

METABOLITE ANALYSIS FOR DISEASE RISK ASSESSMENT

TARGETED AND NON-TARGETED METABOLITE ANALYSIS FOR
DISEASE RISK ASSESSMENT: MEASURING BIOMARKERS OF SMOKE
EXPOSURE AND HABITUAL DIET

By

Nadine L. Wellington, H.B.Sc.

A Thesis Submitted to the School of Graduate Studies in
Partial Fulfillment of the Requirements for the
Degree of Doctor of Philosophy

McMaster University

© Copyright by Nadine L. Wellington, August 2019

Doctor of Philosophy
Chemistry and Chemical Biology

McMaster University
Hamilton, Ontario

TITLE: Targeted and Non-Targeted Metabolite Analysis for Disease Risk
Assessment: Measuring Biomarkers of Smoke Exposure and Habitual Diet

AUTHOR: Nadine L. Wellington, H.B.Sc.

SUPERVISOR: Professor Philip Britz-McKibbin

NUMBER OF PAGES: xxviii, 184

Lay Abstract

Exposomics is an emerging multidisciplinary science aimed at deciphering the complex interactions that impact human health and gene expression, such as lifestyle choices (*i.e.*, habitual diet) and lifelong environmental exposures. There is growing interest in identifying biomarkers that can be readily measured for chronic disease prevention given an alarming global prevalence of obesity and cardiometabolic disorders, including heart disease, type 2 diabetes and cancer. The research in this thesis focuses on developing new analytical methods for identifying and quantifying metabolites that may allow for better assessments of human health, and has contributed to the development of novel biosensors for the targeted analysis of *N*-acetylneuraminic (sialic) acid and related acidic sugars, as well as high resolution methods for broad spectrum analysis of biotransformed organic contaminants from smoke exposure by GC-MS, and plasma and urinary metabolites that differentiate contrasting Prudent and Western diets and correlate well with self-reported diet records.

Abstract

Exposomics applies metabolomics methods and technologies to the comprehensive analysis of all low molecular weight molecules (< 1.5 kDa) in complex biological samples to characterize the interaction between cellular metabolism and exogenous lifestyle exposures that determine health and quality of life. To fully access the diverse classes of biological molecules related to an individual's metabolic profile, metabolomics frequently requires the use of complementary analytical platforms, and employs targeted and untargeted molecular profiling strategies to identify biomarkers that are clinically relevant to an individual's health status. *Chapter 2* describes a quinoline-based boronic acid biosensor for *N*-acetylneuraminic acid that undergoes a striking binding enhancement under strongly acidic conditions. For the first time, this work allows for direct analysis of acidic sugars with high selectivity when using UV absorbance or fluorescence detection based on formation of a highly stable boronate ester complex with metabolites containing an α -hydroxycarboxylate moiety. *Chapter 3* describes a targeted analysis of 24 different organic contaminants using GC-MS that can serve as biomarkers of recent smoke exposure following search-and-rescue training exercises by firefighters located at three different sites across the province of Ontario. Importantly, skin and possible respiratory uptake of various polycyclic aromatic hydrocarbons, methoxyphenols, and resin acids was confirmed by peak excretion of several wood smoke biomarkers in urine within 6 h following acute exposure. *Chapter 4* applied a cross-platform metabolomics strategy based on CE-MS and GC-MS in order to identify and validate dietary biomarkers in matching plasma and urine samples collected from healthy participants in the pilot Diet and Gene Interaction Study (DIGEST). For the first time, we demonstrate that a panel of metabolites can serve as reliable biomarkers following contrasting Prudent and Western diets over 2 weeks of food provisions, which correlated well with self-reported diet records. This work paves

the way for the development of objective biomarkers for accurate assessment of wood smoke exposures, as well as complex dietary patterns as required for new advances in occupational health and nutritional epidemiology.

Acknowledgments

My supervisor, Philip Britz-McKibbin, has been a consistent source of optimism, advice, guidance and constructive criticism throughout the highs and lows of this doctorate. Britz lab members present and past are a wonderfully kind and diverse group of people who have been a great source of support, both academically and personally. I would like to say a special thank you to Dr. Meera Shanmuganathan, Dr. Alicia DiBattista, Dr. Karen P. Lam, Dr. Adriana Nori De Macedo, Dr. Mai Yamamoto, Dr. Michelle Saoi, Jonathon Bloomfield, Sandi Azab, Jennifer Wild, Stellena Mathiapparanam, and Ritchie Ly.

I would like to thank my committee members Dr. David Emslie and Dr. José Moran-Mirabal for their support and unique perspectives on my research. This thesis also would not have been possible without the collaboration and support of several individuals and I would like to give a special thanks to Dr. Hilary Jenkins, Dr. Russell De Souza, Dr. Mateen Shaikh, Dr. Sonia Anand, Lorraine Shaw, Don Shaw, Dr. Sujana Fernando, and Dr. Michael Zulyniak.

Thank you to the experts, especially the many I have met over the years at conferences and professional events, who share wisdom, encouragement, and advice with eager, and occasionally apprehensive, graduate students. These interactions always served as a reminder of my passion and motivations to perform translational chemistry research.

To my angel, Mia: my challenges during this period have been your challenges as well. You have grown into a magnificent young girl and have been my foremost motivation in this pursuit. As we move on to the next chapter, I am so proud and happy to have you by my side. Thank you for your patience, laughter, smiles, love and hugs. Love you beyond the sun, moon, and stars.

To Ricardo, you have given me strength and love beyond what I could have ever imagined, and I am incredibly thankful.

To my mother, I love you and thank you for everything. I am immensely thankful to have the support and encouragement of my loving family and friends. It has been fundamental to every positive change I have made over these years. I will simply say:

Thank You, I Love You



Table of Contents

Lay Abstract	iii
Abstract.....	iv
Acknowledgments	vi
Table of Contents.....	viii
List of Figures.....	x
List of Tables.....	xxi
Index of Abbreviations and Symbols.....	xxiii
Declaration of Academic Achievement.....	xxix
Chapter 1: Introduction.....	1
1.1 The Evolution of Risk Assessment in Human Health.....	1
1.1.1 Biomarkers in Ancient History.....	1
1.1.2 Creatinine and its Ubiquity in Clinical Assessment.....	2
1.1.3 Metabolism, Molecules, and the Evolution of Omics and Precision Medicine.....	4
1.2 Individual and Epidemiological Assessments of Risk.....	7
1.2.1 Traditional Patient Assessment Standards and Protocols.....	7
1.2.2 Biological Surveillance and the Challenges of Precision Medicine....	9
1.3 The Workflow and Statistical Methodologies of Molecular Biomarker Discovery.....	12
1.3.1 What Exactly is a Molecular Biomarker?	12
1.3.2 Human Biological Matrices Used in Biomarker Discovery	14
1.3.3 Major Instrumental Platforms in Metabolomics.....	15
1.3.4 Data Workflows and Statistical Methodologies in Molecular Biomarker Discovery.....	22
1.4 Multi-segment Injection Capillary Electrophoresis (MSI-CE-MS)	30
1.4.1 Advantages and Applications of MSI.....	31
1.5 Thesis Contributions to Targeted and Untargeted Metabolite Profiling for Disease Risk Assessment	34
1.5.1 Developing Chemical Sensors for Rapid Assessment of Health Status	36
1.5.2 Revealing Metabolic Impacts from Wood Smoke Components in Firefighting.....	37
1.5.3 Identifying Nutrimetabolomic Markers of Food Intake and Diet Quality	38
1.6 References.....	39

Chapter 2: The Anomalous Stability of Isoquinoline Boronate Ester Complexes Under Acidic Conditions: A Biosensor for Sialic Acid Determination	50
2.1 <i>Abstract</i>	51
2.2 <i>Introduction</i>	51
2.3 <i>Results and Discussion</i>	53
2.4 <i>References</i>	60
2.5 <i>Supplemental Information</i>	62
Chapter 3: Elucidating the Mechanisms of Smoke Exposures in Municipal Firefighters: A Multi-Centre Training Exercise Study.....	68
3.1 <i>Abstract</i>	69
3.2 <i>Introduction</i>	70
3.3 <i>Results and Discussion</i>	72
3.4 <i>References</i>	88
3.5 <i>Supplemental</i>	91
Chapter 4: Metabolic Trajectories Following Contrasting Western and Prudent Diets from Food Provisions: Robust Biomarkers of Short-term Changes in Habitual Diet.....	106
4.1 <i>Abstract</i>	107
4.2 <i>Introduction</i>	108
4.3 <i>Results</i>	110
4.4 <i>Discussion</i>	129
4.5 <i>References</i>	138
4.6 <i>Supplemental</i>	142
4.7 <i>Supplemental References</i>	175
Chapter 5: Towards Accessing the Exposome for Precision Interventions	177
5.1 <i>Thesis Discoveries and Contributions</i>	178
5.2 <i>References</i>	184

List of Figures

Chapter 1

- Figure 1.1:** *The omics cascade and the biological processes that influence end point metabolites* 7
- Figure 1.2:** *The exogenous and endogenous determinants of phenotype relevant to human health.* 11
- Figure 1.3:** *Projection of the 3 dimensions of information provided when chromatographic/electrophoretic methods (CE, LC, GC) are used in tandem with MS. The sample is resolved into a chromatogram in the first dimension that is plotted by mass-to-charge (m/z) ratios in the second-dimension mass spectra. Relative response is shown in the third dimension.* 19
- Figure 1.4:** *Orthogonal coupling between an ambient pressure ESI spray needle (containing coaxial CE capillary) and MS source that reduces ionization suppression by nonionized species by sampling charged species via an applied electric field. Used with permission from the American Chemical Society.¹³⁹* 19
- Figure 1.5:** *Untargeted metabolomics workflow for the acquisition, cleanup, and processing of data in biomarker studies.* 24
- Figure 1.6:** *(A) Schematic showing 7 injections of MSI-CE-ESI method for high-throughput sample analysis. Six randomized analytical samples are run with a pooled QC in a randomized position. (B) All 7 injections resolve in parallel and 4 metabolites are separated by distinct native electrophoretic mobilities. (C) Total ion electropherogram of analytical run in ESI+ mode. Small, charge-dense cations migrate rapidly, followed by larger, less densely charged ions (e.g., creatinine). Neutral compounds elute with the bulk electroosmotic flow (EOF). (D) Extracted ion electropherograms of 4 urine metabolites resolved by mass-to-charge ratio (m/z) in QTOF-MS with 7 distinct peaks corresponding to 6 analytical samples and a randomized QC.* 31
- Figure 1.7:** *(A) 6 injections of a serially-diluted (1, 2, 5 and 10 times) pooled sweat QC sample plus a blank produces (B) an electropherogram that identifies the metabolite citrulline (m/z 176.1030) as an authentic feature with no background (blank) signal and (inset) excellent linearity ($R^2 = 0.989$) and reproducibility (RSD = 4.2%) among triplicate experiments. A dilution trend filter is employed in MSI-CE-MS untargeted*

metabolomics studies as a rapid first-pass exclusion filter of (C) spurious peaks from chemical and technical noise that do not show effects of sample dilution. Adapted from Nori de Macedo et al.⁷³ 33

Chapter 2

Figure 2.1: A 65-fold enhancement in binding affinity of IQBA to Neu5Ac is achieved under strongly acidic as compared to neutral pH conditions as shown in (A) UV absorbance spectra overlay and (B) binding isotherms based on a 1:1 dynamic complexation model. All studies were performed in triplicate with 40 μ M IQBA in 40 mM phosphate buffer with absorbance changes monitored at 335 nm. These conditions are also optimal for IQBA native fluorescence properties (inset of A) with an enhancement at 377 nm ($\Delta\lambda = 42$ nm) upon Neu5Ac binding. (C) ¹³C-NMR spectra of free Neu5Ac (red) and Neu:IQBA complex confirms the importance of the α -hydroxycarboxylate moiety in stable boronate ester formation based on a -2.5 ppm chemical shift of the C1 carbonyl carbon (175.8 ppm), and a -0.80 ppm shift in the C2 resonance (96.1 ppm). 55

Figure 2.2: A comparison of linear calibration curves for Neu5Ac quantification when using IQBA with UV absorbance ($\lambda_{max} = 335$ nm) and native fluorescence detection ($\lambda_{em} = 377$ nm) in 40 mM phosphate, pH 3.0. Both detection methods offered excellent linearity and acceptable technical precision (CV \approx 12%, n = 18), but fluorescence provided greater sensitivity and lower detection limits as compared to UV absorbance. 59

Figure 2S.1: 2D structures and 3D conformations comparing the thermodynamic stability for two different ternary boronate ester complexes involving IQBA-Neu5Ac-phosphate after energy minimization (MM2) with computer molecular modeling. Each structure compares the predominate ligand binding mode for Neu5Ac involving (A) the α -hydroxycarboxylate anion or (B) the terminal vicinal diol (C8, C9) of glycerol moiety at pH 3. In both cases, the quinolinium ion is fully charged and forms a stable ion pair/zwitterion with the tetrahedral (sp^3 -hybridized) boronate anion, where conformer (A) is significantly more

stable than conformer (B) by -29.1 kcal/mol. Additionally, each complex is not stable (total energy > +20 kcal/mol) if forming a ring-strained sp²-hybridized (trigonal planar) boronic ester under acidic conditions, as well as being far less stable when complexation occurs under neutral or alkaline conditions where the quinoline nitrogen is deprotonated (i.e., neutral)..... 66

Figure 2S.2: ¹¹B-NMR spectra overlay comparing the relative distribution of major boron species at equilibrium for IQBA in 40 mM phosphate prepared in D₂O as a function of buffer pH and Neu5Ac. There are three distinct boron chemical environments for IQBA in phosphate buffer under acidic conditions (pH 3.0, red trace), including a predominantly sp² hybridized boron (i.e., trigonal planar; electron deficient/neutral boronic ester) corresponding to a low-field resonance peak (I, δ = 28.7 ppm), and a less abundant sp³-hybridized boron corresponding to a low-field resonance peak (II, δ = -5.5 ppm) indicative of the tetrahedral boronate anion. Additionally, there exists a third boron resonance near δ = 19.8 ppm (III) that likely corresponds to a phosphoboronate ester complex formed between IQBA and excess phosphate in solution. As expected, when acquiring the ¹¹B-NMR spectra for IQBA under neutral pH conditions (pH 7.0), there is a noticeable increase in the relative abundance of sp³-hybridized boron (II, δ = 9.6 ppm) as compared to sp²-hybridized boron (I, δ = 26.0 ppm) in solution with a low field shift in both boron resonances. However, the inset depicts the emergence of a fourth peak upon addition of excess Neu5Ac to IQBA in phosphate at pH 3.0 (IV, δ = 6.2 ppm) likely corresponding to the ternary IQBA-Neu5Ac-phosphate complex. Importantly, Neu5Ac complexation results in a pronounced increase in the relative abundance of the tetrahedral boronate anion (II) and a corresponding decrease in trigonal planar boronate ester species (I) even under acidic conditions that is comparable to IQBA under neutral conditions. This process reflects the increase in the apparent acidity (shift to lower pK_a) of the cyclic boronate ester complex upon Neu5Ac binding as compared to free IQBA..... 67

Chapter 3

- Figure 3.1:** (A) Triple-stacked bar plot of total ambient volatile and semivolatile MP, RA, and PAH (with carcinogen class indicated) wood smoke markers retained by the caged outer Twister (blue) and the personal pump assembly (purple) with a Teflon filter and sorbent XAD-2 resin connected in series (2 L/min) for Site B, H and O. Red circles indicate carcinogenic PAH. 1: guaiacol; 2: methylguaiacol; 3: ethylguaiacol; 4: syringol; 5: propylguaiacol; 6: eugenol; 7: methylsyringol; 8: isoeugenol; 9: ethylsyringol; 10: propylsyringol; 11: naphthalene (Group 2B); 12: acenaphthylene; 13: acenaphthene; 14: fluorene; 15: phenanthrene/anthracene; 16: fluoranthene; 17: pyrene; 18: benz[a]anthracene (Group 2B); 19: chrysene; 20: benzo[b]fluoranthene (Group 2B); 21: benzo[k]fluoranthene (Group 2B); 22: benzo[a]pyrene (Group 1); 23: indeno[1,2,3-cd]pyrene (Group 2B); 24: dibenz[a,h]anthracene (Group 2A); 25: benzo[ghi]perylene; 26: pimaric acid; 27: sandaracopimaric acid; 28: isopimaric acid; 29: abietic acid; 30: dehydroabietic acid; and 31: 7-oxodehydroabietic acid. (B) 2D PCA scores plot showing heterogeneous intercity and interindividual exposures based on active sampling pumps (n = 16). Site B showed relatively low exposures to MP, PAH and RA presumably related to the large volume multi-storey burn house used. Site H was exposed to smoke high in syringols, while Site O recorded higher amounts of guaiacols and RA..... 74
- Figure 3.2:** (A) 2D PCA scores plot showing significantly lower overall amounts of 22 targeted smoke markers recovered from the passive samplers inside the SCBA when compared the caged external Twister (suit). Propylguaiacol (PropG) and ethylsyringol (EthG) showed the highest penetration of the SCBA mask, consistent with high concentrations measured in smoke by passive and active sampling devices (**Figure 3.1**)..... 79
- Figure 3.3:** (A) Schematic detailing urine collection at 4 distinct time intervals to monitor the kinetics of biomarker excretion and identify compounds elevated due to the smoke exposure. (B) Time-sequenced 2D/HCA heat map of creatinine-normalized urinary wood smoke markers detected by GC-MS. Methylguaiacol (MeG, $p < 4.0 \text{ E-}4$), ethylguaiacol (EthG, $p < 4.0 \text{ E-}3$), syringol ($p < 5 \text{ E-}3$), and 1-hydroxynaphthalene (1-OH-Nap, $p < 0.010$) were highly responsive urine markers of

smoke exposure in the 24-hour period after the fire however, elevated concentration of all markers was observed during the immediate 6-hour period post burn, showing significant uptake of ambient chemicals at the fireground despite protective gear. (C) Boxplots (ANOVA) of top discriminating smoke marker MeG showing significant increase in excretion in the 0-6 h period post exposure..... 82

Chapter 4

Figure 4.1: (A) Overview of study design in this parallel two-arm dietary intervention study involving participants from DIGEST (n = 42) who were assigned a Prudent or Western diet over a 2-week period with matching urine and plasma samples collected at baseline and post-intervention. (B) A 2D heat map with hierarchical cluster analysis (HCA) of the plasma metabolome that were consistently measured in the majority of participants, including non-targeted analysis of polar/ionic metabolites by MSI-CE-MS and total fatty acids by GC-MS. (C) A 2D heat map with hierarchical cluster analysis of the urine metabolome that were consistently measured in majority of participants, including non-targeted analysis of polar/ionic metabolites and targeted electrolytes by CE with indirect UV absorbance. A generalized log transformation and autoscaling was performed on metabolomic datasets together with creatinine normalization for single-spot urine specimens. Participants classified as having a predominate Western diet at baseline who were then assigned a Prudent diet are designated as “W-P” (n = 24), whereas “P-W” (n = 18) refers to participants had lower Western diet score at baseline, but were assigned a Western diet. III

Figure 4.2: (A) Metabolomics data workflow for the identification and quantification of biomarkers of a provisional Prudent diet (e.g., proline betaine annotated based on its m/z:RMT) when using full-scan data acquisition. (B) Multiplexed separations by MSI-CE-MS based on serial injection of seven plasma filtrate (or diluted urine) samples within a single run, including paired samples from each DIGEST participant (i.e., baseline/post-treatment) together with a pooled sample as QC for assessing technical precision and long-term signal

drift. High resolution MS under positive ion mode detection allows for determination of most likely molecular formula for unknown cation (i.e., protonated molecular ion), whereas (C) MS/MS spectra is used for its structural elucidation when compared with an authentic standard. (D) Quantification for metabolites is then performed by external calibration when using an internal standard (Cl-Tyr) for data normalization by MSI-CE-MS. (E) A control chart for ProBet from pooled urine samples as QC analyzed in random positions in every run demonstrates acceptable technical precision over 3 days.

..... 114

Figure 4.3: *Paired supervised multivariate data analysis of (A) plasma and (B) creatinine-normalized urine metabolomic data based on orthogonal partial least-squares-discriminant analysis (OPLS-DA) using the ratio of ion responses or concentrations for metabolites following 2 weeks of food provisions to their baseline habitual diet values. 2D scores plot highlight differences in the overall metabolic phenotype from matching biofluids collected from DIGEST participants assigned to a Prudent (W-P) as compared to a Western (P-W) diet based on a sub-set of metabolites identified from S-plots, as well as univariate statistical analysis as shown in box-whisker plots for top-ranked metabolites significantly different between the treatment arms ($p < 0.05$).*..... 118

Figure 4.4: *Metabolic trajectories for two distinctive biomarkers measured consistently in both plasma and urine specimens that increase significantly following a provisional Prudent diet (W-P) as compared to an assigned Western diet (P-W), namely Me-His and ProBet. Both metabolites were not different at baseline but undergo notable changes after two weeks of food provisions ($q < 0.05$, FDR) with concentrations moderately correlated ($r > 0.400$) to self-reported diet involving health-promoting foods from a Prudent diet. Overall, good dietary adherence was demonstrated for the majority of DIGEST participants with few exceptions (labeled on plots) who had metabolite phenotypes inconsistent with their assigned diet group.*..... 129

Figure 4S.1: *A CONSORT flow diagram outlining selection criteria used in a parallel two-arm randomized clinical trial involving participants from the DIGEST study, where metabolomic*

analyses was performed unblinded on paired serum and urine samples collected at baseline and 2 weeks following a provisional Prudent or Western diet. Overall, 73 of the 84 participants who completed DIGEST had available specimens and complete food records. However, in order to maximize the effect size of this short-term dietary intervention, metabolomic analyses was performed only on a subset of participants ($n = 42$) who had contrasting habitual diets at baseline as evaluated based on an aggregate diet quality score. 161

- Figure 4S.2:** Overview of three major analytical platforms for nontargeted and targeted metabolite/electrolyte profiling of matching plasma and urine specimens from DIGEST participants at baseline and 2 weeks following an assigned Prudent or Western diet from food provisions. MSI-CE-MS was used as the major format for metabolomic analyses of a wide range of polar/ionic metabolites from both plasma filtrate and diluted urine samples when using a stringent data workflow for metabolite authentication with quality control. Also, CE with (indirect/direct) UV absorbance detection was used for targeted analysis of electrolytes in urine, including inorganic cations (e.g., sodium) and anions (e.g., nitrate), whereas GC-MS was applied for targeted analysis of total (hydrolyzed) fatty acids as their FAMES from plasma extracts. 162
- Figure 4S.3:** 2D scores plots from PCA and control charts for recovery standards that highlight the good technical precision as compared to biological variance of metabolomic data from three instrumental platforms, including (A) 84 authenticated metabolites in urine after QC-based batch-correction, (B) 80 polar/ionic metabolites from plasma using MSI-CE-MS and (C) 18 plasma (total) fatty acids using GC-MS. 163
- Figure 4S.4:** 2D scores plot from PCA and HCA/heat map summarizing changes in habitual diet (W-P and P-W) based on 20 macro-/micronutrient categories self-reported food records from DIGEST participants following 2 weeks of contrasting food provisions as compared to their baseline diet. 164
- Figure 4S.5:** Series of volcano plots for plasma (A, B) and urine (C, D) metabolome data highlighting few differences in the metabolic phenotype among DIGEST participants ($n = 42$) at baseline as compared to major changes following two weeks of contrasting diets from food provisions based on a

minimum threshold for significance (mean FC > 1.3, $p < 0.05$) between Prudent and Western assigned groups. Abbreviations refer to DHBA (dihydroxybenzoic acid), ProBet (proline betaine), Me-His (3-methylhistidine), 3-OH-BA (3-hydroxybutyric acid), dC0 (deoxycarnitine), Pro (proline), OH-PCA (hydroxypipicolinic acid), ImPA (imidazolepropionic acid), Ent-G (enterolactone glucuronide), DHBA (dihydroxybenzoic acid) and Me-G (methylguanidine), whereas standard notation is used for plasma fatty acids and unknown ions are denoted by their accurate mass (m/z)..... 165

Figure 4S.6: (A) Metabolomics data workflow for the identification and quantification of biomarkers of a provisional Prudent diet (e.g., 3-methylhistidine annotated based on its m/z :RMT) when using full-scan data acquisition. (B) Multiplexed separations by MSI-CE-MS based on serial injection of seven plasma filtrate (or diluted urine) samples within a single run, including paired samples from each DIGEST participant (i.e., baseline/post-treatment) together with a pooled sample as QC for assessing technical precision and long-term signal drift. High resolution MS under positive ion mode detection allows for determination of most likely molecular formula for unknown cation (i.e., protonated molecular ion), whereas (C) MS/MS spectra is used for its structural elucidation when compared with an authentic standard. (D) Quantification of metabolites is then performed by external calibration when using an internal standard (Cl-Tyr) for data normalization by MSI-CE-MS. (E) A control chart for Me-His from pooled urine samples as QC analyzed in random positions in every run demonstrates acceptable technical precision over 3 days..... 166

Figure 4S.7: Putative identification (level 2) of three unknown cationic metabolites ($[M-H]^+$) detected in urine specimens from DIGEST participants that were significantly elevated in assigned Prudent urine samples ($p < 0.05$) as compared to Western diet following 2 weeks of food provisions, namely (A) 5-hydroxypipicolinic acid (OH-PCA, m/z 146.081), (B) imidazole propionic acid (ImPA, m/z 141.067) and (C) Valinyl-valine (Val-Val, m/z 217.155) based on their characteristic MS/MS spectra (optimal collision energy at 20 V) under positive ion mode conditions..... 167

- Figure 4S.8:** Unambiguous (level 1) and putative identification (level 2) of three unknown anionic metabolites ($[M-H]^-$) detected in urine specimens from DIGEST participants that were decreased (acesulfame K) or elevated (enterolactone glucuronide and dihydroxybenzoic acid) in assigned Prudent ($p < 0.05$) as compared to Western diet groups following 2 weeks of food provisions, namely (A) acesulfame K (ASK, m/z 161.987), (B) enterolactone glucuronide (Ent-G, m/z 473.145) and (C) dihydroxybenzoic acid (DHBA, m/z 153.020) based on their characteristic MS/MS spectra under negative ion mode conditions. Identification was derived from comparison of MS/MS spectra with a standard as shown in mirror plot for ASK together with spiking into urine sample to confirm co-migration. The likely stereochemistry for DHBA was deduced from comparison of *in silico* MS/MS spectra when using HMDB, whereas Ent-G was tentatively identified based on comparison with published MS/MS spectra [Johnson et al. *Metabolites* **2013**, 3: 658-672]. 168
- Figure 4S.9:** Unambiguous identification (level 1) of total (hydrolyzed) fatty acids as their FAME derivatives from plasma extracts when using GC-MS, including resolution of low abundance *trans*-isomer (linoelaidic acid, C18:2n-6*trans*) from major *cis*-isomer (linoleic acid, C18:2n-6*cis*), including detection of minor saturated fatty acids (myristic acid, C14:0). Mirror plots for EI-MS spectra show excellent matches when comparing FAMEs detected in plasma extracts as compared to their references in the NIST database, where the base peak ions correspond to the quantifier ions monitored for saturated and polyunsaturated fatty acids. Overall, plasma total C18:2n-6*trans* was only 0.34% of its major stereoisomer C18:2n-6*cis* which also represents the most abundant fatty acid measured in circulation. 169
- Figure 4S.10:** Top-ranked single and ratiometric metabolites for differentiation of DIGEST participants assigned a Prudent (W-P, $n = 24$) or Western (P-W, $n = 18$) diet following 2 weeks of food provisions when using receiver operating characteristic (ROC) curves. All metabolites were log-transformed, whereas urine metabolites were normalized to creatinine following a QC-based batch correction. Overall, there was good discrimination of contrasting dietary patterns ($AUC > 0.820$; $p < 0.001$) as shown for (A) plasma

	<i>proline betaine (ProBet), (B) plasma 3-methylhistidine to α-linoleic acid (MeHis/C18:3n-6) ratio, (C) urinary hydroxypipicolinic acid (OH-PCA), and (D) urinary hydroxypipicolinic acid to sodium ratio (OH-PCA/Na).</i>	170
Figure 4S.11:	<i>Top-ranked plasma metabolites associated with contrasting Prudent (W-P) and Western (P-W) diets from food provisions when using glog-transformed ion responses measured at baseline (habitual diet, 0) and following food provisions (2 weeks) for DIGEST participants (n = 42). Metabolic trajectories from time course studies were ranked based on Hotelling-T² values when using multivariate empirical Bayes analysis of variance (MEBA) together with student's t-test to confirm statistical significance (p < 0.05).</i>	171
Figure 4S.12:	<i>Top-ranked urinary metabolites associated with contrasting Prudent (W-P) and Western (P-W) diets from food provisions when using glog-transformed ion responses normalized to creatinine measured at baseline (habitual diet, 0) and following food provisions (2 weeks) for DIGEST participants (n = 42). Metabolic trajectories from time course studies were ranked based on Hotelling-T² values when using multivariate empirical Bayes analysis of variance (MEBA) together with student's t-test to confirm statistical significance (p < 0.05), with the exception for DHBA.</i>	172
Figure 4S.13:	<i>Linear correlation plots highlighting the strong association between fasting plasma concentrations of (A) Me-His and (B) ProBet and their creatinine-normalized concentrations measured independently from matching single-spot urine samples for DIGEST participants collected at baseline and then following 2 weeks of food provisions.</i>	173
Figure 4S.14:	<i>2D heat maps and correlation matrices for top-ranked (A) plasma and (B) urinary metabolites associated with contrasting diets from food provisions when using a Pearson correlation analysis on glog-transformed data. Distinctive clusters of metabolite classes suggest common dietary sources and/or biochemical pathways for their regulation, such as circulating ketone bodies (kLeu, kVal, 3-OH-BA), fatty acids (C14:0, C15:0, C18:2, C18:3) and amino acids/carnitines (C0, Pro, Ala) in plasma, as well as plant-derived biotransformed phenols (Ent-G, DHBA) and imidazole metabolites (Me-His, ImPA) in urine. In many cases, urinary metabolites reflective of recent dietary</i>	

patterns were broadly co-linear with other compounds (e.g., OH-PCA, Me-His, ProBet and unknown ion, m/z 276) or had modest correlations to other compounds overall. 175

List of Tables

Chapter 1

- Table 1.1:** *Clinical types and uses of risk-related biomarkers (adapted from Burke, 2016¹¹⁵)* 14
- Table 1.2:** *Metabolomics publications during 2008-2018 by biological matrix as defined by sample type analyzed*..... 15

Chapter 2

- Table 2.1:** *A comparison of the binding affinity and selectivity of IQBA highlighting the anomalous binding enhancement under acidic conditions for Neu5Ac as compared to various other sugars, polyols and sialic acid analogs*..... 57

Chapter 3

- Table 3.1:** *(Mean \pm SD) concentrations (ng/cm²) of MP, PAH, and RA recovered from the outer and inner SCBA lens of firefighters (n = 15)*..... 76
- Table 3.2:** *Top discriminating urine markers of smoke exposure in 24-hr baseline and 24-hr post-exposure urine samples (0-6 h, 6-12 h, 12-24 h) collected from firefighters (n = 15)*..... 86
- Table 3S.1:** *(Mean \pm SD) recovery (ng) of methoxyphenols (MP), polycyclic aromatic hydrocarbons (PAH), and resin acids (RA) recovered from outer caged Twisters (n = 8) attached to the turnout coat compared to within the SCBA (n = 7) of firefighters*..... 100
- Table 3S.2:** *(Mean \pm SD) ambient concentrations (μ g/m³) of methoxyphenols (MP), polycyclic aromatic hydrocarbons (PAH), and resin acids (RA) captured by the personal air pumps (n = 16) of firefighters*..... 101
- Table 3S.3:** *(Mean \pm SD) ambient concentrations (ng/cm²) of methoxyphenols (MP), polycyclic aromatic hydrocarbons (PAH), and resin acids (RA) on the cheeks and arms of firefighters (n = 15)* 102
- Table 3S.4:** *(Mean \pm SD) urine concentrations (ng/g Crn) of methoxyphenols (MP), polycyclic aromatic hydrocarbons (PAH), and resin acids (RA) in 24-hr baseline and 24-hr post-*

exposure urine samples (0-6 h, 6-12 h, 12-24 h) excreted by firefighters (n = 15) after wood smoke exposure..... 104

Chapter 4

Table 4.1:	<i>Major changes in dietary patterns after a 2-week assigned Prudent and Western diet relative to baseline habitual diet of DIGEST participants (n = 42) based on self-reported diet records.</i>	<i>115</i>
Table 4.2:	<i>Top-ranked plasma metabolites associated with a 2 week Prudent or Western provisional diet on healthy participants (n = 42) when using time series MEBA, mixed ANOVA and a partial correlation analysis.</i>	<i>120</i>
Table 4.3:	<i>Top-ranked creatinine-normalized metabolites associated with a 2 week Prudent or Western provisional diet on healthy participants (n = 42) when using time series MEBA, mixed ANOVA and a partial correlation analysis.</i>	<i>124</i>
Table 4S.1:	<i>Baseline group characteristics of a cohort of healthy participants (n = 42) recruited in a two-arm parallel randomized clinical trial to compare the effects of a contrasting diet over 2 weeks (Western or Prudent) from food provisions reflecting changes in habitual dietary patterns.....</i>	<i>153</i>
Table 4S.2:	<i>Summary of representative and reliably measured urinary and plasma metabolites detected in a majority of DIGEST participants patients that are annotated based on their accurate mass (m/z), relative migration time (RMT), ionization mode (p = ESI⁺, n = ESI⁻), most likely molecular formula, compound name, confidence level for identification, and technical precision from QC measurements.</i>	<i>155</i>

Index of Abbreviations and Symbols

Δm	Mass difference
7ODA	7-oxodehydroabiatic acid
3-OH-BA	3-hydroxybutyric acid
2/3/9-OH-Flu	2-, 3-, or 9-hydroxyfluorene
1/2-OH-Nap	1- or 2-hydroxynaphthalene
2/3/4-OH-Phen	2-, 3-, or 4-hydroxyphenanthrene
1-OH-Pyr	1-hydroxypyrene
AA	Abietic acid
ACN	Acetonitrile
ADP	Adenine diphosphate
Ala	Alanine
ANOVA	Analysis of variance
APCI	Atmospheric pressure chemical ionization
APPI	Atmospheric pressure photoionization
ASK	Acesulfame potassium
ATP	Adenine triphosphate
AUC	Area under the receiver-operating characteristic curve
BA	Boronic acid
BaA	Benz[a]anthracene
BaP	Benzo[a]pyrene
BbF	Benzo[b]fluroanthene
BkF	Benzo[k]fluroanthene
BCAA	Branched chain amino acid
BCE	Before common era
BGE	Background electrolyte

BghiP	Benzo[ghi]perylene
BHT	Butylated hydroxytoluene
BMI	Body mass index
BP	Blood pressure
BRCA	Breast cancer gene
C0	<i>L</i> -carnitine
C18	Octadecylsilane
α -CD	Alpha-cyclodextrin
CE	Capillary electrophoresis
CE-MS	Capillary electrophoresis-mass spectrometry
CE-UV	Capillary electrophoresis-ultraviolet detection
CFM-ID	Competitive Fragmentation Modeling for Metabolite Identification
CID	Collision-induced dissociation
Cl-Tyr	3-chloro- <i>L</i> -tyrosine
CONSORT	Consolidated Standards of Reporting Trials
COSY	Correlation spectroscopy
CP	Creatine phosphate
Crn	Creatinine
CV	Coefficient of variance/relative standard deviation
CVD	Cardiovascular disease
CYP	Cytochrome P450 enzyme family
DA	Dehydroabiatic acid
DahA	Dibenz[a,h]anthracene
DASH	Dietary Approaches to Stop Hypertension
dC0	Deoxycarnitine
DCM	Dichloromethane

DHBA	Dihydroxybenzoic acid
DIGEST	Diet and Gene Interaction Study
DMG	Dimethylglycine
DMSO	Dimethylsulfoxide
DNA	Deoxyribonucleic acid
EDTA	Ethylenediaminetetraacetic acid
EI	Electron impact ionization
Ent-G	Enterolactone glucuronide
EOF	Electroosmotic flow
ESI ^{+/-}	Electrospray ionization positive mode/negative mode
EthG	Ethylguaiacol
EthS	Ethylsyringol
FA	Fatty acid
FAME	Fatty acid methyl ester
FC	Fold change
FDR	False discovery rate
FFQ	Food frequency questionnaire
GC	Gas chromatography
GC-MS	Gas chromatography-mass spectrometry
HCA	Hierarchical cluster analysis
HDL	High-density lipoprotein
HEPES	(4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HILIC	Hydrophilic interaction liquid chromatography
HMDB	Human Metabolome Database
HPLC	High-performance liquid chromatography
hsCRP	High-sensitivity C-reactive protein

HSQC	Heteronuclear single quantum coherence spectroscopy
IcdP	Indeno[1,2,3-cd]pyrene
IL-8	Interleukin 8
ImPA	Imidazole propionic acid
IPA	Isopimaric acid
IQBA	Isoquinoline boronic acid
kLeu	Ketoleucine
kVal	Ketovaline
LC	Liquid chromatography
LDL	Low-density lipoprotein
LOD	Limit of detection
LOQ	Limit of quantification
[M-H] ⁺	Molecular ion with positive charge
<i>m/z</i>	Mass-to-charge ratio
Me-His	3-methyl- <i>L</i> -histidine
MEBA	Multivariate empirical Bayes analysis
MeOH	Methanol
MP	Methoxyphenol
MRI	Magnetic resonance imaging
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
MSI-CE-MS	Multi-segment injection-capillary electrophoresis-mass spectrometry
MSTFA	<i>N</i> -methyl- <i>N</i> -(trimethylsilyl)trifluoroacetamide
MW	Molecular weight
MWAS	Metabolome-wide association studies
MWCO	Molecular weight cut-off

NACE	Non-aqueous capillary electrophoresis
NaCl	Sodium chloride
NAM	<i>N</i> -acetylmannosamine
NBA	2-naphthaleneboronic acid
NDS	1,5-naphthalenedisulfonate
Neu5Ac	<i>N</i> -acetylneuraminic acid
NHANES	National Health and Nutrition Examination Survey
NMR	Nuclear magnetic resonance
OH-PCA	Hydroxyproline
OPLS-DA	Orthogonal projections to latent structures - discriminant analysis
OR	Odds ratio
P-W	Prudent to Western diet change
PA	Pimaric acid
PAH	Polycyclic aromatic hydrocarbons
PBA	Phenylboronic acid
PCA	Principle component analysis
PDMS	Polydimethylsiloxane
PF	Protection factor
pK _a	Acid dissociation constant
PLS-DA	Partial least-squares discriminant analysis
PPE	Personal protective equipment
ProBet	Proline betaine
PropG	Propylguaiacol
PropS	Propylsyringol
QC	Quality control sample
QTOF	Quadrupole time-of-flight

R ²	Coefficient of determination
RA	Resin acid
RCT	Randomized controlled trials
RMT	Relative migration time
RNA	Ribonucleic acid
ROC	Receiver operating characteristic
RP	Reversed-phase chromatography
S/N	Signal-to-noise ratio
SCBA	Self-contained breathing apparatus
SIM	Selected ion monitoring
SPA	Sandaracopimaric acid
SPE	Solid phase extraction
TMAO	Trimethylamine <i>N</i> -oxide
TOF-MS	Time-of-flight-mass spectrometry
U(H)PLC-MS	Ultra-(high-)performance-liquid chromatography-mass spectrometry
UV(/Vis)	Ultraviolet(/visible) spectroscopy
Val-Val	Valyl-valine
V _{cap}	Capillary voltage
W-P	Western to Prudent diet change

Declaration of Academic Achievement

All research, data processing, statistical analysis, structural identification, and interpretation described in this thesis was performed by the author, with important experimental contributions from collaborators and colleagues. In *Chapter 2*, ^1H , ^{13}C , and ^{11}B NMR data was acquired by Dr. Hilary Jenkins. In *Chapter 3*, extraction and targeted GC-MS measurements of smoke markers in air, wipe and urine samples was performed by Jonathon Bloomfield. Urine thiocyanate measurements were provided by Ritchie Ly. Jonathon and Alicia Mell contributed GC-MS measurements of total plasma fatty acids in *Chapter 4*. Measurements of height, weight, blood pressure and other clinical data in *Chapter 4* were provided by Drs. Sonia Anand, Russell de Souza, and Michael Zulyniak. Sandi Azab and Ritchie Ly provided quantitative analysis of strong ions in urine by CE-UV. Untargeted data acquisition of plasma samples on a CE-MS platform was performed by Dr. Meera Shanmuganathan and feature extraction was completed by Jennifer Wild.

Chapter 1: Introduction

1.1 The Evolution of Risk Assessment in Human Health

Humanity has triumphed over myriad existential challenges by predicting risk through empirical observation, and the dissemination and adaptation of knowledge.¹ Early cultures sought advantages by interpreting seemingly interconnected natural phenomena: Roman augurs observed avian behaviour for predictions of imminent events,² while Aboriginal Australian cultures read constellations to forecast meteorological changes influencing agricultural yields.³ Precognition of pending crises has been earnestly pursued throughout human history to protect individual health and community longevity;⁴ metabolomics is a modern exemplar of this pursuit for deeper understanding of complex natural systems.

1.1.1 Biomarkers in Ancient History

Initially the purview of seers and religious healers, medical prognostication transitioned from metaphysical to empirical by approximately 5th century BCE.^{4,5} From then until the late 19th century, health assessments were based on subjective patient reports, palpation, and rudimentary auscultation.^{6,7} Through uroscopy (early urinalysis) Classical Greek physicians found the colour, smell, turbidity, viscosity and even taste of urine replete with pathological information.⁸ Amid many incorrect diagnoses and ineffective treatments, systematic analysis provided insights into signs associated with kidney disease (*e.g.*, proteinuric foam and redness due to heme),⁹ jaundice (*e.g.*, dark-pigmented bilirubin),¹⁰ and diabetes (*e.g.*, sweetness due to glycosuria), the earliest diagnoses of chronic disease as a result of aberrant metabolism.¹¹ An ancient Egyptian (~1500 BCE) diagnostic test detected elevated glucose in urine by observing its ability to attract insects,⁶ yet understanding of diabetes pathogenesis did not occur until much later, spurred by a 1775 paper from Matthew Dobson identifying elevated blood sugar as a key feature of diabetes pathology.^{11,12} Today, direct measurements of fasting glucose and glycated hemoglobin (*i.e.*, HgbA1C) remain the clinical gold standards for the diagnosis and

risk assessment of type 2 diabetes worldwide. Individuals at risk of diabetes (*i.e.*, prediabetics) have fasting plasma glucose and %HbA1C of 5.5-7.0 mM and 5.7-6.4%, respectively, where a diagnosis of diabetes occurs beyond these ranges.¹³ Jaundice (*i.e.*, “yellowness”), believed to be a primary illness for over 4000 years, arises from the accretion of bilirubin, a dark-yellow bile pigment produced by hepatic (*i.e.*, liver) metabolism of hemoglobin.¹⁴ Normally, catabolites urobilin and stercobilin are eliminated in the urine and feces, respectively, imparting characteristic yellow and brown colours to these specimens.¹⁰ When this process is impaired, as first described by Hippocratic physicians *ca.* 400 BCE, a person’s blood turns dark with concentrated bilirubin, the stool becomes clay-coloured, and a distinct yellow colour emerges in their skin and sclera.¹⁵ Intense efforts were subsequently required to isolate bilirubin from a complex mixture of liver bile acids before its structure was determined by Hans Fischer in 1942.¹⁰ Bilirubin is now a standard biomarker of liver function to assess the risk of hepatic damage, where the normal clinical reference range for total serum bilirubin is approximately 1-21 μM .¹⁶

1.1.2 Creatinine and its Ubiquity in Clinical Assessment

Separation and analytical chemistry techniques (*e.g.*, distillation, precipitation, titration) of the late 19th century were applied extensively to biofluid analysis to gain further insight into healthy and diseased human physiology.¹⁷ Creatinine (Crn), a highly abundant metabolite of creatine phosphate (CP), was discovered by microscopy of crystallized beef extract in 1838;¹⁸ Max Jaffé later reported a colorimetric alkaline reaction between creatinine and picric acid that produced a brick-red salt in 1886.^{19,20} CP is concentrated in skeletal muscle, donating a high-energy phosphate for ADP regeneration to ATP, a critical reaction for intracellular energy generation.²¹ As Crn is eliminated in urine by renal filtration at moderately-controlled, individually distinct rates regulated by musculature, physical activity, health, and hydration status, it became an important clinical biomarker for assessment of renal function.²² The “Jaffé test” was readily adapted for clinical

assays of Crn,²³ and despite poor specificity and major advancements in clinical chemistry, it endures in modern medicine due to rapid turnaround time and low cost.²⁴ Today, urine and serum creatinine, in conjunction with albumin and urea, are measured in clinics to assess the risk of kidney disease and/or failure. Clinical reference ranges for normal creatinine vary, but are generally fall within 5-23 mM in urine and 44-115 μ M in the serum of adults.^{16,25} Excess serum Crn triggers downstream testing to investigate possible kidney injury or failure, both of which have severe consequences. Several variations on Jaffé's original reaction have been made to improve the specificity and accuracy of creatinine determinations, however significant positive bias still exists between the Jaffé test and more accurate enzymatic and separation methods coupled with mass spectrometry (MS).²⁶⁻²⁸

Urine creatinine is a widely used normalization factor in biological surveillance studies to correct for hydration status and is frequently quantified using the Jaffé test. To standardize the amount of other urine solutes within a sample population, the concentration of each metabolite is calculated as a ratio to the amount of Crn in the sample (*i.e.*, μ mol per mmol Crn).^{25,29} However, a number of exogenous and endogenous factors affect Crn excretion (*e.g.*, age, sex, exercise, diet) possibly propagating significant inaccuracies among scientific investigations, and convoluting useful information within studies, particularly those with diverse study cohorts.^{29,30} Barr et. al.²⁵ reported significantly elevated urine Crn in the spot urine samples of males and individuals under 50 years of age, that were also affected by the time of collection. Given discrepancies in Crn measurements reported in interlaboratory studies,³¹ more accurate methods for quantifying creatinine are needed. Recent alternatives to Crn normalization attempt to remove unwanted interindividual sample variation by correcting to total solute content as measured by specific gravity,^{32,33} osmolality,^{34,35} and total signal of detectable urine compounds in nuclear magnetic resonance (NMR).³⁶⁻³⁸ Additionally, there are approaches to study design and statistical analysis to reduce unwanted variance in biological studies that will be discussed in *Chapter 1.3.4*.

1.1.3 Metabolism, Molecules, and the Evolution of Omics and Precision Medicine

By the early 20th century scientific studies had identified many discrete molecular components present in human bodily fluids, yet their origin and function were largely unknown. Sir Archibald Garrod, a pioneer of biochemical genetics and precision medicine, identified four metabolic disorders with distinct urine phenotypes and familial recurrence, namely alkaptonuria, cystinuria, pentosuria, and albinism.³⁹ He proposed an inherited “structure” affecting individual metabolism acquired through a unique molecular heritability.⁴⁰ The subsequent discoveries of DNA and genes as agents of phenotype initiated exploration for discrete biochemical components that collectively contribute to the phenotype of an organism, and also function as objective biomarkers to predict, prevent, and/or treat human diseases.^{41–43} In 1983, this was first achieved when a mutation in the HTT (huntingtin) gene was linked to Huntington disease,⁴⁴ followed by identification of the cystic fibrosis transmembrane conductance regulator (CFTR) gene in 1989, in which certain disease-causing mutations lead to cystic fibrosis.⁴⁵ Subsequently, the field of genomics,⁴⁶ the structural and functional characterization of genes advanced rapidly: the Human Genome Project, which was completed in 2003, mapped over 25 000 human genes and introduced genetic profiling as a fundamental part of studying disease etiology, such as gene variants associated with both inherited and chronic disease.⁴⁷ Genome-wide association studies (GWAS) are widely used to identify gene loci statistically associated with disease risk in populations;⁴⁸ however, single nucleotide polymorphisms (SNPs), including mutations of unknown or variable consequences often show weak correlations to phenotype. Gene expression is modified by several extrinsic factors, including diet, lifestyle, gut microbiota, and the exposome, elements collectively referred to as “the missing heritability.”⁴⁹ For instance, tumorigenic gene mutations in *BRCA1* and *BRCA2* have a low incidence in the general population (< 10%), but show penetrance of 65-80% (*BRCA1*) and 45-82% (*BRCA2*) in female breast cancer.^{50,51}

However, both environment and diet are known to modulate risk, and not all carriers will develop a related cancer in their lifetime.^{52,53}

The advent of MS-based proteomics provided an unprecedented opportunity to widely probe functional gene expression.^{54,55} As the natural complement to genomics, proteomics heralded the potential to link protein biomarkers to specific alleles related to disease risk, providing novel therapeutic targets and invaluable insight into the complex gene-environment interactions that determine clinical phenotype.⁵⁶ However, MS-based proteomics based on non-targeted protein profiling has faced major technical challenges in translating findings into validated biomarkers that have clinical utility.⁵⁷ Protein biomarkers currently used in clinical assessment rely primarily on classical immunoassay and enzymatic methods, such as serum high sensitivity C-reactive protein (hsCRP) and cardiac troponin T for screening of high risk patients for acute coronary syndrome.⁵⁸ Humans are estimated to have $> 10^4$ proteins, with additional complexity imparted by post-translational modifications (*e.g.*, glycosylation and phosphorylation) that dictate function and activity.⁵⁹ Furthermore, these modifications vary by individual physiology and pathology, generating millions of potential protein targets affecting numerous metabolic pathways.^{60,61} The magnitude of the proteome compounded by analytical hurdles to adequate resolution of low-abundance proteins in complex biological samples make the proteome a much more challenging system to interrogate.⁶²

Metabolites represent the end-products of gene expression and protein activity, that comprise all small molecules (< 1.5 kDa) utilized in cellular metabolism, exogenous compounds, and biotransformed xenobiotics from environmental exposures. Comprehensive metabolite profiling of complex biological samples (*i.e.*, metabolomics) using new advances in separation science, MS and NMR technologies represents the “apogee of the omics trilogy” in molecular biology,^{63,64} which promises new insight into underlying mechanisms of disease (**Figure 1.1**).^{65,66} Since human health is inextricably linked with

environment, metabolomics is a powerful approach to study complex physiological interactions with exogenous determinants of clinical outcomes like habitual dietary intakes and toxicant exposures.⁶⁷⁻⁶⁹ Exposomics, a discipline of metabolomics applied to identifying and understanding the impacts of lifelong exposures on human health offers a more objective appraisal of environmental risk factors that contribute to disparities in population health than can be provided by DNA (genomics), RNA (transcriptomics) or proteins (proteomics).^{70,71} Genes are often poorly correlated to phenotype as they are a near-static assessment of biology that excludes the impact of environmental exposures. For example, a genetic mutation carrier may remain asymptomatic and never express disease.^{72,73} Mutated proteins may confirm a greater risk of developing disease, though many have insufficient specificity and provide too little information regarding the rate of disease progression, or the eventual severity of illness. As the dynamic chemical expression of an individual's unique biochemical interactions with their lifestyle and environment, the exposome is also key to new advances in precision medicine. Wherein all aspects of clinical treatment would be informed by an individual's unique metabolic fingerprint, rather than one-size-fits-most series of clinical guidelines.⁷⁴ Realization of this goal faces significant challenges due to the current dearth of experimental standardization regarding study design, sample preparation, instrumental platforms, and data processing. A large number of dynamic variables that differ widely in polarity and structure define a metabotype (or exposotype),⁷⁵ thus chemical diversity is significantly greater and more difficult to parse than in the human genome, comprised of 2 sets of base pairs, and proteome, built on 21 amino acid building blocks.⁷⁶ Consequently, omics initiatives have struggled with false discoveries due to inadequate study power, poor method validation and/or lack of replication in independent cohorts or laboratories.^{60,61,77,78} Such errors often lead to putative associations to heritable and/or chronic illnesses, that subsequently fail upon replication or validation,^{57,79-81} fundamental to the translation of robust

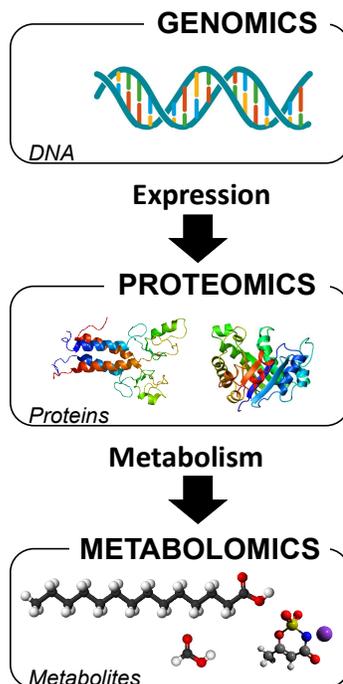


Figure 1.1: *The omics cascade and the biological processes that influence end point metabolites*

biomarkers that have clinical utility to improve patient outcomes, such as screening tools for early detection of high-risk individuals prone to future disease development.

1.2 Individual and Epidemiological Assessments of Risk

1.2.1 Traditional Patient Assessment Standards and Protocols

The most important clinical aspects of assessing an individual's health are a detailed personal and family history, physical examination, and diagnostic tests, if necessary.⁸² Of these, history taking is the most fundamental, and informs results of the latter two.⁸³ Families not only share hereditary disease factors, but also environmental, socioeconomic, and lifestyle factors that contribute to disease risk. Detailed histories probe past and present symptoms, use of prescribed medication,

supplements, tobacco, alcohol, physical activity, and any history of family illness.⁸² Certain circumstances may require more comprehensive patient histories, for instance, occupational risk factors are important to identify for workers routinely exposed to airborne particulate, hazardous chemicals, and repeated physical stress, like firefighters.^{84,85} When diet is a key determinant of health outcomes, detailed nutritional assessments are required to elucidate the effect of diet on phenotype (*e.g.*, heart disease), predict the effect of food intake on medical outcomes (*e.g.*, ulcerative colitis), and monitor dietary adherence (*e.g.*, pancreatic enzymes, fat-soluble vitamins etc.) to ensure normal growth and development in conditions such as cystic fibrosis.⁸⁶ However, the utility of patient history can be challenged by forgetfulness, inaccurate recall, and deceit.⁸⁷ Objective assessment of health status is achieved by the physical exam, employing both qualitative and quantitative methods. Anthropometric measurements (*e.g.*, height, weight, waist circumference) and blood pressure provide insight into cardiovascular disease risk. Distinctive colouring of the skin, eyes, and nails, or odours in breath, can indicate a number of metabolic disorders from chronic liver disease to anemia and alcoholism.⁸⁸

Diagnostic testing generally relies on chemical tests, bioassays or imaging methods (*e.g.*, ultrasound, x-ray, MRI) at a point-of-care facility for diagnosis, treatment monitoring, or as a risk-assessment initiative.⁸⁹ The requisition of diagnostic testing is hypothesis-driven, where a physician may suspect some pathology after completing the patient history and physical exam, and used to determine downstream interventions.⁸⁸ These tests are predominantly performed on serum, plasma and urine, though a number of specialized tests for non-standard excretions have been developed for clinical use, such as pilocarpine-stimulated sweat for confirmatory testing of cystic fibrosis.^{90,91} An extensive panel of 4000 unique laboratory tests are available to clinicians and > 60 billion are performed annually in Canada;⁹² however, routine clinical chemistry testing of validated biomarkers examines only a small fraction of the human metabolome. It is this area of patient care that MS-based clinical care is continuing to revolutionize.

MS-based methods offer a robust and multiplexed instrumental platform for clinical assessments, often having improved throughput, sensitivity and specificity as compared to conventional immunoassay, spectrophotometric, and electrochemical assays performed by automated analyzers. MS also offers the ability to simultaneously quantify a panel of biomarkers, which improves screening accuracy while avoiding the need for a repeat or confirmatory testing.^{73,93,94} Indeed, the advent of MS in newborn screening initiatives enabled rapid multiplexed testing for dozens of rare genetic diseases at incremental costs via analysis of a panel of amino and fatty acids in a heel prick (*i.e.*, single drop) blood spot.⁹⁵ For reasons that will be discussed later, widespread adoption of clinical MS methods has been slow; nevertheless, MS represents a major tool used in metabolomics for biomarker discovery that can also improve the analytical performance of existing clinical tests with better accuracy and fewer interferences.

1.2.2 Biological Surveillance and the Challenges of Precision Medicine

Novel applications of MS-based exposomics simultaneously detect multiple metabolites to identify patterns (*i.e.*, combinations of variables) indicative of specific disease states and their unique subtypes as required for new advances in personalized or precision medicine.⁷⁴ Metabolomics offers cost-effective methods amenable to clinical adaptation, with higher throughput and greater sensitivity to investigate complex human pathology (especially in combination with bioinformatics) than currently offered by the traditional “single” biomarker testing used in contemporary clinical medicine. Correlations found between metabolite patterns and disease in metabolome-wide association studies (MWAS) can be studied on an epidemiological scale to further understand how age, ethnicity, sex, geography, occupation, exposures and myriad other risk factors interact with the metabolome to influence clinical phenotype.^{96,97}

Among the most sought-after goals in exposomics is determining the developmental origins of health and disease;⁹⁸ environmental exposures affect approximately 50 – 60% of the urine and plasma metabolomes,⁹⁹ thus playing a

major role on the initiation and evolution of these processes (**Figure 1.2**). By understanding the processes that induce disease, practical strategies for disease prevention via lifestyle modifications can become the focus of health maintenance, rather than invasive reactionary treatments. Increasingly powerful tools of disease prediction will also allow for presymptomatic detection of diseases (*e.g.*, newborn screening), increasing the probability of better outcomes with safer yet more effective therapeutic interventions. In the past decade a number of collaborative governmental and academic MWAS have begun the complex work of decoding the exposome in large-scale population studies, including EXPOsOMICS (Europe, device-aided personal exposure monitoring), HELIX (Europe, follow-up studies on established birth cohorts), and HERCULES (USA, computational applications applied to the exposome).¹⁰⁰⁻¹⁰² Early MWAS have identified plasma concentrations of fatty acids, and acylcarnitines as significant predictors of mortality (hazard ratio = 1.11-1.18) in patients with suspected coronary artery disease.¹⁰³ Similarly, Chajès *et. al.*¹⁰⁴ reported a 24-39% increase in breast cancer risk in women with elevated serum concentrations of two monounsaturated fatty acids (*i.e.*, oleic and palmitoleic acids), and with significantly higher odds (45-124%) related to circulating trans fatty acids, that is largely derived from intake of hydrogenated oils from processed foods. Preliminary association studies are the first step to identifying putative molecular biomarkers, and related metabolites, that could be adapted for subsequent clinical testing.

