with greater overall study power ($n > 1000$) as well as inclusion of more complicated patient cases,[24] such as ambiguous diagnosis or comorbidities to ensure biomarker use is generalizable to a wider clinical population especially PAD patients with type 2 diabetes and/or chronic kidney disease.[25]

Another objective in future studies will be the investigation of the muscle metabolome of PAD patients to better understand the deleterious impacts of chronic muscle ischemia in localized tissues on the pathophysiology of PAD as compared to systemic changes in circulatory metabolism. Moreover, future longitudinal studies of the serum metabolome of CLTI patients who undergo surgical interventions will also comprise investigating metabolites trajectories over time in search of treatment response biomarkers that reflect improved vascular flow, reduced myopathic ischemia or inflammation. This would greatly assist surgeons in the surveillance of patients following revascularization surgery who may be at risk of relapse and poor survivorship.

Direct application of this research into clinical practice can be foreseen as these potential blood-based biomarkers are non-invasive, easy to measure and implement in primary care settings

as a means to monitor asymptomatic high-risk patients (with a predefined optimal threshold as exemplified in the ROC curves) as well as to monitor progression of symptomatic patients for timely referrals and treatment given the poor survivorship of CLTI (50% mortality rate within 5 years) and the fact that its progression is currently unpredictable.[26] Additionally, when comparing PAD patients and non-PAD controls, we found significant insufficiencies of several metabolites and amino acids (*e.g.,* histidine, lysine and creatine), which could be potentially prescribed in the form of dietary supplements to delay disease progression. Consequently, follow up randomized controlled trials on dietary/supplemental interventions are part of future studies to investigate if a favourable outcome in symptomatic or asymptomatic PAD patients is achieved in support of evidence-based nutrition guidelines and precision medicine.

6.5 Thesis conclusion

In summary, novel MSI-NACE-MS methods have been developed and extensively validated in this thesis for analysis of fatty acids and their analogues, PFASs, to support reliable metabolome measurements for epidemiological and clinical applications. This unique metabolomics platform was devised as one with high-throughput and rigorous quality control measures ideal for biomarker discovery and large-scale studies. Additionally, this thesis identified serum NEFA as an objective measurement of fat for new advances in nutritional epidemiology anticipated to allow for routine clinical testing of dietary fat and ω-3 fatty acids status as potential modifiable cardiovascular risk factors in high-risk populations and as part of personalized diet plans for optimal nutrition. Finally, this thesis presented the first characterization of the serum metabolic signature of CLTI in non-diabetic PAD patients. This pilot study revealed metabolite changes associated with the development of PAD (*susceptibility or risk biomarkers*) as well as the prognosis of CLTI (*prognostic biomarkers*), that could potentially also serve in treatment monitoring and patients'

surveillance following surgical revascularization (*monitoring biomarkers*) to assist vascular surgeons in risk assessment and decision making.

6.6 References

1. Azab S, Ly R, Britz-McKibbin P. Robust method for high-throughput screening of fatty acids by multisegment injection-nonaqueous capillary electrophoresis–mass spectrometry with stringent quality control. *Anal Chem*. 2019;91:2329-2336.

2. Ichihara K, Fukubayashi Y. Preparation of fatty acid methyl esters for gas-liquid chromatography. *J Lipid Res*. 2010;51:635-640.

3. Matyash V, Liebisch G, Kurzchalia TV, Shevchenko A, Schwudke D. Lipid extraction by methyl- *tert* -butyl ether for high-throughput lipidomics. *J Lipid Res*. 2008;49:1137-1146.

4. Kuehnbaum NL, Kormendi A, Britz-McKibbin P. Multisegment injection-capillary electrophoresis-mass spectrometry: a high-throughput platform for metabolomics with high datafidelity. *Anal Chem*. 2013;85:10664-10669.

5. DiBattista A, McIntosh N, Lamoureux M, Al-Dirbashi OY, Chakraborty P, Britz-McKibbin P. Metabolic signatures of cystic fibrosis identified in dried blood spots for newborn screening without carrier identification. *J Proteome Res*. 2019;18:841-854.

6. Saoi M, Li A, McGlory C, et al. Metabolic perturbations from step reduction in older persons at risk for sarcopenia: plasma biomarkers of abrupt changes in physical activity. *Metabolites*. 2019;9:134.

7. Blake BE, Pinney SM, Hines EP, Fenton SE, Ferguson KK. Associations between longitudinal serum perfluoroalkyl substance (PFAS) levels and measures of thyroid hormone, kidney function, and body mass index in the Fernald Community Cohort. *Environ Pollut*. 2018;242:894-904.