Cohort studies, case control studies, and cross-sectional studies use similar approaches to determining putative biomarkers, using tools (*e.g.*, case reports, interviews, literature reviews) for the elucidation of cause-and-effect, without

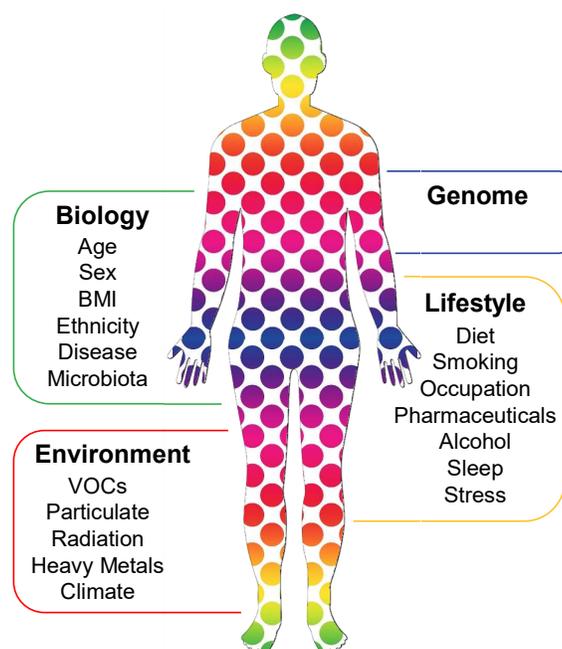


Figure 1.2: The exogenous and endogenous determinants of phenotype relevant to human health.

controlling any aspects of the intervention.¹⁰⁵ Experimentally-controlled studies are typically smaller-scale (*i.e.*, < 300 participants) pilot studies limited to much shorter time periods (*i.e.*, days to months) that contrast two or more groups that receive distinct treatments, or groups where one receives no treatment (*i.e.*, a control) in order to identify biochemical changes in the parallel treatment arms. Randomized controlled trials (RCT) have been the gold standard of clinical inventions for decades as they are ideal for reducing intentional or inadvertent bias by researchers in the selection of study groups, while making a direct comparison between two or more applied treatments, which aids in the interpretation of clinical outcomes.^{106,107} However, due to limitations of funding, ethics, and statistical power findings from RCT are not necessarily directly translatable to real-world scenarios as these studies are generally conducted under well-controlled clinical supervision, reducing the confounding environmental variability found in free-living living populations. Thus, while a rigorous RCT may show efficacy within the study sample, it is not

possible to conclusively determine effectiveness within the general population, without further validation studies in diverse populations.¹⁰⁸

Metabolic differences among minority groups typically poorly represented in experimental trials can lead to a medical interventions causing increased morbidity in those for whom the treatment is less effective, ineffective, or worse, adversely effective within the prescribed therapeutic window.¹⁰⁹ More than 30 medications have been reported to have differential pharmacokinetics affected by factors including genetics and biological sex.¹⁰⁹⁻¹¹¹ For instance, polymorphisms of a xenobiotic-metabolizing liver enzyme (*CYP2D6*) confer varied abilities to metabolize codeine to morphine, for the treatment of pain. A significant proportion of Africans (1.8 - 18.8%) and Europeans (3.2 – 10.4%) are “poor metabolizers” of codeine (low *CYP2D6* activity), that has much lower occurrence in Asian (0 – 2.1%) and Middle Eastern (1.4 – 1.5%) populations.¹⁰⁹ As a consequence, indicated therapeutic dosages are less effective for the two former groups when normalized to body mass alone. To address this, metabolomics is striving towards the study of increasingly stratified populations that reflect the complexity of biochemical pathways in real populations.

1.3 The Workflow and Statistical Methodologies of Molecular Biomarker Discovery

1.3.1 What Exactly is a Molecular Biomarker?

A molecular biomarker may be a small molecule, protein, lipid, carbohydrate, gene, hormone or other molecule of endogenous or exogenous origin that can be objectively measured and associated with a pathological condition or biological change.^{81,112} In exposomics, biomarkers are generally sought to understand functional changes that occur within an individual in response to their environmental exposures, and more specifically to understand how these impacts relate to pathology.⁷⁰ An ideal clinical biomarker must satisfy several analytical and clinical validation criteria, including:¹¹³

- 1) acceptable sensitivity, specificity, accuracy, reproducibility and robustness and other relevant metrics that establish technical performance; and
- 2) an ability to classify or predict the concept of interest with acceptable clinical performance and cost-effectiveness.

Biomarker discovery often begins with an untargeted association study (*e.g.*, metabolomic fingerprinting or profiling), where 10^1 - 10^3 molecules from distinct sample populations are contrasted *post hoc* to identify a small number putative biomarkers that may explain biological differences between each group; however they may also be identified without a preliminary association study based on established scientific knowledge.¹¹⁴ A biomarker may be one or more molecules combined and may be used to make one or more clinical assessments. Some are used as binary determinants of illness (*i.e.*, healthy vs. disease), while others offer more clinical utility in staging the progression of disease or differentiating subtypes of related diseases that is critical for treatment decisions. A review of the available clinical roles for molecular biomarkers is shown in **Table 1.1**.¹¹⁵ The targeted metabolite panel is usually determined by *a priori* knowledge or by an association study that showed significant discrimination between randomized analytical samples.⁶⁴

Table 1.1: *Clinical types and uses of risk-related biomarkers (adapted from Burke, 2016¹¹⁵)*

Type	Role	Relevant Risk
Exposure	Qualitative and/or quantitative measurement of specific chemical exposure	Absorption of hazardous substances, production of hazardous metabolites
Diagnostic	To confirm/rule out the presence of abnormal pathology	Malignant (and possibly benign) disease diminish quality of life and lifespan
Therapeutic	Identify pathological subtype and relevant treatment; monitor effect of treatment on disease (risk)	Inappropriate treatment may be ineffective or harmful
Prognostic	Identify probability of future disease in the absence of intervention	Untreated disease may lead to diminished quality of life and/or lifespan

1.3.2 Human Biological Matrices Used in Biomarker Discovery

For biomarker research, blood and urine remain the most commonly examined biological specimens analyzed in clinical metabolomics studies due to their easy accessibility to adequate sample volumes using well-established clinical sampling procedures and standardized analytical protocols. However, there are several alternative biospecimens which may offer unique specificity for studying aberrant metabolism relevant to specific human diseases, while avoiding the confounding of circulatory metabolites in blood or excretory metabolites in urine. **Table 1.2** lists a selection of biological matrices analyzed in bioanalytical studies of the last decade and an estimation of recent publications.

Table 1.2: *Metabolomics publications during 2008-2018 by biological matrix as defined by sample type analyzed*

Matrix	Access	Invasiveness^a	No. Recent Publications^b
Blood (Serum, Plasma)	Venipuncture, skin prick	M	4699
Other tissue (skin, organ, muscle, bone)	Biopsy	H	4316
Urine	Excretion	N	2159
Feces or Stool	Excretion	N	299
Cerebrospinal Fluid	Lumbar puncture	H	288
Blood Tissue	Venipuncture, skin prick	M	252
Breath Exhale	Excretion	N	236
Saliva	Excretion	N	142
Dried Blood Spot	Skin prick	M	77
Sweat	Excretion	N	69
Amniotic Fluid	Amniocentesis	H	65
Hair	Cutting	N	40
Breast milk	Excretion	N	32

^a(H) High, (M) Moderate, (N) None; ^b2008-2018; searched as “[matrix] and metabolomics”

1.3.3 Major Instrumental Platforms in Metabolomics

The advent of instrumental analysis methods in the latter half of the 20th century greatly expanded the capability to resolve, detect, and identify low molecular weight molecules in human biological specimens. NMR and MS platforms are most commonly used in metabolomic profiling and metabolite identification as they are capable of sensitive, high-throughput and reproducible determination of a wide array of metabolites with excellent selectivity.¹¹⁶ Due to the high chemical diversity and wide dynamic range of largely unknown molecules in the human metabolome, complementary analytical strategies for achieving adequate metabolite coverage are required. NMR allows for non-destructive biological sample analysis, particularly structural determination, with well-known applications to liquid samples and more

recently, solids (*e.g.*, tissue).¹¹⁷ While not commonly coupled to on-line chromatographic instruments for component separation, cryoprobes, unique pulse sequences, two-dimensional (2D) experiments, and sensitivity to multiple nuclei (*e.g.*, ^1H , ^{13}C , and ^{11}B) with little sample preparation make NMR a robust platform for identifying complex sample components, particularly those incompatible with common soft ionization methods in mass spectrometry.^{118,119} In NMR, targeted nuclei resonate in an applied magnetic field between binary energy states defined by the strength of an applied external magnetic field. Small differences (*i.e.*, chemical shifts) in the resonance energies between identical nuclei are caused by unique magnetic effects from nearby atoms, which is translated into structural information based on well-established reference ranges for chemical shifts.¹²⁰ ^1H -NMR is most commonly used in structural elucidation of unknown compounds as it is a significant component of organic molecules and 99.98% of natural hydrogen isotopes. Commonly, ^{13}C -NMR is also performed to effectively parse complex structural problems of structural elucidation but suffers from much higher detection limits due to a low natural isotope abundance.¹²¹

In practice, the dynamic range of NMR is limited by poor resolution when analyzing complex biological samples that often require large sample volumes (> 0.5 mL). These concerns also limit detectable features and quantification is challenged by overlapping signals, low sensitivity, and the wide range of concentrations in biological samples.^{122,123} Two-dimensional NMR yields more information by resolving overlapping signals on the linear axis into two-dimensional space with powerful electronic (*i.e.*, "through-bond") or geometric (*i.e.*, "through-space") homonuclear (*e.g.*, ^1H vs ^1H) interactions (*e.g.* correlation spectroscopy (COSY) or heteronuclear (*e.g.*, ^1H vs ^{11}B) interactions (*e.g.*, heteronuclear single quantum correlation (HSQC));¹²⁴ though these 2D experiments are time-consuming (*i.e.*, low throughput).^{116,125} Additionally, acquisition of an NMR instrument can be cost-prohibitive as high-field modern

instruments range exceed well over a million dollars while requiring high maintenance costs, including cryogenics.¹²⁶

MS has become the most widely used instrument metabolomics because of its lower infrastructure and operating costs, high sensitivity and direct coupling to various ion sources and separation technologies as compared to NMR.⁶⁴ Mass spectrometers employ electric fields to sort ionized gas-phase sample components by their mass-to-charge ratio (m/z) with resolving power of $> 10^6$ (*i.e.*, high mass accuracy) in modern instruments, therefore it is necessary that all compounds be ionized or ionizable to enter the instrument.¹³⁷ Analytes with polar functional groups yield optimal relative response ratios with ESI-MS techniques, whereas nonpolar compounds are better suited for analysis with atmospheric pressure chemical ionization (APCI) or atmospheric pressure photoionization (APPI).¹²⁷ MS is compatible with several high-efficiency separation methods for resolving complex sample mixtures, including high-pressure liquid chromatography (HPLC), gas chromatography (GC), and capillary electrophoresis (CE). techniques are important for expanding resolution and metabolome coverage while also providing complementary information to support unknown identification when using tandem mass spectrometry (MS/MS) via collision-induced dissociation experiments.¹²⁸ The accurate mass of the intact molecular ion (*e.g.*, $[M+H]^+$), together with its charge state and isotopic pattern (*e.g.*, ^{13}C , ^{15}N , ^{34}S and ^{37}Cl) allow for determination of the most likely molecular formulae for unknown compounds with structural information dependent on MS/MS spectral matching at different collision energies relative to known standards.^{129,130}

Polar/ionic molecules in MS are ionized to largely intact gas-phase ions under positive and negative ion mode, yielding a spectrum where signal intensity is plotted against m/z , to depict the relative abundance of sample components.¹³¹ High efficiency separations of complex sample mixtures prior to ionization facilitates isomeric resolution and isobaric interferences while minimizing matrix-induced ion suppression or enhancement effects that contributes to bias in ESI-MS (**Figure**

1.3).¹³² Further, a perpendicular angle between the ion source and MS inlet reduces background noise and increases sensitivity to low abundance compounds by preventing sampling of nonionized material that could foul the instrument and reduce sensitivity (**Figure 1.4**). Detection limits in MS can reach picogram levels depending on ion source and sample injection volume, however, MS sensitivity is still challenged by sample matrix effects that can contribute to ion suppression or enhancement effects, which is a major challenge for quantitative measurements in metabolomics without use of stable-isotope internal standards.¹³³ Nevertheless, the high versatility, selectivity, and sensitivity of MS are the reason for its widespread application in cutting-edge metabolomics research.

Gas chromatography (GC) is favoured for resolution of volatile, low molecular weight, thermally stable metabolites due to excellent reproducibility, high peak capacity, and sensitivity allows for the resolution of thousands of unique peaks.¹³⁴ Furthermore, extensive electron impact (EI-MS) ionization spectral libraries are available for rapid, unambiguous structural identification of metabolites in biological samples.^{135,136} As it relies on the gas-phase partitioning of compounds, extensive sample workup and pre-column chemical derivatization are often needed to analyze polar/ionic metabolites as their trimethylsilyl derivatives. This prevents the study of several thermally labile metabolites and limits the usefulness of GC to lipidomic and low-polarity molecular analysis in human biomonitoring studies.^{137,138}

High-performance liquid chromatography (HPLC) is a widely-used separation technique capable of simultaneously resolving low- and high-polarity compounds, offering broader metabolome coverage than GC, which is limited to volatile, nonpolar, low molecular weight and thermally stable metabolites.¹⁴⁰ Specificity is achieved by modifying analytical conditions like pH, ionic strength, mobile phase composition, column dimensions, stationary phase particle size, and

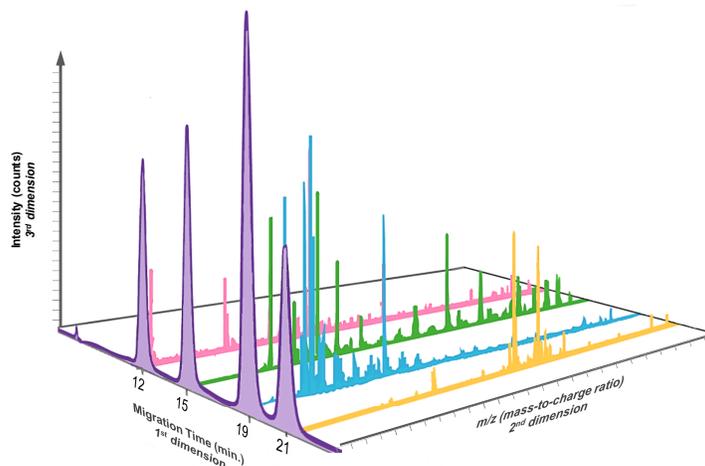


Figure 1.3: Projection of the 3 dimensions of information provided when chromatographic/electrophoretic methods (CE, LC, GC) are used in tandem with MS. The sample is resolved into a chromatogram in the first dimension that is plotted by mass-to-charge (m/z) ratios in the second-dimension mass spectra. Relative response is shown in the third dimension.

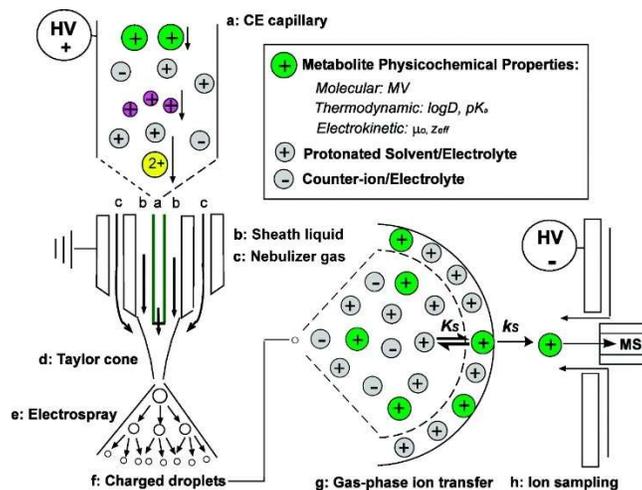


Figure 1.4: Orthogonal coupling between an ambient pressure ESI spray needle (containing coaxial CE capillary) and MS source that reduces ionization suppression by nonionized species by sampling charged species via an applied electric field. Used with permission from the American Chemical Society.¹³⁹

porosity. Instrumental settings like temperature and internal column pressure are also used to influence analyte resolution by differential partition equilibria into the stationary solid phase from the mobile liquid eluent.¹⁴¹ Recent technologies have

greatly enhanced the separation efficiency achievable in LC, particularly the advent of ultra-performance LC (*i.e.*, UPLC) which uses high internal column pressures (approximately twice that of HPLC) and small particles sizes ($< 2 \mu\text{m}$) to significantly increase peak capacity. Smaller particles allow faster mass transfer of analytes between mobile and stationary phases, producing narrower peak widths with greater resolution achieved in much shorter run times when compared to conventional LC.¹⁴² Similar optimizations can be achieved using traditional LC infrastructure with the use of core-shell particle stationary phase, where analyte diffusion are limited to a porous outer layer by a solid core, resulting in rapid mass transfer, faster runs, and efficient resolution comparable to UPLC while avoiding the excessive back pressures of UPLC that may cause leaks.¹⁴³

Reversed-phase (RP) HPLC is the most widely used separation mode in metabolomics as it is possible to separate a wide range of molecular classes while yielding highly reproducible retention times.¹⁴⁴ The selectivity of non-polar (*e.g.*, C8, C18) RP columns is modifiable by the use of derivatized stationary phases to improve retention of compounds based on physicochemical properties like charge, size, or polarity.¹⁴⁵ C18 columns best separate non-polar or weakly polar sample constituents like oils and pesticides, as high polarity molecules are not effectively retained by the stationary phase and co-elute in the void volume. In metabolomics analysis, LC faces technical challenges resolving strongly polar and ionic species prevalent in certain biological samples such as urine and sweat.¹⁴⁶ These are poorly retained by low polarity RP columns which limits the number of detectable features in important molecular classes associated with protein (*i.e.*, amino acids) and fat (*i.e.*, ketones and organic acids) metabolism.^{147,148} Hydrophilic interaction chromatography (HILIC) provides a complementary separation mechanism for the analysis of polar and ionic metabolites as compared to RP-HPLC, which uses volatile acetonitrile-rich eluent solvents more amenable to ESI-MS. Here, analytes partition between a stationary liquid phase adsorbed onto porous (un)modified silica, achieving separations with a combination of hydrophilic, hydrogen-bonding

and electrostatic interactions depending on column type and mobile phase conditions (*e.g.*, pH and ionic strength).^{149,150} A HILIC-UPLC method found over 1100 highly polar compounds in murine urine samples in an approximately 10 min. method with excellent reproducibility.¹⁵¹ However, HILIC has not yet exhibited the ease-of-use and reproducibility necessary for large-scale metabolomics and currently remains an orthogonal technique in small-scale studies.¹⁵²

Capillary electrophoresis (CE) is an electrokinetic separation technique with resolving power that can surpass that offered by HPLC and GC separations in untargeted metabolomics. Uncharged and low polarity compounds can be analyzed using micellar electrokinetic chromatography (MEKC) or a more recently developed technique, non-aqueous capillary electrophoresis (NACE),^{153,154} but it is more widely used for resolving ionic and weakly ionic polar compounds that have a discrete electrophoretic mobility under an external applied voltage.^{147,155} For broad coverage of the metabolome and detection of low-abundance metabolites, CE and HPLC are complementary techniques.¹⁵⁶ A significant portion of the metabolome is ionized at physiological pH and best resolved by CE, however, a low on-column sample volume (~10 nL) limits the detection of low-abundance sample components when compared to conventional LC. This can be problematic for detecting low abundance exposome-related compounds which can be orders of magnitude lower in concentration than endogenous metabolites in human biofluids, that may also co-migrate with signal-suppressing high abundance metabolites (*e.g.*, creatinine). However, low sample volumes are advantageous in studies where samples are inherently volume-limited, such as in forensics applications or biological surveillance studies using non-traditional biofluids.^{73,157,158}

Separations are performed in narrow fused-silica capillaries allowing low consumption of sample and analytical solvents, and comparably lower infrastructure costs. Furthermore, in the absence of a solid phase, sample preparation is rapid and uncomplicated without concerns of column ‘fouling’.¹⁵⁹ CE has also faced challenges in large-scale implementation due to analytical

challenges presented by pH and ionic strength instability, injection volume variation and technically inexperienced users, which can cause significant variability in absolute migration times.¹⁶⁰ This is addressed by relying on distinct intrinsic electrophoretic mobilities of analytes during the isocratic separation that maintain reliable relative migration times to other analytes and internal standards, which are used to standardize the migration times of all detectable compounds.¹⁶¹ The recent development of a multiplexed CE method by our lab greatly increased sample throughput (*i.e.*, up to 13 injections per run) and made CE a significantly better option for analysis of polar/ionic species compared to single-injection methods like HILIC.^{161,162}

1.3.4 Data Workflows and Statistical Methodologies in Molecular Biomarker Discovery

There are two methodological approaches to the measurement and characterization of biological samples in metabolomics: targeted and untargeted analysis. Targeted investigations rely on *a priori* knowledge of the identity and properties of 1 or more (usually < 50) molecules or molecular classes of interest, which are measured exclusively in samples to address a specific hypothesis relating metabolism and the molecules of interest. Since properties of the compounds of interest are known, selective methods of isolation (*e.g.*, extraction) and analysis are optimized for their detection (*i.e.*, more robust and reproducible results) while calibration curves may also be prepared for facile absolute quantification if chemical standards are available.¹⁶³ Using targeted protocols are beneficial for the low computational demands of data processing and the ability to use less resource-intensive analytical platforms if adequate analysis can be achieved by simpler technologies like thin-layer chromatography (TLC), ultraviolet/visible spectroscopy (UV/Vis), and fluorescence. The overall intention of targeted analysis is sample profiling, not discovery, to associate an experimental classification (*e.g.*, disease state) with known metabolite concentrations.¹⁶⁴

Untargeted analysis is also used to analyze stratified samples to address a specific hypothesis but seeks to identify molecules related to the sample classifications and analytical question. As the specific molecular targets are not known before analysis, sample preparation techniques and analytical platforms are selected to maximize the number of detectable signals, which can be maximized using sensitive, high-resolution analytical instruments like NMR and MS.¹⁶³ As a number of these signals are unknown, labour-intensive efforts are required to validate and annotate each signal as a unique feature of interest (*i.e.*, with a unique analytical identifier like migration time and mass-to-charge ratio, m/z :RMT) and not a contaminant, adduct, fragment, or other artifact. Untargeted analysis often yields thousands of raw signals, that are filtered to only hundreds after cleanup. Patterns in the abundance of annotated metabolites impart a molecular fingerprint to each sample, which may be contrasted to determine molecule features related to the sample classification, even when their identities are unknown. However, compromises in analytical sensitivity for global metabolome coverage can mean poorer precision of analytical measurements, and greater difficulty elucidating class membership using untargeted metabolomics data.¹⁶³ Systematic data processing workflows are used to reduced extensive feature lists to set of reproducibly-detected metabolites for subsequent statistical analysis, a typical workflow in untargeted CE-MS analysis is shown in **Figure 1.5**.

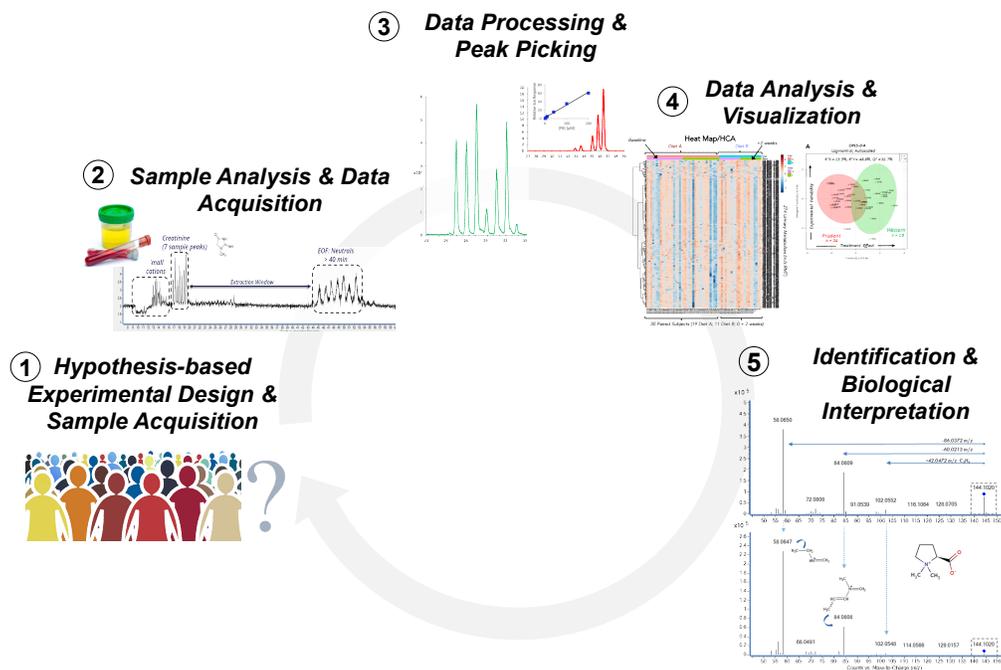


Figure 1.5: Untargeted metabolomics workflow for the acquisition, cleanup, and processing of data in biomarker studies.

Increasing analytical sensitivity and throughput has vastly increased the number of detectable features (and confounding noise) when using high resolution MS methods produced from a medley of experimental designs, data processing, and statistical analyses.¹⁶⁵ The range of available statistical packages for data analysis, combined with increased use of these tools by novice users means that two analysts can generate different outcomes from the exact same data set.^{166–168} For metabolomics to facilitate the transition from discovery to clinical application interdisciplinary collaboration on study design, sample collection, sample analysis, and metabolite identification is essential to produce meaningful, robust, and reproducible data that avoids false discoveries when using untargeted methods for molecular biomarker discovery. As a result, a number of automated, user-interactive online software pipelines for the processing of large datasets have been developed to aid in structural ID (e.g., HMDB,¹⁶⁹ Metlin¹⁷⁰), statistical analysis (e.g.,

XCMS,¹⁷¹ Metaboanalyst¹⁷²), and biochemical interpretation (*e.g.*, HMDB, Metaboanalyst).¹⁷³ Further recommendations have been made to abide by transparent reporting standards in metabolomics for improved replication and translation of studies, and for the public provision of raw data for parallel analysis.^{174,175}

The greatest bottleneck in modern omics analysis involves the storage, processing, cleaning, and analysis of large amounts of data. There are several transformations that may be necessary dependent on the nature of the recorded data, but generally recorded signals are converted into useable data matrices by iterative steps involving:¹⁷⁶

- 1) Cleaning, noise reduction and baseline verification;
- 2) Peak detection and deconvolution;
- 3) Feature alignment among runs and batches (including batch correction);
and
- 4) Application of normalization algorithms (*i.e.*, to correct for hydration status and large dynamic ranges).

As discussed above, even with the most judicious applications of these steps two analysts may yield significantly different results for the same samples. It is therefore of the utmost importance that all data handling methods be clearly reported with the data, including the software used to perform them.¹⁷⁷

Quality control samples (QCs) are ideally technical replicates of a pooled analytical sample, or otherwise standard mixture of representative metabolites, run at pre-defined regular intervals between analytical samples.¹⁷⁸ When a pooled sample is all or a portion of analytical batches, QCs are indispensable in data cleanup by confirming the veracity and reproducibility of low-abundance features that may be easily be mistaken for spurious noise.¹⁷⁹ However, unwanted features including fragments and adducts will also appear reproducibly in QCs, and are easily mistaken for unique features. In-source fragmentation of compounds is an unavoidable consequence of energy transferred to molecules in the ionization

process, though the incidence of fragmentation can be reduced by employing gentler parameters in the MS source.¹⁸⁰ Because the thermodynamics of ionization may differ between labs, the inclusion of fragments and adducts may exacerbate interlaboratory data reproduction issues. Further, inclusion of these features can easily inflate data matrices with highly collinear data that inappropriately increase the classification power of one or more metabolites.¹⁸¹ This problem can be mitigated by comparing the alignment, peak shape, and statistical correlation between compounds with suspected additive or dissociative relationships. Additionally, comparison to collaborative databases of known fragments and adducts, as well as maintenance of similar in-house databases, can make the elimination of these features quick and accurate.¹⁸² Further, consolidation of information is challenged by differing, and sometimes inappropriate, uses of analytical methodology, statistical analysis and data reporting in a relatively young field.^{177,178,183,184} As QCs are identical aliquots with theoretically identical performance characteristics, gradual instrumental changes can be reflected in changing retention times, lowered sensitivity (broadening), and deteriorating peak shape. They also provide guidance on the appropriateness of data extraction and processing steps.¹⁶⁷ Data processing steps that increase variance among QCs, particularly in excess of the general standard of 30% CV, must be examined with prejudice.¹⁸⁵ QCs are also fundamental to identifying low-quality samples and batches to be re-run or discarded and this is most easily achieved when QCs are regularly inserted among analytical samples.¹⁸⁶

When sample acquisition, instrumental analysis, and data extraction are properly implemented, the result is a representative multidimensional spectrum for each sample (*i.e.*, retention time x intensity x m/z). An important task before analysis is then to ensure as much appropriate homogeneity between QCs and unchanged features as possible, to more easily identify the true differences between analytical samples during statistical analysis. This will involve data transformations that include batch (intensity) correction, data normalization, and in urine

metabolomics, correcting for inherent interindividual differences in hydration status. There are numerous data pretreatments available (*i.e.*, log transformation, centering, and scaling) and it is the prerogative of the analyst to adjudicate which functions are necessary for accurate analytical performance from their data.¹⁶⁷ Though, it is worth repeating that the selection of an analytical strategy at this stage may lead to significantly different outcomes upon analysis, highlighting the importance of making informed selections.¹⁸⁷ Batch correction has been increasingly automated with the provision of publicly accessible tools for data alignment, however, the user's skillset is still fundamental to attaining proper output.^{188–190}

Inadequate study power is a funding-correlated problem rampant in untargeted studies. Improperly balanced study designs (variables > samples) comprising highly correlated and highly dynamic variables produced in metabolomics often leads to data overfitting and bias when using multivariate statistical methods.¹⁹¹ It is within this framework that Ioannidis wrote that “Most published research findings are false,” leading to what Forstmeier et. al. describe as a “crisis of confidence.”^{77,165} Systematic error and bias can be introduced at every step in the metabolomics pipeline, and important tools to counter unwanted variation are quality control (QC) and quality assurance (QA) practices, such as following a standard operating procedure (SOP) on rigorously validated methodologies.¹⁹² Inconsistencies in sample handling, such as prolonged exposure to room temperature, may cause significant differences in metabolite profiles due to ongoing biochemical reactions and bacterial metabolism.¹⁹³ Furthermore, delays in collecting biological samples can introduce bias from diurnal variations (*Chapter 1.1.2*) that may be misinterpreted as a class-specific differences in metabolism and promote false discoveries.¹⁹²

Methods to reduce variation differ by sample type and study design. Generally, if urine is not flash frozen soon after collection, filtration, cooling, or preservation (*e.g.*, sodium azide) may be used to reduce bias from biotic (*e.g.*,

bacterial) or abiotic (*e.g.*, oxidation) processes to better stabilize metabolites in the sample.¹⁹⁴ Blood is a versatile fluid that may be collected whole, or partitioned to contain only a portion of blood materials in the form of serum (cells and clotting proteins removed) or plasma (cells removed). Hemolysis, red blood cell rupture caused by high-vacuum collection or rough handling, perturbs metabolite profiles, causing significant changes to lipids and amino acids such as tryptophan and glutathione.¹⁹⁵ While plasma is collected into prepared tubes to prevent clotting (*e.g.*, containing heparin, citrate or EDTA anticoagulants), serum samples are exposed to ambient temperatures for periods of 30 minutes to under 2 hours to allow clotting processes to remove red blood cells and fibrinogen clotting factors.¹⁹³ It is imperative collection processes are standardized, so that temperature and time effects are consistent among samples as it represent a common source of pre-analytical bias.¹⁹⁶ When preparing samples for storage, it is then prudent to portion samples into separate aliquots to permit repeated analysis while avoid repeated exposure of a single sample to room temperatures and multiple free-thaw cycles, which degrade less stable metabolites.¹⁹²

In metabolomics, especially in untargeted studies, it is common that the identity of features will consist of a simple class designation or retention time, and elemental composition (accurate mass) in place of a chemical name.¹⁹⁷ In the absence a molecular identity and very often an analytical standard, it is not possible to quantify features of interest. Here, internal standards (distinct, non-interfering reference molecules added to every sample in identical concentration) serve as references for the ionization response, permitting the use of relative peak areas (RPAs) as stand-ins for absolute concentrations.¹⁵⁵ Internal standards also provide more robust quantification than offered by external QCs as they are identically affected by intrasample variations in injection volume, applied pressures, ambient temperature, and electrospray integrity.¹⁹⁸

While solute concentrations in blood are homeostatically controlled, normalization of urine metabolite data is particularly important to compare highly

variable solute concentrations that fluctuate due to water consumption and biological eliminations.³⁵ In this work, 4- to 20-fold differences in urinary output amongst participants were observed. As described in *Chapter 1.1.2*, spectrophotometric measurements of creatinine are a common normalization factor in urine data, but with the advent of high-resolution instrumental methods it has become clear this is no longer appropriate.^{24,25} Accordingly, the inclusion of urine Crn concentration as an independent variable in regression models, rather than a pretreatment correction factor, was proposed to remove the dependence of statistical significance on Crn abundance.²⁵

Once the highest data integrity has been achieved, the work to identify classifying features (*i.e.*, putative biomarkers of the biological challenge in question) is pursued via statistical analysis. In small-dataset targeted studies, particularly those where the identities and biological relationship between variables is already known, univariate testing (*i.e.*, Student's t-test, Kruskal-Wallis, ANOVA) is suitable and even desirable, provided a method is applied to correct p-values for multiple hypothesis testing.¹⁹⁹⁻²⁰¹ However, in practice, metabolomics datasets initially contain hundreds to thousands of metabolites containing many unknowns, which are unsuitable for univariate statistics. Multivariate classification methods, including principal component analysis (PCA), partial least squares discriminant analysis (PLS-DA), and random forest analysis, are then used for (un)supervised identification of putative biomarkers.²⁰²⁻²⁰⁴ Despite the ability of these instruments to handle (and output) large numbers of metabolites significant to the challenge, the ultimate goal is reduction of data dimensionality and identification of the smallest number of features that can provide acceptable discriminatory power.¹¹⁴ This is imperative for the development of low-cost, high-throughput testing amenable to clinical implementation. Of critical importance as technological capabilities and our knowledge base rapidly increase within the era of "big data" is the need for transparent, rigorously validated protocols and data workflows, applied to well-designed studies that are replicated independently across different laboratories, as

well as open-access resources for collaboration and the dissemination of information within the broader research community.^{175,205}

A candidate feature must achieve validation, via replication, to be classified as a viable biomarker once its clinical utility is evaluated, but this is practiced too infrequently.¹⁷⁷ Novelty is held in higher regard than accuracy, suppressing the capability and incentive for research labs to participate in cross-laboratory validation studies for independent testing of the significance of major findings.^{154,206} To address this in-house, improved standards of testing and reporting have been urged from investigators including redundant sampling, maintaining training and test cohorts, crossover studies, and reporting the discrimination performance of a feature (*i.e.*, sensitivity/specificity, ROC curves, effect size, and performance on class label permutation).¹¹⁴ Additionally, producers and consumers of scientific information we must also learn to appreciate the value of confirming the null hypothesis ($p > 0.05$) and place greater value on reaching the correct conclusion, rather than a novel conclusion.^{169,212} In spite of these challenges, to fully characterize the causes of human illness and develop risk assessment strategies would be an exceptional achievement in human health and preventative medicine, and it is this motivation that drives the bulk of the work in this thesis.

1.4 Multi-segment Injection Capillary Electrophoresis (MSI-CE-MS)

For maximum metabolome coverage, LC and GC (and related chromatographic methods) employ gradient elution (*i.e.*, low to high organic content in RP) or temperature programming (*i.e.*, low to high temperature), while CE analysis is performed isothermally and isocratically in an aqueous buffer system to ensure a homogenous environment for all solutes to maintain consistent electromigration (*i.e.*, EOF + analyte migration) and ionization properties.²⁰⁸ Multi-segment injection mass spectrometry (MSI-CE-MS) is amenable to the sequential injection

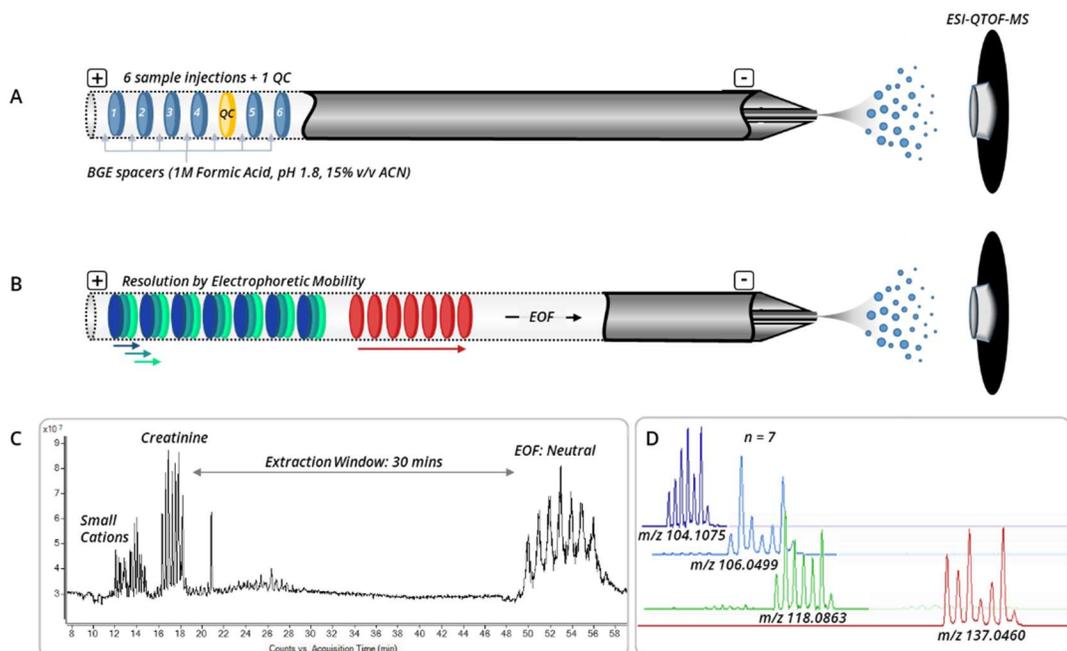


Figure 1.6: (A) Schematic showing 7 injections of MSI-CE-ESI method for high-throughput sample analysis. Six randomized analytical samples are run with a pooled QC in a randomized position. (B) All 7 injections resolve in parallel and 4 metabolites are separated by distinct native electrophoretic mobilities. (C) Total ion electropherogram of analytical run in ESI+ mode. Small, charge-dense cations migrate rapidly, followed by larger, less densely charged ions (e.g., creatinine). Neutral compounds elute with the bulk electroosmotic flow (EOF). (D) Extracted ion electropherograms of 4 urine metabolites resolved by mass-to-charge ratio (m/z) in QTOF-MS with 7 distinct peaks corresponding to 6 analytical samples and a randomized QC.

of up to 13 samples (inclusive of QC) with simultaneous resolution, thereby increasing the throughput of a standard CE analysis up to 10-fold while improving data fidelity through shorter analysis times.^{93,161} The migration of metabolites within an MSI run is equivalent to a singular injection, resulting in similar resolution of all sample components (**Figure 1.6**).

1.4.1 Advantages and Applications of MSI

For quantification of metabolites, calibration curves are necessary to identify the linear response range of a compound, including the limit of detection (LOD), and limit of quantification (LOQ).²⁰⁹ With triplicate analysis for the development of

performance metrics (SD, %CV), this would involve 12 - 15 runs (4 - 5 data points), with a cumulative run time in excess of 12 hrs. (30 min. run time and 25 min. conditioning). Triplicate multiplexed runs offer more accuracy with 6 - 7 data points per regression line, resolved under near-identical analytical conditions (with near-uniform variance) for improved data fidelity in less than 3 hours. Reduced technical and instrumental variability provides greater precision around data points for smaller confidence intervals in the R^2 (coefficient of determination), slope, LOD ($S/N \approx 3$) and LOQ ($S/N \approx 10$).²¹⁰ MSI has been used in several novel high-throughput applications: determining novel markers of enhanced stamina in skeletal muscle of subjects who had supplemented with bicarbonate prior to high-demand aerobic and anaerobic exercise;²¹¹ rapid screening of urine samples for metabolites of controlled common drugs of abuse with online secondary confirmation of ID using tandem MS;⁹³ identifying a suite of lysine analogs related to symptomatic inflammatory bowel syndrome (IBS); and discovering an impaired pathway in the catabolism of two xenobiotic compounds (*i.e.*, pilocarpic acid and mono(2-ethylhexyl)phthalic acid) in juveniles with cystic fibrosis (CF) that differentiated them from children designated CF-positive with greater sensitivity than the current clinical test.⁷³

To expedite peak picking a rapid feature verification method, a dilution trend filter, uses multiplexed serial dilutions of a single analytical sample.¹⁶¹ Simultaneous recording of all mass channels when analyzing 7 serially diluted samples provides a rapid method of feature reduction during pre-processing (**Figure 1.7**). Dilution trend filters produce regression lines for each feature, which are evaluated for linearity and variance to determine their acceptability. Features with inconsistent behavior (*i.e.*, constant or increasing peak area, or sporadic appearance) can be eliminated or labeled for secondary inspection before inclusion in the analytical dataset.¹⁶¹ In the event of missing signals. DiBattista et. al. applied dilution patterning in 7-injection MSI to a trio of paired samples for facile

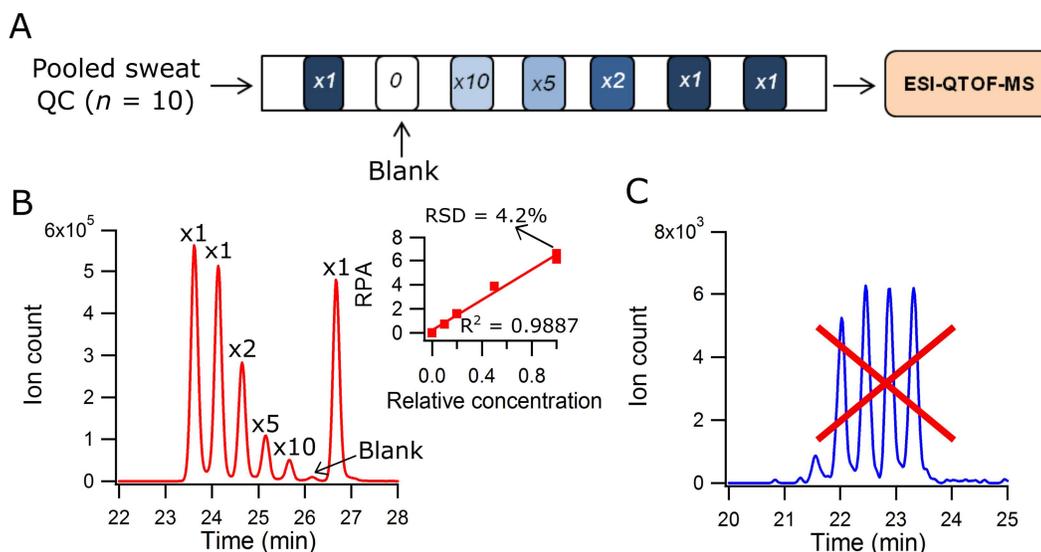


Figure 1.7: (A) 6 injections of a serially-diluted (1, 2, 5 and 10 times) pooled sweat QC sample plus a blank produces (B) an electropherogram that identifies the metabolite citrulline (m/z 176.1030) as an authentic feature with no background (blank) signal and (inset) excellent linearity ($R^2 = 0.989$) and reproducibility ($RSD = 4.2\%$) among triplicate experiments. A dilution trend filter is employed in MSI-CE-MS untargeted metabolomics studies as a rapid first-pass exclusion filter of (C) spurious peaks from chemical and technical noise that do not show effects of sample dilution. Adapted from Nori de Macedo et al.⁷³

recognition of individual samples during data extraction, where each duplicate also functioned as a within-run replicate.⁶⁸

Another novel application is the ability to encode sample ID information into a series of multiplexed samples to reduce ambiguity in sample identification. With each of 3 individuals assigned a distinct dilution pattern (*i.e.*, 1:2, 1:1: or 2:1), 5 previously unidentified features were elevated in a 1:2 pattern in an individual with elevated citrulline in galactose 1-phosphate uridylyl-transferase deficiency. The use of dilution patterning permitted faster visual recognition of aberrant metabolite concentrations for accelerated identification of possibly classifying MS-based

biomarkers in diseases already part of a clinical laboratory-based screening protocol.

While the improvements in data reproducibility, analysis time statistical performance are evident, significant increases in ionic strength when analyzing biological fluids within a single run substantially increases the amount (and length of the migration window) of fast-moving salts and non-ionizable neutral compounds. Suppression effects from highly abundant metabolites (*i.e.*, urine creatinine in positive mode, uric acid in negative mode) may also increase. Increased suppression of small, rapidly migrating cationic organic features that overlap with the suppression band of ions that open the analytical window may reduce the reproducibility of leading peaks, preventing inclusion of these features. Additionally, the low sensitivity of CE for low-abundance metabolites by virtue of the small injection volumes is further reduced by a lower effective capillary length (*i.e.*, less resolution time) and increased solute concentration. Thus, while MSI is an excellent method for high-throughput analysis (targeted and untargeted) of features with easily detectable signals, it presents challenges for the detection and identification of novel low-abundance metabolites characteristic of exogenous exposures pursued in exposomics studies. This is partly addressed by moderately lengthening the CE capillary and run time for improved resolution, while using narrow diameter capillaries to mitigate broadening due to thermal or dispersion effects. By using 10 to 20-fold dilute samples large increases in ionic strength are avoided which maintains electrophoretic mobilities and reduces competitive suppression effects in ESI, without significantly compromising sensitivity due to the low on-column volumes used in CE.

1.5 Thesis Contributions to Targeted and Untargeted Metabolite Profiling for Disease Risk Assessment

Metabolites represent molecular endpoints of functional biology and offer critical insights into the phenotype of an organism relevant to new advances in precision medicine and population health. When applied to the exposome, a vast amount of

information on the identity, absorption, and metabolism of myriad chemicals in our environment is available. However, the effects of these exposures on human health remain poorly understood due to complex interactions with genes, microbiome, diet and lifestyle. However, well-controlled studies are needed to discover viable biomarkers that are both analytically and clinically validated that would directly benefit patient care while reducing healthcare costs.

Respecting the scope and complexity of the human metabolome, work in this thesis first details the development of a novel arylboronic acid biosensor for direct quantification by absorbance or native fluorescence for micromolar levels of *N*-acetylneuraminic acid (sialic acid, Neu5Ac), a ubiquitous cell-surface sugar linked to severe inborn errors of metabolism, chronic and acute illness (*Chapter 2*). An expanded targeted analysis of a panel of smoke biomarkers, including a series of polycyclic aromatic hydrocarbons when using GC-MS was next analyzed from a cohort of firefighters following search-and-rescue training exercises in order to identify chemical exposures for improved risk assessment as outlined in *Chapter 3*. A targeted and nontargeted cross-platform metabolomics study was then explored for identification of dietary biomarkers that reflect two contrasting eating patterns (*e.g.*, Western and Prudent diet) in healthy participants following two weeks of assigned food provisions in *Chapter 4*; this work was also used to verify self-reported diet records in a significant step towards objective biomarkers of habitual diet for new advances in nutritional epidemiology. These studies are contributions to efforts to develop rapid, minimally invasive, and specific assays of the human exposome (*i.e.*, chemical and diet exposures) applicable to population health for more accurate risk assessment of future illness.

1.5.1 Developing Chemical Sensors for Rapid Assessment of Health Status

Neu5Ac is a ubiquitous acidic ($pK_a = 2.6$) sugar essential to mediating aggregation and recognition interactions between red blood cells, proteins and lipid glycoconjugates. Widely distributed in the body, Neu5Ac is primarily found as a terminal sugar conjugated to a diversity of glycan chains, though a homeostatic equilibrium maintains a low concentration of the monomer at sub-micromolar levels in biological fluids and tissues. Already well-characterized clinically, inborn errors of sialic acid metabolism are known to cause severe congenital defects in infants which may not be properly identified until the most severe symptoms are apparent. Furthermore, the extent of glycoprotein sialylation in the absence of genetic causes is associated with both chronic and acute illnesses, including cancer, autoimmune diseases, and the common cold, and is an important biomarker for therapeutic monitoring. *Chapter 2* describes the unusually strong binding affinity between a zwitterionic arylboronic acid, 4-isoquinoline boronic acid (IQBA) and Neu5Ac, which may be exploited for the development of a rapid and inexpensive optical assay analysis of biological samples. IQBA and Neu5Ac binding is greatly enhanced under acidic buffer conditions mediated via a specific covalent interaction involving an α -hydroxycarboxylate moiety lacking in abundant competitive neutral sugars/polyols, such as D-glucose. Many methods used to quantify Neu5Ac lack the sensitivity to independently measure the free monomer, necessitating enzymatic pre-treatment to deconjugate the sugar. Current methods to quantify micromolar concentrations of Neu5Ac rely on a multi-step colorimetric derivation reaction involving thiobarbituric acid and acidic digestion that lacks precision and specificity. The results discussed here are exciting preliminary steps towards the development of optical biosensor for rapid determination of Neu5Ac in biological samples.

1.5.2 Revealing Metabolic Impacts from Wood Smoke Components in Firefighting

The exposome of the occupational firefighter is more complex than the general public, by virtue the number synthetic chemicals contained in household products and myriad toxic and mutagenic combustion byproducts produced during the burning of structural materials. Undetected exposures to these chemicals (*e.g.*, polyaromatic hydrocarbons, hydrogen cyanide, and halogenated organic compounds) and poor adherence to mitigation practices are suspected in anomalously high incidences of cancer, cardiovascular disease, and cardiac arrest in comparison to the general public. Approximately 50% of career firefighters are expected to develop a chronic disease or die before 70 years of age. The hazardous mixture of ambient aerosols, particulate, and gases at the fireground have been well-characterized, fewer have endeavoured to characterize the metabolism of these toxins, and to our best knowledge none have made efforts to identify chemically-permeable weak spots in the fire-protection gear or identify novel markers related to the exposure using a targeted metabolomics method. The diversity of burn sites (*e.g.*, rural or urban) and structures (*e.g.*, residences, industry, and transport) imparts extensive variability of exposure among members of the same station. *Chapter 3* describes a rigorous controlled study to characterize the exposome of occupational firefighters exposed to wood smoke during a controlled burn in a training structure. Analysis of chemical distribution on the skin showed a generalized permeability of the protective clothing and mask that lead to deposition on the skin of the arm and cheek, with unintentional inhalation of ambient chemicals as a likely consequence. Accordingly, urine from male participants from the municipalities of Burlington ($n = 5$), Ottawa ($n = 8$), Hamilton ($n = 5$) was collected in 4 chronological pooled portions over a 48-hour period, where each individual served as their own control by collecting a 24-hour pooled baseline sample for comparison. Exposure and metabolism of ambient toxins was confirmed by several-fold urinary excretion of several methoxyphenols (*e.g.*, guaiacols, syringols) as well as several metabolites

of polyaromatic hydrocarbons that are associated with carcinogenicity due to the formation of reactive oxygen species that can modify cellular DNA.

1.5.3 Identifying Nutrimental Markers of Food Intake and Diet Quality

Dietary intake is a major determinant of health status, life expectancy, and is integral to obesity and chronic disease risk. Epidemiological studies have shown diets rich in fruits and vegetables, lean meats and unsaturated fats and low in processed foods, red meat and saturated fat (*i.e.*, Prudent, Mediterranean) offer optimal nutritional quality with anti-inflammatory effects that support cardiometabolic health. Thus, an ability to accurately assess the habitual diet of individuals is fundamental to provide evidence-based nutritional policies for chronic disease prevention, which is critical in an era of unproven fashionable diets. Current methods of diet assessment in free-living populations rely heavily on self-reporting based on diet records or food frequency questionnaires (FFQ), which are prone to bias and selective reporting. In this context, *Chapter 4* details a multi-platform metabolomics study for assessment of dietary exposure was designed to identify robust dietary biomarkers from healthy participants ($n = 42$) who were assigned either a Prudent or Western diet over a 2-week period. Use of metabolomics platforms permitted the simultaneous measurement and comparison of hundreds of metabolites in matching plasma and urine samples at baseline and following assigned provisional diets, which were then correlated with major nutrient categories from self-reported diet records. For the first time, we revealed that a panel of dietary biomarkers were identified as robust biomarkers reflecting contrasting diets (*i.e.*, metabolic trajectories) when using GC-MS, CE-MS and CE-UV platforms. As expected, many saturated and unsaturated fatty acids were elevated in plasma following a Western diet. Importantly, both plasma and urinary concentrations of proline betaine and 3-methylhistidine dietary biomarkers were positively correlated in participants following a Prudent diet and associated with increases in average fruit and protein intake, respectively. Several organic acids and

amino acids were also elevated in fasting plasma samples when following a Prudent diet, including biotransformed plant-derived metabolites co-metabolized by gut microbiota.

1.6 References

1. R. Boyd and P. J. Richerson, *The Origin and Evolution of Cultures*, Oxford University Press, New York, 2005.
2. D. M. Dunlop, *Proj. Apprais.*, 1987, **2**, 67–68.
3. C. Williams, *Australas. Sci.*, 2015, **36**, 16–19.
4. G. Rosen, *A History of Public Health*, Johns Hopkins University Press, Baltimore, MD, 1993.
5. C. Yapijakis, *In Vivo (Brooklyn)*, 2009, **23**, 507–14.
6. D. Berger, *Med. Lab. Obs.*, 1999, **31**, 28–30, 32, 34–40.
7. B. V. Subbarayappa, *J. Biosci.*, 2001, **26**, 135–143.
8. M. H. Haber, *Clin. Lab. Med.*, 1988, **8**, 415–30.
9. M. J. Geller and S. L. Cohen, *Kidney Int.*, 1995, **47**, 1811–1815.
10. D. A. Lightner, in *Bilirubin: Jekyll and Hyde Pigment of Life*, Springer-Verlag Wien, 2013, pp. 433–464.
11. K. M. King and G. Rubin, *Br. J. Nurs.*, 2003, **12**, 1091–5.
12. M. Dobson and J. Fothergill, *Experiments and observations on the urine in a diabetes*, London, 1776.
13. American Diabetes Association, *Diabetes Care*, 2016, **39**, S13–S22.
14. C. D. Poduri, *Apollo Med.*, 2016, **13**, 76–79.
15. N. Papavramidou, E. Fee and H. Christopoulou-Aletra, *J. Gastrointest. Surg.*, 2007, **11**, 1728–1731.
16. R. A. Mcpherson and M. R. Pincus, *Henry's Clinical Diagnosis and Management by Laboratory Methods*, 2007.
17. L. Rosenfeld, *Clin. Chem.*, 2002, **48**, 186–197.
18. Arthur Hill Hassall, *The Urine in Health and Disease*, John Churchill and Sons, London, 1863.
19. M. Jaffe, *Z.Physiol.Chem.*, 1886, **10**, 391–400.
20. K. . Blass, *Clin. Biochem.* , 2000, **33**, 225–226.
21. M. Wyss and R. Kaddurah-Daouk, *Physiol. Rev.*, 2000, **80**, 1107–1213.
22. J. C. Verhave, P. Fesler, J. Ribstein, G. Du Cailar and A. Mimran, *Am. J. Kidney Dis.*, 2005, **46**, 233–241.
23. O. Folin and H. Wu, *J. Biol. Chem.*, 1919, **38**, 81–110.
24. J. R. Delanghe and M. M. Speeckaert, *NDT Plus*, 2011, **4**, 83–86.
25. D. B. Barr, L. C. Wilder, S. P. Caudill, A. J. Gonzalez, L. L. Needham and J. L. Pirkle, *Environ. Health Perspect.*, 2005, **113**, 192–200.
26. E. W. Holmes, T. H. Oeser, S. E. Kahn, L. Bekeris and E. W. Bermes, *Ann. Clin. Lab. Sci.*, 1983, **13**, 503–10.
27. M. Ou, Y. Song, S. Li, G. Liu, J. Jia, M. Zhang, H. Zhang and C. Yu, *PLoS*

- One*, 2015, **10**, e0133912.
28. L. Bernstone, A. Jayanti and B. Keevil, *Clin. Mass Spectrom.*, 2019, **11**, 21–26.
 29. M. F. Boeniger, L. K. Lowry and J. Rosenberg, *Am. Ind. Hyg. Assoc. J.*, 1993, **54**, 615–627.
 30. L. Alessio, A. Berlin, A. Dell’Orto, F. Toffoletto and I. Ghezzi, *Int. Arch. Occup. Environ. Health*, 1985, **55**, 99–106.
 31. M. R. Munafò, B. A. Nosek, D. V. M. Bishop, K. S. Button, C. D. Chambers, N. Percie du Sert, U. Simonsohn, E.-J. Wagenmakers, J. J. Ware and J. P. A. Ioannidis, *Nat. Hum. Behav.*, 2017, **1**, 0021.
 32. D. L. Heavner, W. T. Morgan, S. B. Sears, J. D. Richardson, G. D. Byrd and M. W. Ogden, *J. Pharm. Biomed. Anal.*, 2006, **40**, 928–942.
 33. E. J. Cone, Y. H. Caplan, F. Moser, T. Robert, M. K. Shelby and D. L. Black, *J. Anal. Toxicol.*, 2009, **33**, 1–7.
 34. V. Chadha, U. Garg and U. S. Alon, *Pediatr. Nephrol.*, 2001, **16**, 374–382.
 35. B. M. Warrack, S. Hnatyshyn, K. H. Ott, M. D. Reily, M. Sanders, H. Zhang and D. M. Drexler, *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.*, 2009, **877**, 547–552.
 36. F. Dieterle, A. Ross, G. Schlotterbeck and H. Senn, *Anal. Chem.*, 2006, **78**, 4281–4290.
 37. S. M. Kohl, M. S. Klein, J. Hochrein, P. J. Oefner, R. Spang and W. Gronwald, *Metabolomics*, 2012, **8**, 146–160.
 38. C. M. Slupsky, K. N. Rankin, J. Wagner, H. Fu, D. Chang, A. M. Weljie, E. J. Saude, B. Lix, D. J. Adamko, S. Shah, R. Greiner, Brian D. Sykes and T. J. Marrie†, 2007, **79**, 6995–7004.
 39. C. Prasad and P. Galbraith, *Clin. Genet.*, 2005, **68**, 199–203.
 40. P. S. Harper, *Landmarks in Medical Genetics: Classic Papers with Commentaries*, Oxford University Press, 2004.
 41. H. J. Muller, *Am. J. Hum. Genet.*, 1949, **1**, 1–18.
 42. M. R. Zavan, *Principles and Practice of Screening for Disease.*, World Health Organization, Geneva, 1969, vol. 123.
 43. C. P. Lorentz, E. D. Wieben, A. Tefferi, D. A. H. Whiteman and G. W. Dewald, *Mayo Clin. Proc.*, 2002, **77**, 773–782.
 44. J. F. Gusella, N. S. Wexler, P. M. Conneally, S. L. Naylor, M. A. Anderson, R. E. Tanzi, P. C. Watkins, K. Ottina, M. R. Wallace, A. Y. Sakaguchi, A. B. Young, I. Shoulson, E. Bonilla and J. B. Martin, *Nature*, 1983, **306**, 234–238.
 45. J. R. Riordan, J. M. Rommens, B. Kerem, N. Alon, R. Rozmahel, Z. Grzelczak, J. Zielenski, S. Lok, N. Plavsic and J. L. Chou, *Science*, 1989, **245**, 1066–73.
 46. B. Kuska, *J. Natl. Cancer Inst.*, 1998, **90**, 93.
 47. J. C. Venter, M. D. Adams, E. W. Myers, P. W. Li, R. J. Mural, G. G. Sutton, H. O. Smith, M. Yandell, C. A. Evans, R. A. Holt, J. D. Gocayne, P. Amanatides, R. M. Ballew, D. H. Huson, J. R. Wortman, Q. Zhang, C. D. Kodira, X. H. Zheng, L. Chen, M. Skupski, G. Subramanian, P. D. Thomas,

- J. Zhang, G. L. Gabor Miklos, C. Nelson, S. Broder, A. G. Clark, J. Nadeau, V. A. McKusick, N. Zinder, A. J. Levine, R. J. Roberts, M. Simon, C. Slayman, M. Hunkapiller, R. Bolanos, A. Delcher, I. Dew, D. Fasulo, M. Flanigan, L. Florea, A. Halpern, S. Hannenhalli, S. Kravitz, S. Levy, C. Mobarry, K. Reinert, K. Remington, J. Abu-Threideh, E. Beasley, K. Biddick, V. Bonazzi, R. Brandon, M. Cargill, I. Chandramouliswaran, R. Charlab, K. Chaturvedi, Z. Deng, V. Di Francesco, P. Dunn, K. Eilbeck, C. Evangelista, A. E. Gabrielian, W. Gan, W. Ge, F. Gong, Z. Gu, P. Guan, T. J. Heiman, M. E. Higgins, R. R. Ji, Z. Ke, K. A. Ketchum, Z. Lai, Y. Lei, Z. Li, J. Li, Y. Liang, X. Lin, F. Lu, G. V Merkulov, N. Milshina, H. M. Moore, A. K. Naik, V. A. Narayan, B. Neelam, D. Nusskern, D. B. Rusch, S. Salzberg, W. Shao, B. Shue, J. Sun, Z. Wang, A. Wang, X. Wang, J. Wang, M. Wei, R. Wides, C. Xiao, C. Yan, A. Yao, J. Ye, M. Zhan, W. Zhang, H. Zhang, Q. Zhao, L. Zheng, F. Zhong, W. Zhong, S. Zhu, S. Zhao, D. Gilbert, S. Baumhueter, G. Spier, C. Carter, A. Cravchik, T. Woodage, F. Ali, H. An, A. Awe, D. Baldwin, H. Baden, M. Barnstead, I. Barrow, K. Beeson, D. Busam, A. Carver, A. Center, M. L. Cheng, L. Curry, S. Danaher, L. Davenport, R. Desilets, S. Dietz, K. Dodson, L. Doup, S. Ferreira, N. Garg, A. Gluecksmann, B. Hart, J. Haynes, C. Haynes, C. Heiner, S. Hladun, D. Hostin, J. Houck, T. Howland, C. Ibegwam, J. Johnson, F. Kalush, L. Kline, S. Koduru, A. Love, F. Mann, D. May, S. McCawley, T. McIntosh, I. McMullen, M. Moy, L. Moy, B. Murphy, K. Nelson, C. Pfannkoch, E. Pratts, V. Puri, H. Qureshi, M. Reardon, R. Rodriguez, Y. H. Rogers, D. Romblad, B. Ruhfel, R. Scott, C. Sitter, M. Smallwood, E. Stewart, R. Strong, E. Suh, R. Thomas, N. N. Tint, S. Tse, C. Vech, G. Wang, J. Wetter, S. Williams, M. Williams, S. Windsor, E. Winn-Deen, K. Wolfe, J. Zaveri, K. Zaveri, J. F. Abril, R. Guigó, M. J. Campbell, K. V Sjolander, B. Karlak, A. Kejariwal, H. Mi, B. Lazareva, T. Hatton, A. Narechania, K. Diemer, A. Muruganujan, N. Guo, S. Sato, V. Bafna, S. Istrail, R. Lippert, R. Schwartz, B. Walenz, S. Yooseph, D. Allen, A. Basu, J. Baxendale, L. Blick, M. Caminha, J. Carnes-Stine, P. Caulk, Y. H. Chiang, M. Coyne, C. Dahlke, A. Mays, M. Dombroski, M. Donnelly, D. Ely, S. Esparham, C. Fosler, H. Gire, S. Glanowski, K. Glasser, A. Glodek, M. Gorokhov, K. Graham, B. Gropman, M. Harris, J. Heil, S. Henderson, J. Hoover, D. Jennings, C. Jordan, J. Jordan, J. Kasha, L. Kagan, C. Kraft, A. Levitsky, M. Lewis, X. Liu, J. Lopez, D. Ma, W. Majoros, J. McDaniel, S. Murphy, M. Newman, T. Nguyen, N. Nguyen, M. Nodell, S. Pan, J. Peck, M. Peterson, W. Rowe, R. Sanders, J. Scott, M. Simpson, T. Smith, A. Sprague, T. Stockwell, R. Turner, E. Venter, M. Wang, M. Wen, D. Wu, M. Wu, A. Xia, A. Zandieh and X. Zhu, *Science*, 2001, **291**, 1304–51.
48. C. Kooperberg, M. LeBlanc and V. Obenchain, *Genet. Epidemiol.*, 2010, **34**, 643–652.
49. S. Sandoval-Motta, M. Aldana, E. Martínez-Romero and A. Frank, *Front. Genet.*, 2017, **8**, 80.

50. J. M. Hall, M. K. Lee, B. Newman, J. E. Morrow, L. A. Anderson, B. Huey and M. C. King, *Science*, 1990, **250**, 1684–9.
51. R. Wooster, S. L. Neuhausen, J. Mangion, Y. Quirk, D. Ford, N. Collins, K. Nguyen, S. Seal, T. Tran, D. Averill, P. Fields, G. Marshall, S. Narod, G. M. Lenoir, H. Lynch, J. Feunteun, P. Devilee, C. J. Cornelisse, F. H. Menko, P. A. Daly, W. Ormiston, R. McManus, C. Pye, C. M. Lewis, L. A. Cannon-Albright, J. Peto, B. A. J. Ponder, M. H. Skolnick, D. F. Easton, D. E. Goldgar and M. R. Stratton, *Science (80-.)*, 1994, **265**, 2088–2090.
52. A. J. Papoutsis, S. D. Lamore, G. T. Wondrak, O. I. Selmin and D. F. Romagnolo, *J. Nutr.*, 2010, **140**, 1607–1614.
53. J. G. Brody, K. B. Moysich, O. Humblet, K. R. Attfield, G. P. Beehler and R. A. Rudel, *Cancer*, 2007, **109**, 2667–2711.
54. T. Shi, E. Song, S. Nie, K. D. Rodland, T. Liu, W. J. Qian and R. D. Smith, *Proteomics*, 2016, **16**, 2160–2182.
55. A. Pandey and M. Mann, *Nature*, 2000, **405**, 837–846.
56. R. E. Banks, M. J. Dunn, D. F. Hochstrasser, J. C. Sanchez, W. Blackstock, D. J. Pappin and P. J. Selby, *Lancet*, 2000, **356**, 1749–1756.
57. E. P. Diamandis, *BMC Med.*, 2012, **10**, 87.
58. R. J. de Winter, J. Fischer, R. Bholasingh, J. P. van Straalen, T. de Jong, J. G. Tijssen and G. T. Sanders, *Clin. Chem.*, 2000, **46**, 1597–603.
59. I. Ezkurdia, D. Juan, J. M. Rodriguez, A. Frankish, M. Diekhans, J. Harrow, J. Vazquez, A. Valencia and M. L. Tress, *Hum. Mol. Genet.*, 2014, **23**, 5866–78.
60. A. G. Paulovich, J. R. Whiteaker, A. N. Hoofnagle and P. Wang, *Proteomics - Clin. Appl.*, 2008, **2**, 1386–1402.
61. H. Mischak, J. P. A. Ioannidis, A. Argiles, T. K. Attwood, E. Bongcam-Rudloff, M. Broenstrup, A. Charonis, G. P. Chrousos, C. Delles, A. Dominiczak, T. Dylag, J. Ehrich, J. Egido, P. Findeisen, J. Jankowski, R. W. Johnson, B. A. Julien, T. Lankisch, H. Y. Leung, D. Maahs, F. Magni, M. P. Manns, E. Manolis, G. Mayer, G. Navis, J. Novak, A. Ortiz, F. Persson, K. Peter, H. H. Riese, P. Rossing, N. Sattar, G. Spasovski, V. Thongboonkerd, R. Vanholder, J. P. Schanstra and A. Vlahou, *Eur. J. Clin. Invest.*, 2012, **42**, 1027–1036.
62. Q. Tian, N. D. Price and L. Hood, *J. Intern. Med.*, 2012, **271**, 111–121.
63. G. J. Patti, O. Yanes and G. Siuzdak, *Nat. Rev. Mol. Cell Biol.*, 2012, **13**, 263–9.
64. M. Suárez, A. Caimari, J. M. del Bas and L. Arola, *TrAC - Trends Anal. Chem.*, 2017, **96**, 79–88.
65. C. B. Newgard, *Cell Metab.*, 2017, **25**, 43–56.
66. C. H. Johnson, J. Ivanisevic and G. Siuzdak, *Nat. Rev. Mol. Cell Biol.*, 2016, **17**, 451–459.
67. A. O. Sullivan, M. J. Gibney and L. Brennan, *Am J Clin Nutr*, 2011, **93**, 314–321.
68. A. Dibattista, N. McIntosh, M. Lamoureux, O. Y. Al-Dirbashi, P.

- Chakraborty and P. Britz-McKibbin, *Anal. Chem.*, 2017, **89**, 8112–8121.
69. C. Caux, C. O'Brien and C. Viau, *Appl. Occup. Environ. Hyg.*, 2002, **17**, 379–386.
 70. C. P. Wild, *Int. J. Epidemiol.*, 2012, **41**, 24–32.
 71. P. D. Juarez, P. Matthews-Juarez, D. B. Hood, W. Im, R. S. Levine, B. J. Kilbourne, M. A. Langston, M. Z. Al-Hamdan, W. L. Crosson, M. G. Estes, S. M. Estes, V. K. Agboto, P. Robinson, S. Wilson and M. Y. Lichtveld, *Int. J. Environ. Res. Public Health*, 2014, **11**, 12866–12895.
 72. A. Burga and B. Lehner, *FEBS J.*, 2012, **279**, 3765–3775.
 73. A. N. Macedo, S. Mathiaparanam, L. Brick, K. Keenan, T. Gonska, L. Pedder, S. Hill and P. Britz-McKibbin, *ACS Cent. Sci.*, 2017, **3**, 904–913.
 74. R. D. Beger, W. Dunn, M. A. Schmidt, S. S. Gross, J. A. Kirwan, M. Cascante, L. Brennan, D. S. Wishart, M. Oresic, T. Hankemeier, D. I. Broadhurst, A. N. Lane, K. Suhre, G. Kastenmüller, S. J. Sumner, I. Thiele, O. Fiehn and R. Kaddurah-Daouk, *Metabolomics*, 2016, **12**, 149.
 75. C. L. Gavaghan, E. Holmes, E. Lenz, I. D. Wilson and J. K. Nicholson, *FEBS Lett.*, 2000, **484**, 169–174.
 76. G. Srinivasan, C. M. James and J. A. Krzycki, *Science*, 2002, **296**, 1459–1462.
 77. J. P. A. Ioannidis, *PLoS Med.*, 2005, **2**, 696–701.
 78. A. C. Schrimpe-Rutledge, S. G. Codreanu, S. D. Sherrod and J. A. McLean, *J. Am. Soc. Mass Spectrom.*, 2016, **27**, 1897–1905.
 79. A. Scalbert, L. Brennan, O. Fiehn, T. Hankemeier, B. S. Kristal, B. van Ommen, E. Pujos-Guillot, E. Verheij, D. Wishart and S. Wopereis, *Metabolomics*, 2009, **5**, 435–458.
 80. C. Wingren, A. Sandström, R. Segersvärd, A. Carlsson, R. Andersson, M. Löhr and C. A. K. Borrebaeck, *Cancer Res.*, 2012, **72**, 2481–2490.
 81. A. J. Atkinson, W. A. Colburn, V. G. DeGruttola, D. L. DeMets, G. J. Downing, D. F. Hoth, J. A. Oates, C. C. Peck, R. T. Schooley, B. A. Spilker, J. Woodcock and S. L. Zeger, *Clin. Pharmacol. Ther.*, 2001, **69**, 89–95.
 82. N. J. Talley and S. O'Connor, *Clinical Examination: A Systematic Guide To Physical Diagnosis*, Elsevier Canada, 7th edn., 2013.
 83. J. C. Muhrer, *Nurse Pract.*, 2014, **39**, 30–35.
 84. E. Weir, *Cmaj*, 2002, **166**, 1041–1043.
 85. M. Fingerhut, D. I. Nelson, T. Driscoll, M. Concha-Barrientos, K. Steenland, L. Punnett, A. Prüss-Ustun, J. Leigh, C. Corvalan, G. Eijkemans and J. Takala, *Med. del Lav.*, 2006, **97**, 313–321.
 86. T. E. Edes, in *Clinical Methods: The History, Physical, and Laboratory Examinations.*, eds. W. HK, W. Hall and J. Hurst, Butterworths, Boston, 3rd edn., 1990, p. 4.
 87. G. Cohen and R. Java, *Appl. Cogn. Psychol.*, 1995, **9**, 273–288.
 88. L. Bickley, P. G. Szilagli and B. Bates, *Bates guide to physical examination and history taking*, Wolters Kluwer Health/Lippincott Williams & Wilkins, 2012.

89. F. H. Wians, *Lab. Med.*, 2009, **40**, 105–113.
90. J. K. Lee, E. G. Liles, S. Bent, T. R. Levin and D. A. Corley, *Ann. Intern. Med.*, 2014, **160**, 171–181.
91. L. E. Gibson and R. E. Cooke, *Pediatrics*, 1959, **23**, 545–549.
92. J. Wolcott, A. Schwartz and C. Goodman, *Laboratory Medicine: A National Status Report*, 2008.
93. A. DiBattista, D. Rampersaud, H. Lee, M. Kim and P. Britz-McKibbin, *Anal. Chem.*, 2017, **89**, 11853–11861.
94. J. W. Honour, *Ann. Clin. Biochem.*, 2011, **48**, 97–111.
95. B. L. Therrell and J. Adams, *J. Inherit. Metab. Dis.*, 2007, **30**, 447–465.
96. E. Holmes, I. D. Wilson and J. K. Nicholson, *Cell*, 2008, **134**, 714–717.
97. M. Ezzati, A. D. Lopez, A. Rodgers, S. Vander Hoorn and C. J. L. Murray, *Lancet*, 2002, **360**, 1347–1360.
98. S. M. Rappaport, *J. Expo. Sci. Environ. Epidemiol.*, 2011, **21**, 5–9.
99. G. Nicholson, M. Rantalainen, A. D. Maher, J. V. Li, D. Malmödin, K. R. Ahmadi, J. H. Faber, I. B. Hallgrímsdóttir, A. Barrett, H. Toft, M. Krestyaninova, J. Viksna, S. G. Neogi, M. E. Dumas, U. Sarkans, B. W. Silverman, P. Donnelly, J. K. Nicholson, M. Allen, K. T. Zondervan, J. C. Lindon, T. D. Spector, M. I. McCarthy, E. Holmes, D. Baunsgaard and C. C. Holmes, *Mol. Syst. Biol.*, 2011, **7**, 525.
100. P. Vineis, M. Chadeau-Hyam, H. Gmuender, J. Gulliver, Z. Herceg, J. Kleinjans, M. Kogevinas, S. Kyrtopoulos, M. Nieuwenhuijsen, D. H. Phillips, N. Probst-Hensch, A. Scalbert, R. Vermeulen and C. P. Wild, *Int. J. Hyg. Environ. Health*, 2017, **220**, 142–151.
101. M. Vrijheid, R. Slama, O. Robinson, L. Chatzi, M. Coen, P. van den Hazel, C. Thomsen, J. Wright, T. J. Athersuch, N. Avellana, X. Basagaña, C. Brochot, L. Bucchini, M. Bustamante, A. Carracedo, M. Casas, X. Estivill, L. Fairley, D. van Gent, J. R. Gonzalez, B. Granum, R. Gražulevičiene, K. B. Gutzkow, J. Julvez, H. C. Keun, M. Kogevinas, R. R. C. McEachan, H. M. Meltzer, E. Sabidó, P. E. Schwarze, V. Siroux, J. Sunyer, E. J. Want, F. Zeman and M. J. Nieuwenhuijsen, *Environ. Health Perspect.*, 2014, **122**, 535–544.
102. Hercules Exposome Research Center, Overview - Hercules Exposome Research Center, <https://emoryhercules.com/about/overview/> (accessed 3 June 2018).
103. S. H. Shah, J. L. Sun, R. D. Stevens, J. R. Bain, M. J. Muehlbauer, K. S. Pieper, C. Haynes, E. R. Hauser, W. E. Kraus, C. B. Granger, C. B. Newgard, R. M. Califf and L. K. Newby, *Am. Heart J.*, 2012, **163**, 844–850.
104. V. Chajès, A. C. M. Thiébaud, M. Rotival, E. Gauthier, V. Maillard, M. C. Boutron-Ruault, V. Joulin, G. M. Lenoir and F. Clavel-Chapelon, *Am. J. Epidemiol.*, 2008, **167**, 1312–1320.
105. M. Porta, *A Dictionary of Epidemiology*, 2014.
106. B. Sibbald and M. Roland, *BMJ*, 1998, **316**, 201.
107. A. Saint-Raymond, S. Hill, J. Martines, R. Bahl, O. Fontaine and L. Bero,

- Lancet*, 2010, **376**, 229–230.
108. T. R. Frieden, *N. Engl. J. Med.*, 2017, **377**, 465–475.
 109. V. J. Burroughs, R. W. Maxey and R. A. Levy, *J. Natl. Med. Assoc.*, 2002, **94**, 1.
 110. S. K. Tate and D. B. Goldstein, *Nat. Genet.*, 2004, **36**, S34–S42.
 111. A. Black, E. Guilbert, D. Costescu, S. Dunn, W. Fisher, S. Kives, M. Mirosh, W. V. Norman, H. Pymar, R. Reid, G. Roy, H. Varto, A. Waddington, M. S. Wagner, A. M. Whelan, C. Ferguson, C. Fortin, M. Kielly, S. Mansouri and N. Todd, *J. Obstet. Gynaecol. Canada*, 2015, **37**, 936–938.
 112. G. B. Louis, T. Damstra, F. Díaz-Barriga, E. Faustman, U. Hass, R. Kavlock, C. Kimmel, G. Kimmel, K. Krishnan, U. Luderer and L. Sheldon, *Environmental Health Criteria 237: Principles for Evaluating Health Risks in Children Associated with Exposure to Chemicals*, 2006.
 113. FDA-NIH Biomarker Working Group, *BEST (Biomarkers, EndpointS, and other Tools)*, Food and Drug Administration (US), Silver Spring, MD, 2017.
 114. J. Xia, D. I. Broadhurst, M. Wilson and D. S. Wishart, *Metabolomics*, 2013, **9**, 280–299.
 115. Burke, *Biomark. Cancer*, 2016, **8**, 89.
 116. K. Bingol and R. Brüsweiler, *Anal. Chem.*, 2014, **86**, 47–57.
 117. A. Paul, S. Kumar, A. Raj, A. A. Sonkar, S. Jain, A. Singhai and R. Roy, *Metabolomics*, 2018, **14**, 119.
 118. Z. Pan and D. Raftery, *Anal. Bioanal. Chem.*, 2007, **387**, 525–527.
 119. G. A. Nagana Gowda and D. Raftery, *J. Magn. Reson.*, 2015, **260**, 144–160.
 120. R. K. Harris, E. D. Becker, S. M. Cabral de Menezes, R. Goodfellow and P. Granger, *Pure Appl. Chem.*, 2001, **73**, 1795–1818.
 121. E. E. Kwan and S. G. Huang, *European J. Org. Chem.*, 2008, **2008**, 2671–2688.
 122. A. H. M. Emwas, R. M. Salek, J. L. Griffin and J. Merzaban, *Metabolomics*, 2013, **9**, 1048–1072.
 123. D. D. Marshall and R. Powers, *Prog. Nucl. Magn. Reson. Spectrosc.*, 2017, **100**, 1–16.
 124. M. Elyashberg, *TrAC - Trends Anal. Chem.*, 2015, **69**, 88–97.
 125. J. L. Markley, R. Brüsweiler, A. S. Edison, H. R. Eghbalian, R. Powers, D. Raftery and D. S. Wishart, *Curr. Opin. Biotechnol.*, 2017, **43**, 34–40.
 126. A. Constans, *Sci.*, 2000.
 127. B. Zhou, J. F. Xiao, L. Tuli and H. W. Ransom, *Mol. BioSyst.*, 2012, **8**, 470–481.
 128. G. N. Gowda, S. Zhang, H. Gu, V. Asiago, N. Shanaiah and D. Raftery, *Expert Rev. Mol. Diagn.*, 2008, **8**, 617–633.
 129. H. Budzikiewicz and R. D. Grigsby, *Mass Spectrom. Rev.*, 2006, **25**, 146–157.
 130. Y. Wang, S. Liu, Y. Hu, P. Li and J. B. Wan, *RSC Adv.*, 2015, **5**, 78728–78737.
 131. C. Junot, G. Madalinski, J. C. Tabet and E. Ezan, *Analyst*, 2010, **135**, 2203–2219.

132. T. M. Annesley, *Clin. Chem.*, 2003, **49**, 1041–1044.
133. J. P. Antignac, K. De Wasch, F. Monteau, H. De Brabander, F. Andre and B. Le Bizec, *Anal. Chim. Acta*, 2005, **529**, 129–136.
134. A. Garcia and C. Barbas, in *Methods in molecular biology (Clifton, N.J.)*, 2011, vol. 708, pp. 191–204.
135. National Institute of Standards and Technology, NIST Mass Spectrom. Data Cent., <https://chemdata.nist.gov/> (accessed 30 June 2019).
136. Wiley, Wiley Registry of Mass Spectral Data, 11th Edition, <https://www.wiley.com/en-ca/Wiley+Registry+of+Mass+Spectral+Data%2C+11th+Edition-p-9781119171010> (accessed 30 June 2019).
137. O. Fiehn, *Curr. Protoc. Mol. Biol.*, 2016, **2016**, 30.4.1-30.4.32.
138. M. M. Koek, R. H. Jellema, J. van der Greef, A. C. Tas and T. Hankemeier, *Metabolomics*, 2011, **7**, 307–328.
139. K. R. Chalcraft, R. Lee, C. Mills and P. Britz-McKibbin, *Anal. Chem.*, 2009, **81**, 2506–2515.
140. I. D. Wilson, R. Plumb, J. Granger, H. Major, R. Williams and E. M. Lenz, *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.*, 2005, **817**, 67–76.
141. M. W. Dong, *Modern HPLC for Practicing Scientists*, John Wiley and Sons, 2006.
142. R. S. Plumb, P. Rainville, B. W. Smith, K. A. Johnson, J. Castro-Perez, I. D. Wilson and J. K. Nicholson, 2006, **78**, 7278–7283.
143. A. E. Ibrahim, H. Hashem, M. Elhenawee and H. Saleh, *J. Sep. Sci.*, 2018, **41**, 1734–1742.
144. H. J. Issaq, E. Abbott and T. D. Veenstra, *J. Sep. Sci.*, 2008, **31**, 1936–1947.
145. P. Jandera, *Anal. Chim. Acta*, 2011, **692**, 1–25.
146. H. G. Gika, I. D. Wilson and G. A. Theodoridis, *J. Chromatogr. B*, 2014, **966**, 1–6.
147. R. Ramautar, A. Demirci and G. J. de Jong, *TrAC - Trends Anal. Chem.*, 2006, **25**, 455–466.
148. D. Rojo, C. Barbas and F. J. Rupérez, *Bioanalysis*, 2012, **4**, 1235–1243.
149. H. P. Nguyen and K. A. Schug, *J. Sep. Sci.*, 2008, **31**, 1465–1480.
150. B. Buszewski and S. Noga, *Anal. Bioanal. Chem.*, 2012, **402**, 231–47.
151. K. Spagou, I. D. Wilson, P. Masson, G. Theodoridis, N. Raikos, M. Coen, E. Holmes, J. C. Lindon, R. S. Plumb, J. K. Nicholson and E. J. Want, *Anal. Chem.*, 2011, **83**, 382–390.
152. W. Jian, Y. Xu, R. W. Edom and N. Weng, *Bioanalysis*, 2011, **3**, 899–912.
153. R. Ramautar, G. W. Somsen and G. J. de Jong, *Electrophoresis*, 2015, **36**, 212–224.
154. S. Azab, R. Ly and P. Britz-McKibbin, *Anal. Chem.*, 2019, **91**, 2329–2336.
155. M. R. N. Monton and T. Soga, *J. Chromatogr. A*, 2007, **1168**, 237–246.
156. R. Ramautar, E. Nevedomskaya, O. A. Mayboroda, A. M. Deelder, I. D. Wilson, H. G. Gika, G. A. Theodoridis, G. W. Somsen and G. J. De Jong, *Mol. Biosyst.*, 2011, **7**, 194–199.

157. I. G. Arcibal, M. F. Santillo and A. G. Ewing, *Anal. Bioanal. Chem.*, 2007, **387**, 51–57.
158. G. de Jong, Ed., *Capillary electrophoresis-mass spectrometry (CE-MS) : principles and applications*, Weinheim, Germany, 2016.
159. J. Godzien, A. Garcia, A. López-Gonzalvez and C. Barbas, in *Capillary Electrophoresis–Mass Spectrometry for Metabolomics*, Royal Society of Chemistry, 2018, pp. 161–183.
160. T. Faller and H. Engelhardt, *J. Chromatogr. A*, 1999, **853**, 83–94.
161. N. L. Kuehnbaum, A. Kormendi and P. Britz-Mckibbin, *Anal. Chem.*, 2013, **85**, 10664–10669.
162. M. Saoi, A. Li, C. McGlory, T. Stokes, M. T. von Allmen, S. M. Phillips and P. Britz-McKibbin, *Metabolites*, 2019, **9**, 134.
163. W. B. Dunn, A. Erban, R. J. M. Weber, D. J. Creek, M. Brown, R. Breitling, T. Hankemeier, R. Goodacre, S. Neumann, J. Kopka and M. R. Viant, *Metabolomics*, 2013, **9**, 44–66.
164. L. D. Roberts, A. L. Souza, R. E. Gerszten and C. B. Clish, *Curr. Protoc. Mol. Biol.*, 2012, **1**, 1–24.
165. W. Forstmeier, E. J. Wagenmakers and T. H. Parker, *Biol. Rev.*, 2017, **92**, 1941–1968.
166. R. Baran, *Metabolomics*, 2017, **13**, 1–4.
167. R. A. van den Berg, H. C. J. Hoefsloot, J. A. Westerhuis, A. K. Smilde and M. J. van der Werf, *BMC Genomics*, 2006, **7**, 142.
168. C. De Felice, A. Cortelazzo, S. Leoncini, C. Signorini, J. Hayek and L. Ciccoli, *J. Siena Acad. Sci.*, 2015, **7**, 15–22.
169. D. S. Wishart, Y. D. Feunang, A. Marcu, A. C. Guo, K. Liang, R. Vázquez-Fresno, T. Sajed, D. Johnson, C. Li, N. Karu, Z. Sayeeda, E. Lo, N. Assempour, M. Berjanskii, S. Singhal, D. Arndt, Y. Liang, H. Badran, J. Grant, A. Serra-Cayuela, Y. Liu, R. Mandal, V. Neveu, A. Pon, C. Knox, M. Wilson, C. Manach and A. Scalbert, *Nucleic Acids Res.*, 2018, **46**, D608–D617.
170. C. Guijas, J. R. Montenegro-Burke, X. Domingo-Almenara, A. Palermo, B. Warth, G. Hermann, G. Koellensperger, T. Huan, W. Uritboonthai, A. E. Aisporna, D. W. Wolan, M. E. Spilker, H. P. Benton and G. Siuzdak, *Anal. Chem.*, 2018, **90**, 3156–3164.
171. H. Gowda, J. Ivanisevic, C. H. Johnson, M. E. Kurczy, H. P. Benton, D. Rinehart, T. Nguyen, J. Ray, J. Kuehl, B. Arevalo, P. D. Westenskow, J. Wang, A. P. Arkin, A. M. Deutschbauer, G. J. Patti and G. Siuzdak, *Anal. Chem.*, 2014, **86**, 6931–6939.
172. J. Chong, O. Soufan, C. Li, I. Caraus, S. Li, G. Bourque, D. S. Wishart and J. Xia, *Nucleic Acids Res.*, 2018, **46**, W486–494.
173. A. Alonso, S. Marsal and A. Julià, *Front. Bioeng. Biotechnol.*, 2015, **3**, 23.
174. K. Haug, R. M. Salek, P. Conesa, J. Hastings, P. De Matos, M. Rijnbeek, T. Mahendraker, M. Williams, S. Neumann, P. Rocca-Serra, E. Maguire, A. González-Beltrán, S. A. Sansone, J. L. Griffin and C. Steinbeck, *Nucleic*

- Acids Res.*, 2013, **41**, D781–D786.
175. M. Sud, E. Fahy, D. Cotter, K. Azam, I. Vadivelu, C. Burant, A. Edison, O. Fiehn, R. Higashi, K. S. Nair, S. Sumner and S. Subramaniam, *Nucleic Acids Res.*, 2016, **44**, D463–D470.
176. L. Yi, N. Dong, Y. Yun, B. Deng, D. Ren, S. Liu and Y. Liang, *Anal. Chim. Acta*, 2016, **914**, 17–34.
177. R. Goodacre, D. Broadhurst, A. K. Smilde, B. S. Kristal, J. D. Baker, R. Beger, C. Bessant, S. Connor, G. Capuani, A. Craig, T. Ebbels, D. B. Kell, C. Manetti, J. Newton, G. Paternostro, R. Somorjai, M. Sjöström, J. Trygg and F. Wulfert, *Metabolomics*, 2007, **3**, 231–241.
178. W. B. Dunn, I. D. Wilson, A. W. Nicholls and D. Broadhurst, *Bioanalysis*, 2012, **4**, 2249–2264.
179. H. G. Gika, G. A. Theodoridis, M. Earll and I. D. Wilson, *Bioanalysis*, 2012, **4**, 2239–2247.
180. Y. F. Xu, W. Lu and J. D. Rabinowitz, *Anal. Chem.*, 2015, **87**, 2273–2281.
181. C. Junot, F. Fenaille, B. Colsch and F. Bécher, *Mass Spectrom. Rev.*, 2014, **33**, 471–500.
182. T. Kind, Mass Spectrometry Adduct Calculator — Metabolomics Fiehn Lab, <http://fiehnlab.ucdavis.edu/staff/kind/metabolomics/ms-adduct-calculator>, (accessed 3 October 2017).
183. O. Fiehn, D. Robertson, J. Griffin, M. vab der Werf, B. Nikolau, N. Morrison, L. W. Sumner, R. Goodacre, N. W. Hardy, C. Taylor, J. Fostel, B. Kristal, R. Kaddurah-Daouk, P. Mendes, B. van Ommen, J. C. Lindon and S. A. Sansone, *Metabolomics*, 2007, **3**, 175–178.
184. G. Poste, *Trends Mol. Med.*, 2012, **18**, 717–722.
185. W. B. Dunn, D. Broadhurst, P. Begley, E. Zelena, S. Francis-Mcintyre, N. Anderson, M. Brown, J. D. Knowles, A. Halsall, J. N. Haselden, A. W. Nicholls, I. D. Wilson, D. B. Kell and R. Goodacre, *Nat. Protoc.*, 2011, **6**, 1060–1083.
186. N. L. Kuehnbaum, J. B. Gillen, M. J. Gibala and P. Britz-McKibbin, *Sci. Rep.*, 2014, **4**, 6166.
187. J. Godzien, M. Ciborowski, S. Angulo and C. Barbas, *Electrophoresis*, 2013, **34**, 2812–2826.
188. J. Xia, I. V. Sinelnikov, B. Han and D. S. Wishart, *Nucleic Acids Res.*, 2015, **43**, W251–W257.
189. R. Wehrens, J. A. Hageman, F. van Eeuwijk, R. Kooke, P. J. Flood, E. Wijnker, J. J. B. Keurentjes, A. Lommen, H. D. L. M. van Eekelen, R. D. Hall, R. Mumm and R. C. H. de Vos, *Metabolomics*, 2016, **12**, 1–12.
190. C. Brunius, L. Shi and R. Landberg, *Metabolomics*, 2016, **12**, 173.
191. D. I. Broadhurst and D. B. Kell, *Metabolomics*, 2006, **2**, 171–196.
192. V. V. Hernandez, C. Barbas and D. Dudzik, *Electrophoresis*, 2017, **38**, 2232–2241.
193. P. Bernini, I. Bertini, C. Luchinat, P. Nincheri, S. Staderini and P. Turano, *J. Biomol. NMR*, 2011, **49**, 231–243.

194. J. Laparre, Z. Kaabia, M. Mooney, T. Buckley, M. Sherry, B. Le Bizec and G. Dervilly-Pinel, *Anal. Chim. Acta*, 2016, **951**, 99–107.
195. P. Yin, A. Peter, H. Franken, X. Zhao, S. S. Neukamm, L. Rosenbaum, M. Lucio, A. Zell, H.-U. Häring, G. Xu and R. Lehmann, *Clin. Chem.*, 2013, **59**, 833–845.
196. B. Kamlage, S. Neuber, B. Bethan, S. G. Maldonado, A. Wagner-Golbs, E. Peter, O. Schmitz and P. Schatz, *Metabolites*, 2018, **8**, ii.
197. K. Dettmer, P. A. Aronov and B. D. Hammock, *Mass Spectrom. Rev.*, 2007, **26**, 51–78.
198. J. P. Schaeper and M. J. Sepaniak, *Electrophoresis*, 2000, **21**, 1421–1429.
199. M. Marusteri and V. Bacarea, *Biochem. Medica*, 2010, **13**, 15–32.
200. M. Chadeau-Hyam, G. Campanella, T. Jombart, L. Bottolo, L. Portengen, P. Vineis, B. Liqueur and R. C. H. Vermeulen, *Environ. Mol. Mutagen.*, 2013, **54**, 542–557.
201. M. E. Glickman, S. R. Rao and M. R. Schultz, *J. Clin. Epidemiol.*, 2014, **67**, 850–857.
202. K. F. Vajargah, R. Mehdizadeh and H. Sadeghi-bazargani, 2014, **8**, 411–422.
203. B. Worley and R. Powers, *Curr. Metabolomics*, 2016, **4**, 97–103.
204. B. Worley and R. Powers, *Curr. Metabolomics*, 2012, **1**, 92–107.
205. D. S. Wishart, T. Jewison, A. C. Guo, M. Wilson, C. Knox, Y. Liu, Y. Djoumbou, R. Mandal, F. Aziat, E. Dong, S. Bouatra, I. Sinelnikov, D. Arndt, J. Xia, P. Liu, F. Yallou, T. Bjorn Dahl, R. Perez-Pineiro, R. Eisner, F. Allen, V. Neveu, R. Greiner and A. Scalbert, *Nucleic Acids Res.*, 2013, **41**, D801–7.
206. A. N. de Macedo, J. Macri, P. L. Hudecki, M. Saoi, M. J. McQueen and P. Britz-McKibbin, *J. Appl. Lab. Med.*, 2017, **1**, 649–660.
207. V. Amrhein, S. Greenland and B. McShane, *Nature*, 2019, **567**, 305–307.
208. C. A. Monnig and R. T. Kennedy, *Anal. Chem.*, 1994, **66**, 280R–314R.
209. L. W. Sumner, A. Amberg, D. Barrett, M. H. Beale, R. Beger, C. a Daykin, T. W.-M. Fan, O. Fiehn, R. Goodacre, J. L. Griffin, T. Hankemeier, N. Hardy, J. Harnly, R. Higashi, J. Kopka, A. N. Lane, J. C. Lindon, P. Marriott, A. W. Nicholls, M. D. Reily, J. J. Thaden and M. R. Viant, *Metabolomics*, 2007, **3**, 211–221.
210. A. Biesemeier, U. Schraermeyer and O. Eibl, *Micron*, 2011, **42**, 461–470.
211. M. Saoi, M. Percival, C. Nemr, A. Li, M. Gibala and P. Britz-McKibbin, *Anal. Chem.*, 2019, **91**, 4709–4718.

Chapter 2: The Anomalous Stability of Isoquinoline Boronate Ester Complexes Under Acidic Conditions: A Biosensor for Sialic Acid Determination

Nadine Wellington and Philip Britz-McKibbin

All research, data processing, statistical analysis, and data interpretation described in this manuscript was performed by the author, with important experimental contributions from collaborators and colleagues. Dr. Hilary Jenkins kindly provided NMR spectra, Sabrina Macklai provided thermodynamic binding calculations, and Philip Britz-McKibbin is the principal investigator of the study.

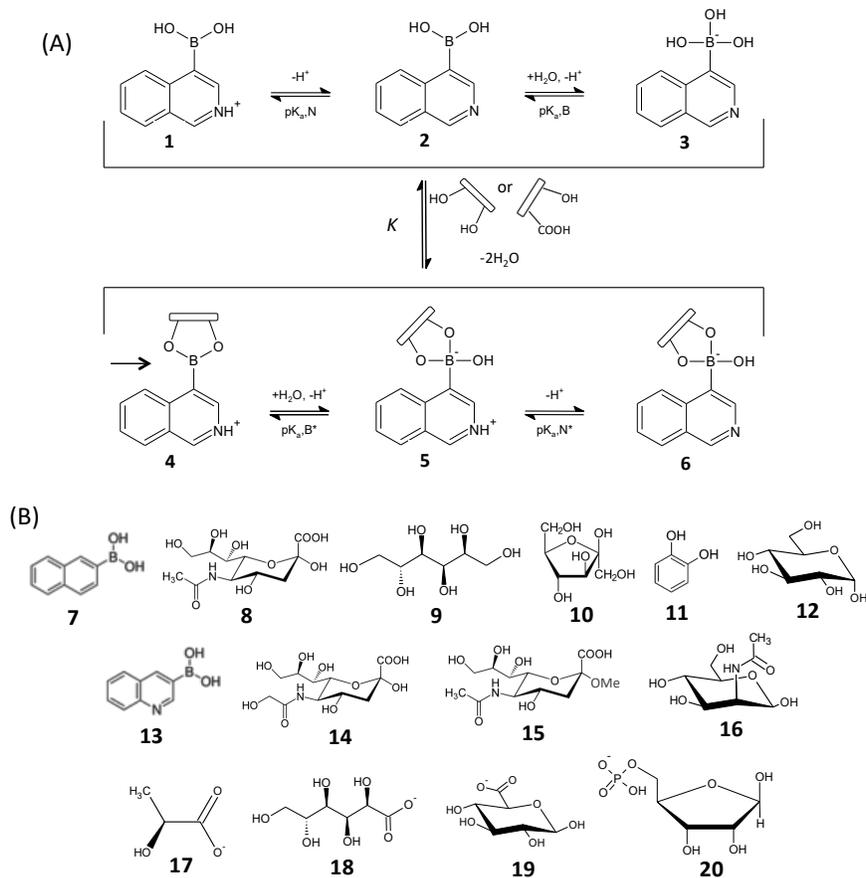
Chapter 2: The Anomalous Stability of Isoquinoline Boronate Ester Complexes Under Acidic Conditions: A Biosensor for Sialic Acid Determination

2.1 Abstract

4-isoquinolineboronic acid (IQBA) exhibits the highest reported binding affinity for sialic acid (Neu5Ac, $K = 5390 \pm 190 \text{ M}^{-1}$) via formation of a stable boronate ester complex under acidic conditions. This anomalous pH-dependent binding enhancement enables the selective analysis of sialic acid with low micromolar detection limits when using UV absorbance or native fluorescence.

2.2 Introduction

Boronic acids (BA) are used in a diverse range of devices and biomimetic materials for molecular sensing applications due to their strong affinity for binding to biomolecules containing vicinal diols, including saccharides, nucleosides and glycoprotein.¹⁻³ BA form reversible covalent complexes primarily with *syn*-periplanar (*cis*) 1,2-diols to generate 5- membered cyclic esters (**4-6** in **Scheme 2.1**). Ligand affinity and specificity can be tuned by functionalizing BA⁴ under optimal buffer pH conditions in conjunction with colorimetric arrays,⁵ as well as their incorporation within nanoscale materials, such as molecular imprinted polymers⁶ or via surface-oriented imprinting.⁷ BA exist as electron-deficient trigonal planar sp^2 -hybridized Lewis acids (**2**) and tetrahedral sp^3 -hybridized boronate anions (**3**), when $\text{pH} > pK_a$ of the boron centre ($pK_{a,B}$). Complexation to high affinity ligands lowers the boron $pK_{a,B}$, yielding a cyclic tetrahedral anionic boronate ester (**5** or **6**) with improved thermodynamic stability as compared to acidic conditions (**4**), where the complex suffers from angle strain and is prone to hydrolysis.⁸ As a result, neutral and alkaline pH conditions generally favor binding of BA with saccharides having suitable vicinal diols.¹



Scheme 2.1: (A) Equilibrium species present for zwitterionic IQBA in aqueous solution as a function of buffer pH and ligand concentration. The thermodynamic stability of the boronate ester complex is dependent on the charge state of IQBA based on reversible complexation with vicinal diols and/or α -hydroxycarboxylates, such as Neu5Ac (14). (B) Model BA and ligands used to assess the unique binding affinity enhancement of IQBA-Neu5Ac under acidic conditions.

Diol functional groups exist in a myriad of biologically important molecules, however BA may also bind strongly to ligands containing other functional groups, such as α -hydroxycarboxylates (**Scheme 2.1**).¹⁰ Nevertheless, BA-based biosensors have been largely focused on glucose biosensing applications as it allows for treatment monitoring of diabetic patients.¹¹⁻¹³ However, apparent binding constants at neutral pH are weak for phenylboronic acid (PBA (7), $K \approx 5 \text{ M}^{-1}$),¹⁴ but can be

greatly improved with the design of a bisboronic acid scaffold for selective trapping of glucopyranose, the major (> 99.5%) species in aqueous solution.¹⁵

Despite their useful optical or electrochemical properties, low toxicity and versatile chemical functionalization, BA probes often suffer from inadequate sensitivity and/or selectivity as compared to enzyme-based electrochemical detectors widely used globally in personal glucose meters.¹⁶ Also, there is urgent need for the design of new BA probes that can target other low abundance sugar metabolites of biological significance beyond glucose. Recently, various heterocyclic BA have been reported to show higher affinity towards neutral monosaccharides at physiological pH relative to traditional PBA analogues.¹⁷⁻¹⁹ In this work, we characterized the pH-dependent binding interactions of 4-isoquinolineboronic acid (IQBA, **1** or **2**) for biorecognition of sialic acid or *N*-acetylneuraminic acid (Neu5Ac, **8**), which serves as the terminal sugar moiety in most glycoproteins mediating critical roles related to cell recognition, virulence and immunological response.²⁰ Neu5Ac proliferation in glycoprotein isoforms also varies in the progression of cardiovascular disease, cancer, and alcoholism, thus represents an important molecular target for routine screening.²¹⁻²³

2.3 Results and Discussion

IQBA is a zwitterionic arylboronic acid probe that undergoes a hyperchromic shift ($\lambda_{max} = 335$ nm) and fluorescence enhancement ($\lambda_{em} = 377$ nm) upon binding to increasing Neu5Ac concentrations (**Figure 2.1A**). Importantly, there is a 65-fold greater binding affinity under acidic conditions (pH 3.0) with a $K = (5390 \pm 190)$ M⁻¹ as compared to neutral (pH 7.0) conditions (**Figure 2.1B**). In contrast, other commonly measured neutral saccharides at pH 7.0 (**Table 2.1**) had far weaker stabilities under acidic conditions, such as sorbitol (**9**), fructose (**10**), catechol (**11**) and glucose (**12**). In fact, the formation constant of IQBA with Neu5Ac ($K = 84 \pm 23$ M⁻¹) is about 10-fold greater than glucose at neutral pH, and 2-fold stronger than reported for 3-(propionamido)-PBA.²⁴ As expected, the binding affinity of naphthalene boronic acid (NBA, **7**) to Neu5Ac revealed far weaker interactions

($K = 6.8 \pm 0.1 \text{ M}^{-1}$) that was independent of buffer pH as it lacks a weakly basic quinolinium ion ($pK_a \approx 5.0$). Additionally, a comparison of the binding of Neu5Ac with a positional isomer of IQBA, namely 3-quinoline boronic acid (QBA, **13**) revealed similar enhancement of binding affinity under acidic conditions albeit with lower overall complex stability ($K = 2600 \pm 150 \text{ M}^{-1}$). Further studies were next aimed to elucidate the exact mechanism of the unusually high binding affinity of IQBA for Neu5Ac under acidic conditions, which is more than 5-fold greater than recently reported pyridine-based boronic acid derivatives ($K \approx 700\text{-}1100 \text{ M}^{-1}$ at pH 5).²⁵ A ^{13}C -NMR spectral overlay confirms that binding of Neu5Ac to IQBA results in a pronounced upfield chemical shifts notably of the carboxylate ($\Delta\delta \approx 2.5 \text{ ppm}$) followed by its adjacent α -hydroxy moiety ($\Delta\delta \approx 0.8 \text{ ppm}$) (**Figure 2.1C**). This is consistent with previous multinuclear NMR studies of PBA interactions with Neu5Ac, which demonstrated involvement of the α -hydroxycarboxylate pendant in boronate ester formation (pH < 9.0) in contrast to binding of the terminal glycerol moiety (C8, C9) that predominates under higher pH conditions.²⁶ Computer molecular modeling confirmed the greater relative stability ($\approx -29 \text{ kcal/mol}$) of Neu5Ac complexation to IQBA when binding via its α -hydroxycarboxylate (C1, C2) as compared to vicinal diol (C8, C9) conformer (**Figure 2S.1** of Supplemental). Additionally, ^{11}B -NMR studies indicated a ternary IQBA-Neu5Ac-phosphate complex occurs in solution that exists in equilibrium primarily as its sp^3 -hybridized tetrahedral boronate anion under acidic conditions (**Figure 2S.2** of Supplemental). Also, we confirmed that optimal binding of IQBA to Neu5Ac occurs at pH ≈ 3 as further acidification ($K \approx 567 \text{ M}^{-1}$, pH = 2) results in protonation of the Neu5Ac carboxylate thereby lowering its nucleophilicity ($pK_a = 2.6$). The binding affinity

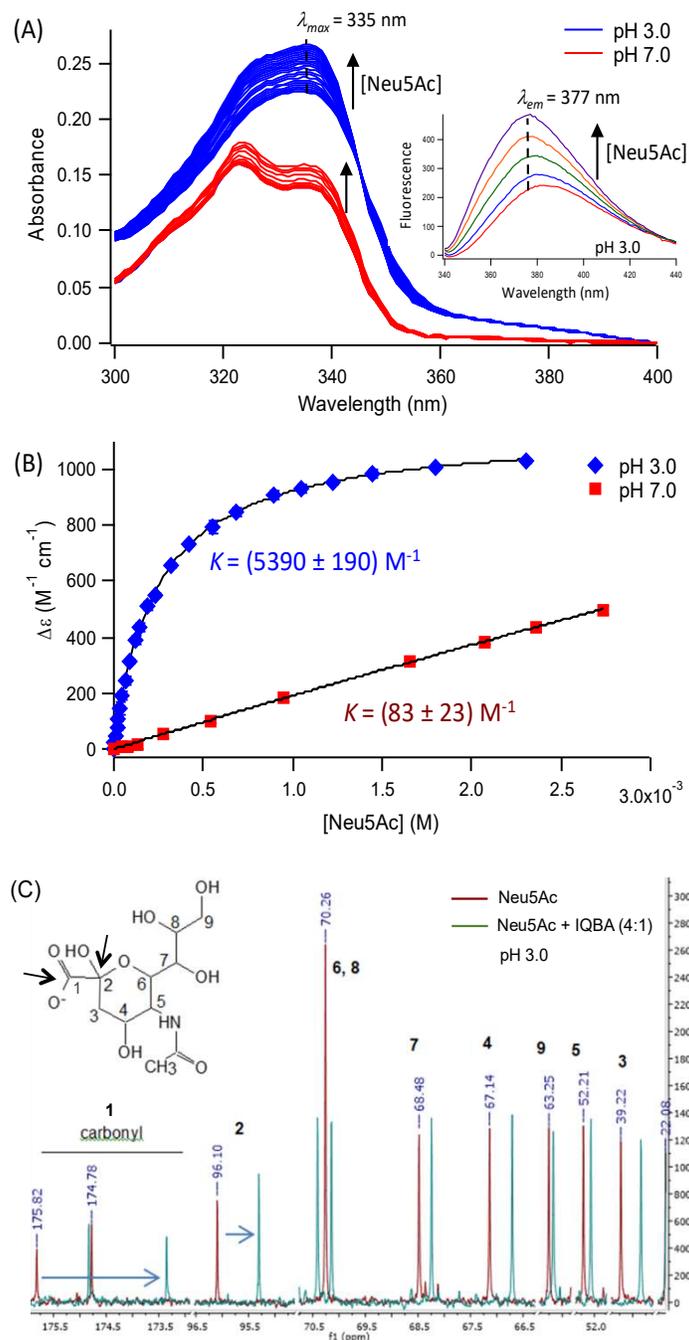


Figure 2.1: A 65-fold enhancement in binding affinity of IQBA to Neu5Ac is achieved under strongly acidic as compared to neutral pH conditions as shown in (A) UV absorbance spectra overlay and (B) binding isotherms based on a 1:1 dynamic complexation model. All studies were performed in triplicate with 40 μ M IQBA in 40 mM phosphate buffer with absorbance changes monitored at 335 nm. These conditions are also optimal for IQBA native fluorescence properties (inset of

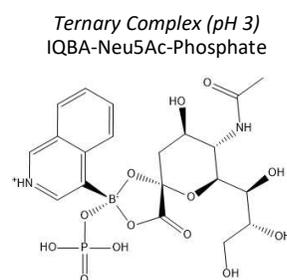
A) with an enhancement at 377 nm ($\Delta\lambda = 42$ nm) upon Neu5Ac binding. (C) ^{13}C -NMR spectra of free Neu5Ac (red) and Neu:IQBA complex confirms the importance of the α -hydroxycarboxylate moiety in stable boronate ester formation based on a -2.5 ppm chemical shift of the C1 carbonyl carbon (175.8 ppm), and a -0.80 ppm shift in the C2 resonance (96.1 ppm).

and selectivity of IQBA to different Neu5Ac analogs, as well as other acidic sugar metabolites (**Scheme 2.1**) were also measured in this study (**Table 2.1**). For example, the mammalian sialic acid derivative that has recently been shown to be expressed in humans during malignant transformation, namely *N*-glycolylneuraminic acid (Neu5Gc, **14**)²⁷ was found to share a similar binding enhancement ($K = 2450 \pm 140 \text{ M}^{-1}$) under acidic conditions, whereas 2-*O*-methylation of Neu5Ac (2-OMe-Neu5Ac, **15**) completely abolished binding affinity ($K < 0.1 \text{ M}^{-1}$) highlighting the key role of both the α -hydroxy moiety and carboxylate anion in stable cyclic boronate ester formation. As expected, *N*-acetylmannosamine (NAM, **16**), a precursor for biosynthesis of Neu5Ac lacking an α -hydroxycarboxylate functionality, has negligible interactions with IQBA at pH 3 ($K \approx 0.29 \text{ M}^{-1}$) similar to other neutral monosaccharides, such as glucose.

In order to better characterize the selectivity of IQBA for biorecognition of targeted analytes in complex biological samples, formation constants were also measured to other classes of acidic sugars and organic acids. For example, lactic acid (**17**) and gluconic acid (**18**), which both contain an α -hydroxycarboxylate moiety were found to have strong binding affinity to IQBA under acidic conditions, whereas glucuronic acid (**19**) had much weaker interactions ($K \approx 70 \text{ M}^{-1}$) given its less flexible pyranose conformation in aqueous solution. Also, the 500-fold enhancement of binding of IQBA to gluconic acid (**18**) as compared to sorbitol (**9**) highlights the critical role of the α -hydroxycarboxylate in greatly enhancing the stability of IQBA complex formation that is resistant to hydrolysis even under strongly acidic conditions. The apparent binding affinity of IQBA to acidic metabolites having α -hydroxycarboxylate moieties are likely dependent on their

Table 2.1: A comparison of the binding affinity and selectivity of IQBA highlighting the anomalous binding enhancement under acidic conditions for Neu5Ac as compared to various other sugars, polyols and sialic acid analogs.

Ligand	pH 3		pH 7	
	K (M^{-1}) ^a	K_{Neu5Ac}/K_{sugar}	K (M^{-1}) ^a	K_{Neu5Ac}/K_{Sugar}
Neu5Ac	5387 ± 192		83 ± 23	
Sorbitol	32 ± 2	168	2576 ± 78	0.03
Fructose	ND ^b	> 10 ³	1372 ± 64	0.06
Glucose	0.13 ± 0.03	> 10 ⁴	8.9 ± 1.4	9.3
Catechol	40.8 ± 1.8	132	3434 ± 183	0.02
2-O-Me Neu5Ac	ND ^b	> 10 ⁷		
Neu5Gc	2450 ± 140	2.2		
NAM	0.29 ± 0.04	> 10 ⁴		
Lactic Acid	2170 ± 605	2.5		
Gluconic Acid	3434 ± 180	1.5		
Glucuronic Acid	70 ± 4	77		
Ribose-5-phosphate	4940 ± 340	1.1		



^aBinding isotherms conducted at ambient temperature with 5 minutes equilibration. Binding constant K determined by fitting the curve $f(c) \propto 1/(1 + Kc)$, where c is concentration of the boronic acid probe. Phosphate buffer concentration held constant at 40 mM and ionic strength at 100 mM.

^bND = Not determined due to very low binding affinity over concentration range studies.

relative acidities, with lower stability measured for the series, lactic acid ($pK_a = 3.86$) < gluconic acid ($pK_a = 3.40$) < Neu5Ac ($pK_a = 2.60$). Moreover, other types of acidic metabolites may also strongly interact with IQBA as reflected by the unexpected high binding affinity measured for ribose-5-phosphate (**20**, $K = 4940 \pm 340 M^{-1}$), which was comparable to Neu5Ac under acidic conditions. In this case, coordination of the strongly acidic phosphate moiety likely occurs in concert with the 1-hydroxy group to form a putative 7-membered cyclic boronate ester complex. This reflects the growing interest in exploiting boron and nucleoside chemistry in medicinal and analytical applications given the structural homology of borate and phosphate anions.²⁸ For instance, formation of a ternary complex between 3-nitrophenylboronic acid, phosphate and a polyol was demonstrated to enhance absorbance responses as a function of phosphate concentration when resolving and detecting UV-transparent, neutral polyol stereoisomers in free solution by capillary

electrophoresis.^{29,30} As a result, BA equilibria processes depicted in **Scheme 2.1** neglect the role of phosphate in solution that takes part in BA complexation besides its function as a buffer for pH control as highlighted in ¹¹B-NMR spectra (**Figure 2S.2** of Supplemental).

A comparison of the analytical performance of IQBA for direct analysis of micromolar concentrations of Neu5Ac was next performed in this study when using UV absorbance and native fluorescence detection under strongly acidic conditions (pH 3) since Neu5Ac binding affinity, as well as IQBA molar absorptivity and fluorescence quantum efficiency (**Figure 2.1A**) is superior than neutral pH conditions. Overall, both methods provided acceptable technical precision (mean CV < 12%, $n = 18$) for Neu5Ac quantification over a 10-fold linear dynamic range ($R^2 > 0.995$). However, sensitivity (*i.e.*, slope, m) was more than 10-fold greater when using native fluorescence as compared to UV absorbance with a corresponding 70% decrease in method detection limit for Neu5Ac from 2.2 μM to 0.63 μM , respectively (**Figure 2.2**). Native fluorescence also likely offers better selectivity in Neu5Ac biosensing applications when analyzing non-invasive human biofluids (*e.g.*, saliva) or biologic samples (*e.g.*, glycoprotein) with fewer spectral or chemical interferences with the exception of certain acidic metabolites identified in this work.

In summary, we report the strongest binding interaction of Neu5Ac with a BA that is greatly enhanced under strongly acidic conditions unlike most saccharides that optimally bind under neutral or alkaline conditions. Previous applications of

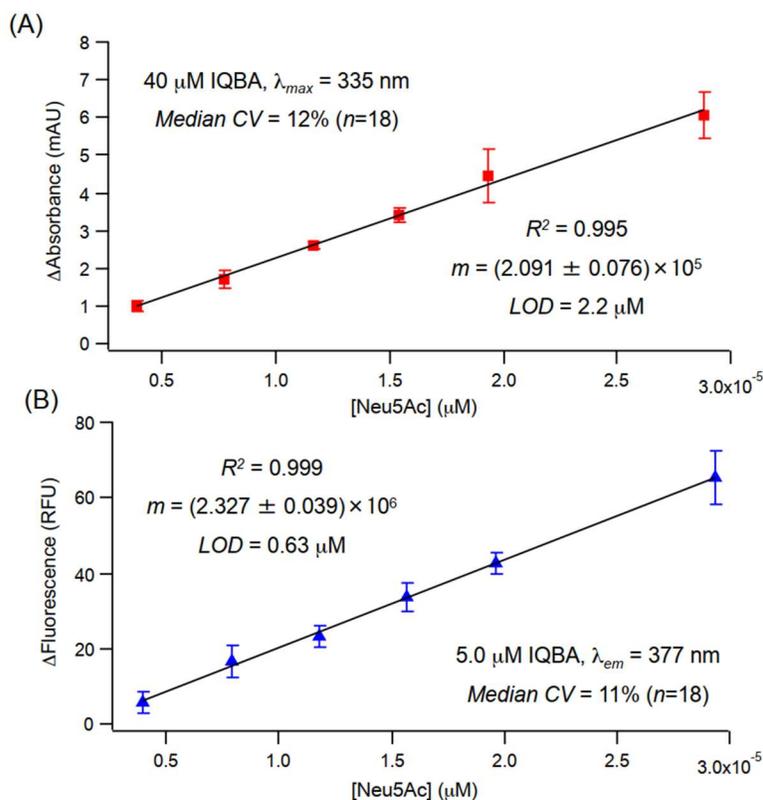


Figure 2.2: A comparison of linear calibration curves for Neu5Ac quantification when using IQBA with UV absorbance ($\lambda_{\text{max}} = 335 \text{ nm}$) and native fluorescence detection ($\lambda_{\text{em}} = 377 \text{ nm}$) in 40 mM phosphate, pH 3.0. Both detection methods offered excellent linearity and acceptable technical precision (CV \approx 12%, n = 18), but fluorescence provided greater sensitivity and lower detection limits as compared to UV absorbance.

water-soluble quinoline-based boronic acids have largely focused on glucose biosensing under physiological conditions.^{17,18} However, binding enhancement for Neu5Ac occurs at pH 3 since it is needed to ensure protonation of nitrogen heterocycle in IQBA while maintaining adequate ionization of acidic sugars/organic acids having α -hydroxycarboxylate (e.g., lactic acid, gluconic acid) or hydroxyphosphate (e.g., ribose-5-phosphate) bidentate functional groups. Additionally, Neu5Ac complexation to IQBA lowers the effective pK_a of the boronate ester resulting in a stable ion pair between the boronate anion and quinolinium heterocycle via a ternary IQBA-Neu5Ac-phosphate complex (insert of

Table 2.1) that is supported by experimental binding studies, computer modeling and ^{11}B -NMR spectra. This anomalous binding mechanism forms a stable zwitterionic boronate ester complex resistant to hydrolysis even under acidic conditions unlike the sp^2 -hybridized neutral boronate ester (**4**) depicted in **Scheme 2.1**. Either direct UV absorbance or native fluorescence measurements can be applied to quantify Neu5Ac with good technical precision ($CV = 12\%$) with low micromolar detection limits. Future work will explore the potential of IQBA to expand biorecognition to other molecular targets beyond conventional neutral saccharides with vicinal diols, including better understanding of synergistic interactions involving ternary complex formation in solution. Additionally, direct characterization of glycoproteins expressing different sialic acid residues will also be examined given their key roles in controlling protein stability and biological activity when using IQBA fluorescence-based biosensors.