8. Sunderland EM, Hu XC, Dassuncao C, Tokranov AK, Wagner CC, Allen JG. A review of the pathways of human exposure to poly- and perfluoroalkyl substances (PFASs) and present understanding of health effects. *J Expo Sci Environ Epidemiol*. 2019;29:131-147.

9. Lindstrom AB, Strynar MJ, Libelo EL. Polyfluorinated compounds: past, present, and future. *Environ Sci Technol*. 2011;45:7954-7961.

10. Azab SM, de Souza RJ, Teo KK, Anand SS, Williams NC, Holzschuher J, McGlory C, Phillips SM; Britz-McKibbin P. Serum non-esterified fatty acids have utility as dietary biomarkers of fat intake from fish, fish oil and dairy in women. *J. Lipid Res*. 2020, *61*, 933–944.

11. Brenna JT, Plourde M, Stark KD, Jones PJ, Lin Y-H. Best practices for the design, laboratory analysis, and reporting of trials involving fatty acids. *Am J Clin Nutr*. 2018;108:211-227.

12. Stark KD, Van Elswyk ME, Higgins MR, Weatherford CA, Salem N. Global survey of the omega-3 fatty acids, docosahexaenoic acid and eicosapentaenoic acid in the blood stream of healthy adults. *Prog Lipid Res*. 2016;63:132-152.

13. Ismaeel A, Franco ME, Lavado R, et al. Altered metabolomic PProfile in patients with peripheral artery disease. *J Clin Med*. 2019;8:1463.

14. Roncal C, Martínez-Aguilar E, Orbe J, et al. Trimethylamine-N-Oxide (TMAO) predicts cardiovascular mortality in peripheral artery disease. *Sci Rep*. 2019;9:15580.

15. Bowden JA, Heckert A, Ulmer C, et al. Harmonizing lipidomics: NIST interlaboratory comparison exercise for lipidomics using SRM 1950-metabolites in frozen human plasma. *J Lipid Res*. 2017;58:2275-2288.

16. Chang Y-H, Gregorich ZR, Chen AJ, et al. New mass-spectrometry-compatible degradable surfactant for tissue proteomics. *J Proteome Res*. 2015;14:1587-1599.

17. Zhu G, Sun L, Dovichi NJ. Thermally-initiated free radical polymerization for reproducible production of stable linear polyacrylamide coated capillaries, and their application to proteomic analysis using capillary zone electrophoresis–mass spectrometry. *Talanta*. 2016;146:839-843.

18. Next-generation capillary electrophoresis–mass spectrometry approaches in metabolomics. https://journals-scholarsportal info.libaccess.lib.mcmaster.ca/details/09581669/v43icomplete/1_ncesaim.xml. Accessed May 15, 2020.

19. Morrison KM, Atkinson SA, Yusuf S, et al. The Family Atherosclerosis Monitoring In earLY life (FAMILY) study. *Am Heart J*. 2009;158:533-539.

20. de Souza RJ, Zulyniak MA, Desai D, et al. Harmonization of food-frequency questionnaires and dietary pattern analysis in 4 ethnically diverse birth cohorts. *J Nutr*. 2016;146:2343-2350.

21. Fekete K, Marosvolgyi T, Jakobik V, Decsi T. Methods of assessment of n-3 long-chain polyunsaturated fatty acid status in humans: a systematic review. *Am J Clin Nutr*. 2009;89:2070S-84S.

22. Abdelmagid SA, Clarke SE, Nielsen DE, et al. Comprehensive profiling of plasma fatty acid concentrations in young healthy canadian adults. Aspichueta P, ed. *PLOS ONE*. 2015;10:e0116195.

23. Dunn WB, Broadhurst DI, Atherton HJ, Goodacre R, Griffin JL. Systems level studies of mammalian metabolomes: the roles of mass spectrometry and nuclear magnetic resonance spectroscopy. *Chem Soc Rev*. 2011;40:387-426.

24. Obuchowski NA, Lieber ML, Wians FH. ROC curves in clinical chemistry: uses, misuses, and possible solutions. *Clin Chem*. 2004;50:1118-1125.

25. Clark N. American Diabetes Association. Peripheral arterial disease in people with diabetes. *Diabetes Care*. 2003;26:3333–3341.

26. Varu VN, Hogg ME, Kibbe MR. Critical limb ischemia. *J Vasc Surg*. 2010;51:230-241.