2.4 References

1. X. Wu, Z. Li, X. X. Chen, J. S. Fossey, T. D. James, and Y. B. Jiang *Chem. Soc. Rev.*, 2013, **42**, 8032–8048.
2. S. D. Bull, M. G. Davidson, J. M. van den Elsen, J. S. Fossey, A. T. Jenkins, Y. B. Jiang, Y. Kubo, F. Marken, K. Sakurai, J. Zhao and T. D. James, *Acc. Chem. Res.* 2013, **46**, 312–326.
3. G. Fang, H. Wang, Z. Bian, J. Sun, A. Liu, H. Fang, B. Liu, Q. Yao and Z. Wu, *RCS Adv.* 2018, **8**, 29400–29427.
4. M. Dowlut and D. G. Hall, *J. Am. Chem. Soc.*, 2006, **128**, 4226–4227.
5. C. J. Musto and K. S. Suslick, *Curr. Opin. Chem. Biol.* 2010, **14**, 758–766.
6. Z. Bie, Y. Chen, J. Ye, S. Wang and Z. Liu, *Angew. Chem. Int. Ed. Engl.* 2015, **54**, 10211–10215.
7. Z. Liu and H. He, *Acc. Chem. Res.* 2017, **50**, 2185–2193.
8. J. Yan, G. Springsteen, S. Deeter and B. Wang, *Tetrahedron*, 2004, **60**, 11205–11209.
9. D. D. James, M. D. Phillips and S. Shinkai, *Boronic Acids in Saccharide Recognition*, RCS Publishing, Volume 12, 2006.
10. C. W. Gray and T. A. Houston, *J. Org. Chem.* 2002, **67**, 5426–5428.
11. D. M. Kim, J. M. Moon, W. C. Lee, J. H. Yoon, C. S. Choi and Y. B. Shim, *Biosens. Bioelectron.*, 2017, **91**, 276–283.

12. Y. J. Huang, W. J. Ouyang, X. Wu, Z. Li, J. S. Fossey, T. D. James and Y. B. Jiang, *J. Am. Chem. Soc.*, 2013, **135**, 1700–1703.
13. Y. Yoshimi, A. Narimatus, K. Nakayama, S. Sekine, K. Hattori and K. Sakai, *J. Artif. Organs* 2009, **12**, 264–270.
14. John P. Lorand; John O. Edwards, *J. Org. Chem.*, 1959, **24**, 769–774.
15. W. Yang, H. He and D. G. Drueckhammer, *Angew. Chemie - Int. Ed.*, 2001, **40**, 1714–1718.
16. X. Sun and T. D. James, *Chem. Rev.*, 2015, **115**, 8001–8037.
17. Y. Cheng, N. Ni, W. Yang, B. Wang, *Chemistry*, 2010, **16**, 13528-13538.
18. Q. J. Shen and W. J. Jin, *Luminescence*, 2011, **26**, 494-499.
19. Y. Cheng, N. Ni, W. Yang and B. Wang, *Chem. - A Eur. J.*, 2010, **16**, 13528–13538.
20. A. Varki, *Trends Mol. Med.*, 2008, **14**, 351–360.
21. K. P. Gopaul and M. A. Crook, *Clin. Biochem.*, 2006, **39**, 667–681.
22. L. Chrostek, B. Cylwik, A. Krawiec, W. Korcz and M. Szmitkowski, *Alcohol Alcohol.*, 2007, **42**, 588–592.
23. E. Rodrigues and M. McCauley, *Cancers*, 2018, **10**, 207.
24. H. Otsuka, E. Uchimura, H. Koshino, T. Okano and K. Kataoka, *J. Am. Chem. Soc.*, 2003, **125**, 3493–3502.
25. A. Matsumoto, A. J. Stephenson-Brown, T. Khan, T. Miyazawa, H. Cabral, K. Kataoka and Y. Miyahara, *Chem. Sci.*, 2017, **8**, 6165–6170.
26. K. Djanashvili, L. Frullano and J. A. Peters, *Chem. Eur. J.*, 2005, **11**, 4010–4018.
27. O. M. T. Pierce and H. Läubli, *Glycobiol.* 2016, **26**, 111-128.
28. A. R. Martin, J. J. Vasseur, and M. Smietana, *Chem. Soc. Rev.* 2013, **42**, 5684-5713.
29. C. Kaiser, G. Segui-Lines, J. C. D'Amaral, A. S. Ptolemy and P. Britz-McKibbin, *Chem. Commun.*, 2008, **21**, 338-340.
30. F. Fei and P. Britz-McKibbin, *Anal. Bioanal. Chem.*, 2010, **398**, 1349-1356.

2.5 Supplemental Information

Experimental

Chemicals and Reagents. All chemicals were purchased from Sigma Aldrich (St. Louis, MO, USA) with the exception of the following reagents, *N*-acetylneuraminic acid, Neu5Ac, *N*-glycolylneuraminic acid, Neu5Gc (Toronto Research Chemicals Toronto, ON, Canada) and catechol (Avocado Research Chemicals, Heysham, Lancashire, UK). All chemicals were used as received without further purification. Buffer solutions were prepared using a monosodium dihydrogen phosphate dihydrate salt ($\text{NaH}_2\text{PO}_4 \cdot 2\text{H}_2\text{O}$) that was prepared in deionized water from a Barnstead EasyPure II LF Ultrapure system. Stock solutions for 3-quinolineboronic acid (QBA) and 2-naphthylboronic acid (NBA) were prepared in 1:1 DMSO:H₂O, while 4-isoquinolineboronic acid (IQBA) were prepared in DMSO. Also, 1.0 M NaOH or HCl was used for pH adjustment for 40 mM phosphate buffers (pH 3 or 7), where ionic strength was corrected using 1 M NaCl to maintaining a constant total ionic strength (100 mM) unless otherwise stated.

UV absorbance spectrophotometry. All ligand binding studies to IQBA, QBA and NBA were performed using UV/Vis absorbance spectrophotometry in 40 mM phosphate buffer (adjusted to 100 mM ionic strength) at pH 3 or pH 7.0 unless otherwise stated. A Cary 50 UV-Vis spectrophotometer (Agilent Technologies Inc., Santa Clara, USA) at ambient temperature ($\sim 25\text{ }^\circ\text{C}$) was used for the collection of absorbance spectra, and all experiments were performed in triplicate ($n = 3$) using a fixed concentration of IQBA (40 μM) as a function of ligand concentration (0.01 to 20 mM). Scanning was performed in high resolution mode from 200 to 400 nm, and absorbance measurements were taken at the peak absorbance for IQBA upon ligand complexation ($\lambda_{max} = 335\text{ nm}$) after a blank correction of a buffer solution devoid of IQBA. A hyperchromic effect was evident with IQBA as a function of ligand concentration at $\lambda_{max} = 335\text{ nm}$, which was used for determination of apparent binding constants from binding isotherms using non-linear least squares

regression analysis, or calibration curves with linear least-squares regression based on changes in absorbance responses as a function of ligand concentration.

Fluorescence spectrophotometry. All native fluorescence measurements were performed at ambient temperature on a Cary Eclipse Fluorescence Spectrophotometer (Agilent Technologies Inc., Santa Clara, USA) with excitation (λ_{ex}) at 335 nm for IQBA under acidic buffer conditions (40 mM phosphate, pH 3.0). Emission spectra for IQBA as a function of ligand concentration were scanned from 340 to 450 nm, where a fluorescence enhancement was evident for IQBA upon ligand complex formation with a peak emission (λ_{em}) monitored at 377 nm. Data collection was performed in the same manner as with UV-vis absorbance with measurements performed in triplicate ($n = 3$), including calibration curves for Neu5Ac determination by native fluorescence when using 5.0 μ M IQBA in 40 mM phosphate at pH 3.0.

Binding Isotherms and Neu5Ac Calibration Curves. Binding isotherms and calibration curves performed using Igor Pro 5.0 software (Wavemetrics Inc., Lake Oswego, OR, USA) when using non-linear and linear least squares regression analyses, respectively. Binding isotherms for various ligands (**Scheme 2.1**) with 40 μ M IQBA (also QBA or NBA as controls) were measured by changes in UV absorbance response at 335 nm under constant ionic strength (100 mM) in aqueous 40 mM phosphate buffers at pH 3.0 or pH 7.0. A 1:1 dynamic complexation model was used to plot changes in molar absorptivity ($\Delta\epsilon$ at 335 nm) for IQBA as a function of ligand concentration based on the following equation (1):

$$\Delta\epsilon = K*\Delta\epsilon_{sat}*c/(1 + K*c) \quad (1)$$

where, K (M^{-1}) is the apparent formation constant of the boronate ester complex involving IQBA and a specific ligand (refer to **Scheme 2.1**), c is the molar concentration of IQBA (M), and $\Delta\epsilon_{sat}$ ($M^{-1}cm^{-1}$) is the asymptotic limit upon binding

saturation where the fraction of complex ≈ 1 . This equation was solved simultaneously using non-linear least squares regression with initial parameters estimated for K and $\Delta\epsilon_{sat}$. Calibration curves for Neu5Ac were performed independently within the linear range (from 3 to 30 μM) of binding isotherms in triplicate ($n = 3$) when using UV absorbance (40 μM IQBA with $\lambda_{max} = 335$ nm) and native fluorescence (5.0 μM IQBA with $\lambda_{ex} = 355$ nm and $\lambda_{em} = 377$ nm) in 40 mM phosphate pH 3.0 when using linear least-squares regression analysis. Method sensitivity for Neu5Ac determination using UV absorbance or native fluorescence was determined from the slope (m) of the calibration curve, whereas limits of detection (LOD, $S/N = 3$, μM) were calculated by the slope and y-intercept from the line of best fit using the following equation (2):

$$\text{LOD} = (3 * \text{y-intercept}) / m \quad (2)$$

Computer Molecular Modeling of Complex Stability. Chem 3D Professional software, version 12, (CambridgeSoft Inc., Cambridge, MA, USA) was used to estimate the relative thermodynamic stability of different binding modes involving Neu5Ac (*i.e.*, via α -hydroxycarboxylate, C1, C2 vs. terminal vicinal diol, C8, C9) to IQBA after energy minimization using molecular mechanics 2 (MM2) algorithm with molecular dynamics with 10,000 iterations. IQBA-Neu5Ac complexes with and without phosphate interactions from buffer at different pH conditions (pH 3.0 and 7.0) that impact the ionization state of quinolinium ion ($pK_a = 5$) were compared in terms of their total energy based on non-covalent interactions (*i.e.*, ion pair, ion-dipole etc.) and steric factors (*e.g.*, ring strain of cyclic boronate ester).

Nuclear Magnetic Resonance (NMR). ^{11}B -NMR and ^{13}C -NMR spectra were acquired using a Bruker (Bruker, Billerica, USA) AVANCE 600 MHz NMR spectrometer equipped with a B-ACS 60 autosampler at ambient temperature. IconNMR software operated autotune and autoshim features for optimal peak

resolution. Samples were run in 5 mm diameter NMR tubes of borosilicate glass. D₂O and DMSO-d₆ were used as solvents to ensure adequate solubilization of IQBA (18 mM) that was prepared in 40 mM phosphate as buffer (pH 3.0 or pH 7.0) with or without addition of excess NeuAc in a 3:1 (Neu5Ac:IQBA) concentration ratio. NMR spectral annotation of carbon resonances for Neu5Ac by ¹³C-NMR [*Chem. Eur. J.*, 2005, 11, 4010] and boron resonances for IQBA by ¹¹B-NMR [*Chem. Eur. J.* 2010, 16, 13528] were consistent with literature reports.

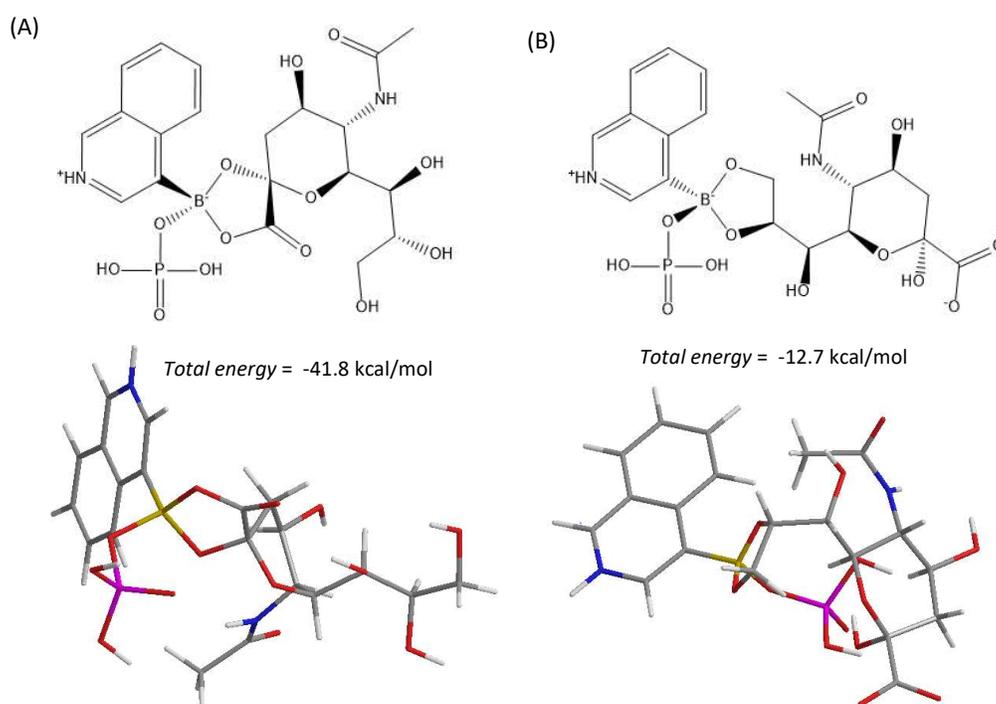


Figure 2S.1: 2D structures and 3D conformations comparing the thermodynamic stability for two different ternary boronate ester complexes involving IQBA-Neu5Ac-phosphate after energy minimization (MM2) with computer molecular modeling. Each structure compares the predominate ligand binding mode for Neu5Ac involving (A) the α -hydroxycarboxylate anion or (B) the terminal vicinal diol (C8, C9) of glycerol moiety at pH 3. In both cases, the quinolinium ion is fully charged and forms a stable ion pair/zwitterion with the tetrahedral (sp^3 -hybridized) boronate anion, where conformer (A) is significantly more stable than conformer (B) by -29.1 kcal/mol . Additionally, each complex is not stable (total energy $> +20 \text{ kcal/mol}$) if forming a ring-strained sp^2 -hybridized (trigonal planar) boronic ester under acidic conditions, as well as being far less stable when complexation occurs under neutral or alkaline conditions where the quinoline nitrogen is deprotonated (i.e., neutral).

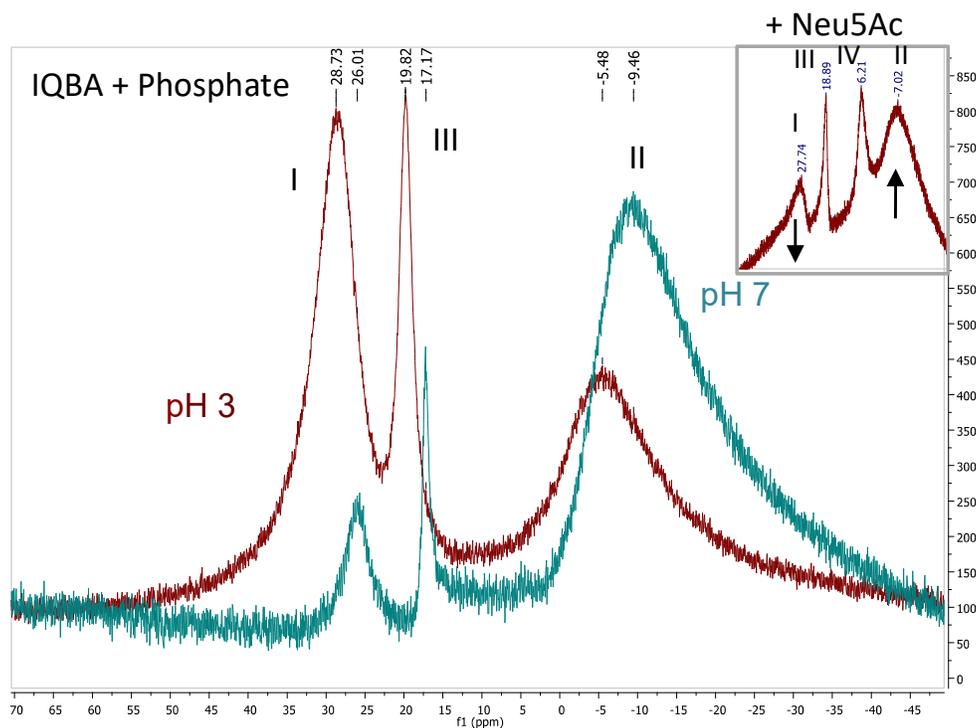


Figure 2S.2: ^{11}B -NMR spectra overlay comparing the relative distribution of major boron species at equilibrium for IQBA in 40 mM phosphate prepared in D_2O as a function of buffer pH and Neu5Ac. There are three distinct boron chemical environments for IQBA in phosphate buffer under acidic conditions (pH 3.0, red trace), including a predominantly sp^2 hybridized boron (i.e., trigonal planar; electron deficient/neutral boronic ester) corresponding to a low-field resonance peak (I, $\delta = 28.7$ ppm), and a less abundant sp^3 -hybridized boron corresponding to a low-field resonance peak (II, $\delta = -5.5$ ppm) indicative of the tetrahedral boronate anion. Additionally, there exists a third boron resonance near $\delta = 19.8$ ppm (III) that likely corresponds to a phosphoboronate ester complex formed between IQBA and excess phosphate in solution. As expected, when acquiring the ^{11}B -NMR spectra for IQBA under neutral pH conditions (pH 7.0), there is a noticeable increase in the relative abundance of sp^3 -hybridized boron (II, $\delta = 9.6$ ppm) as compared to sp^2 -hybridized boron (I, $\delta = 26.0$ ppm) in solution with a low field shift in both boron resonances. However, the inset depicts the emergence of a fourth peak upon addition of excess Neu5Ac to IQBA in phosphate at pH 3.0 (IV, $\delta = 6.2$ ppm) likely corresponding to the ternary IQBA-Neu5Ac-phosphate complex. Importantly, Neu5Ac complexation results in a pronounced increase in the relative abundance of the tetrahedral boronate anion (II) and a corresponding decrease in trigonal planar boronate ester species (I) even under acidic conditions that is comparable to IQBA under neutral conditions. This process reflects the increase in the apparent acidity (shift to lower pKa) of the cyclic boronate ester complex upon Neu5Ac binding as compared to free IQBA.

Chapter 3: Elucidating the Mechanisms of Smoke Exposures in Municipal Firefighters: A Multi-Centre Training Exercise Study

Nadine Wellington, Jonathon Bloomfield, Sujan Fernando, Ritchie Ly, Lorraine Shaw, Don Shaw, and Philip Britz-McKibbin

All research, data processing, statistical analysis, tandem mass spectrometry and interpretation described in this manuscript was performed by the author, with important experimental contributions from collaborators and colleagues. Johnathon Bloomfield performed all GC-MS analysis and assisted with skin wipe and urine sample collection. Sujan Fernando assisted with GC-MS analysis and skin wipe and urine sample collection. Ritchie Ly performed CE-MS urine analysis. Lorraine Shaw and Don Shaw were responsible for logistics, study design, and assisted with sample collection. Christopher D. Simpson kindly provided PAH standards for this study. Philip Britz-McKibbin was responsible for study design and is the principal investigator on the study.

Chapter 3: Elucidating the Mechanisms of Smoke Exposures in Municipal Firefighters: A Multi-Centre Training Exercise Study

3.1 Abstract

Occupational firefighting is linked with a higher risk for cardiovascular disease and various cancers due to chronic exposure to a complex mixture of toxic chemicals in smoke formed by the combustion of natural and synthetic materials, such as wood. Despite protective equipment to mitigate burns and smoke inhalation, firefighters are still prone to whole body skin absorption that complicates risk assessment and mitigation efforts. Herein, air samples, skin wipes, and urine samples collected at four time intervals (T-24 h, 0-6 h, 6-12 h, and 12-24 h) were used to determine the impact of acute wood smoke exposure on 18 career firefighters who conducted 30 min. search-and-rescue training exercises within burn houses located at three different sites in Ontario. Gas chromatography-mass spectrometry (GC-MS) was used to analyze 25 chemicals as tracers of wood smoke absorption including various methoxyphenol (MP) derivatives, polycyclic aromatic hydrocarbons (PAH), and resin acids (RA), as well as their corresponding metabolites excreted in urine following enzyme deconjugation. Substantial contamination was found to penetrate the protective clothing (“turnout gear”) and, to our knowledge, for the first time also the mask of each firefighter, introducing the likelihood of inadvertent respiratory exposures. After burn trials, each inner self-contained breathing apparatus (SCBA) lens showed chemical contamination with 89% of samples having a least 1 contaminant detected with > 100% penetration, where a greater amount was recovered from inside the protective mask than outside. An optimal sampling window of up to 6 h post-exposure is identified by time-dependent urine sampling after smoke exposure where three high abundance guaiacol analogs, as well as hydroxynaphthalene and hydroxyfluorene isomers were significantly elevated in creatinine-normalized urine. Baseline cheek and arm soiling measured prior to the exercise was equivalent to amounts found

after smoke exposure, likely a result of poor hygienic practices and workplace contamination. Findings in this study provide an empirical basis for revising the standards of hygiene and protective garments for firefighters to reduce incidences of chronic illness.

3.2 Introduction

Structural firefighters endure chronic exposure to smoke formed by combustion of numerous home construction materials including synthetic polymers, tar, and wood. Aerosolized wood smoke including toxic particulate (*e.g.* PM_{2.5}) and vapours (*e.g.* benzene, formaldehyde), contain a plethora of toxic inorganic and organic chemicals, such as polycyclic aromatic hydrocarbons (PAHs) some of which are classified as known carcinogens.¹ Chronic smoke exposure in firefighters is associated with long-term respiratory illnesses, cardiovascular disease risk, various cancers, and reduced average life spans as compared to the general population due to cytotoxic and mutagenic chemicals and/or their downstream metabolites.^{2,3} As a result, firefighting is widely considered an occupational health hazard, and was deemed carcinogenic by the International Agency on Cancer Research in 2010.⁴ Accordingly, the province of Ontario recently expanded presumptive legislation benefits to firefighters in 2018 to recognize 17 different cancers and heart failure as occupation-related chronic diseases eligible for workers' compensation.⁵

Wood smoke is most abundant in polycyclic aromatic hydrocarbons (PAH), methoxyphenols (MP), and resin acids (RA) produced by combustion of celluloses and lignin, structural biopolymers of sugars and phenol that constitute > 80% of wood mass.⁶ Several PAH are characterized as known (Group 1) and probable/possible (Group 2 A/B) human carcinogens following bioactivation by specific cytochrome P450 isoforms in the liver that result in formation of cytotoxic reactive oxygen species (*e.g.* epoxides, diols) that damage DNA.⁷ Chronic exposures are also strongly linked to cardiovascular disease related to arterial stiffening and heart attack.⁸ Although MPs and RAs exhibit less toxic effects than PAHs despite their higher concentrations in wood smoke, several studies have

shown that they can activate immunosuppressive and inflammatory pathways in respiratory epithelial and fibroblast cells.⁹⁻¹¹ Also, synergistic effects between compounds in smoke mixtures may enhance the toxicity of certain carcinogenic PAHs (e.g. benzo[a]pyrene) whose exact mechanisms remain poorly understood.¹²

Field monitoring of firefighter exposures is challenging due to unpredictable and highly variable sampling times, and inconsistent sample types (*i.e.*, structural/urban, wildland, industrial, automotive, etc.). Urine PAH metabolites, such as 1-hydroxypyrene (1-OH Pyr), are commonly used as indices of recent smoke exposure human biomonitoring applications,^{13,14} however accurate risk assessment is difficult due to between-subject variability in xenobiotic metabolism that is dependent on several confounding factors including genes, diet, lifestyle and lifelong exposures.^{5,15} Firefighter personal protective equipment (“turnout gear”) protects against heat injury and physical impacts while reducing chemical exposures from smoke inhalation via a self-contained breathing apparatus (SCBA), if properly used and maintained. Nevertheless, several recent studies have demonstrated extensive “whole body” smoke exposure to firefighters by identifying chemicals adsorbed on the skin,¹⁶⁻¹⁸ exhaled on the breath,¹⁹ and biomarker of recent smoke exposure excreted in urine.^{14,16,20} Firefighters use SCBAs to avoid inhaling acutely hazardous smoke during fire suppression, but face substantial risk from PPE permeability and inadequate non-standardized hygiene protocols that leave chronic percutaneous absorption an important route of exposure to firefighters.^{16,21} Specifically, Keir et al. report that skin absorption is likely of greater risk than inhalation, a problematic finding in light of the extensive skin contamination of firefighters reported by Fernando and coworkers, the large surface area of the dermis, and relatively fast rates of transdermal absorption by low polarity lipophilic contaminants (*e.g.* PAH).^{16,22,23} Accordingly, it is also necessary to understand how the kinetic excretion of smoke markers in urine can be used to determine the extent and proximity of smoke adsorption to the time of sampling.

Herein, wood smoke exposure to 18 firefighters wearing full turnout gear during live-burn training exercises conducted at 3 municipal fire stations in the province of Ontario are described. This work is an independent follow-up study to Fernando et. al.¹⁶ that first reported extensive skin contamination with urine excretion of several wood-smoke biomarkers in firefighters wearing full protective gear.¹⁶ In our study, urine samples were pooled over four time intervals (T-24 h, 0-6 h, 6-12 h, and 12-24 h) in order to better ascertain the delay between peak excretion of wood-smoke biomarkers and time of exposure. Two devices were used for personal exposure monitoring, an active sampling apparatus with a filter, XAD-2 tube, and personal air pump connected in series to collect volatiles and solid particulate, and a sorbent Twister for passive sampling of gases. SCBA facepieces were equipped with Twisters to assess the mask as a point of weakness in regulatory PPE that may lead to inhalation or dermal absorption of chemicals. Results confirm deficiencies in the standard PPE used by volunteers in this study and highlights the urgent need for new, scientifically informed performance and hygiene standards in all aspects of firefighting to improve the health and safety of individuals in this profession.

3.3 Results and Discussion

Personal Exposure to Wood Smoke. A total of 18 firefighters from 3 municipal fire houses in Burlington (Site B), Hamilton (Site H), and Ottawa (Site O) in the province of Ontario underwent acute exposure to wood smoke for 30 min within an on-site training facility (“burn house”) during which time they were tasked with fire control or conducting a search-and-rescue training exercise. The non-polar PDMS coating of the Twister sampler retained approximately 5-80 times greater amounts of MP smoke markers compared to PAH and RA due to the high abundance of the lignin precursor found in wood and plant tissue. As large semivolatile compounds partition preferentially to particulate,^{6,16} Twister sampling of PAH and RA was limited; however the pump assembly was able to more effectively retain compounds with MW > 202 g/mol. Good correlation ($r = 0.71$, $p = 0.042$) was found between

the amount of MP, PAH, and RA recovered from the caged outer Twister and active pump apparatus of 6 firefighters who completed the exercise with both devices operational, indicating comparable sampling between apparatus, particularly for volatile low-molecular-weight (LMW) compounds (**Figure 3.1A**; **Table 3S.1**; **Table 3S.2**). Naphthalene was well-adsorbed by the Twister, however large differences in boiling points between naphthalene and heavy PAH resulted in poor correlations ($r < 0.35$) challenging its use as an indicator of exposure to heavier carcinogens such as benzo[b]fluoranthene (BbF) and benzo[a]pyrene (BaP).

Concentrations of PAH in the burn houses determined by the personal air pumps did not exceed established regulatory occupational exposure limits;^{24,25} however, carcinogens BbF, BaP, and benz[a]anthracene (BaA) have no safe exposure limits due to their significant toxicity and were present in all 3 burn houses (**Table 3S.2**). Ambient concentrations of BaP determined by personal air pumps exceeded the highest concentrations encountered by coke oven workers in recent studies (median 880 ng/m³, range: 70 - 13 450 ng/m³), and 12/18 firefighters exceeded ambient levels at which chronic exposure has been linked to neurodegenerative processes and an approximately 60% increased risk of heart disease.²⁶⁻²⁸

The internal volume of the burn houses varied between municipalities, where Site B used a large 4-storey brick building with 2-storey additions, while Site H and Site O used smaller 2- and 1-storey concrete buildings, respectively (**Supplemental, Appendix 3.1**). The exercise was initiated when smoke saturated the structure, however fire intensity and apparent smoke concentration was not controlled for. Different wood smoke exposures were observed between the 3 sites (**Figure 3.1B**; **Table 3S.1**; **Table 3S.2**), likely due to variable wood types. Gaseous MPs syringol and isoeugenol were high in all burn houses, while propylsyringol (PropS) was relatively higher at Sites B and O. Site H had greater concentrations of guaiacols and aerosolized RAs abietic acid and dehydroabietic acid, characteristic of

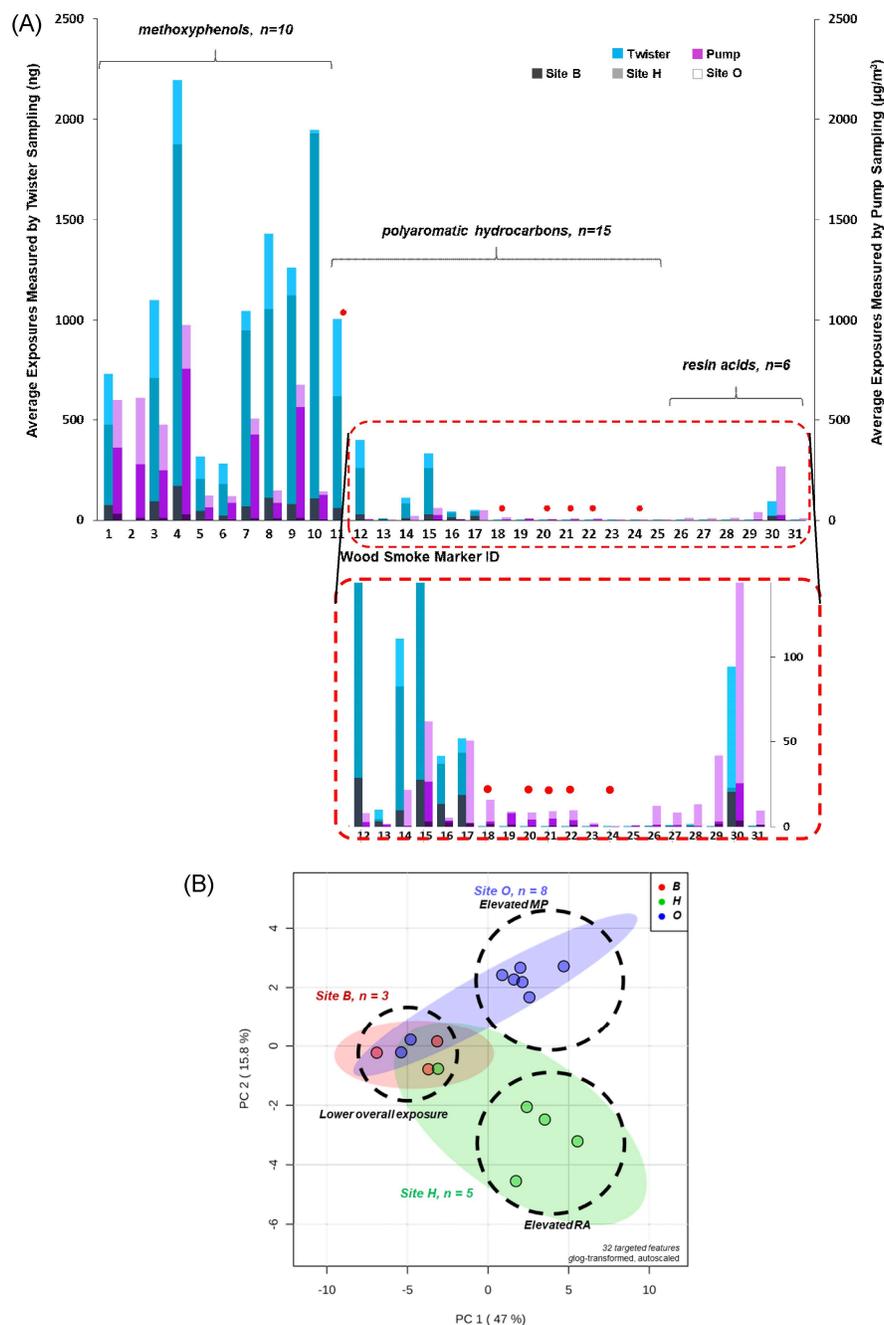


Figure 3.1: (A) Triple-stacked bar plot of total ambient volatile and semivolatile MP, RA, and PAH (with carcinogen class indicated) wood smoke markers retained by the caged outer Twister (blue) and the personal pump assembly (purple) with a Teflon filter and sorbent XAD-2 resin connected in series (2 L/min) for Site B, H and O. Red circles indicate carcinogenic PAH. 1: guaiacol; 2: methylguaiacol; 3: ethylguaiacol; 4: syringol; 5: propylguaiacol; 6: eugenol; 7: methylsyringol; 8: isoeugenol; 9: ethylsyringol; 10: propylsyringol; 11: naphthalene (Group 2B); 12:

acenaphthylene; 13: acenaphthene; 14: fluorene; 15: phenanthrene/anthracene; 16: fluoranthene; 17: pyrene; 18: benz[a]anthracene (Group 2B); 19: chrysene; 20: benzo[b]fluoranthene (Group 2B); 21: benzo[k]fluoranthene (Group 2B); 22: benzo[a]pyrene (Group 1); 23: indeno[1,2,3-cd]pyrene (Group 2B); 24: dibenz[a,h]anthracene (Group 2A); 25: benzo[ghi]perylene; 26: pimaric acid; 27: sandaracopimaric acid; 28: isopimaric acid; 29: abietic acid; 30: dehydroabietic acid; and 31: 7-oxodehydroabietic acid. (B) 2D PCA scores plot showing heterogeneous intercity and interindividual exposures based on active sampling pumps ($n = 16$). Site B showed relatively low exposures to MP, PAH and RA presumably related to the large volume multi-storey burn house used. Site H was exposed to smoke high in syringols, while Site O recorded higher amounts of guaiacols and RA.

coniferous wood (e.g. pine).²⁹ Site B (red) demonstrated significantly lower exposures to 26 of 33 (79%, $p < 0.05$) smoke compounds measured compared to Sites H and O whose exercises were conducted in smaller structures where smoke concentrations were expected to be higher. The classification of 1 individual from Site H and 2 from Site O with Site B ($n = 3$) reflects similarly low overall exposures when compared to their original cohorts which may reflect larger average displacements from the fire.³⁰

Firefighting self-contained breathing apparatus (SCBA) are positive-pressure devices meant to prevent intrusion by ambient chemicals in the event of a leak.³¹ Sample wipes of the outer and inner side of each facepiece lens ($n = 15$) revealed significantly lower amounts of all smoke markers inside the mask; however, EthS ($p = 6.0 \text{ E-}4$) and syringol ($p = 1.2 \text{ E-}4$) showed significant increases on the inner lens when compared to baseline. Methylguaiacol (MeG), dehydroabietic acid (DA), 7-oxodehydroabietic acid (7ODA) and fluorene did not significantly increase, but were recovered in statistically equivalent amounts from both sides of the SCBA lens. The SCBA consistently reduced all exposures for only 2/18 individuals, and fluorene was reduced by a median 10.1% median between the two sides (**Table 3.1**). Fluorene was recovered in greater amounts on the inner lens of 8/15 firefighters, where one individual had a 330-fold higher inner lens concentration. Relatively high amounts on were also found on the cheek ($18.4 \pm 25.2 \text{ ng/cm}^2$), primarily at Sites H and O, though there were no significant

Table 3.1: (Mean \pm SD) concentrations (ng/cm²) of MP, PAH, and RA recovered from the outer and inner SCBA lens of firefighters (n = 15)

	Site B (n = 5)	Site H (n = 5)	Site O (n = 8)	Site B (n = 5)	Site H (n = 5)	Site O (n = 8)	Site B (n = 5)	Site H (n = 5)	Site O (n = 8)	Site B (n = 5)	Site H (n = 5)	Site O (n = 8)
	SCBAO - Pre			SCBAO - Post			SCBAI - Pre			SCBAI - Post		
MP												
Guaiacol	0.51 \pm 0.51	1.14 \pm 1.72	4.94 \pm 4.22	18.3 \pm 19.6	2.0 \pm 1.1	26.3 \pm 7.7	0.59 \pm 0.74	0.75 \pm 1.04	8.22 \pm 7.37	6.59 \pm 8.87	0.77 \pm 1.24	3.81 \pm 4.82
MeG	55.6 \pm 68.5	24.7 \pm 11.2	3.58 \pm 3.66	20.7 \pm 21.1	10.5 \pm 15.2	251 \pm 97	21.1 \pm 11.7	32.3 \pm 7.6	3.92 \pm 2.04	9.6 \pm 11.2	12.9 \pm 16.5	3.67 \pm 3.24
EthG	0.62 \pm 0.40	2.23 \pm 3.54	4.13 \pm 5.15	175 \pm 194	7.44 \pm 8.02	314 \pm 124	1.55 \pm 1.88	0.91 \pm 0.45	3.88 \pm 2.56	7.9 \pm 13.3	3.36 \pm 4.84	4.42 \pm 4.58
PropG	1.36 \pm 2.43	0.16 \pm 0.24	1.28 \pm 2.09	34.3 \pm 62.8	1.17 \pm 0.81	118 \pm 47	1.05 \pm 2.16	0.11 \pm 0.09	1.14 \pm 0.77	0.63 \pm 0.69	0.25 \pm 0.35	2.08 \pm 1.95
Syringol	0.48 \pm 0.44	24.1 \pm 52.7	73 \pm 123	203 \pm 216	3.4 \pm 1.9	1056 \pm 661	0.57 \pm 0.6	4.25 \pm 8.51	71.4 \pm 78.1	4.99 \pm 4.97	0.85 \pm 0.84	130 \pm 130
MeS	0.58 \pm 0.57	9.8 \pm 21.6	41.9 \pm 83.9	279 \pm 303	1.68 \pm 1.43	1188 \pm 576	0.19 \pm 0.3	0.19 \pm 0.15	50.3 \pm 65	2.07 \pm 1.56	0.9 \pm 1.33	87.5 \pm 88.9
EthS	0.86 \pm 0.51	7.6 \pm 16.3	35.5 \pm 76.3	363 \pm 419	4.17 \pm 3.19	1117 \pm 528	3.2 \pm 6.29	0.51 \pm 0.52	44.4 \pm 53.3	2.91 \pm 2.39	0.63 \pm 0.48	83.1 \pm 85.5
Eugenol	0.42 \pm 0.26	0.95 \pm 2.02	1.04 \pm 1.81	2.55 \pm 2.14	0.97 \pm 1.05	17 \pm 9	0.21 \pm 0.16	0.06 \pm 0.05	1.31 \pm 1.52	1.74 \pm 2.17	0.62 \pm 1.01	1.8 \pm 2.3
PAH												
Naphthalene	10.7 \pm 8.4	11.9 \pm 14.3	5.6 \pm 14.8	39.7 \pm 40.5	18.3 \pm 16	6.48 \pm 3.99	6.04 \pm 2.94	21.1 \pm 18.8	3.89 \pm 7.02	16.6 \pm 5.8	22.9 \pm 19.9	3.7 \pm 4.95
Fluorene	0.23 \pm 0.19	0.88 \pm 1.25	8.12 \pm 7.39	1.17 \pm 1.51	0.91 \pm 0.45	1.28 \pm 1.66	0.21 \pm 0.17	0.89 \pm 1.09	4.23 \pm 4.67	3.57 \pm 6.09	2.53 \pm 4.12	9.3 \pm 17.8
Phenanthrene	0.83 \pm 0.80	1.45 \pm 0.91	2.75 \pm 5.45	19.0 \pm 24.4	1.55 \pm 0.73	7.95 \pm 4.75	0.86 \pm 0.85	1.46 \pm 0.86	2.17 \pm 0.92	3.83 \pm 5.09	1.7 \pm 0.51	3.39 \pm 3.26

Fluoranthene	0.66 ± 0.64	0.38 ± 0.20	1.39 ± 3.34	53.6 ± 64.2	3.28 ± 1.04	17.9 ± 14.2	0.49 ± 0.42	0.62 ± 0.3	1.32 ± 1.23	2.92 ± 3.2	2.3 ± 2.45	2.25 ± 2.41
Pyrene	0.77 ± 0.81	0.38 ± 0.21	1.05 ± 2.52	65.2 ± 78.2	4.15 ± 1.67	18.8 ± 13.7	1.09 ± 1.07	0.56 ± 0.44	1.15 ± 1.02	7.24 ± 8.16	1.76 ± 1.34	1.78 ± 1.91
RA												
7ODA	0.41 ± 0.26	1.16 ± 1.37	0.33 ± 0.31	14.2 ± 16.5	1.91 ± 1.87	0.74 ± 0.39	0.31 ± 0.28	1.22 ± 0.85	0.79 ± 0.86	2.4 ± 3.87	1.5 ± 1.57	0.36 ± 0.4
AA	0.75 ± 1.31	0.17 ± 0.15	0.03 ± 0.04	19.0 ± 28.0	3.51 ± 2.8	1.21 ± 0.57	0.25 ± 0.29	0.24 ± 0.22	0.11 ± 0.05	0.84 ± 0.67	3.53 ± 6.43	0.15 ± 0.15
DA	13.7 ± 12.6	18.3 ± 13.0	3.19 ± 2.56	317 ± 423	94.4 ± 65.7	28.5 ± 14.1	9.63 ± 6.01	21.5 ± 15.4	7.46 ± 4.34	52.1 ± 30.4	64 ± 78.4	7.09 ± 7.68
IPA	0.44 ± 0.31	0.81 ± 0.65	0.36 ± 0.50	14.3 ± 16.1	2.4 ± 2.04	1.02 ± 0.64	10.4 ± 13.9	1.23 ± 0.31	0.76 ± 0.49	28.3 ± 18.6	10.3 ± 17.6	0.86 ± 0.56
PA	0.78 ± 0.75	0.33 ± 0.31	0.19 ± 0.35	5.53 ± 6.43	2.72 ± 1.95	1.38 ± 0.85	0.32 ± 0.26	0.58 ± 1.06	0.62 ± 0.83	0.32 ± 0.2	1.44 ± 1.35	0.53 ± 0.56
SPA	0.22 ± 0.23	0.10 ± 0.06	0.05 ± 0.08	5.40 ± 6.51	0.98 ± 1.18	0.81 ± 0.93	0.16 ± 0.15	0.26 ± 0.21	0.34 ± 0.46	0.32 ± 0.2	0.42 ± 0.24	0.31 ± 0.61

correlations to fluorene recovered from the pump or the dorsal forearm (**Table 3S.3**).

All RA were recovered at higher concentrations within the SCBA mask (mean 28.0% increase) when wiping the inner lens (**Table 3.1**). SCBA-contained Twisters ($n = 7$) showed significantly lower smoke concentrations when compared to smoke-exposed caged samplers (suit) for all except 1 firefighter ($p < 5.7 \text{ E-}3$; **Figure 3.2**; **Table 3S.1**); however, abietic acid (AA), dehydroabietic acid (DA), pimaric acid (PA), and sandaracopimaric acid (SPA) were found at concentrations 6-to-84 times higher on the inner Twister when compared to the exposed outer sampler (**Table 3S.1**). Highly abundant smoke compounds propylguaiacol (PropG, 52.5 ± 40.6) ng and ethylsyringol (EthS, 47.0 ± 44.4) ng were recovered from the Twisters in the greatest amounts overall, and found at higher concentrations within the SCBA for 3-4 individuals.

Industry standards require an assigned protection factor (PF) of $> 10\,000$ ($[X_i]_{\text{out}}/[X_i]_{\text{in}}$) for fit-tested SCBA, however when comparing amounts captured by the SCBA Twister the median PF calculated was 0.19 (maximum 5404).³² Notably, rapid, intermittent episodes of negative-pressure within the SCBA due to accelerated respiration during physical exertion has been reported to induce leaks of possibly contaminated air,³³ while moderate to intense sweating can also impair the seal of the SCBA, promoting mask slippage and leaks.³⁴ Furthermore, in a work history survey of 16 participants, 5 respondents reported regularly finding soot on their skin and in expelled mucous after fire suppression activities. Overall, the SCBA appears to provide substandard protection to these firefighters which suggests the urgent need for more frequent fit testing, equipment inspection, and standards of care.

Post-exposure cheek wipes were evaluated as a method for assessing undetected respiratory exposures after fire suppression (**Table 3S.3**), and were collected with paired arm wipe samples after exposure. Gaseous compounds MeG

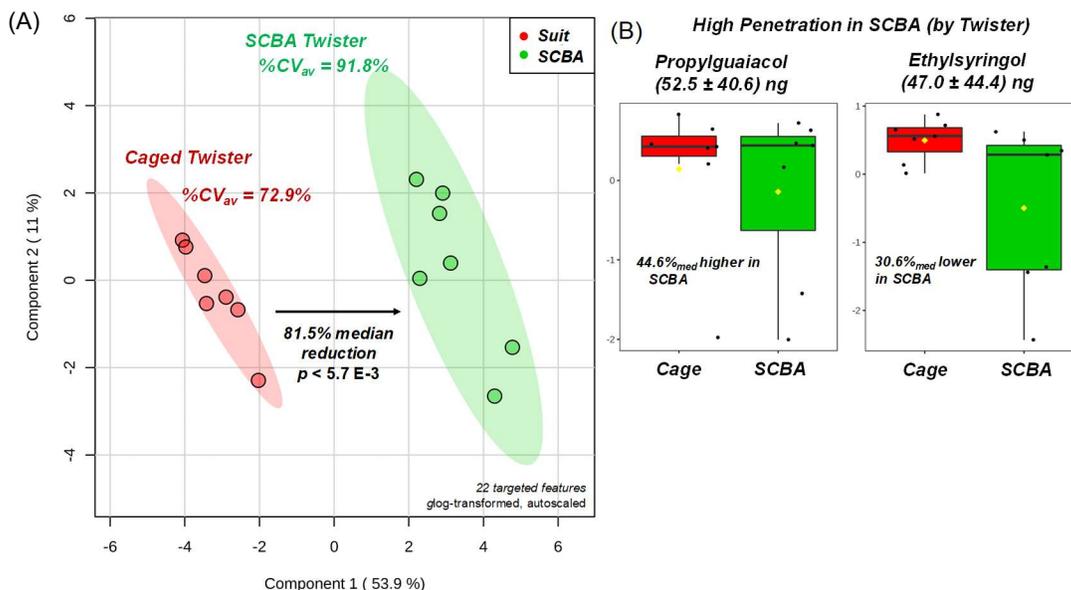


Figure 3.2: (A) 2D PCA scores plot showing significantly lower overall amounts of 22 targeted smoke markers recovered from the passive samplers inside the SCBA when compared the caged external Twister (suit). (B) Propylguaiacol (PropG) and ethylsyringol (EthG) showed the highest penetration of the SCBA mask, consistent with high concentrations measured in smoke by passive and active sampling devices (Figure 3.1).

(mean $19.9 \pm 13.6 \text{ ng/cm}^2$) and naphthalene ($9.8 \pm 14.4 \text{ ng/cm}^2$) were retained in the greatest amounts on the cheek and showed moderate penetration of the SCBA (Table 3S.1). Recovery of MeG from the cheek was not well-correlated to deposition on the inner SCBA lens but was strongly correlated (Spearman) to amounts found in soot on the arm ($r = 0.579$, $p = 0.024$) but naphthalene did not exhibit these associations. Naphthalene amounts deposited on SCBA-covered cheeks of exposed firefighters was approximately 10-fold greater than mean concentrations found on the gloved hands of structural firefighters tasked with search-and-rescue in a similar exercise,²¹ suggesting unusually high exposures via the SCBA. Both compounds were also significantly elevated on the dorsal forearm ($p < 1.0 \text{ E-5}$) protected by gloves, wrist covers and the bunker coat, confirming

another weak point in firefighter PPE that likely contributes to the “whole body exposure” described previously.^{16,18,22} Furthermore, we observed that one fire station used fire hoods made from highly porous heat-resistant knit material, that was also used for protective wrist cuffs of the turnout coat. Indeed, for 75% of participating firefighters surveyed the neck and wrists are the areas of greatest concern and most-cited as under-protected from contamination during regular knockdown duties. Predictably, Site O firefighters who conducted their exercise in the smallest building (Supplemental, **Appendix 3.1**) and were exposed to the highest concentrations of vapour and particulate (**Table 3.1**; **Table 3S.1**; **Table 3S.2**), had the highest accumulation of soot on the cheek. Soot composition was strongly correlated (Spearman) to the smoke profile in the burn house ($r > 0.60$, $p < 0.007$), but similar comparisons were weak for Site B and Site H ($r < 0.3$).

Most firefighters are limited to 1 set of PPE subject to regular inspection and cleaning, yet no standardized protocols exist for the removal and cleaning of turnout gear after smoke exposure and the efficacy of some modern cleaning methods have only recently been found to be substandard.^{21,35} There are numerous recommendations for hygiene optimization between the fireground and fire house, including on-site gear removal and decontamination, promptly bathing and laundering of gear after an exposure, and regular decontamination of vehicles and workspaces.^{17,36,37} Pre-burn baseline wipes of the inner and outer SCBA lens ($n = 18$) revealed soiling on SCBA masks with statistically equivalent amounts of soot found outside and inside the mask (**Table 3.1**). Site O masks were significantly contaminated with all MP (except MeG) and fluorene ($p < 0.010$) when compared to Sites B and H. The latter two groups had higher concentrations of MeG ($q < 0.060$), dehydroabietic acid (DA), naphthalene, and isopimaric acid (IPA, Site H only) that were not significant after FDR correction. In several cases, the soot deposited on the outer SCBA lens after 30 minutes of exposure was a moderate fraction of the amounts recovered at baseline, where MeG (median FC = +34.9%)

and fluorene (median FC = +18.6%) were found at higher amounts prior to wood smoke exposure, suggesting ineffective cleaning procedures may allow long-term accumulation of soot. Concerningly, baseline chemical contamination found on the SCBA was statistically equivalent to amounts recovered at baseline on the skin for 14 of 19 compounds targeted in the sample wipes. Of the 5 significantly differentiated compounds, 4 were found in greater amounts on the cheek when compared to the SCBA and arm (ANOVA, $p < 0.022$, SPA, PA, pyrene, and fluorene). These results further suggest accumulation of chemical exposures within work and living spaces due to substandard hygiene protocols leading to chronic exposures to all personnel and increases the likelihood of exposure by unintentional ingestion.

The excretion of 25 targeted unaltered or deconjugated metabolites (creatinine normalized) was examined across 4 distinct time periods as shown in **Figure 3.3A**. Baseline urine was high in guaiacols and dehydroabiatic acid (**Table 3S.4**) consistent with the most abundant chemicals in wood burn house smoke. In particular, baseline urine samples were high in eugenol (118-343 $\mu\text{g/g}$ creatinine); however, levels of more abundant smoke MPs were comparably low and urine eugenol is affected by common dietary exposures,³⁸ thus it is unlikely to be related to soot contamination. Urine PAH alcohols of naphthalene (1-OH-Nap, 2-OH-Nap), fluorene (2-, 3-, and 9-OH-Flu), and phenanthrene (2-, 3-, and 4-OH-Phen), were low and consistent with concentration ranges reported in the urine of non-exposed controls and pre-shift firefighters in previous studies.³⁹⁻⁴¹ After 30 minutes of exposure to wood smoke the highest rate of chemical excretion occurred between 0-6 hours for all 3 groups and returned to median baseline concentrations or lower

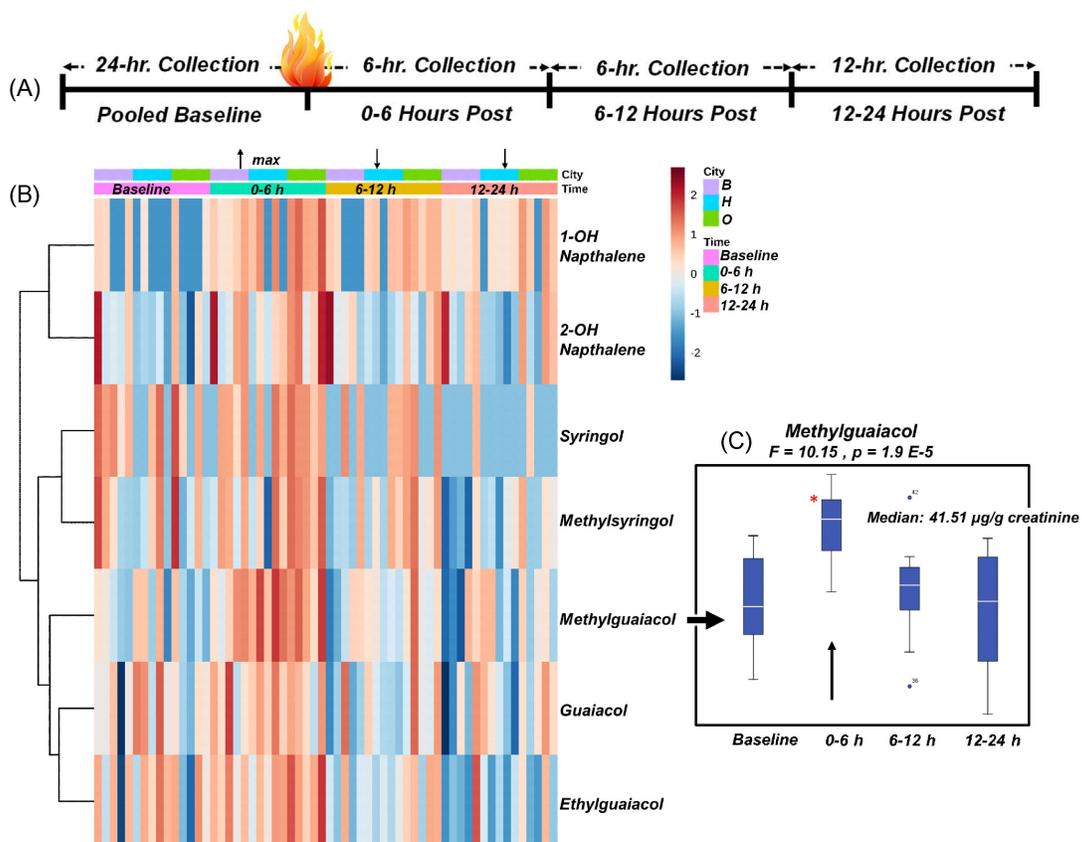


Figure 3.3: (A) Schematic detailing urine collection at 4 distinct time intervals to monitor the kinetics of biomarker excretion and identify compounds elevated due to the smoke exposure. (B) Time-sequenced 2D/HCA heat map of creatinine-normalized urinary wood smoke markers detected by GC-MS. Methylguaiacol (MeG, $p < 4.0 \text{ E-}4$), ethylguaiacol (EthG, $p < 4.0 \text{ E-}3$), syringol ($p < 5 \text{ E-}3$), and 1-hydroxynaphthalene (1-OH-Nap, $p < 0.010$) were highly responsive urine markers of smoke exposure in the 24-hour period after the fire however, elevated concentration of all markers was observed during the immediate 6-hour period post burn, showing significant uptake of ambient chemicals at the fireground despite protective gear. (C) Boxplots (ANOVA) of top discriminating smoke marker MeG showing significant increase in excretion in the 0-6 h period post exposure.

by 12-24 hours post-burn (**Figure 3.3B**). MeG, naphthalene, fluorene, guaiacol and EthG were highly responsive ($p < 0.041$) urine markers of smoke exposure in the 24-hour period after the fire consistent with compounds found in the greatest

concentration on the SCBA lens and cheek, where MeG and naphthalene remained significant ($p < 0.02$) after a stricter Bonferroni correction (**Table 3.2**). Though diet can confound measurements related to urine methoxyphenol concentrations,^{42,43} based on our findings they remain strong candidates as potentially quantitative markers to assess recent exposures due to their short half-lives and rapid clearance and could be effective when included in a multivariate stratification model to improve specificity.⁴⁴ Notably, significant changes in the top 7 urine metabolites are due to decreases in concentration in the 18 h period after the wood smoke exposure. Increases from baseline to 6 h after the exercise (T1-T2) were only significant for MeG and 1-OH-Nap due to small differences between high baseline concentrations and peak elimination in the 0-6 h pooled samples, again reflecting the possibility of chronic inadvertent contamination. Interestingly, all 4 RAs reliably detected (IPA, AA, DA, and 7ODA) reached peak excretion later than MPs in the 6-12 h period, though none were significantly higher in concentration when compared to baseline. IPA, which exhibited high penetration of the SCBA mask, had the greatest increase in urine concentration relative to baseline in Site H (mean FC = 553) and Site O (mean FC = 111), but was not reliably detected in urine samples from the less-exposed group at Site B. It also showed a moderate correlation (Spearman) to the amount of isopimaric acid deposited on the cheek ($r = -0.486$, $p = 0.078$), which may be due to an inverse relationship between respiratory uptake and material remaining to deposit. To our best knowledge this is the first study to examine the feasibility of RAs as urine markers of exposure to firefighters, and it is possible that RAs may be specific longer-term indicators of wood smoke exposures that could permit qualitative assessment of wood smoke exposure when sampling within a 6 h window is not feasible.

The greatest proportion of PAHs (> 75%) was recovered as 1- and 2-OH-Nap in the 0-6 h collections of pooled urine (**Figure 3.3**; **Table 3S.4**) which were found at a lower median concentration of 10.7 ng/mL (0.53 - 64.00 ng/mL) than found in

post-shift urine samples of coke oven workers 81.1 ng/mL (21.1 - 1225.7 ng/mL).⁴⁵ Similar to MPs, median OH-PAH concentrations were highest for site O which peaked at 6 h (9339.8 ng/g Crn) but remained elevated in the 12-24 h pooled samples when compared to baseline (FC = 2.2), an effect that was also observed in the 12-24 h samples of site B firefighters (FC = 1.9). Conversely, the less experienced firefighters at Site H showed rapid clearance of all PAH in the 6 h period post-exposure, where 2-OH-Nap reached levels equivalent to baseline by the 12-24 h collection. This observation could reflect a metabolic effect due to significantly ($p < 0.039$) lower cumulative years of firefighter experience at Site H (1.9 ± 1.6 years) when compared to Site B (13 ± 9 years) and Site O (18 ± 5 years), related to the high lipophilicity of these smoke markers and a reduced rate of excretion with increasing bioaccumulation.⁴⁶ A similar effect is observed in **Table 3.2** when comparing the rapid excretion of more polar MPs to the slower kinetic excretion of naphthalene (as 1-OH-Nap) which partitions favourably into adipose tissue and undergoes biotransformation before urinary excretion.⁴⁷

Pearson correlation analysis revealed cheek deposits of the PAH fluorene (3-OH metabolite, $r = 0.56$, $p = 0.036$) and EthG ($r = 0.55$, $p = 0.041$) recovered after intrusion into the SCBA were consistent with concentrations found in urine in the first 6 h period when controlling for baseline levels of each metabolite. Interestingly, despite the substantial partitioning of ambient MeG to both the cheek and arm, there was no correlation between either deposit and amounts excreted in urine, which could be a result of exposure by direct inhalation due loss of vacuum integrity in the SCBA. Method detection limits were not sufficiently low (≥ 10 ng/g creatinine) to detect some RA and PAH metabolites, notably 1-hydroxypyrene, the standard urine marker of PAH exposure.^{14,48,49} Heavy multi-ring PAH primarily undergo biliary (*i.e.* liver) metabolism and are excreted with bile in the feces, making metabolites of procarcinogens like benzo[a]pyrene difficult to measure in free-living subjects.⁵⁰ Poor suppression of ambient phenanthrene by the SCBA

(median 16.2% reduction, **Table 3.1**) resulted in only trace amounts on the cheek (**Table 3S.3**), however concentrations were strongly correlated to those found on the inner SCBA lens ($r = 0.72$, $p = 0.028$) and moderately correlated (Spearman) to concentrations of related metabolites among 0-6 h pooled samples (2-OH-Phen, $r = 0.524$, $p = 0.045$; 4-OH-Phen, $r = 0.521$, $p = 0.045$). Distribution of fluorene metabolites in the 0-6 h period after the fire showed strong correlations to amounts recovered from the cheeks (3-OH-Flu, $r = 0.536$, $p = 0.036$) and arms (9-OH-Flu, $r = 0.585$, $p = 0.028$) of all firefighters. Urine concentrations of 2-OH-Phen and 9-OH-Flu have been positively associated with pro-inflammatory effects that are linked to the etiology of heart disease and diabetes. In a review of NHANES (2003-2004) data, Everett et. al. found a moderate positive relationship (OR = 1.09; CI = 1.06-3.42) between amounts of 2-OH-Phen in excess of 49 g/g Crn and concentrations of inflammation marker high-sensitivity C-reactive protein (hsCRP) ≥ 3 mg/L, which reached an OR of 3.17 if urine excretion exceeded 148 ng/g creatinine.⁵¹ Similar but weaker effects were calculated for 9-OH-Flu (OR = 2.28, > 749 ng/g Crn); however, several low-level persistent environmental PAH are implicated.⁴⁸ HsCRP is a modern marker of systemic health that is a more robust predictor of cardiovascular disease and metabolic syndrome than traditional anthropometric measurements (e.g. waist-to-hip ratio and blood lipids/cholesterol).⁵² The physiological stress of chronic inflammation has been conclusively linked to arterial calcification (i.e. atherosclerosis), obesity, and cancer, which are notable concerns among long-term career firefighters. Within each time interval, 4 firefighters had urinary 2-OH-Phen exceeding the reported upper limit of 148 ng/g Crn, while 72.2% of pooled urine samples contained 9-OH-Flu in concentrations exceeding 161 ng/g Crn, a concentration associated with 1.89-times increased odds of elevated hsCRP. In fact, during the 24-hr collection window metabolites of fluorene persistently exceeded urine concentrations correlated to

Table 3.2: Top discriminating urine markers of smoke exposure in 24-hr baseline and 24-hr post-exposure urine samples (0-6 h, 6-12 h, 12-24 h) collected from firefighters ($n = 15$)

Marker	F	p ^a	q ^a	Effect Size ^b	Pairwise Comparisons				
					T1-T2	T2-T3	T3-T4	T1-T4	T2-T4
Guaiacol	2.83	0.046	0.040	0.132		* ↓			* ↓
MeG	10.15	1.9 E-5	4.1 E-4	0.352	* ↑	* ↓			* ↓
EthG	7.13	3.9 E-4	4.1 E-3	0.276		* ↓			* ↓
Syringol	6.62	0.001	0.005	0.262		* ↓	* ↓	* ↓	* ↓
MeS	4.19	0.010	0.040	0.183			* ↓	* ↓	* ↓
PropS	3.01	0.038	0.040	0.139					* ↓
1-OH-Nap	5.38	0.003	0.013	0.224	* ↑	* ↑		* ↑	

a p-adjusted value (q-value) based on False Discovery Rate (FDR) using Benjamini-Hochberg procedure;

b Effect size measured using Partial Eta Square.

inflammation risk in $\geq 47\%$ of samples. Urine samples were also surveyed for cyanide, another smoke component of concern, however no increase in cyanide excretion due to smoke exposure was found in this study. A recent acceleration of research into the occupational exposures to firefighters have sought to confirm a causal link between the complex hazardous aerosolized mixtures of chemicals at the fireground, surreptitious exposures and high rates of cancer and cardiovascular disease among firefighters. In this study we observed that heterogeneity in the chemical profile and magnitude of wood smoke exposures to firefighters varied, in part due to the type of structure containing the fire, in addition to known effects determined by the proximity to wood smoke. Personal air pumps were more effective at capturing information regarding the carcinogenic and mutagenic potential of these exposures than passive Twister sorbents, as heavier semivolatiles partition favourably to particulate, which was not efficiently sampled using the sorbent device. Protective clothing was confirmed to mitigate, but not prevent, the

deposition of soot under clothing, where the SCBA was particularly deficient in achieving the indicated protection factor (PF = 10 000). In addition, this led to fluorene, several MPs, and all RAs depositing on the cheek under the SCBA mask.

Urine MPs (MeG, guaiacol, EthG, syringol, MeS, and PropS) and 1-OH-Nap increased significantly within 6 h following acute smoke exposure due to a high concentration in the wood smoke, which establishes an optimal time window for future urine sample collection for firefighters in the field. However, identifying the immediate effect of an acute smoke exposure was challenged due to a moderate background at baseline that mitigated significant increases in urine concentration. We suggest this is due to the cumulative effects of cross-contamination from handling turnout gear and inadequate hygienic practices, which represents a non-negligible risk of chronic exposure, including by ingestion, to firefighters and coworkers. There were some challenges in this study regarding sample collection, including the lack of chemical tracers to confirm whether all urine voids within the 48 h were reliably captured, which is a source of potential bias. Furthermore, no wipes were taken after baseline sampling to confirm complete removal of contaminants before the fire exercise, thus it is possible that some remaining sediment was captured in the post-exposure wipes. However, as several post-exposure samples yielded higher concentrations on the skin after the fire, the determination of deficiencies in the turnout gear remains unchanged. Lastly, this cross-sectional study included both experienced firefighters and new recruits that was a source of variation between training sites, including fire intensity and dimensions of burn houses impacting local concentrations of contaminants measured in air samples and absorbed onto skin. Moreover, our study design cannot directly correlate the smoke exposure with chronic disease risk assessment as other serum biomarkers of inflammation were not measured, which optimally requires a longitudinal study. As an acute exposure study, the extrapolation of these findings to the characterization of all smoke exposures to firefighters is challenged by the

high variability in the frequency and type of exposures firefighters regularly endure. Nonetheless, the findings discussed here are reproduced among all firefighters across multiple sites and supported by similar observations in the literature. Additionally, the exposure determined here occurred in a controlled environment, with PPE used under ideal circumstances, which is not likely to be the case during a major structural, urban or industrial fire. In conclusion, there is urgent need for the design of improved equipment that mitigates smoke exposure from skin absorption, as well as better training/maintenance of SCBA usage, and the implementation of more effective hygiene/cleaning practices of turnout gear (e.g., access to secondary bunker gear) in order to improve the occupational health of structural firefighters and trainees.

3.4 References

1. M. Ezzati, A. D. Lopez, A. Rodgers, S. Vander Hoorn and C. J. L. Murray, *Lancet*, 2002, **360**, 1347–1360.
2. A. F. Scott and C. A. Reilly, *Chem. Res. Toxicol.*, 2019, **32**, 219–221.
3. A. Muala, G. Rankin, M. Sehlstedt, J. Unosson, J. A. Bosson, A. Behndig, J. Pourazar, R. Nyström, E. Pettersson, C. Bergvall, R. Westerholm, P. I. Jalava, M. S. Happo, O. Uski, M.-R. Hirvonen, F. J. Kelly, I. S. Mudway, A. Blomberg, C. Boman and T. Sandström, *Part. Fibre Toxicol.*, 2015, **12**, 33.
4. A. Blair, D. Blask, M. Bråtveit, T. Brock, J. L. Burgess, G. Costa, S. Davis, P. A. Demers, J. Hansen, E. Haus, P. J. Landrigan, G. K. Lemasters, F. Lévi, F. Merletti, C. J. Portier, E. Pukkala, E. Schernhammer, K. Steenland, R. Stevens, R. Vermeulen, T. Zheng, Y. Zhu, J. Arendt, C. Austin, J. Cherrie, A. Huici-Montagud, K. Mundt, A. Altieri, R. Baan, A. Altieri, R. Baan, V. Bouvard, V. James Coglianò, F. Giannandrea, F. El Ghissassi, Y. Grosse, J. Heck, J. Mitchell, N. Napalkov, B. Secretan, K. Straif, S. Egraz, M. Javin, B. Kajo, M. Lézère, H. Lorenzen-Augros, C. Freeman, N. Guha, L. Galichet, A. S. Hameau, S. Moutinho and D. Russell, *IARC Monogr. Eval. Carcinog. Risks to Humans*, 2010, **98**, 804.
5. Government of Ontario, *Ontario Regulation 253/07: FIREFIGHTERS*, 2018.
6. R. C. Petterson, *Chem. Solid Wood*, 1984, **207**, 57–126.
7. T. Shimada, *Drug Metab. Pharmacokinet.*, 2006, **21**, 257–76.

8. I. Shiue, *Environ. Sci. Pollut. Res.*, 2015, **22**, 16962–16968.
9. T. A. Söderberg, A. Johansson and R. Gref, *Toxicology*, 1996, **107**, 99–109.
10. M. Cordella, C. Torri, A. Adamiano, D. Fabbri, F. Barontini and V. Cozzani, *J. Hazard. Mater.*, 2012, **231–232**, 26–35.
11. J. L. Burkey, J.-M. Sauer, C. A. McQueen and I. Glenn Sipes, *Mutat. Res. Mol. Mech. Mutagen.*, 2000, **453**, 25–33.
12. J. Laitinen, M. Mäkelä, J. Mikkola and I. Huttu, *Toxicol. Lett.*, 2012, **213**, 129–33.
13. T. Shimada and Y. Fujii-Kuriyama, *Cancer Sci.*, 2004, **95**, 1–6.
14. B. Gill, A. Mell, M. Shanmuganathan, K. Jobst, X. Zhang, D. Kinniburgh, N. Cherry and P. Britz-McKibbin, *Anal. Bioanal. Chem.*, 2019, **411**, 1397–1407.
15. J. Yang, D. Teehan, A. Farioli, D. M. Baur, D. Smith and S. N. Kales, *Am. J. Cardiol.*, 2013, **112**, 1962–1967.
16. S. Fernando, L. Shaw, D. Shaw, M. Gallea, L. Vandenenden, R. House, D. K. Verma, P. Britz-McKibbin and B. E. McCarry, *Environ. Sci. Technol.*, 2016, **50**, 1536–1543.
17. N. Cherry, Y. Aklilu, J. Beach, P. Britz-McKibbin, R. Elbourne, J.-M. Galarneau, B. Gill, D. Kinniburgh and X. Zhang, *Ann. Work Expo. Heal.*, 2019, **63**, 448–458.
18. B. Strandberg, A. Julander, M. Sjöström, M. Lewné, K. A. Hatice and C. Bigert, *Chemosphere*, 2018, **198**, 274–280.
19. M. A. G. Wallace, J. D. Pleil, K. D. Oliver, D. A. Whitaker, S. Mentese, K. W. Fent and G. P. Horn, *J. Occup. Environ. Hyg.*, 2019, **16**, 355–366.
20. N. K. Jayatilaka, P. Restrepo, L. Williams, M. Ospina, L. Valentin-Blasini and A. M. Calafat, *Anal. Bioanal. Chem.*, 2017, **409**, 1323–1332.
21. K. W. Fent, B. Alexander, J. Roberts, S. Robertson, C. Toennis, D. Sammons, S. Bertke, S. Kerber, D. Smith and G. Horn, *J. Occup. Environ. Hyg.*, 2017, **14**, 801–814.
22. J. L. A. Keir, U. S. Akhtar, D. M. J. Matschke, T. L. Kirkham, H. M. Chan, P. Ayotte, P. A. White and J. M. Blais, *Environ. Sci. Technol.*, 2017, **51**, 12745–12755.
23. A. R. Katritzky, D. A. Dobchev, E. Hür, K. Tämm, L. Kurunczi, M. Karelson, A. Varnek and V. P. Solov’ev, 2006, **49**, 3305–3314.
24. M. of L. Government of Ontario, Current Occupational Exposure Limits for Ontario Workplaces Required under Regulation 833, http://www.labour.gov.on.ca/english/hs/pubs/oel_table.php, (accessed 1 July 2019).
25. Centers for Disease Control and Prevention (CDC) - The National Institute for Occupational Safety and Health (NIOSH), NIOSH Pocket Guide to Chemical Hazards, <https://www.cdc.gov/niosh/npg/npgsyn-a.html#a>,

- (accessed 18 July 2019).
26. C. Qiu, B. Peng, S. Cheng, Y. Xia and B. Tu, *Am. J. Ind. Med.*, 2013, **56**, 347–355.
 27. I. Burstyn, H. Kromhout, T. Partanen, O. Svane, S. Langård, W. Ahrens, T. Kauppinen, I. Stücker, J. Shaham, D. Heederik, G. Ferro, P. Heikkilä, M. Hooiveld, C. Johansen, B. G. Randem and P. Boffetta, *Epidemiology*, 2005, **16**, 744–50.
 28. J. G. VanRooij, M. M. Bodelier-Bade and F. J. Jongeneelen, *Occup. Environ. Med.*, 1993, **50**, 623–652.
 29. B. R. T. Simoneit, W. F. Rogge, M. A. Mazurek, L. J. Standley, L. M. Hildemann, G. R. Cass, B. R. T. Simonelt, W. F. Rogge, M. A. Mazurek, L. J. Standley, L. M. Hildemann and G. R. Cass, *Environ. Sci. Technol.*, 1993, **27**, 2533–2541.
 30. A. I. Miranda, V. Martins, P. Cascão, J. H. Amorim, J. Valente, R. Tavares, C. Borrego, O. Tchepel, A. J. Ferreira, C. R. Cordeiro, D. X. Viegas, L. M. Ribeiro and L. P. Pita, *Environ. Int.*, 2010, **36**, 736–745.
 31. N. Fire, P. Association, B. Park and P. O. Box, *Policy*.
 32. U.S. Occupational Safety and Health Administration, *Assigned Protection Factors for the Revised Respiratory Protection Standard*, Washington, D.C., 2009.
 33. R. A. Bryant and A. Mensch, *J. Occup. Environ. Hyg.*, 2011, **8**, 437–446.
 34. A. T. Johnson, *J. Biol. Eng.*, 2016, **10**, 4.
 35. A. C. Mayer, K. W. Fent, S. Bertke, G. P. Horn, D. L. Smith, S. Kerber and M. J. La Guardia, *J. Occup. Environ. Hyg.*, 2019, **16**, 129–140.
 36. Ontario Ministry of Labour, Hygiene and Decontamination, <https://www.ontario.ca/document/firefighter-guidance-notes/6-1-hygiene-and-decontamination>, (accessed 14 May 2019).
 37. NFPA (National Fire Protection Association), .
 38. M. Wagenstaller and A. Buettner, *Metabolomics*, 2014, **10**, 225–240.
 39. O. Adetona, C. D. Simpson, G. Onstad and L. P. Naeher, *Ann. Occup. Hyg.*, 2013, **57**, 979–991.
 40. M. Oliveira, K. Slezakova, C. P. Magalhães, A. Fernandes, J. P. Teixeira, C. Delerue-Matos, M. do Carmo Pereira and S. Morais, *J. Hazard. Mater.*, 2017, **334**, 10–20.
 41. Z. Li, C. D. Sandau, L. C. Romanoff, S. P. Caudill, A. Sjodin, L. L. Needham and D. G. Patterson, *Environ. Res.*, 2008, **107**, 320–331.
 42. R. L. Dills, X. Zhu and D. A. Kalman, *Environ. Res.*, 2001, **85**, 145–158.
 43. D. H. Phillips, *Mutat. Res. Toxicol. Environ. Mutagen.*, 1999, **443**, 139–147.
 44. R. L. Dills, M. Paulsen, J. Ahmad, D. A. Kalman, F. N. Elias, Simpson and Christopher D., 2006, **40**, 2163–2170.
 45. F. Rossella, L. Campo, S. Pavanello, L. Kapka, E. Siwinska and S. Fustinoni,

- Occup. Environ. Med.*, 2009, **66**, 509–16.
46. M. Bouchard, R. Thuot, G. Carrier and C. Viau, *J. Toxicol. Environ. Heal. - Part A*, 2002, **65**, 1195–1209.
 47. T. Fukami, M. Katoh, H. Yamazaki, T. Yokoi and M. Nakajima, *Chem. Res. Toxicol.*, 2008, **21**, 720–725.
 48. O. Alshaarawy, M. Zhu, A. Ducatman, B. Conway and M. E. Andrew, *Environ. Res.*, 2013, **126**, 98–104.
 49. C. Caux, C. O'Brien and C. Viau, *Appl. Occup. Environ. Hyg.*, 2002, **17**, 379–386.
 50. A. Ramesh, S. A. Walker, D. B. Hood, M. D. Guillén, K. Schneider and E. H. Weyand, *Int. J. Toxicol.*, 2004, **23**, 301–333.
 51. C. J. Everett, D. E. King, M. S. Player, E. M. Matheson, R. E. Post and A. G. Mainous, *Environ. Res.*, 2010, **110**, 79–82.
 52. P. M. Ridker, J. E. Buring, N. R. Cook and N. Rifai, *Circulation*.
 53. N. L. Kuehnbaum, A. Kormendi and P. Britz-Mckibbin, *Anal. Chem.*, 2013, **85**, 10664–10669.
 54. J. Chong, O. Soufan, C. Li, I. Caraus, S. Li, G. Bourque, D. S. Wishart and J. Xia, *Nucleic Acids Res.*, 2018, **46**, W486-494.
 55. WFBC, Women Firefighters Biomonitoring Collaborative, <http://womenfirefighterstudy.com/> (accessed 2 August 2019).

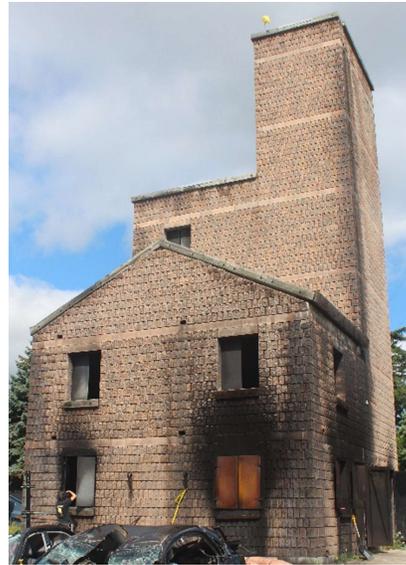
3.5 Supplemental

Appendix

3.1. Photos of Burn Houses at Site B, Site H, and Site O

3.2. Urine Collection Protocol for Study Participants

1. Photos of Burn Houses at Site B, Site H, and Site O



Site H



Site O



2. Urine Collection Protocol for Study Participants

General Guidelines (48 hr)

It is strongly recommended that you follow these instructions over a 48-hour time period. Sudden changes in behaviour may introduce interferences that will bias this study:

- Avoid barbecued and charred foods
- Avoid strenuous exercise
- Avoid alcohol or increases in alcohol consumption
- Maintain your usual diet and exercise routine
- Ensure the correct name is on all of your sample collection containers
- Store all sealed/labeled containers in a cooler that is provided

Urine Collection Instructions

I. Pre-Exposure Urine Collection (24 hr)

Please use the provided sample container(s) (3-4) to collect *all urine* over the 24-hr pre-exposure period. Depending on hydration, total urine volume for 24 hrs is on average over 2 L, but can vary significantly between individuals. You will be provided several 1L wide-mouth plastic containers for collection and storage of urine.

- Discard your first morning urine at the start of Day 1. Record this as the start date and time of your 24-hour collection on your sample bottle.
- For *every* following urination during the 24-hour period, please collect the sample in the provided containers. Please use additional containers as needed depending on total volume of urine.
- Ensure that the lid is secure on the sample container when not in use and placed in a sealed cooler during storage.
- At 24 hours please attempt a final urination. Record this as the stop date and time on your container and secure the lid tightly. If another urination is not possible, record the stop date and time at

the end of the 24-hour period and secure the lid tightly.

II. Post-Exposure Urine Collection (24 hr)

Use the provided sample containers (3-4) to collect *all urine* over the 3 stages of the 24-hour post-exposure period. ***Please ensure that you are using the appropriately labelled sample container at each urination.***

- Confirm that you have been provided 3 sample containers labeled “0-6 hr”, “6-12 hr”, and “12-24 hr”.
- At the first urination following fire exposure, please urinate into the container labelled “0-6 hr” and record this as the start date and time on your container.
- For *every* following urination, please collect the sample in the “0-6 hours” sample container.
- At 6 hours following fire exposure, record the *stop* date and time on your container and secure the lid tightly. Record the *same date and time* as the *start* date and time on your “6-12 hr” sample container and follow the same procedure as before.
- At 12-hours following fire exposure, label the stop date and time on the container. Begin using the “12-24 hr” sample container and label the start date and time as previously instructed. Collect *all* urine.
- At 24 hours following fire exposure, attempt a final urination. Record this as the stop date and time on your container and secure the lid tightly. If another urination is not possible, record the stop date and time at the end of the 24-hour period and secure the lid tightly.

Please note any deviations from the above procedure or any concerns, if any. Submit this sheet with your samples. (If none, then you may keep this sheet.)

Supplemental Information

Experimental

Materials

Ultra HPLC grade LC-MS solvents were used to make all buffers and sheath liquids, unless otherwise stated. Ultra HPLC grade LC-MS solvents water, methanol (MeOH), acetonitrile (ACN), isopropanol, and dichloromethane (DCM) were obtained from Caledon Laboratories Ltd. (Georgetown, ON, Canada). HPLC grade solvents (DCM, water, methanol, ACN, and toluene) were used for GC-MS sample preparation and purchased from Caledon Laboratories Ltd. N-methyl-N-(trimethylsilyl)trifluoroacetamide (MSTFA) >98.5%, β -glucuronidase (from *Helix Pomatia*, Type Hp-2, aqueous solution, $>1.0 \times 10^5$ units/mL) and sodium acetate (NaAc, reagent plus 99%), and 3-chloro-L-tyrosine (Cl-Tyr) were purchased from Sigma-Aldrich Inc. (Canada & USA). PAH standards were purchased from Chiron AS (Norway). Phenanthrene-d₁₀ (98%), pyrene-d₁₀ (98%), and chrysene-d₁₂ (98%) were purchased from Cambridge Isotope Laboratories (USA). Guaiacol, methylguaiacol, ethylguaiacol, propylguaiacol, syringol, methylsyringol, eugenol, isoeugenol, and the hydroxyl-PAHs were purchased from MRI (USA). Ethylsyringol, propylsyringol, guaiacol-d₄ (98%), syringol-d₆ (98%), and acetosyringone-d₆ (98%) were generously provided by Christopher D. Simpson.

Study Design

Eighteen full-time occupational firefighters from the municipalities of Burlington (Site B), Hamilton (Site H), and Ottawa (Site O) in the province of Ontario volunteered to participate in a controlled burn exercise where they conducted a mock search-and-rescue in an on-site training structure (“burn house”) while a localized active fire burned within the structure (pictures of each burn house are provided for reference on Supplemental **Appendix 3.1**). This study was approved by the McMaster Research Ethics Manager (MREB #2070). All volunteers were

healthy males with a median age of 40 years and median BMI of 27.0 (22.5-32.3). All participants identified as Caucasian except 1, and all identified as non-smokers except 1. Site B and site O had a mean 13 ± 9 and 18 ± 5 years of firefighting experience, respectively, while Group H firefighters were relatively recent recruits with 1.9 ± 1.6 years of experience. None had attended a fire within the past 3 days and remained off duty for the duration of the study while refraining from smoke exposures (*e.g.* smoking, barbecuing, fires), alcohol, and changes to their regular daily routines (**Appendix 3.2**). Each consented to collect four batches of pooled urine samples over a 48-h period, including baseline (T-24 h prior to burn trial), and 0-6 h, 6-12 h and 12-24 h following smoke exposure. Urine samples from each firefighter were collected in four pre-rinsed (with methanol) 1 L polyethylene containers, and personal coolers containing ice packs were used for sample storage, which were returned the following day for pickup. Baseline samples were taken from 24-hr pooled urine the day before the burn. Pre- and post-burn sample wipes were taken of the inner and outer SCBA lens, 1 forearm, and both cheeks using pre-extracted (in DCM) 7-cm diameter Whatman filter papers soaked in isopropanol. Forearm samples were standardized at 3 wipes of 10 strokes each over a foil template with an approximately 30 cm² circular cut-out. Wipes were folded inward and sealed in plastic bags to prevent transfer or contamination. Each participant was also equipped with a personal gas and particulate sampling apparatus with a Teflon filter connected in series to an XAD-2 sorbent tube and air pump calibrated to 2 L/min. Sorbent Gerstel Twisters (Maryland, USA) with 0.5 mm thick polydimethylsiloxane (PDMS) coatings were affixed to the turnout coat in stainless steel mesh for passive ambient air sampling and inside the SCBA to monitor for potential chemical penetration of the mask. Pallets of untreated wood and straw were used for fuel and time was given for smoke to fill the burn house after ignition. Each group entered together in full regulatory turnout gear and SCBA. One to two firefighters controlled the fire while the remainder conducted a search and rescue

exercise for 30 min before extinguishing the fire and exiting the structure as a group. Due to additional activities after the burn exercise that increased their handling of soiled bunker gear, 3 firefighters at Site O were removed from all post-exposure analyses to avoid bias in the study.

Air Sample Extraction

After the exercise 16/18 pumps remained functional and 8/18 caged passive sampling Twisters were recovered. Filters were extracted in 3 mL of DCM with sonication for 20 min. XAD-2 resin was extracted by agitating the contents in ACN for 30 minutes and filtered using a 0.45 μm syringe filter. 1 mL of filtrate was solvent exchanged into 1 mL of ACN by N₂ blow down. A 50 μL aliquot of the combined extract was derivatized with 40 μL MSTFA at 60°C for 30 min. to form trimethylsilyl derivatives of MP. Pyrene-d10 (10 μL , 10 ng/ μL) was added as an internal standard prior to analysis.

Wipe Extraction

Each wipe was spiked with 10 μL of recovery standard prior to accelerated solvent extraction (ASE 300, Dionex, Sunnyvale, CA). The recovery standard consisted of guaiacol-d₄, syringol-d₆, acetosyringone-d₆, phenanthrene-d₁₀, and chrysene-d₁₂ each at a concentration of 10 ng/ μL in ACN. DCM at 110 °C at 10.3 MPa was heated for 5 min then underwent static extraction for 5 min for three extraction cycles. The ASE extract was mixed with 1 mL of toluene then blown to 1 mL using a Rotovap. A 50 μL aliquot of the extract was derivatized with 40 μL MSTFA at 60 °C for 30 min. Pyrene-d10 (10 μL , 10 ng/ μL) was added as an internal standard prior to analysis.

GC-MS Urine Workup

Urine samples were pre-filtered with a 0.45 μm syringe filter. A 3 mL aliquot was then mixed with 5 mL of 0.100 M NaAc buffer (pH 5.5), then 20 μL of a recovery

standard mix consisting of guaiacol-d4, syringol-d6, acetosyringone-d6, and [13C6]-1-hydroxypyrene each at 1 ng/ μ L. Ten microlitres of β -glucuronidase/sulfatase was added to the sample and allowed to incubate at 37 °C for 17-18 h to fully deconjugate hydroxy-PAH metabolites.¹⁴ Varian Focus SPE cartridges (50 mg, 6 mL) were used for sample cleanup and conditioned with 1 mL of MeOH followed by 1 mL of water at a flow rate of 10 mL/min. The sample was then loaded at a flow rate of 1 mL/min. The cartridge was rinsed with 1 mL of water followed by 20% MeOH in 0.100 M NaAc buffer (pH 5.5). The sorbent of the cartridge was dried by aspirating air through it for 5 min. followed by nitrogen blow down for 5 min. to ensure complete removal of water. The cartridge was eluted with 5 mL of DCM at a rate of 0.5 mL/min. The 5 mL DCM extract was run through a sodium sulfate-packed Pasteur pipet to remove residual water then blown down to 50 μ L and derivatized with 40 μ L of MSTFA at 60 °C for 30 min. Pyrene-d10 (10 μ L, 10 ng/ μ L) was added as an internal standard prior to analysis. Pooled urine samples were run intermittently between sample batches to calculate technical precision.

GC-MS and GC-MS/MS Analysis

Air and wipe samples were analyzed on an Agilent 6890N gas chromatograph (GC) coupled to an Agilent 5973N mass selective detector (MSD) in selected ion monitoring (SIM) mode. A suite of 31 PAH, MP, and RA were targeted in both air and skin samples (see Supplemental Information for target compounds) using a method described previously.¹⁶ Skin wipe data is normalized to the total area of all wipes used per sample. Urine samples were analyzed on a Varian CP-3800 GC coupled to a Varian 1200L triple-quadrupole mass spectrometer, operated in multiple-reaction monitoring (MRM) mode. Twenty-five hydroxy-PAH, MP, and RA were targeted in urine samples (Supplemental Information). Complementary univariate and multivariate analysis methods were used to identify key metabolites

associated with the smoke exposure. All quantities reported are relative to the pyrene-d₁₀ internal standard and all urine data is normalized to creatinine reduce bias from hydration status. Individual creatinine was determined using the CE-MS method described below. Non-detects were replaced by LOD/2. Modelled data was log-transformed and autoscaled, unless otherwise stated.

CE-MS Urine Workup

For the determination of urine creatinine samples were thawed on ice and centrifuged for 10 minutes at 14 000 g. Supernatant was diluted 20-fold with pH 4.9 75 mM ammonium acetate (NH₄Ac) buffer and spiked to contain 10 μM of 3-chloro-L-tyrosine as an internal standard. Analysis was performed on an Agilent G7100A CE coupled to an Agilent 6550 quadrupole time-of-flight (QTOF) mass spectrometer using an MSI-CE-MS method.⁵³ Samples were analyzed in positive mode using a 1 M formic acid (pH 1.8) with 15% v/v acetonitrile background electrolyte.

Statistical Data Analysis

Univariate and multivariate statistical analyses, including Mann-Whitney U test, one- and two-factor analysis of variance (repeat measures ANOVA, Kruskal-Wallis), and Pearson correlation were performed using Microsoft Excel and Statistical Package for the Social Sciences (SPSS 23.0). Hierarchical cluster analysis (HCA)/2D heat maps and principal component analysis (PCA) were performed using MetaboAnalyst 4.0.⁵⁴ Creatinine normalization was used to correct for hydration status and non-detects were replaced by a minimal value (LOD/2). Modelled data was log-transformed and autoscaled, unless otherwise stated.

Table 3S.1: (Mean \pm SD) recovery (ng) of methoxyphenols (MP), polycyclic aromatic hydrocarbons (PAH), and resin acids (RA) recovered from outer caged Twisters ($n = 8$) attached to the turnout coat compared to within the SCBA ($n = 7$) of firefighters

	Site B ($n = 5$)	Site H ($n = 2$)	Site O ($n = 1$)	Site B ($n = 5$)	Site H ($n = 2$)
	Caged Twister			SCBA Twister	
MP					
Guaiacol	73.9 \pm 28.4	254 \pm 225	400	0.67 \pm 0.54	0.39 \pm 0.31
EthG	94.0 \pm 26.1	392 \pm 161	615	1.05 \pm 0.69	0.68 \pm 0.16
PropG	46.0 \pm 26.9	110 \pm 25	159	72.9 \pm 25.6	1.50 \pm 1.00
Syringol	171 \pm 24	319 \pm 45	1706	2.04 \pm 2.23	0.55 \pm 0.09
MeS	69.9 \pm 25.7	96 \pm 11	881	1.04 \pm 1.08	0.78 \pm 0.68
EthS	79.3 \pm 36.5	139 \pm 38	1044	65.0 \pm 39.1	1.82 \pm 1.77
PropS	109 \pm 75	18.5 \pm 8.9	1821	2.00 \pm 4.19	6.06 \pm 7.2
Eugenol	24 \pm 6.4	101 \pm 23	156	1.21 \pm 0.33	0.3 \pm 0.08
Isoeugenol	112 \pm 43	373 \pm 79	944	21.3 \pm 12.7	0.136 \pm 0.001
PAH					
Naphthalene	58.8 \pm 12.3	390 \pm 286	557	15.4 \pm 6	7.6 \pm 10.1
Fluorene	9.7 \pm 3.3	28.7 \pm 3.3	72.7	0.23 \pm 0.22	0.18 \pm 0.14
Phenanthrene	27.5 \pm 9.1	72 \pm 66	233	0.49 \pm 0.97	0.1 \pm 0.08
Fluoranthene	13.4 \pm 5.8	4.7 \pm 1.6	23.5	0.96 \pm 0.96	0.14 \pm 0.14
Pyrene	18.8 \pm 4.8	8.4 \pm 8.7	24.9	4.93 \pm 6.12	2.07 \pm 0.32
Acenaphthylene	28.9 \pm 7.5	138 \pm 22	231	3.27 \pm 3.31	0.12 \pm 0.05
Acenaphthene	2.95 \pm 0.72	5.4 \pm 6.7	1.6	0.97 \pm 0.8	0.26 \pm 0.06
RA					
7ODA	0.15 \pm 0.05	0.200 \pm 0.004	0.21	0.18 \pm 0.1	0.12 \pm 0.05
AA	0.16 \pm 0.24	0.14 \pm 0.05	0.06	0.17 \pm 0.16	0.1 \pm 0.05
DA	20.3 \pm 4.46	71.4 \pm 4.0	2.7	20 \pm 4.9	37.8 \pm 10.4
IPA	0.67 \pm 0.25	0.84 \pm 1.03	0.06	0.49 \pm 0.14	0.29 \pm 0.17
PA	0.22 \pm 0.17	0.14 \pm 0.08	0.21	0.23 \pm 0.16	0.204 \pm 0.001
SPA	0.33 \pm 0.24	0.534 \pm 0.004	0.12	0.3 \pm 0.17	0.22 \pm 0.13

Table 3S.2: (Mean \pm SD) ambient concentrations ($\mu\text{g}/\text{m}^3$) of methoxyphenols (MP), polycyclic aromatic hydrocarbons (PAH), and resin acids (RA) captured by the personal air pumps ($n = 16$) of firefighters

	Site B ($n = 3$)	Site H ($n = 5$)	Site O ($n = 8$)
MP			
Guaiacol	31.7 \pm 9.6	9.6 \pm 234	234 \pm 194
MeG	12.4 \pm 14.9	14.9 \pm 330	330 \pm 306
EthG	13.8 \pm 7.1	7.1 \pm 228	228 \pm 207
PropG	2.6 \pm 2.7	2.7 \pm 59.3	59.3 \pm 54.8
Syringol	30.4 \pm 12.2	12.2 \pm 222	222 \pm 195
MeS	10.5 \pm 6.2	6.2 \pm 79.3	79.3 \pm 61.2
EthS	13.3 \pm 9.3	9.3 \pm 112	112 \pm 83.4
PropS	3 \pm 2.3	2.3 \pm 17	17 \pm 11.7
Eugenol	4.5 \pm 2.1	2.1 \pm 35.2	35.2 \pm 49.8
Isoeugenol	10.4 \pm 7.4	7.4 \pm 63.5	63.5 \pm 63.8
PAH			
Naphthalene	0.11 \pm 0.07	0.07 \pm 0.52	0.52 \pm 0.36
Fluorene	0.03 \pm 0.02	0.02 \pm 20.7	20.7 \pm 16.8
Phenanthrene/ Anthracene	3.09 \pm 1.49	1.49 \pm 35.2	35.2 \pm 8.4
Fluoranthene	2.74 \pm 0.91	0.91 \pm 1.76	1.76 \pm 14.27
Pyrene	2.16 \pm 4.69	4.69 \pm 48.2	48.2 \pm 3.6
Acenaphthylene	0.15 \pm 0.09	0.09 \pm 5.01	5.01 \pm 9.28
Acenaphthene	0.22 \pm 0.23	0.23 \pm 0.65	0.65 \pm 0.37
Chrysene	1.41 \pm 0.86	0.86 \pm 0.77	0.77 \pm 7.87
BaA	1.52 \pm 3.7	3.7 \pm 12.8	12.8 \pm 29
BbF	0.55 \pm 1.22	1.22 \pm 3.83	3.83 \pm 1.43
BkF	0.59 \pm 0.27	0.27 \pm 4.4	4.4 \pm 2.3
BaP	0.89 \pm 0.3	0.3 \pm 5.78	5.78 \pm 2.33
IcdP	0.02 \pm 0.45	0.45 \pm 0.9	0.9 \pm 5.44
DahA	0.003 \pm 0.004	0.03 \pm 0.15	0.15 \pm 1.63
BghiP	0.08 \pm 0.12	0.52 \pm 1.49	0.03 \pm 0.02
RA			
7ODA	0.96 \pm 6.31	6.31 \pm 8.1	8.1 \pm 351.4
AA	1.56 \pm 0.3	0.3 \pm 38.7	38.7 \pm 8.6
DA	3.7 \pm 1.71	1.71 \pm 244	244 \pm 29
IPA	0.28 \pm 0.2	0.2 \pm 12.7	12.7 \pm 9.2
PA	0.23 \pm 0.12	0.12 \pm 10.8	10.8 \pm 0.02
SPA	0.22 \pm 0.24	0.24 \pm 7.3	7.3 \pm 7.1

Table 3S.3: (Mean \pm SD) ambient concentrations (ng/cm²) of methoxyphenols (MP), polycyclic aromatic hydrocarbons (PAH), and resin acids (RA) on the cheeks and arms of firefighters (n = 15)

	Site B (n = 5)	Site H (n = 5)	Site O (n = 8)	Site B (n = 5)	Site H (n = 5)	Site O (n = 8)	Site B (n = 5)	Site H (n = 5)	Site O (n = 8)	Site B (n = 5)	Site H (n = 5)	Site O (n = 8)
	Check - Pre			Check - Post			Arm - Pre			Arm - Post		
<i>MP</i>												
Guaiacol	0.51 \pm 0.48	2.04 \pm 3.99	2.73 \pm 2.08	0.98 \pm 0.79	1.33 \pm 2.32	4.29 \pm 2.54	0.81 \pm 0.71	1.94 \pm 3.29	6.65 \pm 4.67	0.81 \pm 0.71	0.38 \pm 0.38	3.46 \pm 2.52
MeG	27.6 \pm 3.4	30.5 \pm 7.5	2.18 \pm 1.89	26.4 \pm 1.3	31.09 \pm 5.71	2.23 \pm 1.1	25.4 \pm 1.5	26.18 \pm 10.4	3.1 \pm 1.84	25.4 \pm 1.5	35 \pm 8.4	9.05 \pm 17.92
EthG	0.62 \pm 0.97	0.44 \pm 0.28	2.61 \pm 2.39	0.23 \pm 0.12	0.58 \pm 0.32	1.74 \pm 1.28	0.23 \pm 0.17	1.59 \pm 2.77	3.11 \pm 2.22	0.23 \pm 0.17	0.37 \pm 0.22	0.57 \pm 0.66
PropG	2.17 \pm 2.85	0.05 \pm 0.07	0.69 \pm 0.97	3 \pm 2.7	0.12 \pm 0.06	0.54 \pm 0.24	1.97 \pm 2.69	0.14 \pm 0.23	0.7 \pm 0.6	1.97 \pm 2.69	0.15 \pm 0.16	0.21 \pm 0.26
Syringol	0.21 \pm 0.19	0.44 \pm 0.33	36.5 \pm 59.2	0.48 \pm 0.41	0.86 \pm 0.68	31.8 \pm 30.1	0.39 \pm 0.3	1.02 \pm 1.39	50.4 \pm 37.8	0.39 \pm 0.3	0.51 \pm 0.19	12.6 \pm 13.4
MeS	1.31 \pm 1.67	0.14 \pm 0.07	19.2 \pm 32.3	0.23 \pm 0.1	0.34 \pm 0.25	14.5 \pm 15.5	0.12 \pm 0.1	1.13 \pm 0.96	22.5 \pm 17.4	0.12 \pm 0.1	0.13 \pm 0.05	4.54 \pm 6.57
EthS	3.65 \pm 7.08	0.39 \pm 0.3	16.01 \pm 25.6	4.94 \pm 9.79	0.63 \pm 0.51	11.1 \pm 12.1	0.33 \pm 0.3	1.19 \pm 1.59	15.8 \pm 12	0.33 \pm 0.3	0.43 \pm 0.44	3.77 \pm 5.45
Eugenol	0.34 \pm 0.26	0.09 \pm 0.07	0.52 \pm 0.6	0.31 \pm 0.29	0.12 \pm 0.15	0.59 \pm 0.62	0.33 \pm 0.26	0.14 \pm 0.08	1.41 \pm 1.05	0.33 \pm 0.26	0.3 \pm 0.33	0.21 \pm 0.25

PAH												
Naphthalene	5.79 ±	0.98 ±	2.57 ±	8.43 ±	8.8 ±	7.7 ±	10.4 ±	13 ± 17.6	1.69 ±	10.4 ±	16.1 ±	0.95 ±
	4.21	0.74	2.53	5.28	11.4	18.9	3.6		1.77	3.6	20.3	1.79
Fluorene	2.62 ±	28.8 ±	38.3 ±	0.46 ±	27.9 ±	17.2 ±	0.25 ±	0.37 ±	5.06 ±	0.25 ±	0.37 ±	2.26 ±
	4.9	37.3	28.2	0.25	33.1	22	0.1	0.15	5.52	0.1	0.13	3.21
Phenanthrene	1.05 ±	0.84 ±	2.77 ±	1.37 ±	0.69 ±	1.44 ±	1.3 ±	1.1 ±	1.29 ±	1.3 ±	1.26 ±	0.82 ±
	0.88	0.63	3.44	0.55	0.51	0.83	0.55	0.68	0.71	0.55	0.38	0.72
Fluoranthene	0.94 ±	0.74 ±	0.86 ±	0.52 ±	1.09 ±	0.69 ±	0.26 ±	0.69 ±	0.31 ±	0.26 ±	0.3 ±	0.25 ±
	1.55	0.53	0.78	0.36	0.95	0.4	0.14	1.03	0.36	0.14	0.1	0.26
Pyrene	1.69 ±	3.02 ±	1.93 ±	0.58 ±	6.11 ±	1.64 ±	0.15 ±	0.61 ±	0.24 ±	0.15 ±	0.23 ±	0.2 ±
	3.37	2.71	2.14	0.37	7.48	1.67	0.05	1.24	0.27	0.05	0.2	0.26
RA												
7ODA	0.24 ±	0.4 ±	0.28 ±	0.19 ±	0.18 ±	0.33 ±	0.18 ±	0.23 ±	0.6 ±	0.18 ±	0.42 ±	0.44 ±
	0.42	0.32	0.19	0.14	0.2	0.32	0.16	0.27	0.73	0.16	0.34	0.12
AA	0.17 ±	0.19 ±	0.09 ±	0.18 ±	0.05 ±	0.05 ±	0.17 ±	0.23 ±	0.06 ±	0.17 ±	0.04 ±	0.06 ±
	0.1	0.27	0.09	0.07	0.01	0.04	0.09	0.37	0.04	0.09	0.02	0.07
DA	8.4 ±	1.37 ±	2.56 ±	2.86 ±	3.46 ±	3.45 ±	2.12 ±	7.07 ±	2.61 ±	2.12 ±	2.26 ±	1.55 ±
	14.9	0.55	1.18	0.9	1.93	2.04	0.36	12.53	1.97	0.36	1.31	0.77
IPA	0.37 ±	1.72 ±	1.19 ±	0.34 ±	1.66 ±	0.62 ±	0.23 ±	0.94 ±	0.55 ±	0.23 ±	1.15 ±	0.17 ±
	0.37	1.45	0.89	0.11	0.9	0.94	0.09	0.24	0.29	0.09	0.32	0.2
PA	2.41 ±	1.21 ±	1.34 ±	1.46 ±	0.97 ±	0.67 ±	0.09 ±	0.75 ±	0.28 ±	0.09 ±	0.03 ±	0.03 ±
	2.87	1.43	1.33	2.93	0.6	0.57	0.09	0.6	0.3	0.09	0.03	0.02
SPA	1.57 ±	1.44 ±	1.52 ±	1.46 ±	1.84 ±	0.55 ±	0.08 ±	0.11 ±	0.11 ±	0.08 ±	0.02 ±	0.03 ±
	3.13	1.46	1.44	2.93	2.61	0.57	0.09	0.12	0.07	0.09	0.01	0.01

Table 3S.4: (Mean \pm SD) urine concentrations (ng/g Crn) of methoxyphenols (MP), polycyclic aromatic hydrocarbons (PAH), and resin acids (RA) in 24-hr baseline and 24-hr post-exposure urine samples (0-6 h, 6-12 h, 12-24 h) excreted by firefighters ($n = 15$) after wood smoke exposure.

	Group B ($n = 5$)	Group H ($n = 5$)	Group O ($n = 5$)	Group B ($n = 5$)	Group H ($n = 5$)	Group O ($n = 5$)	Group B ($n = 5$)	Group H ($n = 5$)	Group O ($n = 5$)	Group B ($n = 5$)	Group H ($n = 5$)	Group O ($n = 5$)
	Baseline			0-6 h			6-12 h			12-24 h		
MP												
Guaiacol	17384 \pm 10762	40395 \pm 24949	16108 \pm 7513	36553 \pm 29853	36201 \pm 13546	31007 \pm 13479	18131 \pm 24720	13991 \pm 8000	31288 \pm 23065	16489 \pm 18909	16095 \pm 12590	24915 \pm 12849.2
MeG	9795 \pm 5354	17320 \pm 12690	9429 \pm 9096	27782 \pm 15866	58273 \pm 31130	49855 \pm 28105	10819 \pm 8993	11179 \pm 3861	23612 \pm 21538	10793 \pm 13005	14424 \pm 12969	12505 \pm 6954.2
EthG	19121 \pm 14810	28449 \pm 20130	18622 \pm 19306	27555 \pm 16010	24355 \pm 13888	56070 \pm 38089	10529 \pm 12040	7415 \pm 2620	22485 \pm 11424	18524 \pm 36437	4497 \pm 2651	9994 \pm 10971.7
PropG	727 \pm 994	1166 \pm 1308	1358 \pm 2044	554 \pm 757	2843 \pm 2605	6525 \pm 5572	625 \pm 1373	4798 \pm 8032	4464 \pm 1835	LOD	55600 \pm 123340	2806 \pm 3790
Syringol	3371 \pm 3398	1755 \pm 3371	3443 \pm 6420	1499 \pm 1259	1883 \pm 2402	3484 \pm 2737	1078 \pm 1588	575 \pm 772	1741 \pm 1634	247 \pm 528	LOD	122 \pm 249
MeS	2982 \pm 4745	1312 \pm 2079	3785 \pm 7051	1669 \pm 1290	1906 \pm 2844	4927 \pm 2611	574 \pm 994	901 \pm 925	2614 \pm 2110	205 \pm 392	271 \pm 221	778 \pm 944.5
EthS	39057 \pm 80624	26046 \pm 42226	5786 \pm 11148	41686 \pm 80922	14248 \pm 12469	5942 \pm 2433	22830 \pm 47489	27668 \pm 39156	6369 \pm 7094	99392 \pm 216110	6929 \pm 6124	6965 \pm 13180.7
PropS	713 \pm 1449	137 \pm 282	1604 \pm 3562	886 \pm 1788	329 \pm 711	1026 \pm 605	73 \pm 138	LOD	510 \pm 645	LOD	121 \pm 247	LOD
Eugenol	159561 \pm 97466	1052402 \pm 1581291	288086 \pm 477704	269026 \pm 270822	432757 \pm 595436	322073 \pm 261184	307789 \pm 411099	703291 \pm 1087857	504524 \pm 245863	163773 \pm 241113	111511 \pm 108937	381247 \pm 331561.9
PAH												
1-OH-Nap	564 \pm 518	107 \pm 215	405 \pm 648	1194 \pm 878	2179 \pm 2432	6673 \pm 5041	380 \pm 515	1023 \pm 606	2050 \pm 854	595 \pm 253	461 \pm 265	1922 \pm 1595
2-OH-Nap	7363 \pm 9513	1586 \pm 556	2922 \pm 2528	9430 \pm 9464	3692 \pm 1608	10228 \pm 7765	8381 \pm 11930	1756 \pm 564	4674 \pm 2831	7310 \pm 8242	1252 \pm 373	5119 \pm 3519
2-OH-Flu	210 \pm 130	115 \pm 100	266 \pm 296	379 \pm 170	280 \pm 181	531 \pm 390	241 \pm 151	194 \pm 203	296 \pm 264	255 \pm 124	192 \pm 213	470 \pm 499.5

3-OH-Flu	63 ± 55	40 ± 66	151 ± 265	92 ± 66	85 ± 69	216 ± 148	83 ± 54	88 ± 72	154 ± 142	90 ± 54	83 ± 59	243 ± 247.6
9-OH-Flu	285 ± 139	281 ± 273	378 ± 424	454 ± 198	763 ± 605	878 ± 532	253 ± 148	195 ± 174	597 ± 233	266 ± 134	207 ± 243	654 ± 496.8
2-OH-Phen	126 ± 258	LOD	89 ± 114	85 ± 71	LOD	183 ± 171	124 ± 254	55 ± 99	96 ± 121	121 ± 102	78 ± 107	LOD
3-OH-Phen	141 ± 114	53 ± 95	111 ± 224	82 ± 69	50 ± 88	193 ± 128	121 ± 103	140 ± 178	102 ± 130	96 ± 119	202 ± 148	422 ± 470.2
4-OH-Phen	38 ± 38	46 ± 78	70 ± 94	84 ± 142	LOD	85 ± 101	53 ± 96	66 ± 87	49 ± 53	49 ± 86	140 ± 183	LOD
RA												
7ODA	2395 ± 1549	6714 ± 8535	2121 ± 1577	1985 ± 1045	7496 ± 7971	2939 ± 1786	4306 ± 2990	4299 ± 4516	2742 ± 2236	3053 ± 2177	3630 ± 4836	3369 ± 1761
AA	LOD	1610 ± 2748	1026 ± 1397	2167 ± 4363	3634 ± 5277	949 ± 1333	908 ± 1177	1358 ± 2094	973 ± 1165	80 ± 154	609 ± 566	639 ± 687.2
DA	5447 ± 2049	12320 ± 8434	5067 ± 3308	4370 ± 2089	16018 ± 11086	11201 ± 13305	9304 ± 5125	13617 ± 3291	12719 ± 7657	6968 ± 5454	11018 ± 6693	14959 ± 8741.9
IPA	615 ± 1104	775 ± 1709	4121 ± 4514	459 ± 1003	2935 ± 3237	6802 ± 6703	2910 ± 5388	4622 ± 3423	12597 ± 11647	LOD	2011 ± 2128	3860 ± 8606.2

Chapter 4: Metabolic Trajectories Following Contrasting Western and Prudent Diets from Food Provisions: Robust Biomarkers of Short-term Changes in Habitual Diet

Nadine Wellington, Meera Shanmuganathan, Russell J. de Souza, Michael A. Zulyniak, Sandi Azab, Jonathon Bloomfield, Alicia Mell, Ritchie Ly, Dipika Desai, Sonia S. Anand, and Philip Britz-McKibbin

Sonia S. Anand, Michael Zulyniak, Russell J. de Souza, and Dipika Desai conceived and supervised the study, including participant recruitment, collection of dietary records, and coordination of biospecimen collection. Nadine Wellington, Meera Shanmuganathan, and Sandi Azab were involved in non-targeted metabolite profiling of urine and plasma samples by MSI-CE-MS, whereas Sandi Azab, Jonathon Bloomfield, Alicia Mell, and Ritchie Ly contributed to targeted analysis of total plasma fatty acids and urinary electrolytes, including sample pretreatment and data acquisition. Nadine Wellington and Philip Britz-McKibbin were involved in data analysis and wrote the manuscript, including statistical analysis, metabolite identification and data interpretation with Sonia S. Anand and Russell J. de Souza contributing to the final version of the manuscript.

Chapter 4: Metabolic Trajectories Following Contrasting Western and Prudent Diets from Food Provisions: Robust Biomarkers of Short-term Changes in Habitual Diet

4.1 Abstract

A large body of evidence has linked unhealthy eating with an alarming increase in obesity and chronic disease worldwide. However, existing methods of assessing dietary intake rely on food frequency questionnaires or dietary records that are prone to bias and selective reporting. Herein, metabolic phenotyping was performed on 42 healthy participants from the Diet and Gene Intervention (DIGEST) pilot study, a parallel two-arm randomized clinical trial that provided complete diets to all participants. Matching urine and plasma specimens were collected at baseline and following 2 weeks of either an assigned Prudent or Western diet. Targeted and nontargeted metabolite profiling was conducted using three analytical platforms, where 80 serum metabolites and 84 creatinine-normalized urinary metabolites were reliably measured ($CV < 30\%$) in the majority of participants ($> 75\%$) after implementing a rigorous data workflow for metabolite authentication with quality control. We classified a panel of metabolites with distinctive trajectories following 2 weeks of food provisions when using complementary univariate and multivariate statistical models. Unknown metabolites associated with contrasting dietary patterns were identified (level 1 or 2) with high resolution MS/MS after spiking with authentic standards. Overall, 3-methylhistidine and proline betaine concentrations increased consistently when participants were assigned a Prudent diet ($q < 0.05$) in both plasma and urine samples with a decrease in the Western diet group. Similarly, creatinine-normalized urinary imidazole propionate, hydroxyphenylacetic acid, dihydroxybenzoic acid, and enterolactone glucuronide, as well as plasma ketoleucine and ketovaline increased with a Prudent diet ($p < 0.05$) after adjustments for age, sex and BMI. In contrast, plasma myristic acid, linoelaidic acid, linoleic acid, α -linoleic acid, pentadecanoic

acid, alanine, proline, carnitine and deoxycarnitine, as well as urinary acesulfame K increased among participants following a Western diet. Most metabolites were also correlated ($r > \pm 0.30$, $p < 0.05$) to changes in average intake of specific nutrients from self-reported diet records reflecting good adherence to food provisions. This study revealed robust biomarkers sensitive to short-term changes in habitual diet for monitoring healthy eating patterns that is needed for new advances in nutritional epidemiology and the design of evidence-based public health policies for chronic disease prevention.

4.2 Introduction

A global epidemic of obesity and chronic non-communicable diseases threaten to reduce life expectancy and impose a severe burden on public health.^{1,2} Diet and lifestyle are two key modifiable determinants of human health of particular importance for risk of cardiovascular disease (CVD), type 2 diabetes and some cancers.³ CVD remains the leading cause of death globally⁴ which has been associated with a Western diet. Contemporary Western diets rich in *trans* fats, processed foods and red meat, including regular consumption of sweetened beverages and high glycemic index foods lacking adequate fibre, have been strongly linked to chronic inflammation and metabolic syndrome⁵ that increasingly impacts the metabolic health across the lifespan.⁶ In contrast, a Prudent eating pattern (*e.g.*, DASH, Mediterranean, Nordic diets etc.) that includes greater intake of fruits & vegetables, lean meats and whole grains reduces blood lipids, improves blood sugar homeostasis and lowers blood pressure.^{7,8} However, there is urgent need for more accurate dietary assessment tools for the design of evidence-based nutritional policies that are effective for chronic disease prevention on a population level.⁹

Nutritional epidemiologists face unique challenges in light of the highly complex chemical composition of foods, whose physiological effects are often confounded by interactions of diet with genes, lifestyle, microbiome and other

environmental exposures.¹⁰ To date, observational studies in nutrition mainly rely on self-reported measures of dietary intake, including methods of recall (*e.g.*, food frequency questionnaires, 24 h dietary recall) or real-time recording (*e.g.*, food diaries), which are prone to bias and selective reporting.¹¹ Alternatively, targeted assays exist for measuring energy expenditure (*e.g.*, doubly-labeled water), as well as specific macronutrients (*e.g.*, protein), electrolytes (*e.g.*, sodium) and micronutrients (*e.g.*, vitamin D) with established reference ranges associated with nutritional status and/or chronic disease risk. However, these methods are not routinely applied in large-scale human studies due to cost barriers while representing only a small fraction of total food exposures.^{12,13} In this context, new advances in high throughput metabolomics offer a holistic approach to measure complex dietary patterns in lieu of specific nutrients based on comprehensive analysis of metabolites in human biofluids, such as urine and plasma.¹⁴ Recent metabolomic studies have identified dietary biomarkers^{15,16} to monitor for dietary adherence, as well as validate or correct standard dietary assessment tools used in nutritional epidemiology.¹⁷⁻²¹ However, few dietary biomarkers are specific to certain foods nor adequately validated as quantitative measures of recent intake or habitual food consumption in well-controlled randomized clinical trials.²²⁻²⁴

Herein, metabolic phenotyping of matching blood and urine specimens were analyzed from healthy participants from the Diet and Gene Intervention (DIGEST) pilot study, which was a randomized controlled trial to explore the short-term effects of a Prudent diet on CVD risk factors where individuals were provided all foods to prepare at home.²⁵ A modest reduction in systolic and diastolic blood pressure and total cholesterol was reported for participants following a Prudent diet for two weeks as compared to a Western diet; however, dietary adherence relied on participant self-reporting, and food preparation methods were not standardized likely contributing to variability in treatment responses.²⁵ In this work, we sought to identify specific metabolic trajectories in plasma and urine that function as

responsive biomarkers reflecting short-term changes in habitual diet, that were measured in free-living individuals outside of a dedicated metabolic ward or hospital stay. These dietary biomarkers not only confirmed good adherence to food provisions but were also associated with healthy eating patterns indicative of a Prudent diet,²⁶ unlike a Western diet that increases overall risk for CVD.²⁷

4.3 Results

Study Design, Baseline Habitual Diet and Metabolomics Workflow. The DIGEST study was a two-arm parallel dietary intervention involving healthy/non-smoking participants recruited from the local community as described elsewhere.²⁵ A CONSORT diagram summarizes eligibility criteria (**Figure 4S.1**), where all participants completed a 7-day prospectively collected diet record and then were randomly allocated to eat a weight-maintaining Prudent or Western diet over 2 weeks. Participants ($n = 42$) with contrasting habitual diets were selected in this unblinded metabolomics study based on availability of matching plasma and urine samples with complete diet records as depicted in **Figure 4.1A**. There were more women (64%) recruited than men, however there were no differences in age (mean age of 47 years ranging from 20 to 69 years), body composition (mean BMI of 27 kg/m² with 26% defined as obese) and average caloric intake (mean of 1940 kcal/day) between assigned diet groups (**Table 4S.1**). Also, no differences in baseline blood lipids, fasting glucose, inflammatory biomarkers and blood pressure were measured between the two treatment arms. In this study, 18 participants were classified as having a Prudent-like diet at baseline (*i.e.*, low Western diet score) who were randomized to a Western diet (referred to as P-W), and 24 participants with a predominate Western diet at baseline were randomized to a Prudent diet (referred

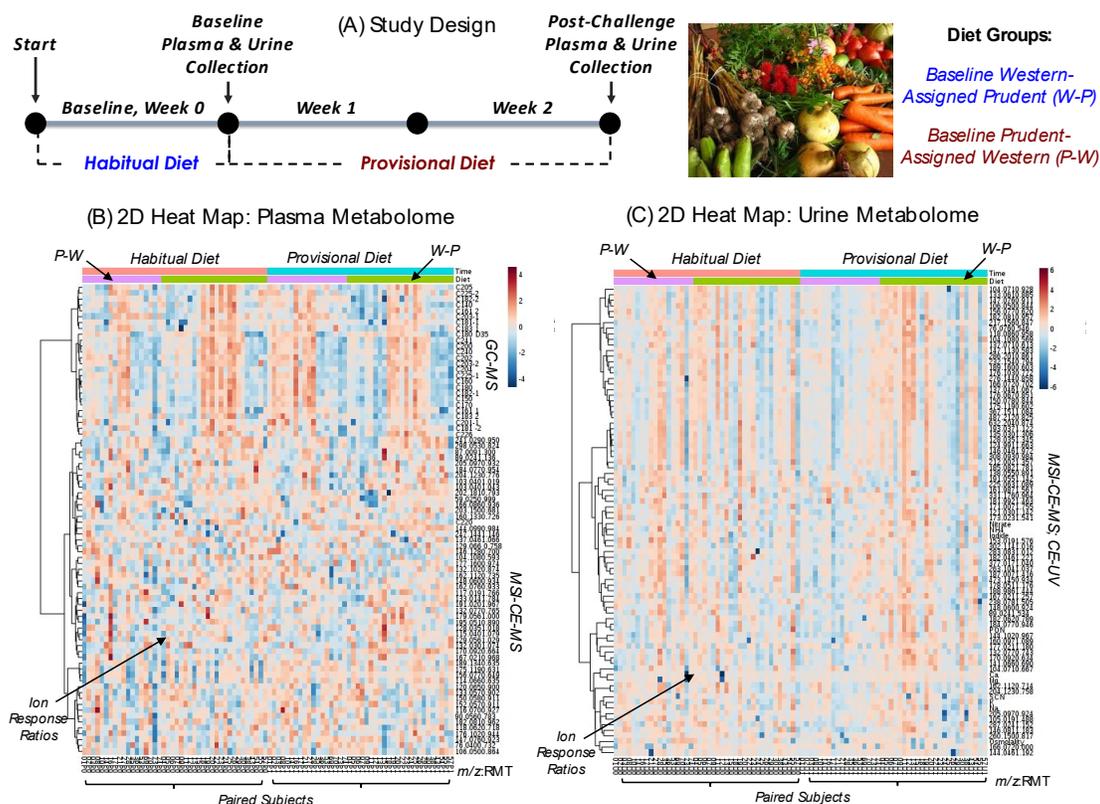


Figure 4.1: (A) Overview of study design in this parallel two-arm dietary intervention study involving participants from DIGEST ($n = 42$) who were assigned a Prudent or Western diet over a 2-week period with matching urine and plasma samples collected at baseline and post-intervention. (B) A 2D heat map with hierarchical cluster analysis (HCA) of the plasma metabolome that were consistently measured in the majority of participants, including non-targeted analysis of polar/ionic metabolites by MSI-CE-MS and total fatty acids by GC-MS. (C) A 2D heat map with hierarchical cluster analysis of the urine metabolome that were consistently measured in majority of participants, including non-targeted analysis of polar/ionic metabolites and targeted electrolytes by CE with indirect UV absorbance. A generalized log transformation and autoscaling was performed on metabolomic datasets together with creatinine normalization for single-spot urine specimens. Participants classified as having a predominate Western diet at baseline who were then assigned a Prudent diet are designated as “W-P” ($n = 24$), whereas “P-W” ($n = 18$) refers to participants had lower Western diet score at baseline, but were assigned a Western diet.

to as W-P). Overall, participants reported excellent adherence to all provided food items.²⁵ **Figure 4.1B** and **C** depict 2D heat maps for matching plasma and urine metabolomes from DIGEST participants ($n = 42$) collected at baseline and following 2 weeks of assigned food provisions. A total of 80 serum metabolites and 84 urinary metabolites were reliably measured ($CV < 30\%$) in the majority of participants ($> 75\%$) when using a validated data workflow for nontargeted metabolite profiling with stringent quality control (QC).²⁸⁻³⁰ A rigorous approach to metabolite authentication was implemented to reject spurious, redundant and background ions that comprise the majority of molecular features detected in ESI-MS³¹ in order to reduce false discoveries in metabolomics.³² Overall, three orthogonal platforms were used to characterize polar/ionic metabolites in plasma and urine samples using multisegment injection-capillary electrophoresis-mass spectrometry (MSI-CE-MS), as well as total (hydrolyzed) plasma fatty acids as their methylester derivatives (FAMES) by GC-MS, and inorganic urinary electrolytes by CE with indirect UV detection (**Figure 4S.2**). Also, 2D scores plots from principal component analysis (PCA) of plasma and creatinine-normalized urine metabolome demonstrated good technical precision from pooled samples used as QCs (median $CV = 4-12\%$) as compared to the biological variance measured in random/single-spot urine (median $CV = 65-78\%$) and fasting plasma (median $CV = 32-53\%$) metabolomes (**Figure 4S.3**). A batch-correction algorithm was also applied to urine metabolome data to minimize signal drift when using MSI-CE-MS,²⁹ where each run comprised a serial injection of six randomized samples together with a pooled QC. Also, control charts for recovery standards provide further evidence of acceptable intermediate precision (mean $CV < 9\%$) with few outliers (**Figure 4S.3**).

A complete list of authenticated metabolites reliably measured in this study (**Table 4S.2**) is annotated by their accurate mass and relative migration time (m/z :RMT) under positive (+) or negative (-) ion detection mode, as well as their

most likely molecular formula and mass error, level of identification, and compound name. Unambiguous identification of metabolites associated with contrasting diets was performed by spiking with authentic standards (if available) in conjunction with high resolution MS/MS, which were compared to reference spectra available in public databases (HMDB, Metlin); otherwise, spectral annotation was guided by *in silico* fragmentation³³ using recommended reporting standards for metabolite identification.³⁴ An overview of this metabolomics workflow is outlined in **Figure 4.2**, which shows the detection of an unknown protonated molecule ($[M-H]^+$) in plasma by MSI-CE-MS, followed by its annotation by high resolution MS and subsequent identification (level 1) as proline betaine (ProBet) using MS/MS after comparison to an authentic standard at an optimal collision energy. Reliable quantification of ProBet using an external calibration curve is also demonstrated with good technical precision ($CV < 15\%$, $n = 20$) as shown in a control chart based on repeated analysis of a QC in every run over the duration of the study.

Changes in Dietary Intake from Food Provisions and Biomarker Classification. Major changes in self-reported dietary patterns among DIGEST participants were evident after 2 weeks as summarized in **Table 4.1**. Although there were no significant changes in BMI or average caloric intake between the two treatment arms, greater palatability and satiety was previously reported for participants assigned to a Prudent diet.²⁵ As expected, the Prudent diet group (W-P) had higher intake of dietary fibre (total, insoluble, soluble), major electrolytes (K, Mg) fruit and/or vegetable, vitamins, poly:sat, protein, and sugar or total carbohydrates, whereas the P-W group had higher intake of fat (total, saturated, and *trans*), sodium and cholesterol. **Figure 4S.4** illustrates the relationship among 20 of the most significant nutrient categories reflecting contrasting diets when using PCA along with a hierarchical cluster analysis (HCA) and 2D heat map. There was

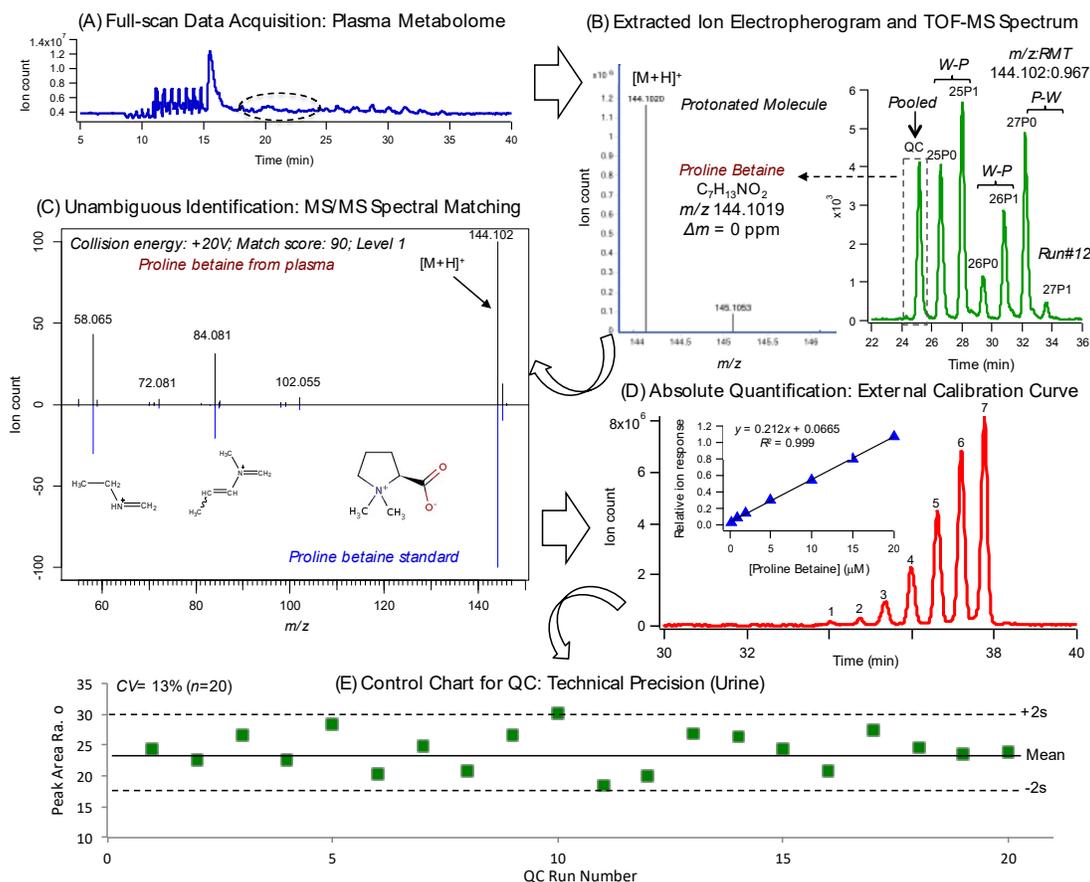


Figure 4.2: (A) Metabolomics data workflow for the identification and quantification of biomarkers of a provisional Prudent diet (e.g., proline betaine annotated based on its m/z :RMT) when using full-scan data acquisition. (B) Multiplexed separations by MSI-CE-MS based on serial injection of seven plasma filtrate (or diluted urine) samples within a single run, including paired samples from each DIGEST participant (i.e., baseline/post-treatment) together with a pooled sample as QC for assessing technical precision and long-term signal drift. High resolution MS under positive ion mode detection allows for determination of most likely molecular formula for unknown cation (i.e., protonated molecular ion), whereas (C) MS/MS spectra is used for its structural elucidation when compared with an authentic standard. (D) Quantification for metabolites is then performed by external calibration when using an internal standard (Cl-Tyr) for data normalization by MSI-CE-MS. (E) A control chart for ProBet from pooled urine samples as QC analyzed in random positions in every run demonstrates acceptable technical precision over 3 days.

Table 4.1: Major changes in dietary patterns after a 2-week assigned Prudent and Western diet relative to baseline habitual diet of DIGEST participants ($n = 42$) based on self-reported diet records.

Diet Category ^a	W-P, $n = 24$	P-W, $n = 18$	p for comparison/outcome
Δ Insoluble fibre intake (g/2000 kcal/day)	(14.0 ± 5.3)	(-5.0 ± 3.5)	$p = 1.4 \text{ E-}15$; Greater insol. fibre intake in Prudent arm
Δ Mg intake (mg/2000 kcal/day)	(189 ± 89)	(-134 ± 70)	$p = 3.5 \text{ E-}15$; Greater Mg intake in Prudent arm
Δ Fruit & veggie intake (servings/2000 kcal/day)	(3.6 ± 1.4)	(-1.8 ± 1.3)	$p = 7.3 \text{ E-}15$; Greater intake in fruit/veggie in Prudent arm
Δ Total fibre intake (g/2000 kcal/day)	(16.6 ± 8.4)	(-13.4 ± 8.1)	$p = 5.2 \text{ E-}14$; Greater total fibre intake in Prudent arm
Δ Energy from sat. fat (%)	(-5.4 ± 3.2)	(4.6 ± 2.4)	$p = 1.8 \text{ E-}13$; Greater intake of sat. fat in Western arm
Δ Vegetable intake (cup eq./2000 kcal/day)	(1.8 ± 0.80)	(-0.91 ± 0.92)	$p = 2.4 \text{ E-}12$; Greater veggie intake in Prudent arm
Δ K intake (mg/2000 kcal/day)	(1338 ± 617)	(-854 ± 667)	$p = 2.5 \text{ E-}13$; Greater K intake in Prudent arm
Δ Vitamin E (mg/2000 kcal/day)	(7.7 ± 5.3)	(-7.0 ± 4.0)	$p = 5.1 \text{ E-}12$; Higher intake of vit. E in Prudent arm
Δ Poly:sat fatty acid (ratio)	(0.47 ± 0.21)	(-0.14 ± 0.18)	$p = 8.2 \text{ E-}12$; Greater intake of poly:sat in Prudent arm
Δ Vitamin C (mg/2000 kcal/day)	(149 ± 69)	(-40 ± 54)	$p = 1.2 \text{ E-}11$; Higher intake of vit. C in Prudent arm
Δ Soluble fibre intake (g/2000 kcal/day)	(3.9 ± 2.1)	(-1.5 ± 1.5)	$p = 2.3 \text{ E-}11$; Greater total fibre intake in Prudent arm
Δ Fruit intake (cup eq./2000 kcal/day)	(1.79 ± 0.93)	(-0.92 ± 0.99)	$p = 5.9 \text{ E-}11$; Greater fruit intake in Prudent arm
Δ Energy from fat (%)	(-7.5 ± 5.6)	(5.6 ± 5.6)	$p = 9.0 \text{ E-}10$; Greater intake of total fat in western arm
Δ Na intake (mg/2000 kcal/day)	(-694 ± 590)	(754 ± 658)	$p = 6.4 \text{ E-}9$; Greater Na intake in western arm
Δ Vitamin A (μg /2000 kcal/day)	(12973 ± 56344)	(-7847 ± 14060)	$p = 1.4 \text{ E-}7$; Higher intake of vit. E in Prudent arm
Δ Energy from sugar (%)	(8.9 ± 5.4)	(-1.5 ± 5.8)	$p = 7.3 \text{ E-}7$; Higher sugar intake in Prudent arm
Δ Energy from protein (%)	(1.9 ± 3.6)	(-3.2 ± 2.7)	$p = 1.5 \text{ E-}5$; Greater intake of protein in Prudent arm
Δ Energy from carbohydrates (%)	(8.5 ± 7.8)	(-0.35 ± 5.7)	$p = 2.9 \text{ E-}4$; Greater intake of total carbs in Prudent arm
Δ Cholesterol ^b (mg/2000 kcal/day)	(-101 ± 140)	(54 ± 110)	$p = 4.8 \text{ E-}4$; Greater intake of cholesterol in Western arm
Δ Energy from trans fat (%)	(-0.26 ± 0.55)	(0.27 ± 0.23)	$p = 6.4 \text{ E-}4$; Greater intake of trans fats in Western arm

^a Mean differences (Δ) in self-reported dietary patterns were evaluated from food records collected twice over a 2-week period at clinical visits as compared to the baseline habitual diet of each participant.

^b There were no significant changes in measured total, LDL and HDL cholesterol based on standard clinical blood measurements when using a two-tailed student's t -test with equal variance.

strong co-linearity ($r > \pm 0.70$) among most nutrient categories with two distinctive clusters reflecting opposing Prudent and Western eating patterns assigned to DIGEST participants.

Volcano plots (**Figure 4S.5**) were initially used to evaluate changes in the metabolic phenotype of participants using minimum cut-off thresholds (*i.e.*, mean fold-change or $FC > 1.3$; $p < 0.05$). Overall, contrasting diets generated pronounced changes in a wide range of plasma and urinary metabolites that was largely absent for the same participants given modest differences in their baseline habitual diets (**Table 4S.1**). For instance, 10 plasma and 16 urinary metabolites were differentially expressed in W-P as compared to P-W diet groups, including four metabolites satisfying a Benjamini-Hochberg/FDR adjustment ($q < 0.05$), including ProBet, 3-methylhistidine (Me-His) and two unknown urinary metabolites subsequently identified (level 2) as hydroxypipicolinic acid (OH-PCA) and imidazole propionic acid (ImPA). The identification and quantification of Me-His was confirmed in both plasma and urine (**Figure 4S.6**), whereas several unknown urinary metabolites were putatively identified (level 1 or 2) based on their characteristic MS/MS spectra, such as OH-PCA (**Figure 4S.7**) and acesulfame K (ASK; **Figure 4S.8**). Similarly, targeted analysis of FAMES from hydrolyzed plasma extracts using GC-MS (**Figure 4S.9**) allows for resolution of low abundance *trans* isomers (linoelaidic acid, C18:2n-6*trans*) and saturated fatty acids (myristic acid, C14:0) from abundant dietary fatty acids (linoleic acid, C18:2n-6*cis*). As expected, several circulating fatty acids (**Figure 4S.5**) were consistently elevated following a Western diet due to higher average consumption of total fats as compared to a Prudent diet.

Biomarkers of Contrasting Diets and Their Correlation to Diet Records.

Complementary statistical methods that take advantage of the repeated-measures study design were used to classify metabolites responses to contrasting dietary patterns. A paired orthogonal partial least-squares–discriminant analysis (OPLS-

DA) model was used to rank metabolites in plasma (**Figure 4.3A**) and urine (**Figure 4.3B**) that were modulated by assigned diets relative to each participant's baseline habitual diet (*i.e.*, ion response ratio). Both OPLS-DA models demonstrated good accuracy ($R^2 > 0.840$) with adequate robustness ($Q^2 > 0.200$) after permutation testing ($p < 0.05$, $n = 1000$). S-plots confirmed that ProBet and Me-His were consistently elevated following a Prudent diet (W-P) in both plasma and urine samples, whereas total plasma C14:0 and C18:2n-6*cis* had the most significant increase following an assigned Western diet (P-W). Additionally, top-ranked creatinine-normalized urinary metabolites excreted at higher levels following a Prudent diet included ImPA, OH-PCA, dihydroxybenzoic acid (DHBA), enterolactone glucuronide (Ent-G), nitrate and an unknown cation (m/z 217.156, $[M-H]^+$) tentatively identified as a dipeptide, valinyl-valine (Val-Val), whereas ASK was only modestly increased ($p = 0.0686$) following a Western diet. Additionally, excellent discrimination among DIGEST participants following a Prudent or Western diet was achieved when using top-ranked single or ratiometric biomarkers from a receiver operating characteristic (ROC) curve ($AUC > 0.820$; $p < 1.0 \text{ E-}5$) for plasma and creatinine-normalized urine samples (**Figure 4S.10**). For instance, plasma ProBet and the ratio of Me-His/C18:3n-6*trans* demonstrated good sensitivity and specificity ($\approx 80\text{-}90\%$) for differentiating DIGEST participants based on their assigned diets similar to urinary OH-PCA and the ratio of OH-PCA/Na. A multivariate empirical Bayes analysis of variance (MEBA)³⁵ was also used to characterize time-dependent metabolite profiles related to assigned diets after two weeks of food provisions. In this case, metabolic trajectories with distinctive time-course profiles following a Prudent or Western diet were ranked based on their Hotelling's T^2 distribution as shown for plasma (**Figure 4S.11**) and urine (**Figure 4S.12**), which were consistent with metabolites identified as dietary biomarkers from volcano plots, ROC curves and OPLS-DA models.

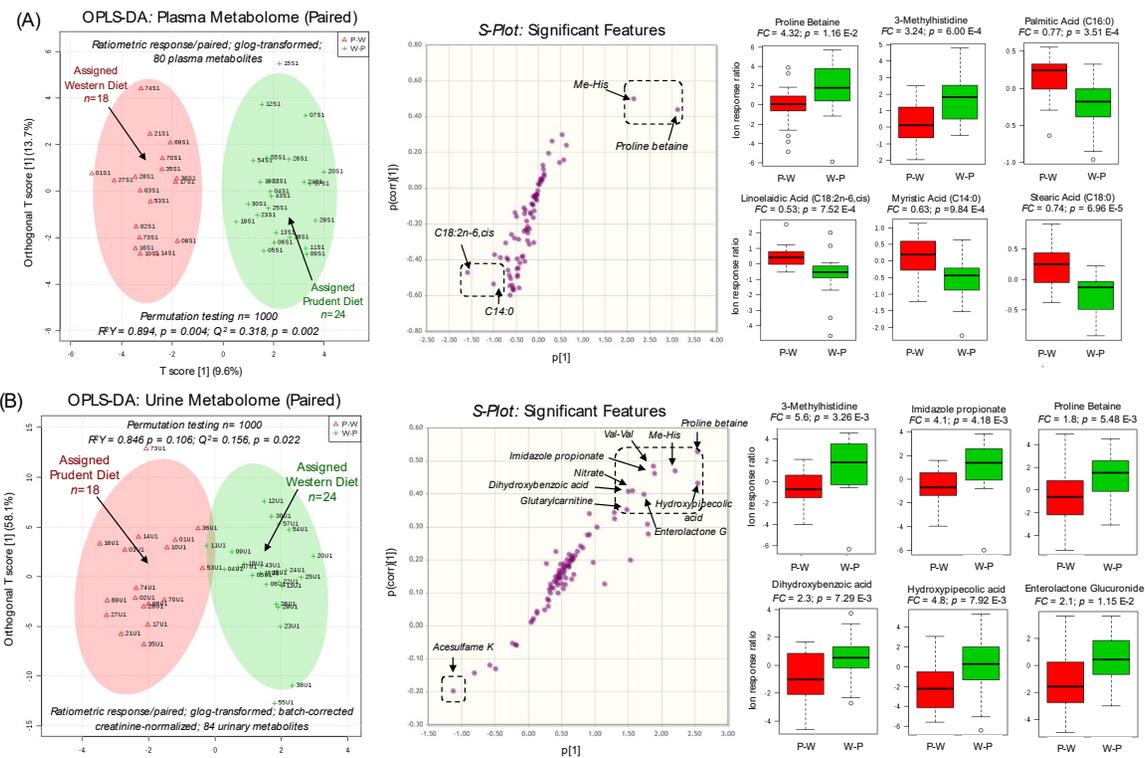


Figure 4.3: Paired supervised multivariate data analysis of (A) plasma and (B) creatinine-normalized urine metabolomic data based on orthogonal partial least-squares-discriminant analysis (OPLS-DA) using the ratio of ion responses or concentrations for metabolites following 2 weeks of food provisions to their baseline habitual diet values. 2D scores plot highlight differences in the overall metabolic phenotype from matching biofluids collected from DIGEST participants assigned to a Prudent (W-P) as compared to a Western (P-W) diet based on a sub-set of metabolites identified from S-plots, as well as univariate statistical analysis as shown in box-whisker plots for top-ranked metabolites significantly different between the treatment arms ($p < 0.05$).

A mixed ANOVA model, and a partial Pearson correlation analysis to self-reported diet records after adjustment for sex, age and BMI were next applied to further validate the relevance of dietary biomarkers identified from multivariate statistical models. **Table 4.2** highlights that ProBet and Me-His were the most robust plasma metabolites associated with a Prudent diet that satisfied several statistical parameters (T^2 , F -value, r , adjusted p -value). For instance, ProBet was positively associated ($r \approx 0.520$, $p = 0.001$) with self-reported intake of fruit (cup eq./2000 kcal), vitamin C (mg/2000 kcal) and fruit & vegetable servings (servings/2000 kcal), as well as negatively associated with total fat intake ($r < -0.530$, $p < 0.001$), including *trans* and saturated fat (% energy). Me-His had strong positive correlations ($r = 0.530-0.570$, $p < 0.001$) with protein (%energy), insoluble fibre (g/2000 kcal), electrolytes (Mg, K; mg/2000 kcal), as well as fruit, and fruit & vegetable intake reflecting a Prudent diet. Other plasma metabolites classified as dietary biomarkers of contrasting diets in this study included two carnitines (*e.g.*, carnitine, C0; deoxycarnitine, dC0), two amino acids (*e.g.*, proline, Pro; alanine, Ala), three ketone bodies/intermediates (*e.g.*, ketoleucine; kLeu; ketovaline, kVal; 3-hydroxybutyric acid, OH-BA), and several long-chain fatty acids (*e.g.*, C14:0, C15:0, C18:2n-6*trans*, C18:3n-6*cis*, C18:2n-6*cis*). Overall, all total hydrolyzed fatty acids were positively correlated to a Western diet with a higher average intake of fats (*trans* fats, saturated fats) and a corresponding lower intake of fruit & vegetable, poly:sat and micronutrients (vitamins A, C and E). Similar outcomes were also measured for plasma carnitines and amino acids, which were positively correlated to a Western diet. In contrast, metabolic intermediates of branched-chain amino acids and energy metabolism, namely plasma kLeu, kVal and OH-BA, were positively associated with a Prudent diet, including higher average intake of protein, fibre, fruit & veggie, poly:sat, and vitamins. **Table 4.2** summarizes 14 plasma metabolites that function as robust biomarkers of contrasting diets since they satisfied at least two of the three statistical models ($p < 0.05$) following adjustment

Table 4.2: Top-ranked plasma metabolites associated with a 2 week Prudent or Western provisional diet on healthy participants ($n = 42$) when using time series MEBA, mixed ANOVA and a partial correlation analysis.

Metabolite/ID	Identifier/MSI	T^{2a}	F -value ^b	p -value ^b	r^c	p -value ^c	Food record ^d
Proline betaine (ProBet) HMDB04827	144.102:0.984 (+) MSI-CE-MS C ₇ H ₁₃ NO ₂ Level 1	24.6	8.7	0.007	-0.601	< 0.001	Change %fat
					-0.544	< 0.001	<i>trans</i> fat %energy
					-0.528	0.001	Sat fat %energy
					0.528	0.001	Fruit/Vitamin C
					0.518	0.001	Fruit & Veggie
3-Methylhistidine (MeHis) HMDB00479	170.092:0.664 (+) MSI-CE-MS C ₇ H ₁₁ N ₂ O ₃ Level 1	24.9	14.0	0.001	0.573	< 0.001	Magnesium
					0.561	< 0.001	Protein %energy
					0.553	< 0.001	Insoluble Fibre
					0.546	< 0.001	Potassium
					0.534	0.001	Fibre/Fruit & Veggie
Proline (Pro) HMDB00162	116.070:0.927 (+) MSI-CE-MS C ₅ H ₉ NO ₂ Level 1	14.6	5.9	0.020	0.495	0.002	<i>trans</i> fat %energy
					-0.412	0.010	Fruit & Veggie (serv)
					-0.378	0.019	Veggie
					-0.373	0.021	Fruit & Veggie
					-0.362	0.026	Fruit
Carnitine (C0) HMDB00062	162.112:0.735 (+) MSI-CE-MS C ₇ H ₁₅ NO ₃ Level 1	12.2	8.9	0.005	-0.464	0.003	Poly:Sat
					0.426	0.008	<i>trans</i> fat %energy
					-0.404	0.012	Fruit & Veggie
					-0.386	0.017	Vitamin E
					-0.368	0.023	Vitamin C
Deoxycarnitine or γ -Butyrobetaine (dC0) HMDB01161	146.128:0.700 (+) MSI-CE-MS C ₇ H ₁₆ NO ₂ Level 2	11.9	7.9	0.008	0.367	0.024	Change %fat
					0.366	0.024	Cholesterol
					-0.352	0.030	Magnesium
					0.340	0.037	Sodium
					-0.336	0.039	Poly:Sat

Linoelaidic acid (C18:2n-6 <i>trans</i>) HMDB06270	294/67.1:15.289 GC-MS C ₁₈ H ₃₂ O ₂ Level 2	10.3	21.5	< 0.001	-0.579 -0.555 -0.486 0.485 0.464	< 0.001 < 0.001 0.002 0.002 0.003	Poly:Sat Fruit & Veggie/Vit. E Vitamin C Sat fat %energy <i>trans</i> fat %energy
Pentadecanoic acid (C15:0) HMDB000673	294/67.1:14.171 GC-MS C ₁₈ H ₃₂ O ₂ Level 2	9.9	16.8	< 0.001	-0.471 0.408 -0.403 -0.379 0.379	0.003 0.011 0.012 0.019 0.019	Poly:Sat Change %fat Fruit & Veggie Vitamin A Change %sat fat
Alanine (Ala) HMDB00161	90.056:0.783 (+) MSI-CE-MS C ₃ H ₇ NO ₂ Level 1	9.6	6.2	0.018	0.452 0.439 0.428 -0.395 0.386	0.004 0.006 0.007 0.014 0.017	Change %sat fat Change %fat <i>trans</i> fat %energy Protein %energy Sat fat %energy
Ketoleucine or 4-Methyl-2- oxopentanoic acid (kLeu) HMDB00695	129.056:1.209 (-) MSI-CE-MS C ₆ H ₁₀ O ₃ Level 2	7.7	4.4	0.043	0.493 -0.459 0.456 0.453 0.452	0.002 0.004 0.004 0.004 0.004	Fruit & Veggie Sat fat %energy Fruit Poly:Sat Protein %energy/ Vitamin C/E
3-Hydroxybutyric acid (OH-BA) HMDB00357	103.040:1.043 (-) MSI-CE-MS C ₄ H ₈ O ₃ Level 1	7.6	2.9	0.097	0.437 -0.429 0.425 0.419 0.415	0.006 0.007 0.008 0.009 0.01	Fruit Sat/trans fat %energy Poly:Sat Vit A Fruit & Veggie

α -Linoleic acid (C18:3n-6 <i>cis</i>) HMDB001388	292/79.1:15.096	7.0	11.6	0.002	-0.441	0.006	Poly:Sat
	GC-MS				-0.397	0.013	Vitamin A
	C ₁₈ H ₃₀ O ₂				-0.391	0.015	Fruit & Veggie
	Level 2				0.391	0.015	<i>trans</i> fat %energy
					-0.387	0.016	Vitamin E
Ketovaline or α -Isovaleric acid (kVal) HMDB00019	115.040:1.079 (-)	6.3	2.4	0.125	0.489	0.002	Protein %energy
	MSI-CE-MS				0.472	0.003	Fibre (kcal)
	C ₅ H ₈ O ₃				0.466	0.003	Fruit & Veggie
	Level 2				0.458	0.004	Vitamin E
					0.451	0.004	Poly:Sat
Myristic acid (14:0) HMDB00826	242/74.1:10.336	5.0	15.2	< 0.001	-0.535	0.001	Poly:Sat
	GC-MS				-0.512	0.001	Fruit & Veggie
	C ₁₅ H ₃₀ O ₂				0.503	0.001	Change %fat
	Level 1				0.465	0.003	Change %sat. fat
					-0.463	0.009	Vitamin A
Linoleic acid (C18:2n-6 <i>cis</i>) HMDB000673	294/67.1:14.171	2.6	16.4	< 0.001	-0.438	0.006	Poly:Sat
	GC-MS				0.420	0.009	Change %fat
	C ₁₈ H ₃₂ O ₂				0.412	0.005	Change %sat. fat
	Level 2				-0.382	0.018	Fruit & Veggie
					-0.370	0.022	Vitamin A

a Hotelling's *T*-squared distribution using MEBA on glog-transformed metabolomic time series data.

b Mixed ANOVA model derived from within-subject (diet x time interaction, $p < 0.05$) contrasts when adjusted for sex, age and BMI.

c Partial Pearson correlation of urinary metabolites to food records with listwise deletion adjusted for sex, age and BMI, where $r > \pm 0.30$ and $p < 0.05$.

d Top-five categories from food records significantly correlated to urinary metabolites following provisional diets.

for covariates between groups while also having a correlation ($r > \pm 0.3$, $p < 0.05$) with at least two nutrient categories from self-reported diet records. An analogous strategy was also used to identify 8 creatinine-normalized urinary metabolites significantly associated with contrasting diets (**Table 4.3**). Urinary Me-His and ProBet were among the top-ranked metabolites sensitive to short-term changes in habitual diet with strong positive associations with healthful eating patterns indicative of a Prudent diet. Additionally, several other urinary metabolites were also associated with a Prudent eating pattern, including OH-PCA and ImPA. Furthermore, two plant-derived phenolic metabolites in urine, namely Ent-G and DHBA were also correlated to healthy eating patterns with a greater intake of fruit and/or vegetable and micronutrients, and a lower intake of total fat. However, creatinine-normalized Val-Val and DMG in urine were weakly correlated with only 2 nutrient categories ($p \approx 0.05$) from self-reported diet records. Interestingly, urinary ASK, nitrate and an unidentified cation (m/z :RMT, 276.144:0.858, $[M-H]^+$) were not correlated to any major nutrient category despite showing treatment responses to contrasting diets.

Metabolic Trajectories from Food Provisions and Metabolite Correlation Analysis. Representative metabolic trajectories are depicted for top-ranked biomarkers of contrasting diets that were measured in plasma (**Figure 4S.11**) and urine specimens (**Figure 4S.12**). In all cases, metabolic phenotype changes were evident following 2 weeks of food provisions with the exception of urinary DHBA, which was the only compound different between assigned diet groups at baseline ($p = 8.03 \text{ E-}3$). The majority of dietary biomarkers underwent an increase in response for participants following a Prudent diet except for several circulating fatty acids, two amino acids (Pro, Ala) and two carnitines (C0, dC0) in plasma, which increased following a Western diet. Metabolic trajectory plots also highlight considerable between-subject variances to assigned diets while also identifying

Table 4.3: Top-ranked creatinine-normalized metabolites associated with a 2 week Prudent or Western provisional diet on healthy participants ($n = 42$) when using time series MEBA, mixed ANOVA and a partial correlation analysis.

Metabolite/ID	Identifier/MSI	T^{2a}	F -test ^b	p -value ^b	r^c	p -value ^c	Food record ^c
3-Methylhistidine (MeHis) HMDB00479	170.092:0.664 (+) MSI-CE-MS C ₇ H ₁₁ N ₂ O ₃ Level 1	17.9	7.8	0.008	0.524	0.001	Fibre (kcal)
					0.517	0.001	Fruit & Veggie
					0.457	0.004	Vitamin E
					-0.432	0.007	trans fat %energy
					0.431	0.007	Protein %energy
5-Hydroxypipelicolic acid (OH-PCA)* HMDB0029246	146.081:1.180 (+) MSI-CE-MS C ₆ H ₁₁ NO ₃ Level 2	16.3	1.1	0.293	-0.468	0.003	Change fat
					0.397	0.013	Fibre (kcal)
					0.390	0.016	Fruit & Veggie
					0.381	0.018	Vitamin E
					0.374	0.021	Poly:Sat
Imidazole propionic acid (ImPA) HMDB02271	141.066:0.690 (+) MSI-CE-MS C ₆ H ₈ N ₂ O ₂ Level 2	16.1	10.8	0.002	0.515	0.001	Fibre (kcal)
					0.511	0.001	Fruit & Veggie
					0.471	0.003	Protein %energy
					0.463	0.003	Vitamin E
					0.444	0.005	Poly:Sat
Proline betaine (ProBet) HMDB04827	144.099:0.984 (+) MSI-CE-MS C ₇ H ₁₃ NO ₂ Level 1	15.5	10.8	0.002	0.487	0.002	Poly:Sat
					-0.487	0.002	trans fat %energy
					0.482	0.002	Fibre (kcal)
					0.480	0.002	Fruit & Veggie/Vit. E
					0.469	0.003	Fibre (insoluble)
Valinyl-valine (Val-Val) HMDB0029140	217.156:0.847 (+) MSI-CE-MS C ₁₀ H ₂₀ N ₂ O ₃ Level 3	10.9	3.8	0.060	0.320	0.050	Poly:Sat
					0.320	0.050	Vitamin E

Enterolactone glucuronide (Ent-G) HMDB ---	473.145:0.934 (-)	8.0	7.3	0.010	-0.434	0.006	Fat (kcal)
	MSI-CE-MS				0.387	0.016	Vitamin C
	C ₂₄ H ₂₅ O ₁₀				0.340	0.037	Fruit (cup eq.)
	Level 2				0.332	0.042	Fruit & Veggie
					0.316	0.054	Veggie (cup eq.)
Dihydroxybenzoic acid (DHBA) or protocatachuic acid* HMDB0001856	153.019:1.576 (-)	7.9	10.3	0.003	-0.403	0.012	Fat (kcal)
	MSI-CE-MS				0.383	0.018	Sugar %energy
	C ₇ H ₆ O ₄				0.355	0.029	Vitamin C
	Level 2				0.324	0.047	Veggie (cup eq.)
					0.310	0.058	Fruit & Veggie
Dimethylglycine (DMG) HMDB0000092	104.108:0.569 (+)	2.9	3.6	0.065	0.356	0.028	Fruit & Veggie (serv.)
	MSI-CE-MS				0.322	0.049	Fibre (kcal)
	C ₄ H ₉ NO ₂						
	Level 1						

a Hotelling's T-squared distribution using MEBA on glog-transformed metabolomic time series data.

b Mixed ANOVA model derived from within-subject (diet x time interaction, $p < 0.05$) contrasts when adjusted for sex, age and BMI.

c Partial Pearson correlation of urinary metabolites to food records with listwise deletion adjusted for sex, age and BMI, where $r > \pm 0.30$ and $p < 0.05$.

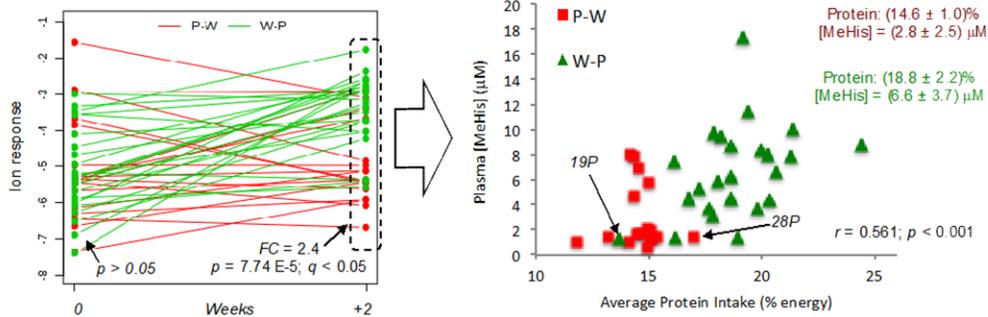
d Top five categories from food records significantly correlated to urinary metabolites following provisional diet.

outliers due to potential dietary non-adherence and/or inaccurate self-reporting. **Figure 4.4** illustrates four metabolic trajectory plots for ProBet and Me-His as they were among the most sensitive biomarkers responsive to contrasting diets measured consistently in both plasma and urine samples. Also, scatter plots show the quantitative relationship between Me-His and ProBet concentrations in plasma as compared to their excreted concentrations in urine with self-reported average intake of protein (% energy) and fruit servings (servings/2000 kcal) over 2 weeks, respectively. For example, there was a 2.4-fold increase in mean plasma Me-His concentration following 2 weeks of food provisions that corresponded to a 28% greater intake of dietary protein when comparing Prudent (W-P, $n = 24$) and Western (P-W, $n = 18$) diet groups. Similar results were also evident when comparing creatinine-normalized concentrations of Me-His in urine, which generated a 4.8-fold higher mean concentration in Prudent relative to Western diet treatment arm. Overall, there was excellent correlation between Me-His concentrations and self-reported dietary protein intake ($r = 0.430$ to 0.560) with few exceptions, such as one participant (W-P, #19) who had consistently low Me-His concentrations in both biofluids consistent with self-reported protein intake that was characteristic of the Western diet group (P-W) indicative of dietary non-adherence. In contrast, a second participant (P-W, #28) had higher than average Me-His concentrations in both plasma and urine samples despite their low self-reported protein intake from diet records suggestive of diet record bias.

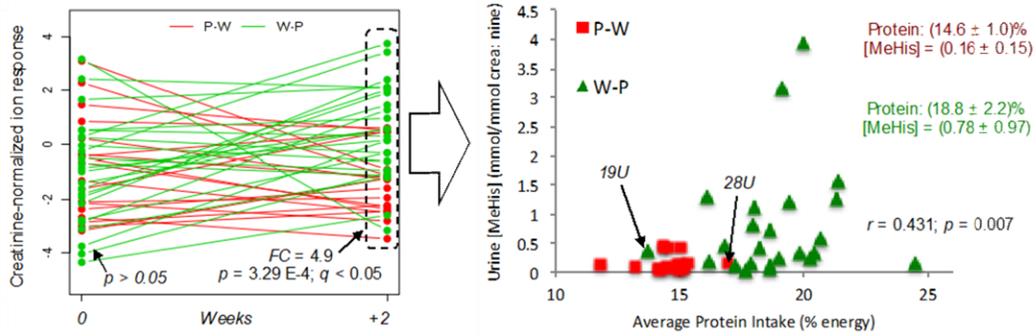
Figure 4.4 also depicts metabolic trajectories for plasma and urinary ProBet concentrations after 2 weeks of food provisions to DIGEST participants as compared to their baseline habitual diet along with scatter plots depicting their correlation ($r = 0.430$ to 0.530) to daily fruit servings. Similar to Me-His, the same participant (W-P, #19) had lower ProBet concentrations in both plasma and urine with diet records reflecting a Western diet low in fresh fruit intake despite being assigned a Prudent diet. Additionally, 3 participants had higher than expected

ProBet concentrations in circulation (P-W, #28, 35, 36) inconsistent with self-reported diet records; interestingly, ProBet concentrations for these same participants were far less elevated in urine likely caused by differences in the detection time window when analyzing these two biofluids for exogenous dietary biomarkers of recent food intake, such as ProBet. Overall, there was a strong linear correlation between Me-His ($r = 0.638$) and ProBet ($r = 0.547$) concentrations measured from matching plasma and urine samples (**Figure 4S.13**) collected at baseline and following assigned diets ($n = 84$). Additionally, 2D heat maps and correlation matrices for top-ranked plasma (14) and urinary (11) metabolites provide insights into their underlying biochemical relationships (**Figure 4S.14**). As expected, urinary imidazole metabolites derived from histidine, Me-His and ImPA ($r = 0.956$), plasma saturated fatty acids, C14:0 and C15:0 ($r = 0.873$), plasma branched-chain amino acid intermediates, kLeu and kVal ($r = 0.705$), as well as plant-derived phenol metabolites in urine, DHBA and Ent-G ($r = 0.662$) were among a group of highly co-linear metabolites correlated to similar nutrient categories from diet records (**Table 4.2**; **Table 4.3**).

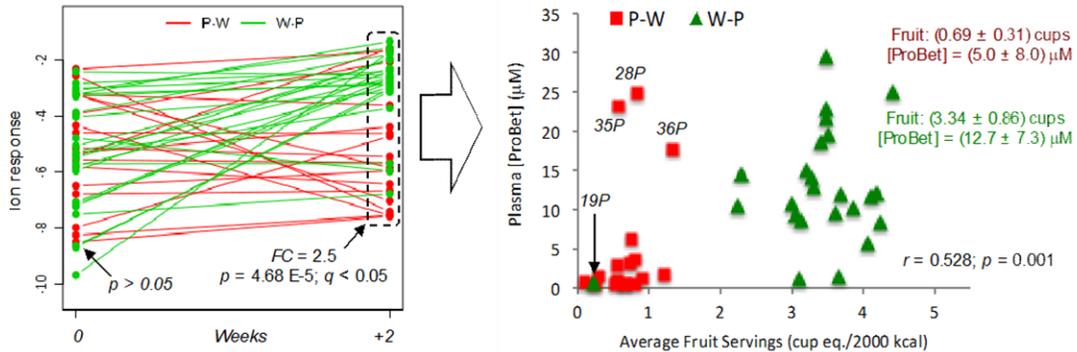
(A) Plasma Methylhistidine Trajectory and Correlation to Self-Reported Average Protein Intake



(B) Urinary Methylhistidine Trajectory and Correlation to Self-Reported Average Protein Intake



(C) Plasma Proline Betaine Trajectory and Correlation to Self-Reported Average Fruit Intake



(D) Urinary Proline Betaine Trajectory and Correlation to Self-Reported Average Protein Intake

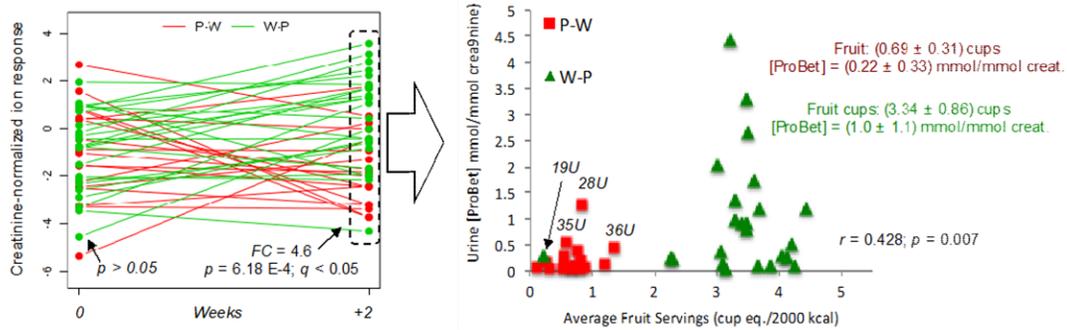


Figure 4.4: *Metabolic trajectories for two distinctive biomarkers measured consistently in both plasma and urine specimens that increase significantly following a provisional Prudent diet (W-P) as compared to an assigned Western diet (P-W), namely Me-His and ProBet. Both metabolites were not different at baseline but undergo notable changes after two weeks of food provisions ($q < 0.05$, FDR) with concentrations moderately correlated ($r > 0.400$) to self-reported diet involving health-promoting foods from a Prudent diet. Overall, good dietary adherence was demonstrated for the majority of DIGEST participants with few exceptions (labeled on plots) who had metabolite phenotypes inconsistent with their assigned diet group.*

4.4 Discussion

Accurate assessment tools of complex dietary patterns are needed to promote human health since sub-optimal diet is responsible for about 20% of preventable deaths from non-communicable diseases worldwide.³⁶ However, few validated biomarkers exist for routine monitoring of habitual diet,³⁷ such as omega-3 fatty acids³⁸ and water insoluble fibre.³⁹ In this work, a panel of metabolites from plasma and urine was demonstrated to respond to short-term dietary changes when applying a cross-platform metabolomics approach with stringent QC (**Figure 4.1**; **Figure 4.2**) and a rigorous data workflow for metabolite authentication (**Figure 4S.2**; **Figure 4S.3**).²⁸⁻³⁰ Since all DIGEST participants had poor Prudent diet eating habits at baseline (**Table 4.1**), we hypothesized that assigning a Prudent diet (W-P) from food provisions would likely induce a more pronounced metabolic phenotype change than a Western diet (P-W); indeed, several top-ranked metabolites ($q < 0.05$, FDR) measured in plasma and urine were largely positively associated with a Prudent diet as shown in volcano plots (**Figure 4S.5**). Unlike controlled feeding studies within a laboratory setting, DIGEST participants were provided cooking suggestions with meal plans by a dietician that still allowed for flexibility in food preparations.²⁵ In this study, short-term dietary changes were found to impact the intake of 20 specific nutrient categories from self-reported diet records (**Table 4.2**;

Table 4S.1; Figure 4S.4). For instance, a Prudent diet was consistent with a higher consumption of dietary fibre, fruit and/or vegetables, electrolytes and vitamins, but with lower intake of dietary fat, sodium and cholesterol in contrast to a Western diet. To the best of our knowledge, this is the first metabolomics study to investigate the impact of contrasting diets using food provisions. As dietary adherence, potential misreporting and variations in food preparations represent uncontrolled variables, we aimed to identify metabolites from plasma and urine that can serve as robust biomarkers of habitual diet applicable to a free-living population.

ProBet (**Figure 4.2**) and Me-His (**Figure 4S.6**) were among the most significant metabolites ($q < 0.05$, FDR) associated with a Prudent diet, an eating pattern that promotes good health while contributing to chronic disease prevention.^{40,41} In this case, ProBet and Me-His displayed opposing metabolic trajectories in both plasma and urine after 2 weeks of an assigned Prudent or Western diet with no differences measured at baseline (**Figure 4.4**). This was a consistent outcome from univariate and multivariate (**Figure 4.3; Figure 4S.5**) statistical methods after adjustments for covariates (sex, age, BMI), including mixed ANOVA and correlation models (**Table 4.2; Table 4.3**). Indeed, plasma ProBet or the ratio of Me-His/C18:3n-6*cis* provided good discrimination ($AUC \approx 0.82$ to 0.87 , $p < 3.0 \text{ E-}5$) of contrasting diets (**Figure 4S.10**). Additionally, ProBet and Me-His concentrations in plasma and urine were positively associated ($r \approx 0.40$ - 0.60 , $p < 0.001$) with eating patterns reflecting a Prudent diet, including a higher intake of fibre, fruit, fruit & vegetable, protein and vitamins/electrolytes, with a lower consumption of *trans* or saturated fats as compared to a Western diet. In fact, ProBet is an exogenous biomarker specific to citrus fruit that has been validated in well-controlled feeding studies⁴² since it is not prevalent in most other foods.⁴³ In fact, ProBet has been replicated in large-scale observational studies as a robust dietary biomarker ($r \approx 0.40$) of recent citrus fruit/juice intake when compared to standardized FFQs, which can be measured in either blood or urine specimens.¹⁸

Me-His has long been reported as an index of myofibrillar muscle protein turn-over under fasting conditions,⁴⁴ whereas it also can serve as a biomarker of recent meat consumption (*e.g.*, chicken) with lower plasma concentrations measured in vegetarians as compared to omnivores.⁴⁵ Consequently, fasting plasma and creatinine-normalized urinary concentrations of ProBet and Me-His were associated with average fruit (servings/2000 kcal) and protein (%energy) as the most likely primary food sources (**Figure 4.4**), which also confirmed excellent dietary adherence with few exceptions. For instance, one participant following a Prudent diet (#19, W-P) had consistently lower than expected concentrations of ProBet and Me-His in both plasma and urine samples, which correctly corresponded to their self-reported diet records. In contrast, three participants following a Western diet (#25, 36, 38, P-W) were found to have higher than expected plasma ProBet when compared to their diet record, but this trend was less apparent in their matching urine samples. These observations are likely due to incidental intake of fruit juice or citrus beverages not included with food provisions that also highlights the different detection windows for dietary biomarkers when relying on “single-spot” plasma or random urine samples.¹⁶ For instance, ingestion of ProBet or orange juice results in a peak concentration in circulation (< 1-2 h) that reflects more recent intake as compared to its later excretion in urine (< 2-24 h).⁴⁶ Nevertheless, there was a strong linear correlation between circulating and excretory concentrations of ProBet ($r = 0.638$) and Me-His ($r = 0.547$) measured in matching plasma and urine samples collected in this study (**Figure 4S.13**).

Two urinary metabolites were also identified by MS/MS (level 2, **Figure 4S.7**) as sensitive dietary biomarkers ($q < 0.05$, FDR) reflecting a Prudent diet, namely OH-PCA and ImPA (**Table 4.3**). Other potential isomeric/isobaric candidates for these metabolites were ruled out by comparing their MS/MS spectra with those predicted *in silico* using CFM-ID³³ in the absence of authentic standards for more confident identification (level 1). Their metabolic trajectories (**Figure**

4S.12) displayed a notable increase ($FC \approx 4$ to 6 , $p < 0.001$) in excretion following a Prudent diet with no differences measured at baseline similar to trends observed for urinary ProBet and Me-His excretion. ImPA is a normal constituent of human urine derived from the metabolism of histidine⁴⁷ which has recently been identified as a product of gut microbiota activity that also regulates insulin sensitivity.⁴⁸ This highlights the fact that many dietary biomarkers are not only dependent on habitual dietary intake and host (liver) metabolism, but are also co-metabolized by gut microbiota with poorly understood effects on human health. Urinary excretion of ImPA was significantly correlated with fibre, fruit & vegetable and protein intake ($r \approx 0.50$, $p \approx 0.001$), which comprise eating patterns consistent with a Prudent diet.⁴⁰ Similarly, urinary OH-PCA was found to have a moderate correlation with fibre and fruit & vegetable intake, and inversely related to total fat. This data indicates that higher excretion of OH-PCA in urine is likely derived from intake of leguminous plants⁴⁹ and citrus fruits⁵⁰ when following a Prudent diet, but represents an endogenous lysine metabolite⁵¹ also produced by gut microbiota.⁵² Indeed, urinary OH-PCA or its ratio to sodium (OH-PCA/Na) discriminated between DIGEST participants from two diet treatment arms (**Figure 4S.10**) with good accuracy ($AUC \approx 0.83$ to 0.88 , $p < 3.0 \text{ E-}4$), as well as sensitivity and specificity ($\approx 90\%$).

Two other metabolites derived from edible plant sources were also identified by MS/MS (**Figure 4S.8**) since they were elevated in urine ($FC \approx 2.5$ to 3.8) when following a Prudent diet as shown by their urinary metabolic trajectory plots (**Figure 4S.12**), namely Ent-G and DHBA. In the case for Ent-G, a MS/MS spectral match based on three characteristic product ions, including a neutral loss of a glucuronide is in close agreement with published data.⁵³ These urinary metabolites were consistently associated ($r \approx 0.30$ - 0.40 , $p < 0.05$) with fruit, vegetable, vitamin C, and/or total sugar intake, and inversely correlated to total fat (**Table 4.3**). Ent-G is a major phytoestrogen from dietary plant lignins, and is excreted in urine as its

monoglucuronide conjugate following biotransformation by human intestinal bacteria.⁵⁴ Even in controlled feeding studies, there is considerable between-subject variation in urinary excretion of enterolignin metabolites due to complex interactions with liver and colonic environments,⁵⁵ which has been reported to possess putative anticancer, antioxidant and/or estrogenic activity.⁵⁶ DHBA is a major phenolic acid constituent from most cereals (*e.g.*, wheat, rye),⁵⁷ which can serve as a biomarker of dietary fibre intake allowing for differentiation of contrasting low and high (> 48 g/day) fibre diets.⁵⁸ In fact, urinary DHBA was the only biomarker differentially excreted at baseline that reflected modest differences in fibre intake between assigned DIGEST participant groups (**Figure 4S.5**). Urinary Val-Val and DMG were also biomarkers related to a Prudent diet, but had weak correlations with only two nutrients (**Table 4.3**), whereas the artificial/low calorie sweetener ASK, and inorganic nitrate were not associated with any nutrient categories from self-reported diet records. ASK was elevated in urine following a Western diet (**Figure 4.3B**), but was rather sporadic with frequent missing data (*i.e.*, below detection limit) since it reflects recent intake of certain sugar-sweetened beverages.⁵⁹ In contrast, nitrate exposure has been reported to be mainly from vegetable consumption due to agricultural fertilizer usage⁶⁰ that is consistent with its increase in urine following a Prudent diet.

The major circulating ketone body, OH-BA and two branched-chain amino acid intermediates, kVal and kLeu also increased in plasma following a Prudent diet as compared to a Western diet (**Table 4.2**). Increases in OH-BA from the liver during ketosis occurs during prolonged fasting or following strenuous exercise,³⁰ as well as abrupt changes in habitual diet, such as adopting a low glycemic index or very low carbohydrate diet.¹⁹ In our work, plasma OH-BA was moderately correlated ($r \approx 0.42$, $p < 0.01$) to increases in fruit, fruit & vegetable and poly:sat consumption and inversely associated with saturated and *trans* fat intake. Since a Prudent diet is characterized by greater consumption of fibre-rich foods with a

lower glycemic index, this may contribute to a mild ketogenic physiological state unlike a Western diet that includes regular consumption of processed foods high in salt and added refined sugar, while low in dietary fibre.^{40,41} Indeed, a Prudent diet composed of whole foods elicits fewer adverse health effects with better adherence than highly restrictive ketogenic diets, which is effective in regulating insulin sensitivity in type 2 diabetes and pre-diabetic patients.⁶¹ Plasma kVal and kLeu were also positively correlated ($r \approx 0.45-0.50$, $p < 0.004$) with key nutrient categories associated with a Prudent diet, including higher intake of protein, fruit and/or vegetable, poly:sat and vitamins. Both plasma metabolites are generated by extra-hepatic branched-chain amino acid transferases prior to oxidative decarboxylation and subsequent utilization as energy substrates within muscle tissue.⁶² The metabolism of branched-chain amino acids plays other critical roles in human health, including ammonia detoxification, protein biosynthesis and insulin sensitivity⁶³ while serving as predictive biomarkers of type 2 diabetes.⁶⁴ A correlation matrix/heat map (**Figure 4S.14**) confirms that plasma kLeu and kVal were strongly co-linear ($r \approx 0.70$, $p = 6.8 \text{ E-}14$) while also being closely associated with OH-BA ($r \approx 0.48$, $p = 3.7 \text{ E-}6$) reflecting common dietary patterns that influence their circulating concentrations. Also, urinary Me-His and ImPA ($r \approx 0.96$, $p < 1.0 \text{ E-}15$), as well as plasma C14:0 and C15:0 ($r \approx 0.87$, $p = 8.4 \text{ E-}15$) were among the most strongly correlated metabolites that originate from consumption of foods rich in dietary histidine and saturated fats, respectively.

Unlike branched-chain amino acid intermediates, two circulating amino acids, Ala and Pro were associated with greater intake of dietary fats (saturated, *trans*, total), and inversely correlated to a Prudent diet due to lower intake of fruit, vegetable or protein (**Table 4.2**). As a result, their plasma metabolic trajectories increased when DIGEST participants were assigned a Western diet for 2 weeks (**Figure 4S.11**). Fasting amino acid concentrations reflect long-term habitual diet rather than recent dietary intake, where Ala has been reported to be inversely

associated to plant-based protein diets.⁶⁵ This is consistent with outcomes in our study, since plasma Ala was negatively correlated to average protein intake ($r \approx -0.40$, $p = 0.014$). Similar to Ala, plasma Pro was reported to be inversely associated with a Prudent diet as measured in a cross-sectional observational study that was adjusted for age, sex and BMI.⁶⁶ This was consistent with our findings since plasma Pro was inversely related to healthy eating patterns, such as lower intake of fruits and/or vegetables and higher consumption of processed foods with *trans* fats (**Table 4.2**). As expected, plasma Pro was correlated ($r \approx 0.46-0.49$, $p < 1.0 \text{ E-}5$) with circulating levels of Ala, as well as C0 indicative of a Western diet (**Figure 4S.14**). Similar outcomes were also measured for two carnitine metabolites (C0 and dC0) since they had metabolic trajectories that increased for DIGEST participants following a Western diet, which were correlated with greater intake of dietary fat, sodium or cholesterol (**Table 4.2**). Although *de novo* synthesis of C0 is derived from dC0 via lysine metabolism, red meat represents a major dietary source of C0 that is also metabolized by gut microbiota with subsequent host hepatic conversion to generate the thrombosis-promoting metabolite, *N*-trimethylamine oxide (TMAO);⁶⁷ however, plasma or urinary TMAO were not modulated by short-term, contrasting diets in our study. In fact, recent studies have shown that anaerobic gut microbiota species can also generate TMAO via its atherogenic intermediate, dC0 due to chronic C0 exposure from the diet.⁶⁸ Nevertheless, C0 is still widely promoted as a nutritional supplement and ergogenic aide to improve fatty acid energy metabolism, as well as alleviate muscle injury from strenuous exercise.⁶⁹ Lastly, a series of plasma total (hydrolyzed) fatty acids had metabolic trajectories that increased when following a Western diet, which were directly associated with greater intake of total, saturated and *trans* fats, but lower consumption of poly:sat, vitamins and fruits & vegetables (**Table 4.2**); these included a low abundance circulating *trans* fatty acid, C18:2n-6*trans*, as well as saturated fats (C14:0, C15:0), and omega-6 fatty acids, namely C18:3n-6*cis* and C18:2n-6*cis*. Indeed, high intake

of omega-6⁷⁰ and saturated⁷¹ fatty acids has long been associated with a Western diet that increases systemic inflammation and chronic disease risk. Nevertheless, there remains on-going controversy regarding the optimal dietary fat composition needed to promote cardiometabolic health.⁷² Recent clinical trials and observational studies have demonstrated that circulating C14:0, C17:0 and notably C15:0 represent dietary biomarkers of dairy fat intake whose impact on cardiometabolic health may likely be beneficial.⁷³ In contrast, greater consumption of processed foods containing vegetable oils rich in C18:2n-6*cis* and other omega-6 fatty acids is hypothesized to be a major dietary culprit for cardiovascular disease prevalence in developed countries.⁷⁴ Public health policies have been far more effective in the past decade to reduce dietary *trans* fat intake to less than 1% energy based on WHO recommendations with animal meats/dairy now being more significant than industrial sources from partial hydrogenation of vegetable oils.⁷⁵ These trends are consistent with data measured in this study, as fasting plasma concentrations of C18:2n-6*trans* were about 0.34% of its stereoisomer and most abundant fatty acid in circulation, C18:2n-6*cis* (**Figure 4S.9**).

In summary, a panel of dietary biomarkers that reflect contrasting Prudent and Western diets were identified based on their distinctive metabolic trajectories measured in matching plasma and urine samples using a cross-platform metabolomics strategy. All DIGEST participants were provided whole foods for consumption over a two-week period while maintaining normal lifestyle habits with no significant changes in their caloric intake, BMI, blood pressure, as well as standard lipid or inflammatory biomarkers as compared to baseline. Me-His and ProBet were the most significant dietary biomarkers associated with a Prudent diet consistently measured in both plasma and urine. Also, urinary ImPA, OH-PCA, Ent-G and DHBA, as well as fasting plasma OH-BA, kVal and kLeu were also positively associated with a Prudent diet. These dietary biomarkers reflect greater consumption of health-promoting foods containing insoluble fibre, protein,

essential nutrients and bioactive phytochemicals with a low glycemic index as compared to highly processed foods in contemporary Western diets. Also, a series of circulating saturated and polyunsaturated fatty acids, as well as plasma Ala, Pro, C0 and dC0 were classified as dietary biomarkers of a Western diet reflecting greater intake of fats, cholesterol and salt, but having lower overall nutrient and fibre quality. Other urinary biomarkers of contrasting diets including ASK, nitrate, DMG and Val-Val, did not have strong associations with any specific nutrient categories from self-reported food records. Strengths of this study include the use of complementary statistical methods with appropriate adjustments, access to matching biospecimens and food records from participants, and use a validated metabolomics data workflow for biomarker discovery and authentication with stringent QC. However, there were several study limitations, including the short duration of the dietary intervention, as well as modest sample size involving participants recruited from a single centre without strict dietary adherence monitoring. Future studies that include multiple time points for biomonitoring of long-term changes in habitual diet with greater study power are recommended. Also, the integration of metabolomics with fecal microbiome data is needed given the important roles of commensal microbiota in nutrient generation and metabolite biotransformation that varies considerably between participants. Also, certain dietary biomarkers tentatively identified in this study still require further structural elucidation to confirm their exact stereoisomer configuration. Overall, our work provides strong corroborating evidence of the utility of food exposure biomarkers to accurately differentiate complex dietary patterns that is generalizable to a free-living, healthy population. This is urgently needed for new advances in nutritional epidemiology and chronic disease prevention, including assessing the impact of maternal nutrition on fetal development early in life and metabolic syndrome risk in childhood.

4.5 References

1. P. Hossain, B. Kawar and P. El Nahas, *New Engl. J. Med.* 2007, **356**, 213–215.
2. Healthy Diet In: Fact Sheet No. 394 [Internet]. World Health Organization; 2017 [cited 23 Oct 2018]. Available: <https://www.who.int/en/news-room/fact-sheets/detail/healthy-diet>
3. A. Mente and S. Yusuf, *Lancet Public Health*, 2018, **17**, e408–e409.
4. Cardiovascular diseases (CVDs). In: Fact Sheets [Internet]. World Health Organization; 2017 [17 May 2017]. Available: <http://www.who.int/mediacentre/factsheets/fs317/en/>
5. L. Cordain, S. B. Eaton, A. Sebastian, N. Mann, S. Lindeberg, B. A. Watkin, J. H. O’Keefe and J. Brand-Miller, *J. Am J Clin Nutr.*, 2005, **81**, 341–354.
6. D. Al-Hamad and D. Raman, *Metabolic syndrome in children and adolescents. Transl. Pediatr.*, 2017, **6**, 397–407.
7. P. D. Wood, M. L. Stefanick, P. T. Williams and W. L. Haskell, *N. Engl. J. Med.*, 1991, **325**, 461–466.
8. M. Dehghan, A. Mente, X. Zhang, S. Swaminathan, W. Li, V. Mohan, R. Iqbal, R. Kumar, E. Wentzel-Viljoen, A. Rosengren A et al., *Lancet*, 2017, **390**, 2050–2062.
9. C. E. Ramsden, D. Zamora, S. Majchrzak-Hong, K. R. Faurot, S. K. Broste, R. P. Frantz, J. M. Davis, A. Ringel, C. M. Suchindran and J. R. Hibbeln, *Brit. Med. J.*, 2016, **353**, i1246.
10. J. P. A. Ioannidis, *JAMA*, 2018, **320**, 969–970.
11. A. Naska, A. Lagiou and P. Lagiou, *Dietary assessment methods in epidemiological research: Current state of the art and future*, *F1000Research*, 2017, **6**, 926.
12. A. Scalbert, L. Brennan, C. Manach, C. Andres-Lacueva, L. O. Dragsted, J. Draper, S. M. Rappaport, J. J. van der Hooft, D. S. Wishart, *The food metabolome: A window over dietary exposure. Am. J. Clin. Nutr.*, 2014, **99**, 1286–1308.
13. N. J. W. Rattray, N. C. Deziel, J. D. Wallach, S. A. Khan, V. Vasiliou, J. P. A. Ioannidis and C. H. Johnson, *Hum. Genomics.*, 2018, **12**, 4.
14. M. Guasch-Ferre, S. N. Bhupathiraiu and F. B. Hu, *Clin. Chem.*, 2018, **64**, 82–98.
15. L. Brennan and F. B. Hu, *Mol. Nutr. Food Res.*, 2018, **170**, e1701064.
16. L. Brennan, *J. Nutr.*, 2018, **148**, 821–822.
17. M. B. Andersen, Å. Rinnan, C. Manach, S. K. Poulsen, E. Pujos-Guillot, T. M. Larsen, A. Astrup and L. O. Dragsted, *L. O. J. Proteome Res.*, 2014, **13**, 1405–1418.
18. M. C. Playdon, J. N. Sampson, A. J. Cross, R. Sinha, K. A. Guertin, K. A. Moy, N. Rothman, M. L. Irwin, S. T. Mayne, R. Stolzenberg-Solomon and S. C. Moore, *Am. J. Clin. Nutr.*, 2016, **104**, 776–789.

19. T. Esko, J. N. Hirschhorn, H. A. Feldman, Y. H. Hsu, A. A. Deik, C. B. Clish, C. B. Ebbeling and D. S. Ludwig, *Am. J. Clin. Nutr.*, 2017, **105**, 547–554.
20. H. Gibbons, E. Carr, B. A. McNulty, A. P. Nugent, J. Walton, A. Flynn, M. J., Gibney and L. Brennan, *L. Mol. Nutr. Food Res.*, 2017, **61**.
21. K. A. Guertin, S. C. Moore, J. N. Sampson, W. Y. Huang, Q. Xiao, R. Z. Stolzenberg-Solomon, R. Sinha and A. J. Cross, *Am. J. Clin. Nutr.*, 2014, **100**, 208–217.
22. K. Hanhineva, M. A. Lankinen, A. Pedret, U. Schwab, M. Kolehmainen, J. Paananen, V. de Mello, R. Sola, M. Lehtonen, K. Poutanen, M. Uusitupa and H. Mykkänen, *J. Nutr.*, 2015, **145**, 7–17.
23. B. Khakimov, S. K. Poulsen, F. Savorani, E. Acar, G. Gürdeniz, T. M. Larsen, A. Astrup, L. O. Dragsted and S. B. Engelsen, *J. Proteome Res.*, 2016, **15**, 1939–1954.
24. I. Garcia-Perez, J. M. Posma, R. Gibson, E. S. Chambers, T. H. Hansen, H. Vestergaard, T. Hansen, M. Beckmann, O. Pedersen, P. Elliott, J. Stamler, J. K. Nicholson, J. Draper, J. C. Mathers, E. Holmes, and G. Frost, *Lancet Diabetes Endocrinol.*, 2017, **5**, 184–195.
25. M. A. Zulyniak, R. J. de Souza, A. Mente, S. Kandasamy, M. Nundy, D. Desai, K. Raman, R. Hasso, G. Paré, J. Beyene and S. S. Anand, *BMC Nutr.*, 2016, **2**, 34.
26. F. B. Hu, E. B. Rimm, M. J. Stampfer, A. Ascherio, D. Spiegelman and W. C. Willett, *Am. J. Clin. Nutr.*, 2000, **72**, 912–921.
27. J. M. Kerver, E. J. Yang, L. Bianchi and W. O. Song, *Am. J. Clin. Nutr.*, 2003, **78**, 1103–1110.
28. A. Nori de Macedo, S. Mathiaparanam, L. Brick, K. Keenan, T. Gonska, L. Pedder, S. Hill and P. Britz-McKibbin, *ACS Cent. Sci.*, 2017, **3**, 904–913.
29. A. DiBattista, N. McIntosh, M. Lamoureux, O. Y. Al-Dirbashi, P. Chakraborty and P. Britz-McKibbin, *J. Proteome Res.*, 2019, **18**, 841–854.
30. M. Saoi, M. Percival, C. Nemr, A. Li, M. J. Gibala and P. Britz-McKibbin, *Anal. Chem.*, 2019, **91**, 4709–4718.
31. N. G. Mahieu and G. J. Patti, *Anal. Chem.*, 2017, **89**, 10397–10406.
32. D. I. Broadhurst and D. B. Kell, *Metabolomics*, 2006, **2**, 171–196.
33. F. Allen, A. Pon, M. Wilson, R. Greiner, and W. Wishart, *Nucleic Acids Res.*, 2014, **42**, W94–W99.
34. R. M. Salek, C. Steinbeck, M. R. Viant, R. Goodacre and W. B. Dunn, *Gigascience*, 2013, **2**, 13.
35. B. Y. Tai and T. P. A. Speed, *Ann. Statistics*, 2006, **34**, 2387–2412.
36. GBD 2017 Diet Collaborators. *Lancet*, 2019, **393**, 1958–1972.
37. L. B. Moore, S. V. Liu, T. M. Halliday, A. P. Neilson, V. E. Hedrick and B. M. Davy *J. Nutr.*, 2017, **147**, 2364–2373.

38. K. D. Stark, M. E. Van Elswyk, M. R. Higgins, C. A. Weatherford and N. Salem, N. Jr. *Prog. Lipid Res.*, 2016, **63**,132–152.
39. Y. Lin, I. Huybrechts, C. Vereecken, T. Mouratidou, J. Valtuena, M. Kersting, M. Gonzalez-Gross, S. Bolca, J. Warmberg, M. Cuenca-Garcia, M. et al., *J. Nutr.*, 2015, **54**, 771–782.
40. L. Marks, *L. Food Policy*, 1985, **10**, 166–174.
41. F. B. Hu, *J. Am. Diet. Assoc.*, 2003, **73**, 61–67.
42. T. T. Fung, E. B. Rimm, D. Spiegelman, N. Rifai, G. H. Tofler, R. Willett Lang, T. Lang, M. Bader, A. Beusch, V. Schlagbauer and T. Hofmann, *T. J. Agric. Food Chem.*, 2017, **65**, 1613–1619.
43. S. S. Heinzmann, I. J. Brown, Q. Chan, M. Bictash, M. E. Dumas, S. Kochhar, J. Stamler, E. Holmes, P. Elliott and J. K. Nicholson, *Am. J. Clin. Nutr.*, 2010, **92**, 436–443.
44. C. L. Long, D. R. Dillard, J. H. Bodzin, J. W. Geiger and W. S. Blakemore, *Metabolism*, 1988, **37**, 844–849.
45. B. Kochlik, C. Gerbracht, T. Grune and D. Weber, *D. Mol. Nutr. Food Res.*, 2018, **62**, e1701062.
46. W. Atkinson, P. Downer, M. Lever, S. T. Chambers and P. M. George, *Eur. J. Nutr.*, 2007, **46**, 446-452.
47. N. P. Sen, P. L. McGeer and R. M. Paul. *Biochem. Biophys. Res. Commun.*, 1962, **9**, 257–261.
48. A. Koh, A. Molinaro, M. Ståhlman, M. T. Khan, C. Schmidt, L. Mannerås-Holm, H. Wu, A. Carreras, H. Jeong, L. E. Olofsson, et al., *Cell*, 2018, **175**, 947–961.
49. Y. Kunii, M. Otsuka, S. Kashino, H. Takeuchi and S. Ohmori, *S. J. Agric. Food Chem.*, 1996, **44**, 483–487.
50. L. Servillo, A. Giovane, M. L. Balestrieri, G. Ferrari, D. Cautela, and D. Castaldo, *J. Agric. Food Chem.*, 2012, **60**, 315–321.
51. J. Dancis and J. Hutzler, *Am. J. Hum. Genet.*, 1986, **38**, 707–711.
52. T. Fujita, T. Hada and K. Higashino, *Clin. Chim Acta.*, 1999, **287**, 145–156.
53. C. H. Johnson, S. K. Manna, K. W. Krausz, J. A. Bonzo, R. D. Divelbiss, M. G. Hollingshead, and F. J. Gonzalez, *Metabolites*, 2013, **3**, 658–672.
54. U. Knust, W. E. Hull, B. Spiegelhalder, H. Bartsch, T. Strowitzki and R. W. Owen. *Food Chem. Toxicol.*, 2006, **44**, 1038–1049.
55. J. W. Lampe C. Atkinson and M. A. Hullar MA. *JAOAC Int.*, 2006, **89**, 1174–1181.
56. C. Rodríguez-García, C. Sánchez-Quesada, E. Toledo, M. Delgado-Rodríguez and J. J. Gaforio, *Molecules*, 2019, **24**, 917.
57. B. Khakimov, B. M. Jespersen and S. B. Engelsen, *Foods* 2014, **3**, 569–585.
58. A. Johansson-Persson, T. Barri, M. Ulmius, G. Onning and L. O. Dragsted, *Anal. Bioanal. Chem.*, 2013, **405**, 4799–4809.

59. C. Logue, L. R. C. Dowey, J. J. Strain, H. Verhagen, S. McClean and A. M. Gallagher. *J. Agric. Food Chem.*, 2017, **65**, 4516–4525.
60. N. G. Hord, Y. Tang and N. S. Bryan, *Am. J. Clin. Nutr.*, 2009, **90**, 1–10.
61. F. A. Spritzler, *Diabetes Spectrum*, 2012, **25**, 238–243.
62. C. Nie, T. He, W. Zhang, G. Zhang and X. Ma, *Int. J. Mol. Sci.*, 2018, **19**, e954.
63. M. Holeček *Nutr. Metab.* 2018, **15**, 33.
64. J. L. Flores-Guerrero, M. C. J. Osté, L. M. Kieneker, E. G. Gruppen, J. Wolak-Dinsmore, J. D. Otvos, M. A. Connelly, S. J. L. Bakker and R. P. F. Dullaart, *J. Clin. Med.*, 2018, **7**, e513.
65. B. Merz, L. Frommherz, M. J. Rist, S. E. Kulling, A. Bub and B. Watzl, *Nutrients*, 2018, **10**, e623.
66. A. Bouchard-Mercier, I. Rudkowska, S. Lemieux, P. Couture and M.-C. Vohl, *Nutrition J.*, 2013, **12**, 158.
67. R. A. Koeth, Z. Wang, B. S. Levison, J. A. Buffa, E. Org, B. T. Sheehy, E. B. Britt, X. Fu, Y. Wu, L. Li, et al., *Nat. Med.*, 2013, **19**, 576–585.
68. R. A. Koeth, B. R. Lam-Galvez, J. Kirsop, Z. Wang, B. S. Levison, X. Gu, M. F. Copeland, D. Bartlett, D. B. Cody, H. J. Dai, et al., *J. Clin. Invest.*, 2019, **129**, 373–387.
69. R. Fielding, L. Riede, J. P. Lugo and A. Bellaminne, *Nutrients*, 2018, **10**, e349.
70. E. Patterson, R. Wall, G. F. Fitzgerald, R. P. Ross and C. Stanton, *J. Nutr. Metab.*, 2012, **2012**, 539426.
71. J. S. Zheng, S. J. Sharp, F. Imamura, A. Koulman, M. B. Schulze, Z. Ye, J. Griffin, M. Guevara, J. M. Huerta, J. Kröger, I. Sluijs, I. et al., *BMC Med.*, 2017, **15**, 203.
72. A. Mente, M. Dehghan, S. Rangarajan, M. McQueen, G. Dagenais, A. Wielgosz, S. Lear, W. Li, H. Chen, S. Yi, S. et al., *Lancet Diabetes Endocrinol.*, 2017, **5**, 774–787.
73. U. Risérus and M. Marklund, *M. Curr. Opin. Lipidol.*, 2017, **28**, 46–51.
74. J. J. DiNicolantonio and J. H. O’Keefe, *Open Heart J.*, 2018, **5**, e000898.
75. A. J. Wanders, P. L. Zock and I. A. Brouwer, *Nutrients*, 2017, **9**, 840.

4.6 Supplemental

Acknowledgements

Philip Britz-McKibbin acknowledges funding from the Natural Sciences and Engineering Research Council of Canada, Genome Canada and Faculty of Science at McMaster University. Sonia S. Anand holds the Heart and Stroke Michael G DeGroot Chair in Population Health Research and a Canada Research Chair in Ethnicity and Cardiovascular Disease. The authors also acknowledge support of Dr. Marcus Kim from Agilent Technologies Inc.

Experimental

Chemicals and Reagents. Ultra HPLC grade LC-MS solvents (water, methanol, acetonitrile) obtained from Caledon Laboratories Ltd. (Georgetown, ON, Canada) were used to prepare all buffer and sheath liquid solutions, unless otherwise stated. Proline betaine was purchased from Toronto Research Chemicals (Toronto, Ontario, Canada). All other chemicals were obtained from Sigma-Aldrich Inc. (St. Louis, MO, USA).

Study Design, Participant Eligibility and Dietary Self-reporting. The Diet and Gene Intervention Study (DIGEST) was a 2-arm, parallel unblinded study to compare the effects of two weeks of a Prudent diet compared with a Western diet on cardiovascular risk factors and gene expression in apparently healthy adults. Healthy participants were recruited using flyers, and self-referral methods from McMaster University and the surrounding areas. Exclusion criteria were an unwillingness to eat an assigned diet, or serious disease or illness. The full inclusion and exclusion criteria are described elsewhere,¹ and participant eligibility for this work is also summarized in the CONSORT chart (**Figure 4S.2**). A subset of 42 participants from the DIGEST study with paired urine and serum samples were selected for targeted and nontargeted metabolomics analysis using three

complementary instrumental platforms (**Figure 4S.1**), where subjects had completed self-reported dietary records at baseline and two weeks following food provisions. All participants from DIGEST picked up their food allotment each week at the grocery store, during clinic visits, or food provisions were delivered to their home by volunteers.¹ Additionally, cooking suggestions with meal plans were provided by a dietician that still allowed for flexibility in food preparations while participants were requested to maintain their normal lifestyle habits (*e.g.*, physical activity). In order to maximize treatment effects in this short-term dietary intervention pilot study, subjects were assigned into two parallel arms of contrasting diets, namely a Western diet reflecting a typical Canadian macronutrient profile with higher intake of processed foods (*e.g.*, burgers, fried chicken, cereals, processed cheeses), and a Prudent diet based around minimally processed foods composed of lean protein (*e.g.*, poultry, fish, legumes), whole grains and a high amount of fresh fruits and vegetables.¹ An aggregate diet quality index score was used to classify participants as having a predominantly Prudent or Western habitual diet at baseline based on five categories assessed from self-reporting questionnaires, including polyunsaturated/saturated fatty acid ratio (poly/sat > 1.0), relative intake of saturated fatty acids (< 7%), total fibre (> 28 g/day), daily servings fruits and vegetables (> 5), and daily potassium intake (> 3500 mg/day). Briefly, habitual dietary patterns at baseline was evaluated by having participants attend a screening visit 1 week prior to beginning DIGEST where they were also provided a diary to record all their foods. Both diet groups were balanced with respect to age and adiposity (BMI), however most participants in this study were female (64%), and a majority were Caucasian (78%) with no self-identified tobacco smokers. Each participant was assigned a 'Prudent score' and 'Western score' from 1 (low) to 4 (high) based on their quartile rankings.¹ Habitual diet was classified as predominantly Western if the difference of [Western score - Prudent score] \geq 2. Challenges in recruiting eligible participants who exclusively followed a Prudent

dietary pattern resulted in no significant differences ($p > 0.05$) in their habitual Prudent scores between the two assigned diet groups at baseline (**Table 4S.1**); thus minor differences in baseline dietary patterns among DIGEST participants were largely related to differences in consumption of foods associated with a Western diet (**Table 4.1**). Participants from DIGEST were informed to maintain their usual lifestyle and physical activity routine during the study period and were provided a 7-day menu plan for each of the 2 weeks, which listed which foods they were to eat for specific meals. Servings of Prudent-type (*e.g.*, fruit & vegetables, lean meats, high fiber) and Western-type (*e.g.*, red meat, salty food, high saturated fats) diet was scored and ranked in quartiles.¹ Dietary adherence was a measure of the % of the foods "prescribed" that they reported eating that was self-reported based on the foods they checked-off from menu list that they consumed, which was $> 95\%$ for both treatment arms.¹ A total of 20 micro- and macronutrient categories (from over 120) from self-reporting dietary records were determined to be significantly ($q < 0.05$; Bonferroni adjustment) different between assigned Prudent and Western diets for DIGEST participants in this pilot study (**Table 4.1**; **Figure 4S.4**), which were subsequently correlated with top-ranked plasma and urinary metabolites when validating putative dietary biomarkers of contrasting diets (**Table 4.2**).

Nontargeted Metabolite Profiling of Plasma and Urine by MSI-CE-MS.

Fasting plasma (EDTA) samples together with matching single-spot urine samples were collected from all DIGEST participants during clinic visits on day 1 and day 14, which were then stored at $-80\text{ }^{\circ}\text{C}$.¹ Multisegment injection-capillary electrophoresis-mass spectrometry (MSI-CE-MS) was the major platform used for nontargeted profiling of polar/ionic metabolites from both plasma and urine samples,² which was performed on an Agilent G7100A CE (Agilent Technologies Inc., Mississauga, ON, Canada) equipped with a coaxial sheath liquid (Dual AJS) Jetstream electrospray ion source coupled to an Agilent 6230 TOF-MS system. All

separations were performed using uncoated fused-silica capillaries (Polymicro Technologies, AZ, USA) with a total length of 120 cm and inner diameter of 50 μm . About 7 mm of polyimide coating was removed from both distal ends to avoid sample carry-over and prevent polyimide swelling upon contact with organic solvent.³ The background electrolyte (BGE) consisted of 1 M formic acid with 15% *v/v* acetonitrile (pH 1.80) under positive ion mode, and 50 mM ammonium bicarbonate (pH 8.50) under negative ion mode for analysis of the ionic metabolome, including cationic and anionic metabolites from matching plasma and urine specimens, respectively. All CE separations were performed under normal polarity with an applied voltage of 30 kV at 25 °C. A pressure gradient of 2 mbar/min from 0 to 40 min was used for MSI-CE-MS analyses under negative ion mode conditions to shorten analysis times for highly charged anionic metabolites (*e.g.*, citrate). The TOF-MS system was operated with full-scan data acquisition over a mass range of *m/z* 50-1700 and an acquisition rate of 500 ms/spectrum. The sheath liquid was comprised of 60% *v/v* MeOH with 0.1% *v/v* formic acid for positive ion mode, and 50% *v/v* MeOH for negative ion mode. The ESI conditions were $V_{\text{cap}} = 2000$ V, nozzle voltage = 2000 V, nebulizer gas = 10 psi, sheath gas = 3.5 L/min at 195 °C, drying gas = 8 L/min at 300 °C and the MS voltage settings were fragmentor = 120 V, skimmer = 65V and Oct1 RF = 750 V.

A seven sample serial injection format was used in MSI-CE-MS²⁻⁶ consisting of a serial injections of six discrete samples together with a pooled quality control (QC) within each experimental run; the latter sample was used to assess technical variance while also allowing for robust batch correction due to long-term signal drift in ESI-MS.^{5,7} Multiplexed separations in MSI-CE-MS was performed by programming a hydrodynamic injection sequence with each sample (5 s at 100 mbar) interspaced with a BGE spacer (40 s at 100 mbar) prior to voltage application as described elsewhere.²⁻⁶ In this study, nontargeted metabolite profiling of plasma or urine specimens by MSI-CE-MS was performed by pairing together matching

baseline and post-treatment samples for three individual DIGEST participants together with a QC in a randomized injection position for each run (**Figure 4.2**). Different serial sample injection configurations in MSI-CE-MS were also applied for rigorous metabolite authentication using a dilution trend filter, as well as acquisition of calibration curves for reliable quantification of metabolites as described elsewhere.²⁻⁶ Briefly, authentic metabolites were defined by their unique accurate mass and relative migration time (m/z :RMT) under positive (+) or negative (-) ion mode detection after filtering out spurious signals, background ions, as well as redundant adducts/in-source fragments and isotopic features that comprise the majority of signals when performing MS-based metabolomics.⁸ Additionally, only frequently detected plasma or urinary metabolites measured in the majority of samples from DIGEST participants (> 75%) with acceptable technical precision based on repeated analysis of QCs (mean CV < 30%) were included in the final metabolomics data matrix (**Table 4S.1**) as a way to reduce false discoveries and data overfitting. Missing (*i.e.*, zero) data below method detection limits were replaced with a minimum value corresponding to $\frac{1}{2}$ of the lowest responses measured for a given metabolite in all samples analyzed.

Unknown Metabolite Identification by MS/MS. In all cases, authenticated metabolites defined by their m/z :RMT under positive (+) or negative (-) ion mode detection were further characterized by their most likely molecular formula, mass error (< 10 ppm) and overall technical precision (%CV) from pooled plasma or urine samples. High resolution tandem mass spectrometry (MS/MS) was employed for structural elucidation of unknown metabolites of biological significance from pooled samples in this study. Targeted MS/MS experiments were performed on an Agilent G7100A CE system with a coaxial sheath liquid Jetstream electrospray ion source connected to an Agilent 6500 iFunnel QTOF instrument. Metabolite identification in this work adopted reporting standards recommended from the

Metabolomics Standards Initiative,⁹ including unambiguous identification (level 1) that is confirmed with matching MS/MS spectra and co-migration by authentic standard acquired on the same instrument, tentative/probable identification (level 2) by comparison of MS/MS spectra from public databases or published literature, partial annotation of MS/MS spectra guided by *in silico* software tools with metabolite class (level 3), and compounds with unknown chemical structure (level 4). The latter case typically occurred for low abundance metabolites that had inadequate responses for their precursor ion when acquiring MS/MS spectra via collision-induced dissociation (CID) experiments. In our case, the electromigration behavior of polar/ionic metabolites (*i.e.*, electrophoretic mobility or RMT) provided additional information that complemented MS/MS when selecting among potential isobaric or isomeric candidate ions, in addition to their likely biochemical relevance in human plasma or urine. MS/MS spectra were acquired from pooled plasma or urine samples that were injected hydrodynamically using a conventional single sample injection plug at 50 mbar for 10 s followed by 5 s with BGE. Precursor ions were selected for CID experiments at 10, 20 and 40 V. Mirror plots comparing MS/MS spectra of unknown metabolites under an optimal collision energy were then compared to their respective authentic reference standard if available, which were generating using the “InterpretMSSpectrum” R Package. Otherwise, MS/MS spectra were annotated based on their characteristic product ions or neutral losses for *de novo* structural elucidation (level 2 or 3), which was guided by *in silico* MS/MS spectra generated by CFM-ID¹⁰ or comparison to MS/MS spectra deposited in open-access public repositories (HMDB, <http://www.hmdb.ca>) or published in literature (if available). The exact stereochemistry of certain metabolites identified in plasma or urine were uncertain if authentic standards were not available for confirmation.

Total Plasma Fatty Acid Determination by GC-MS. A GC-MS method was used for targeted analysis of total (hydrolyzed) fatty acids (FA) and their isomers from plasma extracts on an Agilent 6890 GC coupled to an Agilent 5973 single quadrupole mass spectrometer with electron impact ionization (EI) as described elsewhere with minor modifications.¹¹ Total hydrolyzed plasma FA were analyzed by GC-MS as their methyl ester derivatives (FAME) using *N*-methyl-*N*-(trimethylsilyl)trifluoroacetamide (MSTFA) ($\geq 98.5\%$, GC from Sigma-Aldrich). Isotopically-labelled myristic acid- d_{27} (98%), stearic acid- d_{35} (98%) and pyrene- d_{10} (98%) were obtained from Cambridge Isotope Laboratories (Tewksbury, MA, USA). HPLC grade chloroform ($\geq 99.5\%$, GC), methanol (99.8%, GC), hexanes ($\geq 99.5\%$, GC) and Ultra LC-MS grade water were purchased from Caledon Laboratories Ltd. (Georgetown, ON, Canada). The antioxidant, butylated hydroxytoluene (BHT, 490 μL) was added to all fatty acids calibrant solutions prepared in methanol to prevent autooxidation. Briefly, a 10 μL aliquot of thawed plasma was mixed with 5 μL of a 1.0 mg/mL C18:0- d_{35} recovery standard. Concentrated sulfuric acid (10 μL) was added as a transesterification catalyst and vortexed for 2 min before incubation for 4 h at 80 °C to produce FAMEs for improved volatility. Cooled samples were then mixed with 500 μL of 9 g/L NaCl and 200 μL hexanes and vortexed for 2 min prior to centrifugation at 14,000 g at 4 °C for 5 min. GC inserts were prepared with 45 μL of hexane supernatant with 5 μL of internal standard pyrene- d_{10} and vortexed for 2 min before injection. Total plasma FAMEs were resolved on a Supleco SP-2380 column (30 m x 0.25 mm x 0.20 μM) using an optimal temperature program within 30 min.⁶ Samples were injected in 1.0 μL volumes using a splitless injector held at 250 °C, the carrier gas was helium at 1.0 mL/min and the transfer line was held at 270 °C. Relative abundances were measured by the relative response of a quantifying ion (*e.g.*, [M-15]⁺) relative to the pyrene- d_{10} internal standard. Pooled plasma samples as QC and blank extracts were analyzed together with each batch of 8-10 randomized plasma

samples from DIGEST participants when using GC-MS in order to assess technical precision and monitor for background contamination. A temperature program used for resolution of major trimethylsilylated fatty acids from plasma hydrolysates comprised of a temperature gradient of 20 °C/min starting from 2 min at 80 °C until 20 min at 160 °C, which was further increased to 190 °C for 3 min prior to elution at 300 °C for 5 min with a total run time of 28 min. In most cases, FAMES were quantified in GC-MS based on integration of the relative response ratio of their [M-15]⁺ fragment ion relative to pyrene-d₁₀ as internal standard when using single ion monitoring (SIM) mode detection. Calibration curves from a serial dilution of fatty acid standards were subsequently used for their quantification, as well as their identification when comparing their characteristic EI-MS spectra (70 eV) and elution times (*i.e.*, co-elution with spiking).

Targeted Urinary Electrolyte Analysis. Targeted analysis of inorganic/involatile electrolytes in urine was performed using two complementary CE-UV methods adapted from Nori de Macedo *et al.*¹² and Saoi *et al.*⁵ for anionic (*e.g.*, nitrate) and cationic (*e.g.*, sodium) electrolytes, respectively. Analysis of major cationic electrolytes was performed on diluted urine samples that were thawed, vortexed for 30 s and centrifuged at 14,000 g for 5 min. An aliquot of the supernatant was diluted with deionized water and spiked with 0.5 mM lithium as an internal standard. Samples were analyzed on an Agilent G7100A CE system with UV photodiode array detection with indirect absorbance detection at 214 nm. All samples were injected hydrodynamically for 10 s (at 35 mbar) and separation was performed under normal polarity at 30 kV at 25°C using a 50 µm inner diameter capillary with 60 cm total length. The background electrolyte (BGE) was 5 mM formic acid containing 12.5 mM creatinine (Crn) and 4 mM 18-crown-6 ether at pH 4.0 (adjusted with 1 M sulfuric acid. In this case, ammonium (NH₄⁺), sodium (Na⁺), potassium (K⁺), calcium (Ca²⁺) and magnesium (Mg²⁺) were analyzed in urine

samples.⁵ Additionally, all urine samples were analyzed using a complementary CE assay for UV-absorbing inorganic anions, including nitrate (NO_3^-), iodide (I^-) and thiocyanate (SCN^-) as described elsewhere.¹² In this case, the BGE was comprised of 180 mM lithium hydroxide, 180 mM phosphoric acid, and 10 mM α -cyclodextrin (α -CD) at pH 3.0 (adjusted with 1 M phosphoric acid), where 1,5-naphthalene disulfonate (NDS) was used as an internal standard. Samples were injected hydrodynamically for 80 s at 0.5 psi and analyzed at 25 °C under a reversed polarity at -18 kV with UV absorbance detection at 226 nm (288 nm for NDS). In both CE-UV assays, a pooled urine sample serving as a QC was measured intermittently after every batch of 6 random urine samples. In all cases, creatinine concentrations were measured by MSI-CE-MS under positive ion mode conditions, which were used to reduce biological variance and correct for differences in hydration status for single-spot urine samples analyzed in this study.

Data Preprocessing and Statistical Analysis. All MSI-CE-MS data were integrated and analyzed using Agilent MassHunter Qualitative Analysis B.07.00 and Microsoft Excel and Igor (Wavemetrics Inc., OR, USA). In all cases, the integrated ion response (*i.e.*, peak area) for each metabolite was normalized to an internal standard, 4-chlorotyrosine (Cl-Tyr) migrating from the same sample by MSI-CE-MS. Also, a QC-based batch correction algorithm was applied to creatinine-normalized urine metabolomic data to adjust for long-term signal drift in ESI-MS during data acquisition as outlined in a recent work.⁵ This algorithm is based on an empirical Bayesian frameworks⁷ that takes advantage of the QC samples included in each serial injection run when using MSI-CE-MS, as well as batch and injection sequence information. However, batch correction did not provide any significant improvement in the overall technical precision of plasma metabolome data and thus was not necessary to perform in this case. All non-batch (plasma) and batch-corrected (urine) metabolomic data was pre-processed using

generalized log-transformation and autoscaling prior to multivariate statistical analysis using MetaboAnalyst 4.0 (www.metaboanalyst.ca),¹³ including volcano plots, principle component analysis (PCA), hierarchical cluster analysis (HCA)/2D heat maps, receiver operating characteristic (ROC) curves, orthogonal partial least squares-discriminant analysis (OPLS-DA), as well as multivariate empirical Bayes analysis (MEBA) of variance; the latter method is optimal when analyzing time-series data¹⁴ as related to metabolic trajectories. To validate each OPLS-DA model, cross-validation and permutation testing ($n = 1000$) on paired metabolome data sets (*i.e.*, ratio of metabolite response based on assigned diet/baseline diet for each subject) following glog transformation and autoscaling, whereas Hotelling's T-squared distribution using MEBA was performed on glog-transformed metabolomic time series data from DIGEST participants at baseline and following two weeks of assigned food provisions. These complementary statistical approaches were initially used for unsupervised data exploration to identify overall trends, as well as supervised data analysis for ranking metabolite candidates modulated by contrasting assigned diets among DIGEST participants without adjustments for covariates. Additionally, normality tests, partial Pearson correlation analysis, and mixed model ANOVA were performed on top-ranked dietary biomarker candidates using the Statistical Package for the Social Sciences (SPSS, version 18.0). In this case, a partial listwise Pearson correlation analysis of lead plasma and creatinine-normalized metabolite responses to 20 major nutrient categories from self-reported food records from DIGEST participants ($n = 42$) were adjusted for age, sex, and post-intervention BMI. Only metabolites that had a correlation coefficient of $r > \pm 0.300$ and $p < 0.05$ for more than two nutrient categories were considered significant in this work. A repeat measures general linear mixed ANOVA model was also performed with the number of levels set at 2 for the repeat sampling (*i.e.*, time; baseline habitual diet and assigned diet after 2 weeks) while setting the intervention diet (*i.e.*, treatment arm; P-W or W-P) as the

between-subject factor with age, sex and post-intervention *BMI* as potential covariates. Overall, plasma and urine metabolites reflecting contrasting dietary patterns assigned to DIGEST participants that satisfied MEBA and/or mixed ANOVA, as well as partial correlation analysis to two or more nutrient categories from self-reported diet records were considered as robust dietary biomarkers (**Table 4.2, Table 4.3**).

Table 4S.1: Baseline group characteristics of a cohort of healthy participants ($n = 42$) recruited in a two-arm parallel randomized clinical trial to compare the effects of a contrasting diet over 2 weeks (Western or Prudent) from food provisions reflecting changes in habitual dietary patterns.

Variable	Prudent assigned W-P, $n = 24$	Western assigned P-W, $n = 18$	Significance ^b
Sex (n; %)	--	--	$p > 0.05$; More females than males recruited in each arm
F	66%, $n = 16$	61%, $n = 11$	
M	33%, $n = 8$	39%, $n = 7$	
Age (mean)	(50 ± 18)	(43 ± 20)	$p > 0.05$; Wide disparity in age with no differences between arms
< 50 y	(29 ± 9, $n = 9$)	(28 ± 10, $n = 10$)	
> 50 y	(62 ± 7, $n = 15$)	(63 ± 8, $n = 8$)	
BMI (mean)	(28 ± 6)	(26 ± 6)	$p > 0.05$; Wide disparity in body composition and no differences between arms
Lean (19-24.9 kg/m ²)	(23 ± 2, $n = 7$)	(22 ± 2, $n = 11$)	
Overweight/obese (25-44 kg/m ²)	(30 ± 5, $n = 17$)	(31 ± 5, $n = 7$)	
Habitual baseline diet index ^a	--	--	$p = 0.0037$; Greater western habitual dietary patterns in prudent assigned arm
Prudent diet score (< 0.5)	(0.42 ± 0.93)	(0.92 ± 0.68)	
Western diet score (> 1.0)	(3.4 ± 0.9)	(0.70 ± 1.31)	
Average caloric intake (kcal)	(1985 ± 560)	(1895 ± 640)	$p > 0.05$
Average fiber intake (/2000 kcal)	(21.3 ± 6.3); 24	(26.6 ± 8.5); 18	$p = 0.018$; Higher intake of fiber in western assigned arm
Average poly:sat fatty acid (ratio)	(0.44 ± 0.16); 24	(0.58 ± 0.16); 18	$p = 0.0067$; Higher poly:sat intake in western assigned arm
Average energy from sat. fat (%)	(11.9 ± 3.2); 24	(9.8 ± 2.0); 18	$p = 0.015$; Higher sat. fat intake in prudent assigned arm
Urinary Na/K (ratio)	(1.31 ± 0.72); 24	(0.80 ± 0.55); 18	$p = 0.016$; Higher Na/K intake in prudent assigned arm
Fasting glucose (mM)	(5.1 ± 1.0)	(4.9 ± 0.4)	$p > 0.05$
LDL cholesterol (mM)	(3.1 ± 1.0)	(2.8 ± 0.9)	$p > 0.05$
HDL cholesterol (mM)	(1.55 ± 0.42)	(1.46 ± 0.44)	$p > 0.05$
Total cholesterol (mM)	(5.2 ± 1.3)	(5.0 ± 1.0)	$p > 0.05$
Triglycerides (mM)	(0.63 ± 0.18)	(0.61 ± 0.18)	$p > 0.05$
ApoB/ApoA1 ratio	(1.24 ± 0.77)	(1.0 ± 0.18)	$p > 0.05$
CRP (mg/L)	(2.4 ± 4.1)	(2.2 ± 2.9)	$p > 0.05$
IL-8 (ng/L)	(9.1 ± 6.5)	(6.8 ± 2.1)	$p > 0.05$
Average systolic BP (mmHg)	(121 ± 18)	(114 ± 17)	$p > 0.05$

Average dystolic BP (mmHg)	(78 ± 11)	(74 ± 10)	$p > 0.05$
-------------------------------	-----------	-----------	------------

a Self-reported diet index score at baseline was used as a single aggregate index to categorize dietary patterns as predominantly Prudent or Western in terms of average daily intake of total fiber, fruit & vegetables, potassium, polyunsaturated/saturated fatty acid ratio and % saturated fatty acids.

b There were no significant differences ($p < 0.05$) in classic serum/plasma biomarkers of CVD risk at baseline, as well as following 2-week dietary intervention between treatment arm among subjects in this study.

Table 4S.2: Summary of representative and reliably measured urinary and plasma metabolites detected in a majority of DIGEST participants patients that are annotated based on their accurate mass (m/z), relative migration time (RMT), ionization mode ($p = ESI^+$, $n = ESI^-$), most likely molecular formula, compound name, confidence level for identification, and technical precision from QC measurements.

Biofluid	m/z :RMT:polarity	Molecular Formula ^a	Monoisotopic Mass	Δm (ppm)	Compound ID	HMDB ID	Level of Confidence ^b	%CV ^c
Urine	76.0757 : 0.546:+	C ₃ H ₉ NO	76.0757	0.0	Trimethylamine <i>N</i> -oxide	HMDB00925	2	9.86
Urine	104.0706 : 0.667:+	C ₄ H ₉ NO ₂	104.0706	0.0	γ -Aminobutyric acid	HMDB00112	2	39.69
Urine	104.0706 : 0.928:+	C ₄ H ₉ NO ₂	104.0706	0.0	Dimethylglycine	HMDB0000092	2	33.80
Urine	104.1075 : 0.569:+	C ₅ H ₁₄ NO	104.1075	0.0	Choline	HMDB0000097	2	12.54
Urine	106.0499 : 0.844:+	C ₃ H ₇ NO ₃	106.0498	0.9	Serine	HMDB0000187	2	11.45
Urine	118.0863 : 0.958:+	C ₅ H ₁₁ NO ₂	118.0862	0.8	Betaine	HMDB0000043	2	12.35
Urine	133.0611 : 0.885:+	C ₄ H ₈ N ₂ O ₃	133.0607	3.0	Asparagine	HMDB0000168	2	12.87
Urine	137.0460 : 1.067:+	C ₅ H ₄ N ₄ O	137.0458	1.5	Hypoxanthine	HMDB0000157	2	6.33
Urine	137.0706 : 0.613:+	C ₇ H ₈ N ₂ O	137.0709	2.2	Methylnicotinamide	HMDB0003152	2	16.81
Urine	138.0550 : 0.891:+	C ₇ H ₇ NO ₂	138.0549	0.7	<i>p</i> -Aminobenzoic acid	HMDB0001392	2	13.31
Urine	141.0660 : 0.690:+	C ₆ H ₈ N ₂ O ₂	141.0658	1.4	Imidazole propionic acid	HMDB0002820	2	8.23
Urine	144.1020 : 0.967:+	C ₇ H ₁₃ NO ₂	144.1019	0.7	Proline betaine	HMDB0004827	1	12.88
Urine	146.0812 : 1.183:+	C ₆ H ₁₁ NO ₃	146.0811	0.7	<i>Unknown</i>	HMDB0001263	4	9.13
Urine	147.0764 : 0.911:+	C ₅ H ₁₀ N ₂ O ₃	147.0764	0.0	Glutamine	HMDB0000641	2	8.92
Urine	147.1128 : 0.583:+	C ₆ H ₁₄ N ₂ O ₂	147.1128	0.0	Lysine	HMDB0000182	2	9.50
Urine	148.0604 : 0.924:+	C ₅ H ₉ NO ₄	148.0604	0.0	Glutamic acid	HMDB0003339	2	11.26
Urine	150.0775 : 0.844:+	C ₆ H ₇ N ₅	150.0774	0.7	<i>Unknown</i>		4	25.74
Urine	156.0768 : 0.620:+	C ₆ H ₉ N ₃ O ₂	156.0767	0.6	Histidine	HMDB0000177	2	6.84
Urine	160.0970 : 1.089:+	C ₇ H ₁₃ NO ₃	160.0968	1.2	<i>Unknown</i>		2	9.27
Urine	162.1125 : 0.714:+	C ₇ H ₁₅ NO ₃	162.1124	0.6	Carnitine	HMDB0000062	2	9.60
Urine	166.0724 : 0.702:+	C ₆ H ₇ N ₅ O	166.0723	0.6	Methylguanine	HMDB0001566	2	8.66
Urine	170.0924 : 0.634:+	C ₇ H ₁₁ N ₃ O ₂	170.0924	0.0	3-Methylhistidine	HMDB0000479	1	6.14
Urine	175.1190 : 0.602:+	C ₆ H ₁₄ N ₄ O ₂	175.1189	0.6	Arginine	HMDB0000517	2	8.53
Urine	176.0666 : 0.851:+	C ₅ H ₉ N ₃ O ₄	176.0666	0.0	Guanidinosuccinic acid	HMDB0003157	2	5.62

Urine	176.1030 : 0.772:+	C ₆ H ₁₃ N ₃ O ₃	176.1029	0.6	Citrulline	HMDB0000904	2	6.46
Urine	182.0810 : 0.957:+	C ₉ H ₁₁ NO ₃	182.0811	0.5	Tyrosine	HMDB0000158	2	9.86
Urine	189.1598 : 0.603:+	C ₉ H ₂₀ N ₂ O ₂	189.1597	0.5	Trimethyllysine	HMDB0001325	2	14.16
Urine	204.1230 : 0.758:+	C ₉ H ₁₇ NO ₄	204.123	0.0	Acetylcarnitine	HMDB0000201	2	12.37
Urine	205.0972 : 0.924:+	C ₁₁ H ₁₂ N ₂ O ₂	205.0971	0.5	Tryptophan	HMDB0000929	2	28.26
Urine	217.1560 : 0.847:+	C ₁₀ H ₂₁ N ₂ O ₃	217.1546	6.4	Valylvaline	HMDB0029140	2	25.92
Urine	232.1544 : 0.794:+	C ₁₁ H ₂₁ NO ₄	232.1543	0.4	Butyrylcarnitine	HMDB0002013	2	10.77
Urine	260.1495 : 0.817:+	C ₁₂ H ₂₁ NO ₅	260.1492	1.2	<i>Unknown</i>		4	16.33
Urine	276.1441 : 0.858:+	C ₁₂ H ₂₁ NO ₆	276.1441	0.0	Glutaryl carnitine	HMDB0013130	2	7.71
Urine	286.2013 : 0.861:+	C ₁₅ H ₂₇ NO ₄	286.2013	0.0	<i>Unknown</i>		4	9.81
Urine	367.1509 : 1.084:+	C ₁₇ H ₂₂ N ₂ O ₇	367.15	2.5	Mannosyltryptophan		2	6.60
Urine	487.2120 : 0.825:+	C ₁₈ H ₃₄ N ₂ O ₁₃	487.2134	2.9	Glucosylgalactosyl hydroxylysine	HMDB0000585	2	8.54
Urine	89.0244 : 1.534:-	C ₃ H ₆ O ₃	89.0243	1.1	Lactic acid	HMDB0000190	2	27.69
Urine	105.0193 : 1.488:-	C ₃ H ₆ O ₄	105.0193	0.0	Glycerate	HMDB0000139	2	21.65
Urine	121.0295 : 1.142:-	C ₇ H ₆ O ₂	121.0294	0.8	Benzoic acid	HMDB0001870	2	11.25
Urine	124.9914 : 1.663:-	C ₂ H ₆ O ₄ S	124.9913	0.8	Ethylsulfate	HMDB0031233	2	14.32
Urine	128.0353 : 1.345:-	C ₅ H ₇ NO ₃	128.0352	0.8	Oxo-proline	HMDB0000267	2	14.51
Urine	135.0299 : 1.306:-	C ₄ H ₈ O ₅	135.0298	0.7	Threonic acid	HMDB62620	2	14.99
Urine	144.0458 : 1.192:-	C ₉ H ₇ NO	144.0454	2.8	Indole-3-carboxaldehyde	HMDB0029737	2	19.91
Urine	146.0460 : 1.972:-	C ₅ H ₉ NO ₄	146.0458	1.4	Glutamic acid	HMDB0000148	2	13.12
Urine	153.0193 : 1.576:-	C ₇ H ₆ O ₄	153.0193	0.0	Dihydroxybenzoic acid	HMDB0013677	2	16.22
Urine	161.9869 : 1.561:-	C ₄ H ₅ NO ₄ S	161.9866	1.9	Acesulfame K	HMDB0033585	1	25.13
Urine	167.0211 : 1.257:-	C ₅ H ₄ N ₄ O ₃	167.021	0.6	Uric acid	HMDB0000289	2	12.41
Urine	171.0068 : 1.755:-	C ₃ H ₉ O ₆ P	171.0063	2.9	Glycerol phosphate	HMDB0002520	2	8.54
Urine	177.0229 : 1.180:-	C ₆ H ₁₀ O ₄ S	177.0226	1.7	<i>Unknown</i>		4	26.13
Urine	178.0510 : 1.176:-	C ₉ H ₉ NO ₃	178.0509	0.6	Hippuric acid	HMDB0000714	2	22.57
Urine	181.9917 : 1.463:-	C ₇ H ₅ NO ₃ S	181.9917	0.0	Saccharin	HMDB0029723	1	8.38
Urine	182.0459 : 1.221:-	C ₈ H ₉ NO ₄	182.0458	0.5	Pyridoxic acid	HMDB0000017	2	16.91
Urine	185.0820 : 1.781:-	C ₉ H ₁₄ O ₄	185.0819	0.5	<i>Unknown</i>		4	18.92

Urine	187.0071 : 1.416:-	C ₇ H ₈ O ₄ S	187.007	0.5	<i>p</i> -Cresol sulfate	HMDB0011635	2	14.36
Urine	188.9865 : 1.444:-	C ₆ H ₆ O ₅ S	188.9862	1.6	Pyrocatechol sulfate	HMDB0059724	2	32.92
Urine	191.0552 : 1.142:-	C ₇ H ₁₂ O ₆	191.056	4.2	Quinic acid	HMDB0003072	2	13.44
Urine	193.0373 : 1.122:-	C ₇ H ₆ N ₄ O ₃	193.0366	3.6	<i>Unknown</i>		4	13.12
Urine	212.0023 : 1.357:-	C ₈ H ₇ NO ₄ S	212.0022	0.5	Indoxyl sulfate	HMDB0000682	2	12.04
Urine	225.0631 : 1.089:-	C ₈ H ₁₀ N ₄ O ₄	225.0629	0.9	5-Acetylamino-6-formylamino-3-methyluracil	HMDB0011105	2	21.23
Urine	238.0780 : 1.505:-	-	-	-	<i>Unknown</i> [M-2H] ²⁻		4	14.49
Urine	263.1037 : 1.037:-	C ₁₃ H ₁₆ N ₂ O ₄	263.1037	0.0	Phenylacetylglutamine	HMDB0006344	2	12.07
Urine	283.0827 : 1.012:-	C ₁₃ H ₁₆ O ₇	283.0823	1.4	<i>p</i> -Cresol glucuronide	HMDB0011686	2	13.97
Urine	287.0236 : 1.152:-	C ₁₁ H ₁₂ O ₇ S	287.023	2.1	Dihydroxyphenyl- γ -valerolactone sulfate	HMDB0029191	2	26.79
Urine	302.1138 : 1.016:-	C ₁₅ H ₁₇ N ₃ O ₄	302.1146	2.6	Indoleacetyl glutamine	HMDB0013240	2	7.79
Urine	308.0987 : 0.984:-	C ₁₁ H ₁₉ NO ₉	308.0987	0.0	<i>Unknown</i>		4	12.74
Urine	331.1760 : 0.964:-	C ₁₇ H ₂₄ N ₄ O ₃	331.1776	4.8	<i>Unknown</i>		4	10.67
Urine	377.0170 : 1.040:-	C ₁₇ H ₆ N ₄ O ₇	377.0164	1.6	<i>Unknown</i>		4	20.45
Urine	473.1453 : 0.934:-	C ₂₄ H ₂₅ O ₁₀	473.1448	1.1	Enterolactone glucuronide	--	2	14.37
Urine	632.2044 : 0.874:-	C ₂₃ H ₃₉ NO ₁₉	632.2043	0.2	Sialyllactose	HMDB0000825	2	9.78
Urine	112.0515 : 0.710:-	C ₄ H ₇ N ₃ O	112.0516	0.9	Creatinine	HMDB0000562	2	9.61
Plasma	76.0402 : 0.732:+	C ₂ H ₅ NO ₂	76.0393	11.8	Glycine	HMDB0000123	2	8.17
Plasma	90.0557 : 0.783:+	C ₃ H ₇ NO ₂	90.0549	8.9	Alanine	HMDB0000161	2	4.15
Plasma	104.1075 : 0.592:+	C ₅ H ₁₄ NO	104.1075	0.0	Choline	HMDB0000097	2	30.39
Plasma	106.0500 : 0.864:+	C ₃ H ₇ NO ₃	106.0498	1.9	Serine	HMDB0000187	2	3.23
Plasma	114.0662 : 0.635:+	C ₄ H ₇ N ₃ O	114.0653	7.9	Creatinine	HMDB0000562	2	32.14
Plasma	116.0705 : 0.927:+	C ₅ H ₉ NO ₂	116.0706	0.9	Proline	HMDB0000162	2	3.55
Plasma	118.0618 : 0.718:+	C ₃ H ₇ N ₃ O ₂	118.0611	5.9	Guanidoacetic acid	HMDB0000128	2	9.93
Plasma	120.0654 : 0.900:+	C ₄ H ₉ NO ₃	120.0655	0.8	Threonine	HMDB0000167	2	8.14
Plasma	129.0656 : 0.75:+	C ₅ H ₈ N ₂ O ₂	129.0658	1.5	<i>Unknown</i>		3	6.84
Plasma	132.0766 : 0.765:+	C ₄ H ₉ N ₃ O ₂	132.0767	0.8	Creatine	HMDB0000064	2	4.80
Plasma	132.1017 : 0.873:+	C ₆ H ₁₃ NO ₂	132.1019	1.5	Leucine/isoleucine	HMDB0000687	2	5.20

Plasma	133.0573 : 0.901:+	C ₄ H ₈ N ₂ O ₃	133.0607	25.6	Asparagine	HMDB0000168	2	4.20
Plasma	137.0459 : 1.066:+	C ₅ H ₄ N ₄ O	137.0458	0.7	Hypoxanthine	HMDB0000157	2	12.85
Plasma	144.0988 : 0.984:+	C ₇ H ₁₃ NO ₂	144.1013	17.3	Proline betaine	HMDB0004827	2	45.39
Plasma	146.1182 : 0.699:+	C ₇ H ₁₅ NO ₂	146.1176	4.1	Deoxycarnitine	HMDB0001161	3	26.48
Plasma	147.0761 : 0.922:+	C ₅ H ₁₀ N ₂ O ₃	147.0763	1.4	Glutamine	HMDB0000641	2	3.81
Plasma	148.0603 : 0.934:+	C ₅ H ₉ NO ₄	148.0603	0.0	Glutamic acid	HMDB0003339	2	8.50
Plasma	150.0583 : 0.910:+	C ₅ H ₁₁ NO ₂ S	150.0583	0.0	Methionine	HMDB0000696	2	4.65
Plasma	152.0567 : 0.910:+	C ₅ H ₅ N ₅ O	152.0563	2.6	Guanine	HMDB0000132	2	34.61
Plasma	156.0766 : 0.649:+	C ₆ H ₉ N ₃ O ₂	156.0763	1.9	Histidine	HMDB0000177	2	31.32
Plasma	160.1332 : 0.725:+	C ₈ H ₁₇ NO ₂	160.1323	5.6	2-Aminooctanoic acid	HMDB0000991	2	22.11
Plasma	162.0761 : 0.933:+	C ₆ H ₁₁ NO ₄	162.0753	4.9	Aminoadipic acid	HMDB0000510	2	14.34
Plasma	162.1123 : 0.735:+	C ₇ H ₁₅ NO ₃	162.1123	0.0	Carnitine	HMDB0000062	2	3.89
Plasma	166.086 : 0.9355:+	C ₉ H ₁₁ NO ₂	166.0853	4.2	Phenylalanine	HMDB0000159	2	12.34
Plasma	170.0922 : 0.663:+	C ₇ H ₁₁ N ₃ O ₂	170.0923	0.6	Methylhistidine	HMDB0000479	2	15.47
Plasma	175.1191 : 0.631:+	C ₆ H ₁₄ N ₄ O ₂	175.1183	4.6	Arginine	HMDB0000517	2	37.36
Plasma	176.1025 : 0.943:+	C ₆ H ₁₃ N ₃ O ₃	176.1023	1.1	Citrulline	HMDB0000904	2	4.25
Plasma	182.081 : 0.9616:+	C ₉ H ₁₁ NO ₃	182.0803	3.8	Tyrosine	HMDB0000158	2	3.52
Plasma	189.1337 : 0.635:+	C ₇ H ₁₆ N ₄ O ₂	189.1343	3.2	Monomethylarginine	HMDB0029416	2	44.31
Plasma	202.1807 : 0.793:+	C ₁₁ H ₂₃ NO ₂	202.1802	2.5	<i>Unknown</i>		4	56.73
Plasma	203.1499 : 0.680:+	C ₈ H ₁₈ N ₄ O ₂	203.1493	3.0	Dimethylarginine	HMDB0003334/ HMDB0001539	2	11.69
Plasma	204.1233 : 0.776:+	C ₉ H ₁₇ NO ₄	204.1223	4.9	Acetylcarnitine	HMDB0000201	2	4.50
Plasma	205.0966 : 0.931:+	C ₁₁ H ₁₂ N ₂ O ₂	205.0963	1.5	Tryptophan	HMDB0000929	2	17.46
Plasma	241.0289 : 0.950:+	C ₆ H ₁₂ N ₂ O ₄ S ₂	241.0303	5.8	Cystine (disulfide)	HMDB0000192	2	5.04
Plasma	247.1441 : 1.146:+	C ₁₄ H ₁₈ N ₂ O ₂	247.1433	3.2	Tryptophan betaine	HMDB0061115	2	14.53
Plasma	298.0526 : 0.823:+	C ₈ H ₁₅ N ₃ O ₅ S ₂	298.0523	1.0	Cysteinylglycine disulfide	HMDB0000709	2	7.87
Plasma	87.0087 : 1.301:-	C ₃ H ₄ O ₃	87.00874	0.5	Pyruvic acid	HMDB0000243	2	14.14
Plasma	89.0244 : 1.136:-	C ₃ H ₆ O ₃	89.02439	0.1	Lactic acid	HMDB0000190	2	8.13
Plasma	103.0400 : 1.019:-	C ₄ H ₈ O ₃	103.04	0.0	3-Hydroxybutyric acid	HMDB0000357	2	6.91
Plasma	103.0400 : 1.043:-	C ₄ H ₈ O ₃	103.04	0.0	2-Hydroxybutyric acid	HMDB0000008	2	11.53

Plasma	115.0400 : 1.078:-	C ₅ H ₈ O ₃	115.04	0.0	Alpha-ketoisovaleric acid	HMDB0000019	2	15.83
Plasma	117.0193 : 1.766:-	C ₄ H ₆ O ₄	117.0193	0.0	Succinic acid	HMDB0000254	2	31.01
Plasma	128.0353 : 1.018:-	C ₅ H ₇ NO ₃	128.0352	0.8	Oxo-proline	HMDB0000267	2	11.92
Plasma	129.0557 : 1.029:-	C ₆ H ₁₀ O ₃	129.0556	0.8	3-methyl-2-oxovaleric acid	HMDB0000491	2	13.47
Plasma	132.0302 : 1.073:-	C ₄ H ₇ NO ₄	132.0302	0.0	Aspartic acid	HMDB0006483	2	12.40
Plasma	133.0142 : 1.783:-	C ₄ H ₆ O ₅	133.0142	0.0	Malic acid	HMDB0000744	2	32.51
Plasma	167.021 : 0.969:-	C ₅ H ₄ N ₄ O ₃	167.021	0.0	Uric acid	HMDB0000289	2	11.10
Plasma	179.0561 : 0.999:-	C ₆ H ₁₂ O ₆	179.056	0.6	Glucose	HMDB0000122	2	11.73
Plasma	191.0197 : 1.967:-	C ₆ H ₈ O ₇	191.0197	0.0	Citric acid	HMDB0000094	2	21.51
Plasma	195.051 : 0.889:-	C ₆ H ₁₂ O ₇	195.051	0.0	Gluconic acid	HMDB0000625	2	13.37
Plasma	C14:0 : 10.15	C ₁₄ H ₂₈ O ₂	227.2017	-	Myristic acid	HMDB0000806	2	19.55
Plasma	C15:0 : 10.79	C ₁₅ H ₃₀ O ₂	241.2173	-	Pentadecanoic acid	HMDB0000826	2	13.63
Plasma	C16:0 : 11.43	C ₁₆ H ₃₂ O ₂	255.233	-	Palmitic acid	HMDB0000220	2	12.70
Plasma	C16:1 1 : 11.74	C ₁₆ H ₃₀ O ₂	253.2173	-	Hexadecenoic acid	HMDB0037647	2	13.16
Plasma	C16:1 2 : 11.82	C ₁₆ H ₃₀ O ₂	253.2173	-	Palmitoleic acid	HMDB0003229	2	14.46
Plasma	C17:0 : 12.06	C ₁₇ H ₃₄ O ₂	269.2486	-	Margaric acid	HMDB0002259	2	12.27
Plasma	C18:0 : 12.77	C ₁₈ H ₃₆ O ₂	283.2643	-	Stearic acid	HMDB0000827	2	13.24
Plasma	C18:1 1 : 13.20	C ₁₈ H ₃₄ O ₂	281.2486	-	Elaidic acid	HMDB0000573	2	11.85
Plasma	C18:1 2 : 13.26	C ₁₈ H ₃₄ O ₂	281.2486	-	Oleic acid	HMDB0000207	2	8.50
Plasma	C18:2 1 : 13.92	C ₁₈ H ₃₂ O ₂	279.233	-	Linoleic acid	HMDB0000673	2	13.94
Plasma	C18:2 2 : 15.02	C ₁₈ H ₃₂ O ₂	279.233	-	Linoelaidic acid	HMDB0006270	2	33.45
Plasma	C18:3 1 : 14.47	C ₁₈ H ₃₀ O ₂	277.2173	-	γ-Linolenic acid	HMDB0003073	2	17.16
Plasma	C18:3 2 : 14.82	C ₁₈ H ₃₀ O ₂	277.2173	-	α-Linolenic acid	HMDB0001388	2	14.86
Plasma	C20:0 : 14.37	C ₂₀ H ₄₀ O ₂	311.2956	-	Arachidic acid	HMDB0002212	2	13.34
Plasma	C20:1 : 14.88	C ₂₀ H ₃₈ O ₂	309.2799	-	Eicosenoic acid	HMDB0002231	2	30.24
Plasma	C20:2 : 15.75	C ₂₀ H ₃₆ O ₂	307.2643	-	Eicosadienoic acid	HMDB0005060	2	19.19
Plasma	C20:3 1 : 16.41	C ₂₀ H ₃₄ O ₂	305.2486	-	Eicosatrienoic acid	HMDB0002925	2	11.94
Plasma	C20:3 2 : 17.44	C ₂₀ H ₃₄ O ₂	305.2486	-	Unknown		4	18.70
Plasma	C20:4 : 16.94	C ₂₀ H ₃₂ O ₂	303.233	-	Arachidonic acid	HMDB0001043	2	13.71

Plasma	C20:5 : 18.18	C ₂₀ H ₃₀ O ₂	301.2173	-	Eicosapentaenoic acid	HMDB0001999	2	17.27
Plasma	C22:0 : 16.34	C ₂₂ H ₄₄ O ₂	339.3269	-	Behenic acid	HMDB0000944	2	13.57
Plasma	C22:5 1 : 19.43	C ₂₂ H ₃₄ O ₂	329.2486	-	<i>Unknown</i>		4	13.76
Plasma	C22:5 2 : 20.8	C ₂₂ H ₃₄ O ₂	329.2486	-	Docosapentaenoic	HMDB0006528	2	14.12
Plasma	C22:6 : 21.34	C ₂₂ H ₃₂ O ₂	327.233	-	Docosahexaenoic acid	HMDB0062579	2	12.88
Plasma	C24:0 : 18.63	C ₂₄ H ₄₈ O ₂	367.3582	-	Lignoceric acid	HMDB0002003	2	12.15
Plasma	C24:1 :19.28	C ₂₄ H ₄₆ O ₂	365.3425	-	Nervonic acid	HMDB0002368	2	10.88

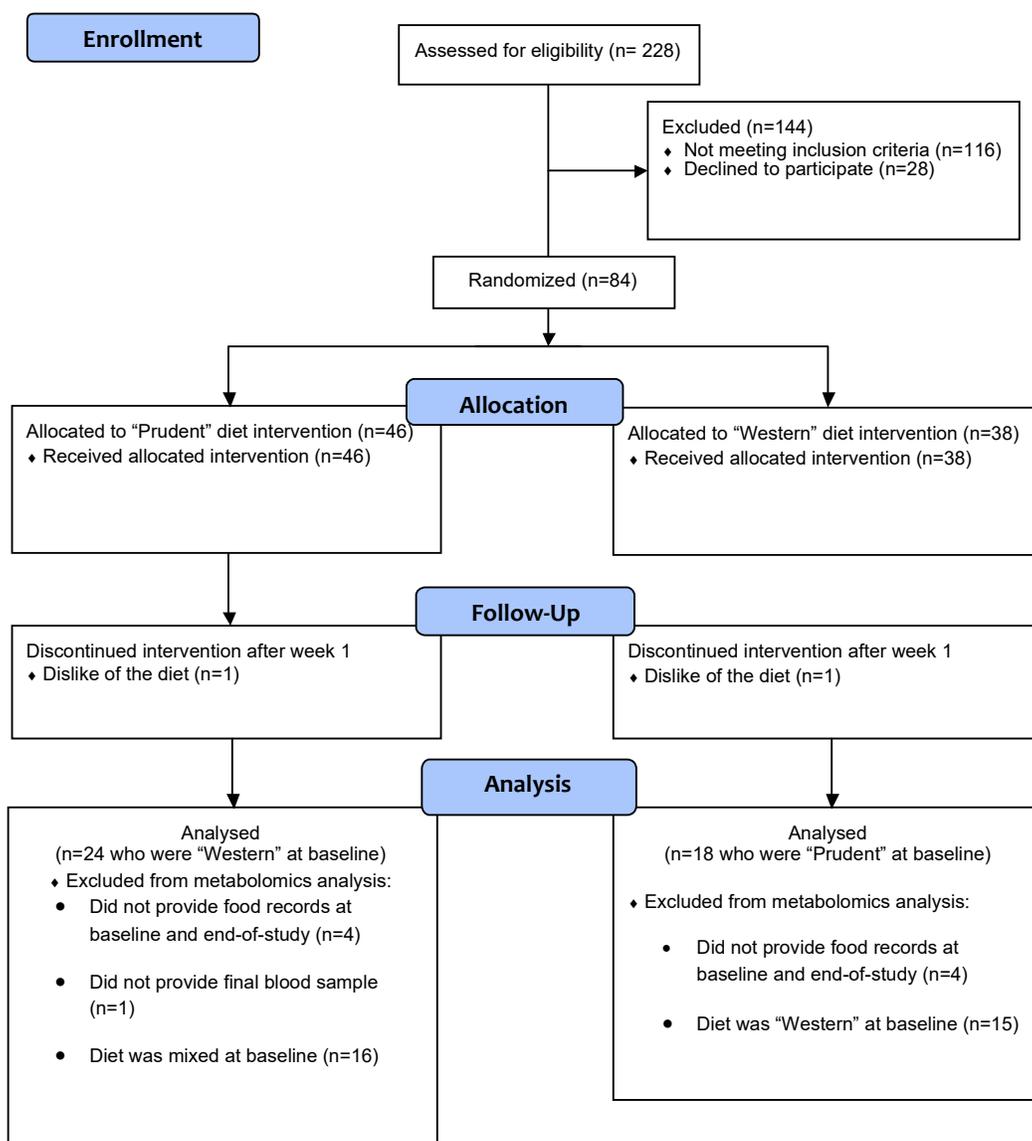


Figure 4S.1: A CONSORT flow diagram outlining selection criteria used in a parallel two-arm randomized clinical trial involving participants from the DIGEST study, where metabolomic analyses was performed unblinded on paired serum and urine samples collected at baseline and 2 weeks following a provisional Prudent or Western diet. Overall, 73 of the 84 participants who completed DIGEST had available specimens and complete food records. However, in order to maximize the effect size of this short-term dietary intervention, metabolomic analyses was performed only on a subset of participants ($n = 42$) who had contrasting habitual diets at baseline as evaluated based on an aggregate diet quality score.

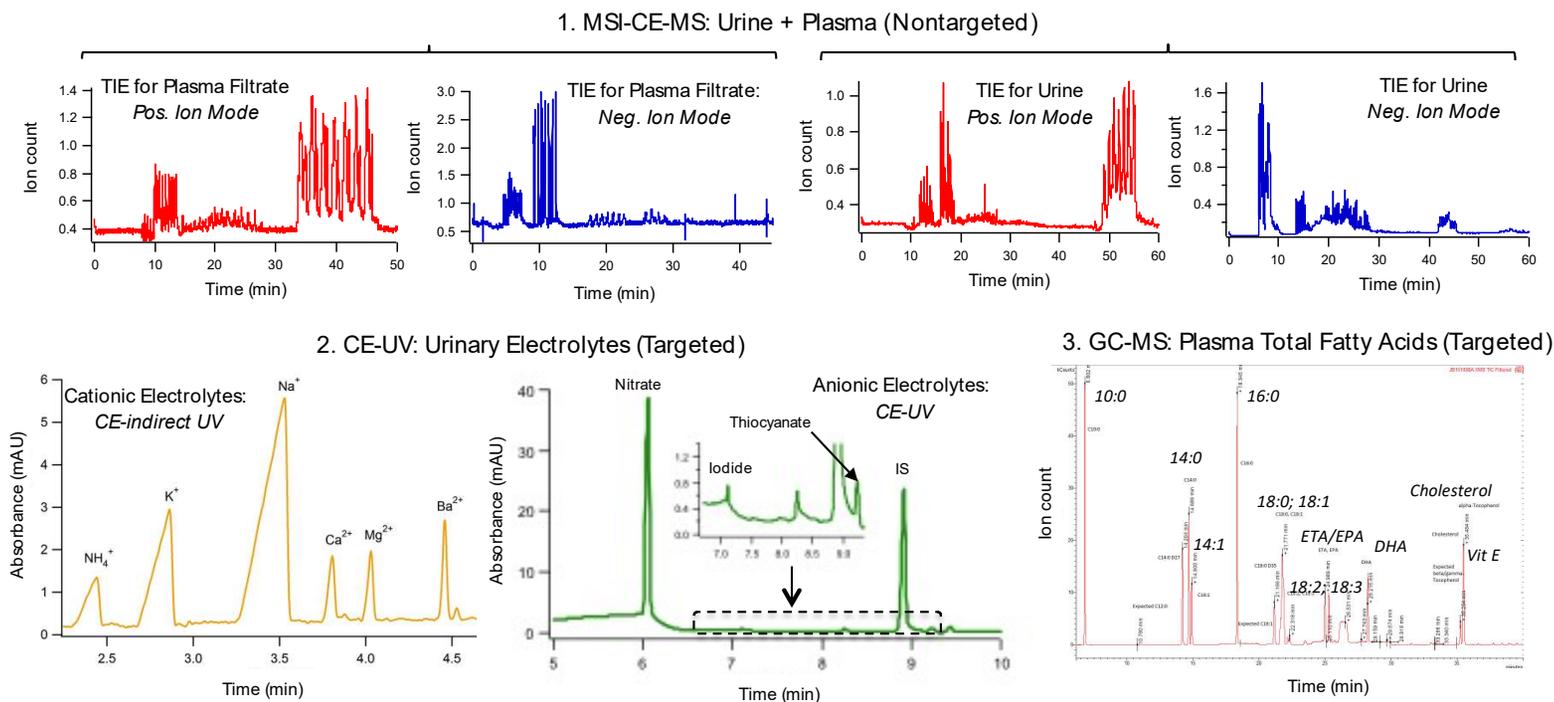


Figure 4S.2: Overview of three major analytical platforms for nontargeted and targeted metabolite/electrolyte profiling of matching plasma and urine specimens from DIGEST participants at baseline and 2 weeks following an assigned Prudent or Western diet from food provisions. MSI-CE-MS was used as the major format for metabolomic analyses of a wide range of polar/ionic metabolites from both plasma filtrate and diluted urine samples when using a stringent data workflow for metabolite authentication with quality control. Also, CE with (indirect/direct) UV absorbance detection was used for targeted analysis of electrolytes in urine, including inorganic cations (e.g., sodium) and anions (e.g., nitrate), whereas GC-MS was applied for targeted analysis of total (hydrolyzed) fatty acids as their FAMES from plasma extracts.

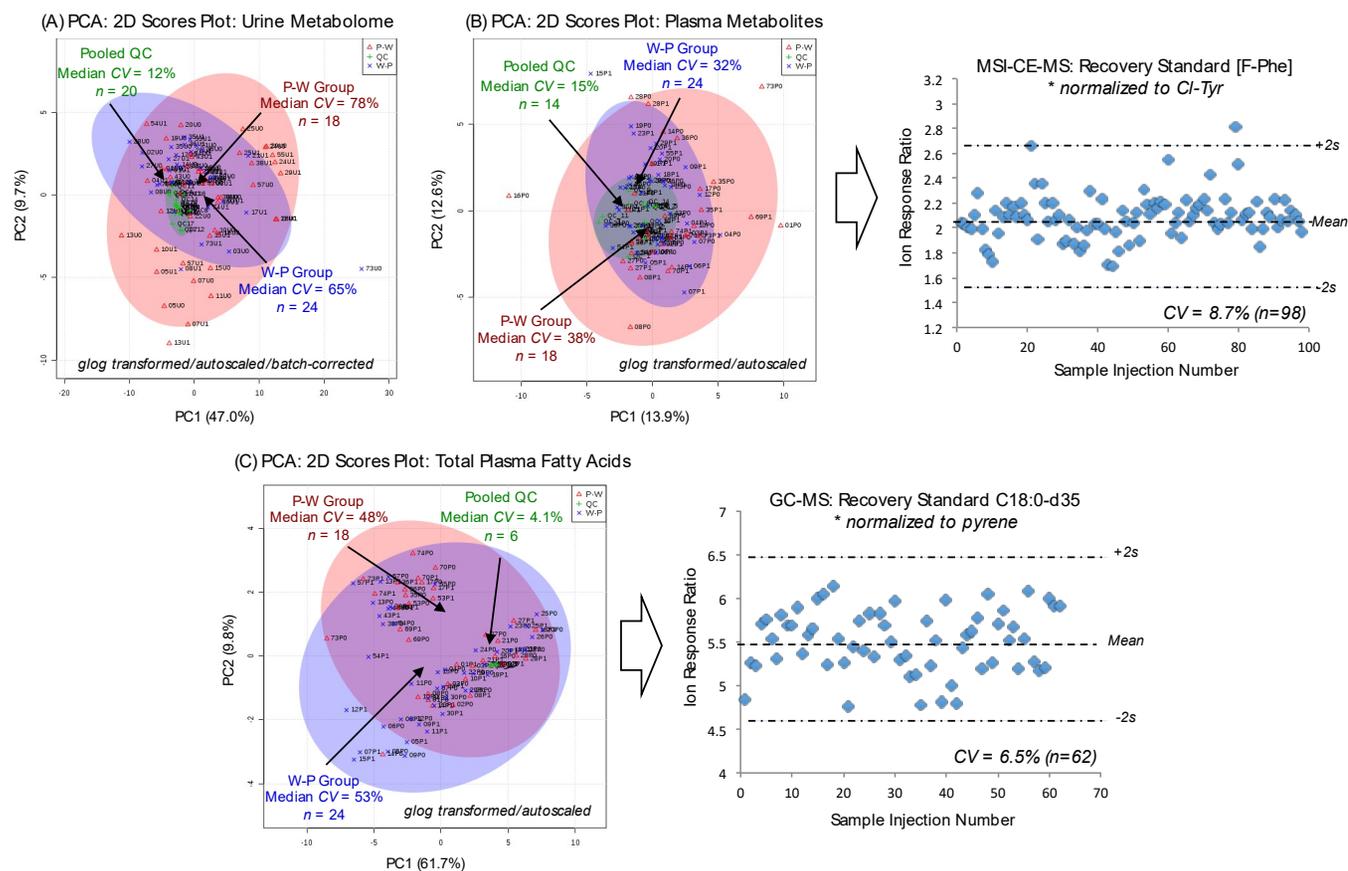


Figure 4S.3: 2D scores plots from PCA and control charts for recovery standards that highlight the good technical precision as compared to biological variance of metabolomic data from three instrumental platforms, including (A) 84 authenticated metabolites in urine after QC-based batch-correction, (B) 80 polar/ionic metabolites from plasma using MSI-CE-MS and (C) 18 plasma (total) fatty acids using GC-MS.

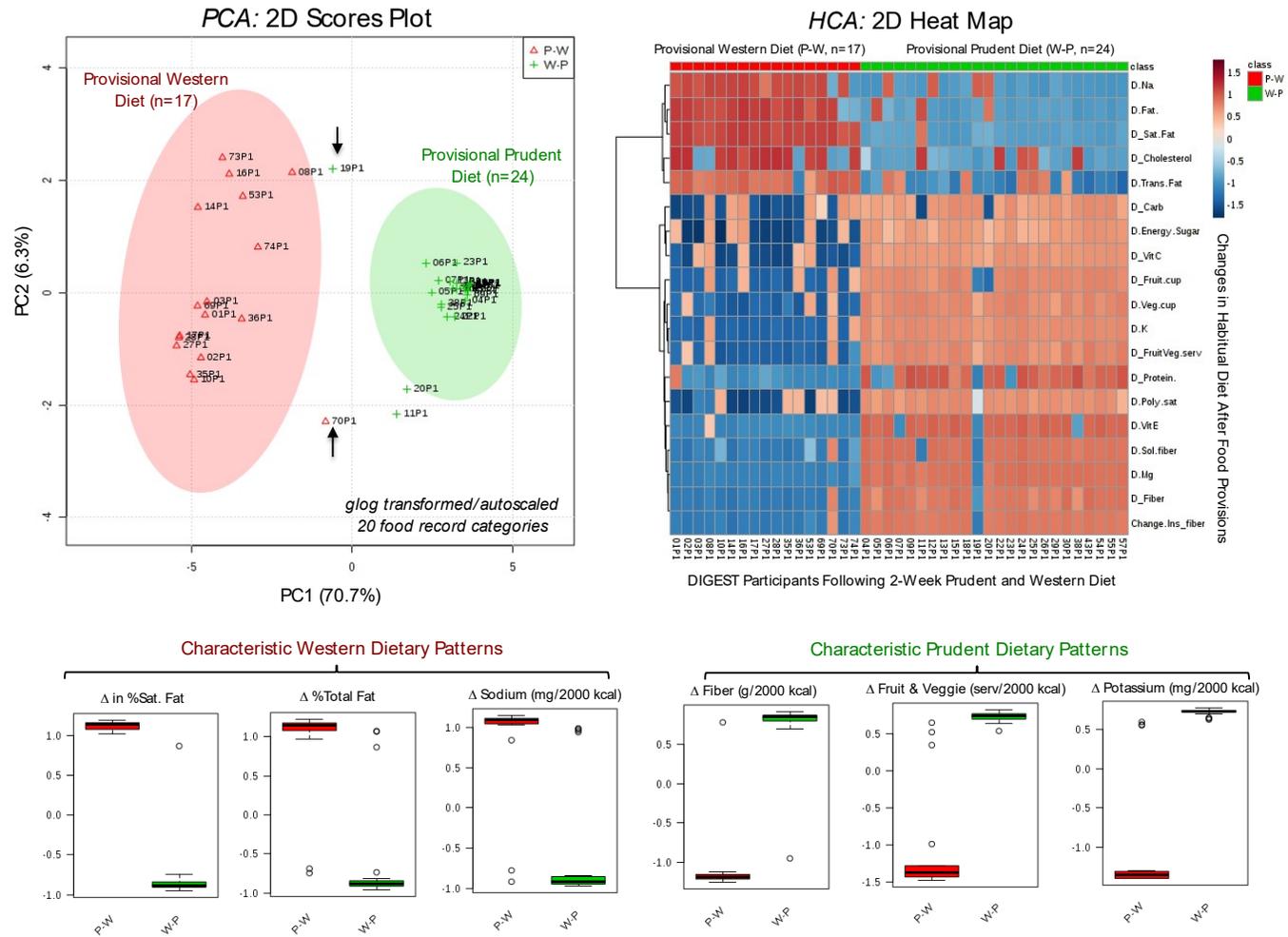


Figure 4S.4: 2D scores plot from PCA and HCA/heat map summarizing changes in habitual diet (W-P and P-W) based on 20 macro-/micronutrient categories self-reported food records from DIGEST participants following 2 weeks of contrasting food provisions as compared to their baseline diet.

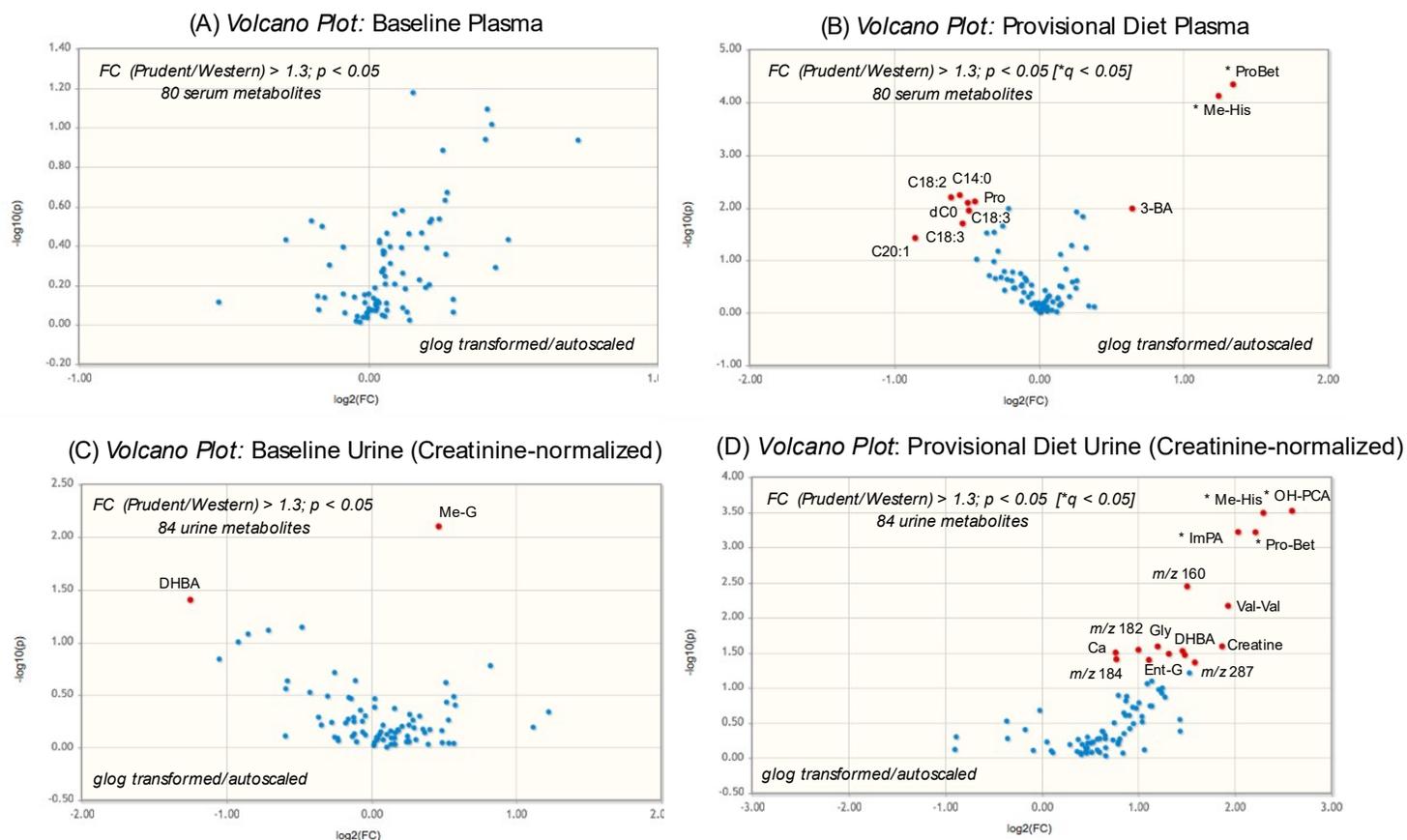


Figure 4S.5: Series of volcano plots for plasma (A, B) and urine (C, D) metabolome data highlighting few differences in the metabolic phenotype among DIGEST participants ($n = 42$) at baseline as compared to major changes following two weeks of contrasting diets from food provisions based on a minimum threshold for significance (mean FC > 1.3, $p < 0.05$) between Prudent and Western assigned groups. Abbreviations refer to DHBA (dihydroxybenzoic acid), ProBet (proline betaine), Me-His (3-methylhistidine), 3-OH-BA (3-hydroxybutyric acid), dC0 (deoxycarnitine), Pro (proline), OH-PCA (hydroxypiperic acid), ImPA (imidazolepropionic acid), Ent-G (enterolactone glucuronide), DHBA (dihydroxybenzoic acid) and Me-G (methylguanine), whereas standard notation is used for plasma fatty acids and unknown ions are denoted by their accurate mass (m/z).

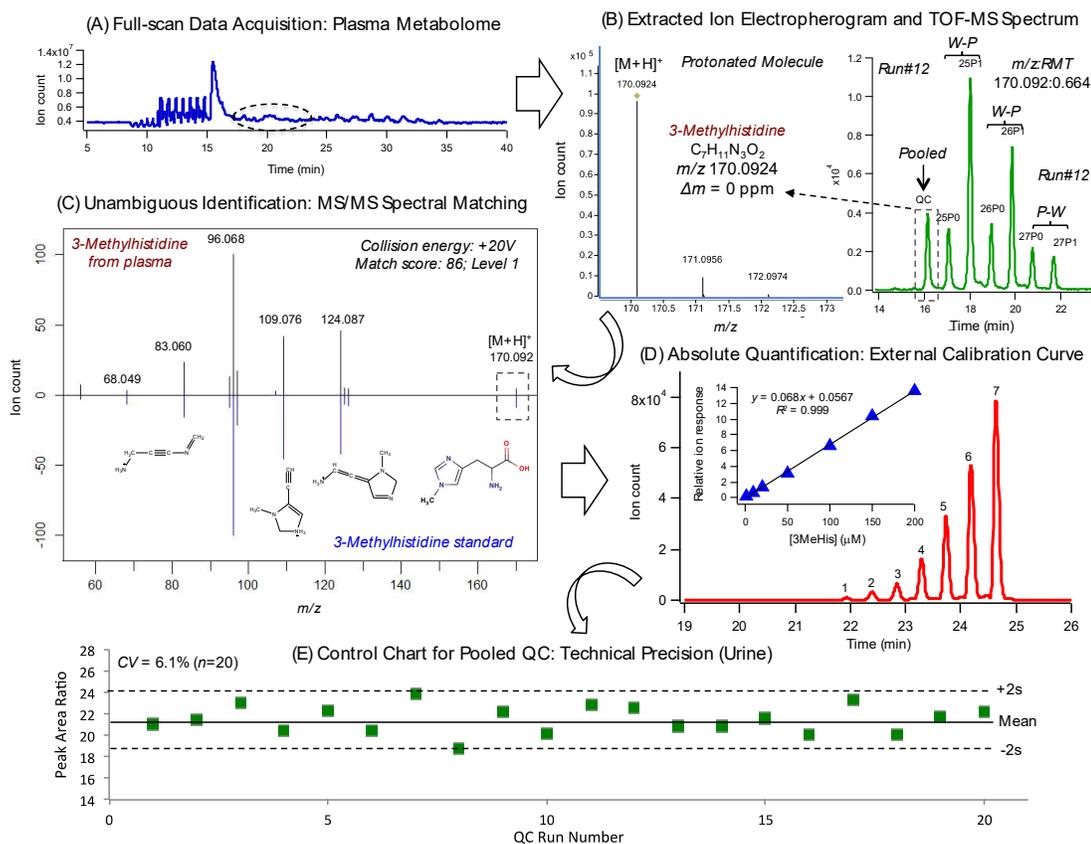


Figure 4S.6: (A) Metabolomics data workflow for the identification and quantification of biomarkers of a provisional Prudent diet (e.g., 3-methylhistidine annotated based on its m/z :RMT) when using full-scan data acquisition. (B) Multiplexed separations by MSI-CE-MS based on serial injection of seven plasma filtrate (or diluted urine) samples within a single run, including paired samples from each DIGEST participant (i.e., baseline/post-treatment) together with a pooled sample as QC for assessing technical precision and long-term signal drift. High resolution MS under positive ion mode detection allows for determination of most likely molecular formula for unknown cation (i.e., protonated molecular ion), whereas (C) MS/MS spectra is used for its structural elucidation when compared with an authentic standard. (D) Quantification of metabolites is then performed by external calibration when using an internal standard (Cl-Tyr) for data normalization by MSI-CE-MS. (E) A control chart for Me-His from pooled urine samples as QC analyzed in random positions in every run demonstrates acceptable technical precision over 3 days.

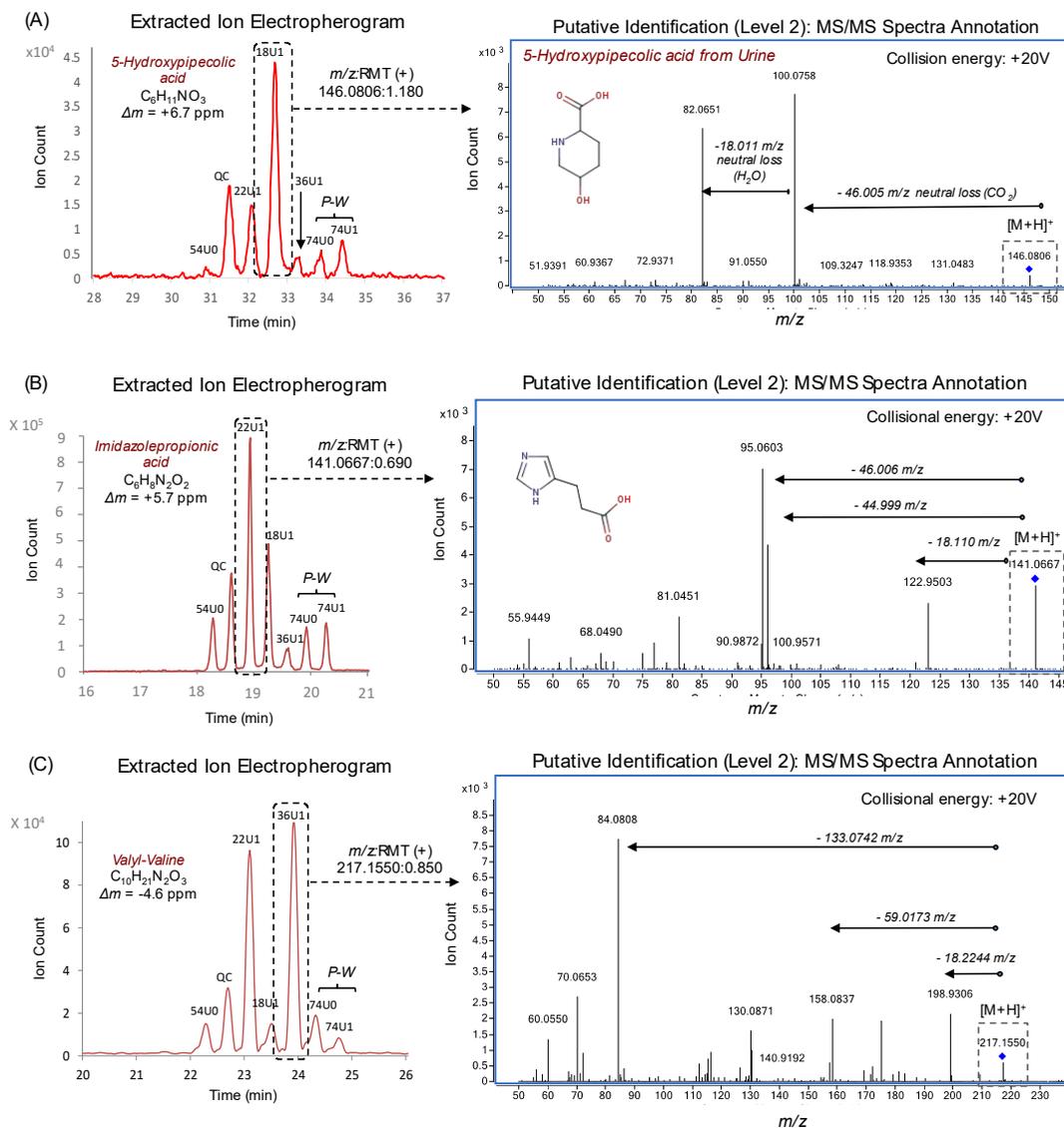


Figure 4S.7: Putative identification (level 2) of three unknown cationic metabolites ($[M-H]^+$) detected in urine specimens from DIGEST participants that were significantly elevated in assigned Prudent urine samples ($p < 0.05$) as compared to Western diet following 2 weeks of food provisions, namely (A) 5-hydroxypipelic acid (OH-PCA, m/z 146.081), (B) imidazole propionic acid (ImPA, m/z 141.067) and (C) Valinyl-valine (Val-Val, m/z 217.155) based on their characteristic MS/MS spectra (optimal collision energy at 20 V) under positive ion mode conditions.

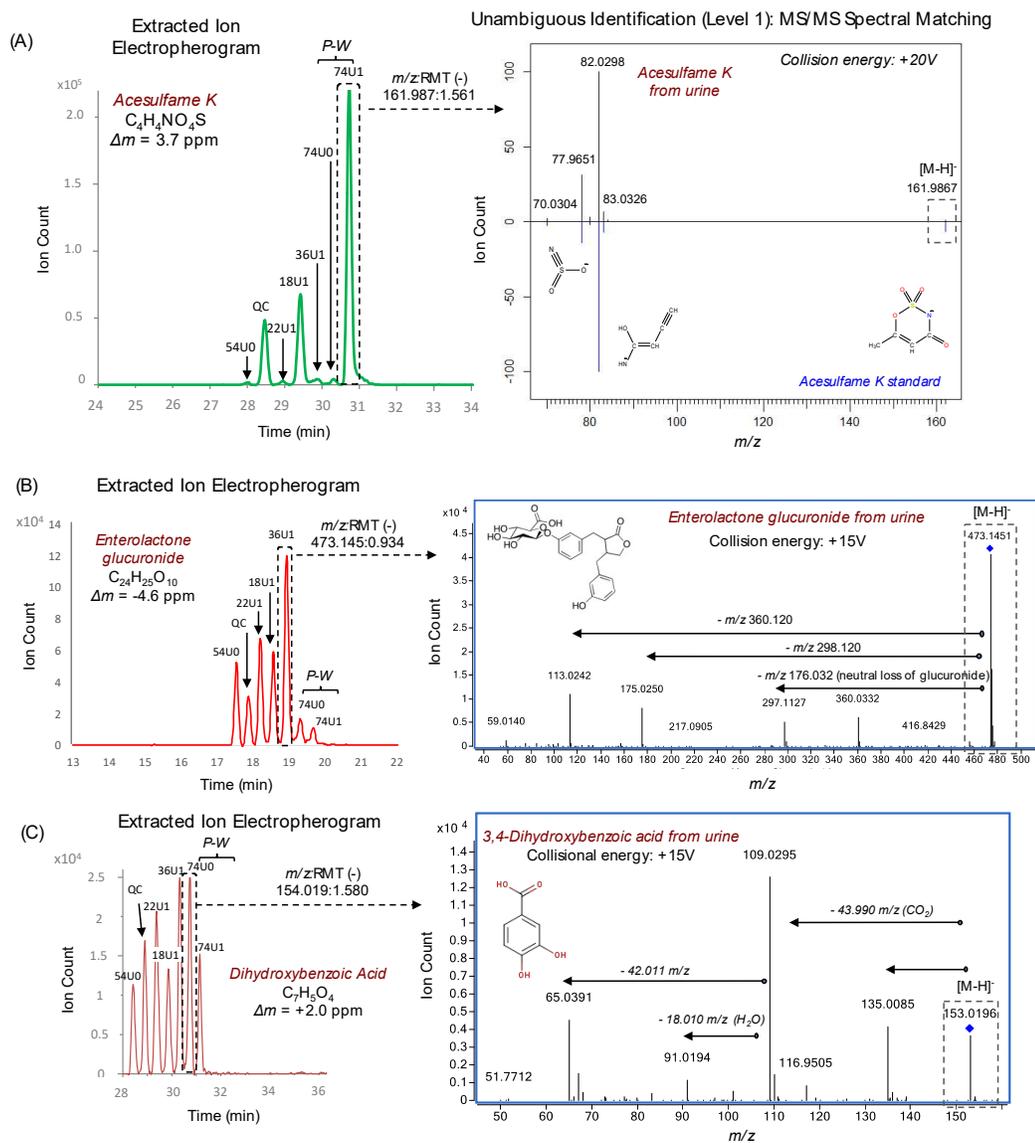


Figure 4S.8: Unambiguous (level 1) and putative identification (level 2) of three unknown anionic metabolites ($[M-H]^-$) detected in urine specimens from DIGEST participants that were decreased (acesulfame K) or elevated (enterolactone glucuronide and dihydroxybenzoic acid) in assigned Prudent ($p < 0.05$) as compared to Western diet groups following 2 weeks of food provisions, namely (A) acesulfame K (ASK, m/z 161.987), (B) enterolactone glucuronide (Ent-G, m/z 473.145) and (C) dihydroxybenzoic acid (DHBA, m/z 153.020) based on their characteristic MS/MS spectra under negative ion mode conditions. Identification was derived from comparison of MS/MS spectra with a standard as shown in mirror plot for ASK together with spiking into urine sample to confirm co-migration. The likely stereochemistry for DHBA was deduced from comparison of *in silico* MS/MS spectra when using HMDB, whereas Ent-G was tentatively identified based on comparison with published MS/MS spectra [Johnson et al. *Metabolites* **2013**, 3: 658-672].

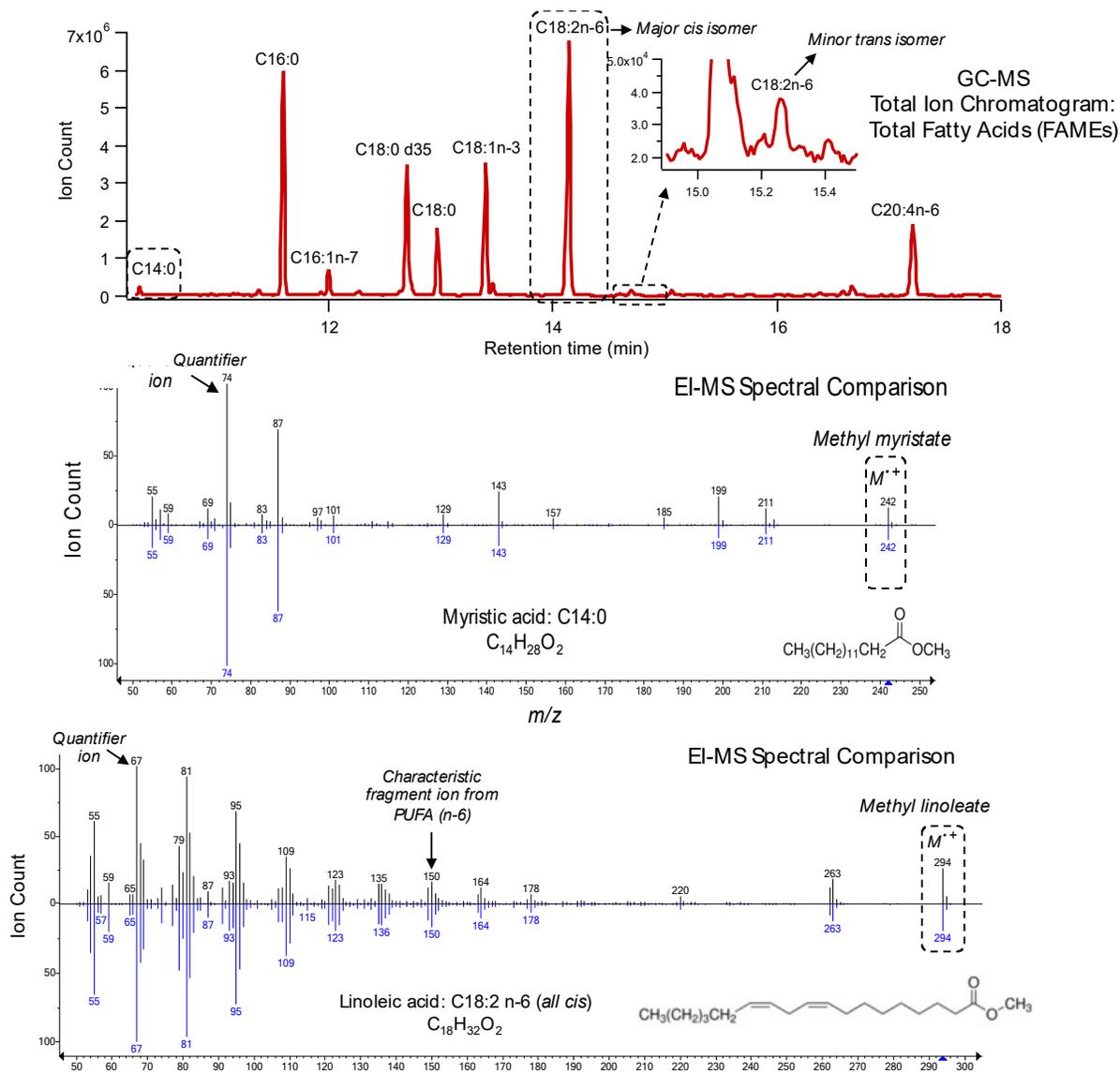


Figure 4S.9: Unambiguous identification (level 1) of total (hydrolyzed) fatty acids as their FAME derivatives from plasma extracts when using GC-MS, including resolution of low abundance trans-isomer (linoelaidic acid, C₁₈:2n-6trans) from major cis-isomer (linoleic acid, C₁₈:2n-6cis), including detection of minor saturated fatty acids (myristic acid, C₁₄:0). Mirror plots for EI-MS spectra show excellent matches when comparing FAMEs detected in plasma extracts as compared to their references in the NIST database, where the base peak ions correspond to the quantifier ions monitored for saturated and polyunsaturated fatty acids. Overall, plasma total C₁₈:2n-6trans was only 0.34% of its major stereoisomer C₁₈:2n-6cis which also represents the most abundant fatty acid measured in circulation.

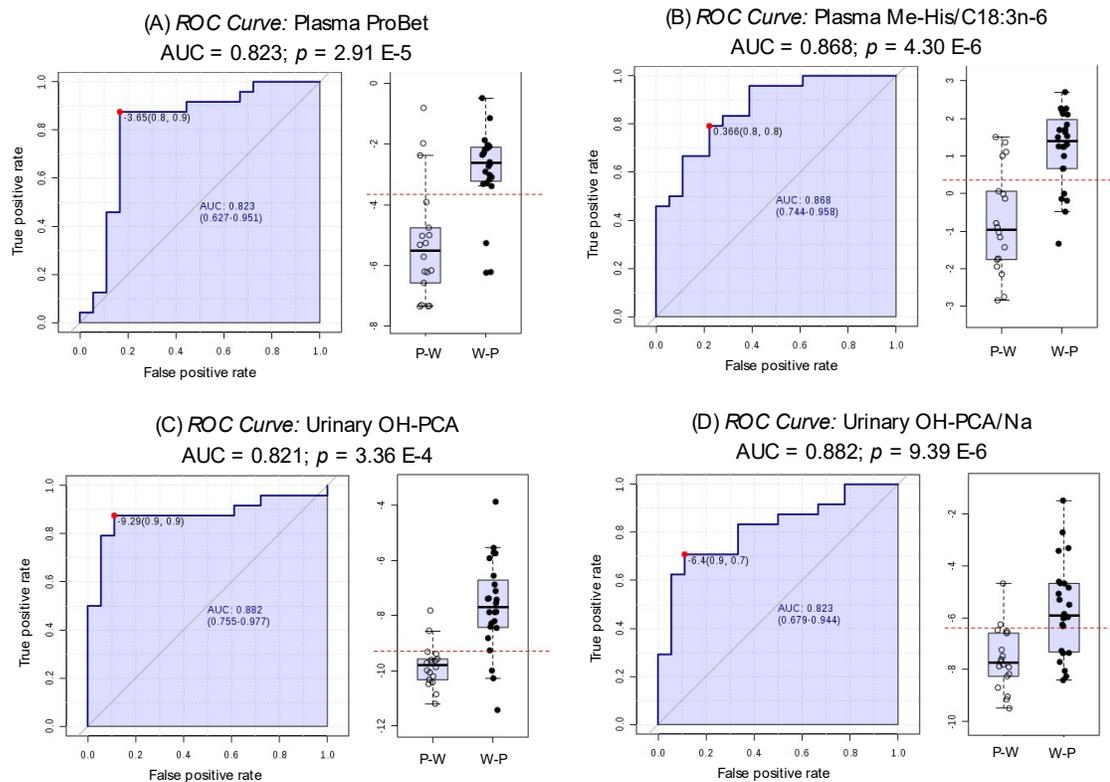


Figure 4S.10: Top-ranked single and ratiometric metabolites for differentiation of DIGEST participants assigned a Prudent (W-P, $n = 24$) or Western (P-W, $n = 18$) diet following 2 weeks of food provisions when using receiver operating characteristic (ROC) curves. All metabolites were glog-transformed, whereas urine metabolites were normalized to creatinine following a QC-based batch correction. Overall, there was good discrimination of contrasting dietary patterns ($\text{AUC} > 0.820$; $p < 0.001$) as shown for (A) plasma proline betaine (ProBet), (B) plasma 3-methylhistidine to α -linoleic acid (MeHis/C18:3n-6) ratio, (C) urinary hydroxypipercolic acid (OH-PCA), and (D) urinary hydroxypipercolic acid to sodium ratio (OH-PCA/Na).

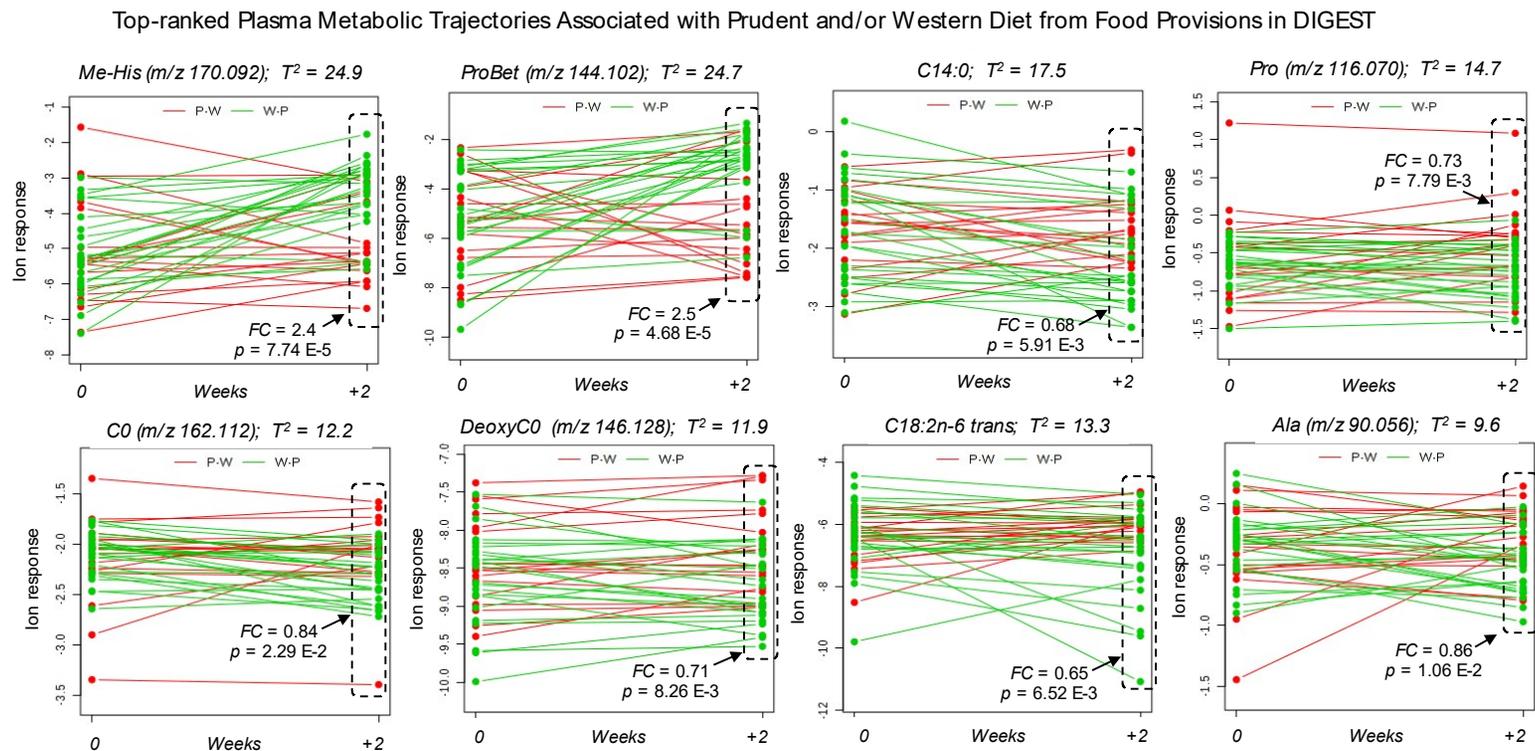


Figure 4S.11: Top-ranked plasma metabolites associated with contrasting Prudent (W-P) and Western (P-W) diets from food provisions when using glog-transformed ion responses measured at baseline (habitual diet, 0) and following food provisions (2 weeks) for DIGEST participants ($n = 42$). Metabolic trajectories from time course studies were ranked based on Hotelling- T^2 values when using multivariate empirical Bayes analysis of variance (MEBA) together with student's t -test to confirm statistical significance ($p < 0.05$).

Top-ranked Urinary Metabolic Trajectories Associated with Prudent and/or Western Diet from Food Provisions in DIGEST

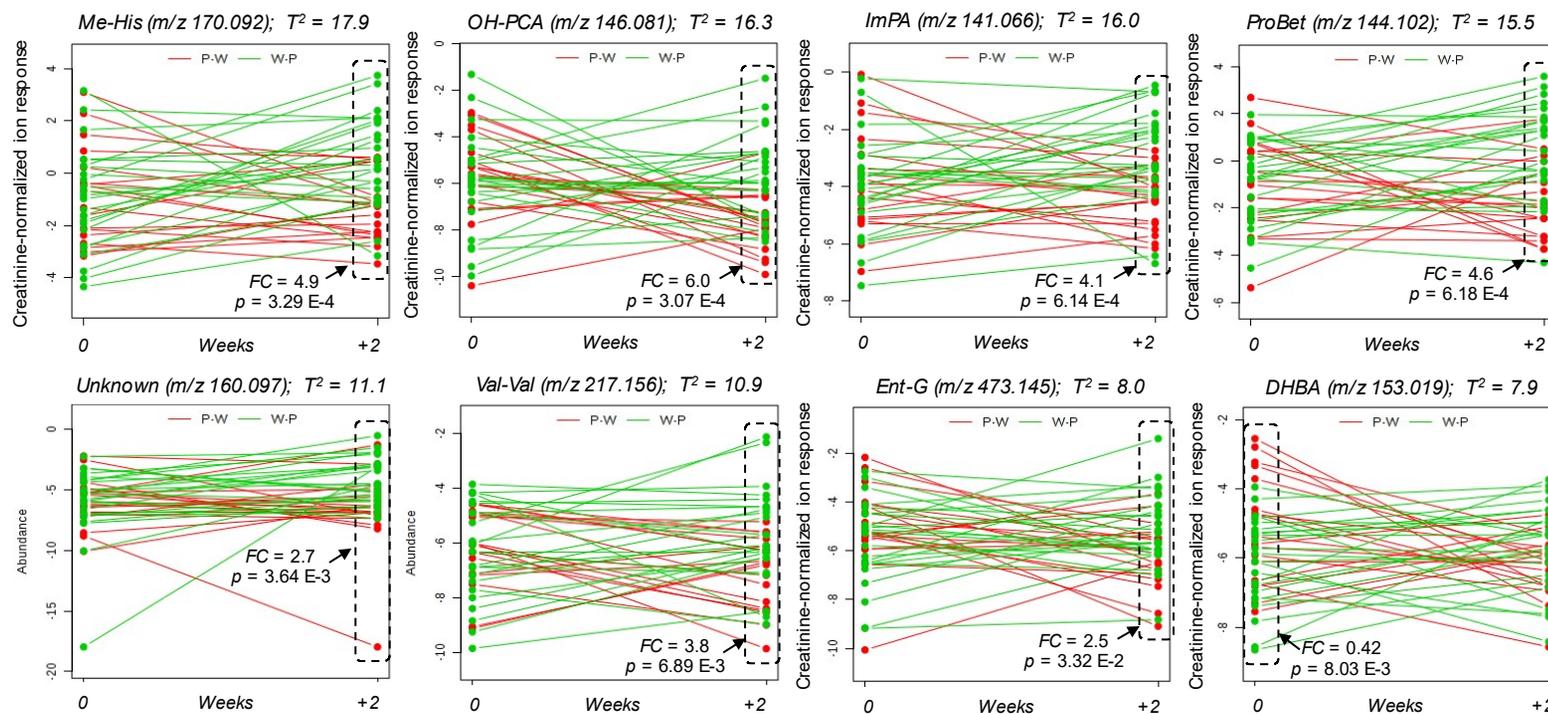


Figure 4S.12: Top-ranked urinary metabolites associated with contrasting Prudent (W-P) and Western (P-W) diets from food provisions when using glog-transformed ion responses normalized to creatinine measured at baseline (habitual diet, 0) and following food provisions (2 weeks) for DIGEST participants ($n = 42$). Metabolic trajectories from time course studies were ranked based on Hotelling- T^2 values when using multivariate empirical Bayes analysis of variance (MEBA) together with student's t -test to confirm statistical significance ($p < 0.05$), with the exception for DHBA.

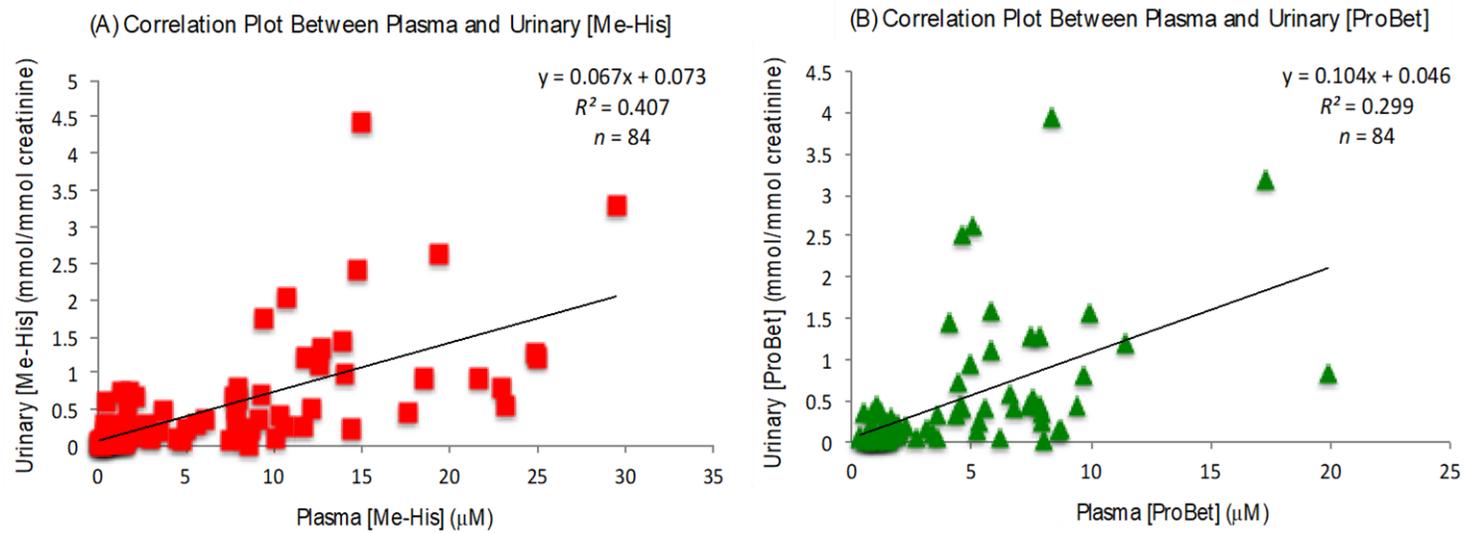
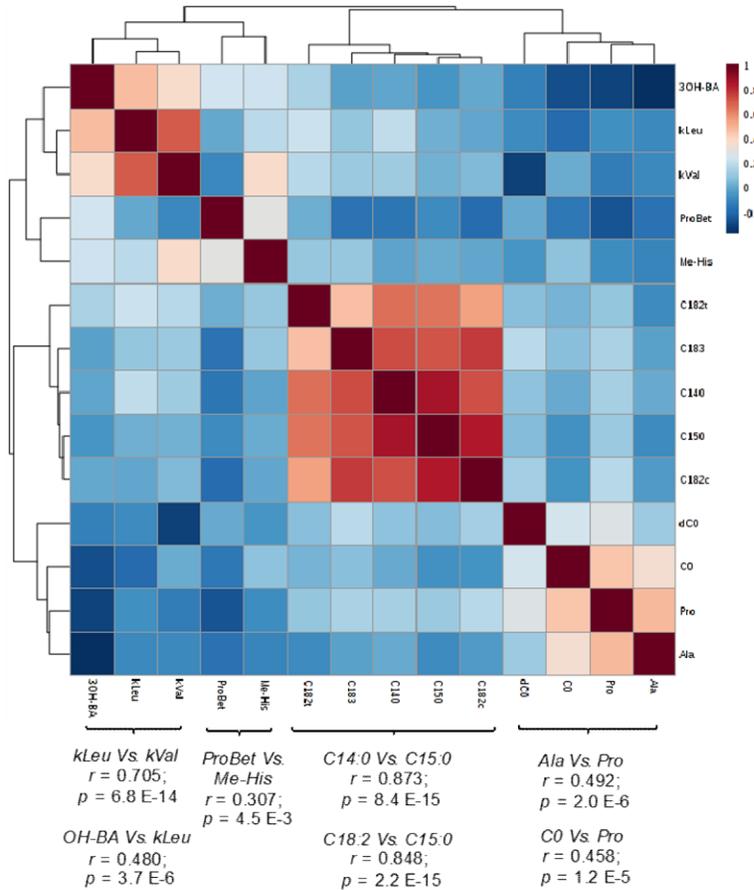


Figure 4S.13: Linear correlation plots highlighting the strong association between fasting plasma concentrations of (A) Me-His and (B) ProBet and their creatinine-normalized concentrations measured independently from matching single-spot urine samples for DIGEST participants collected at baseline and then following 2 weeks of food provisions.

(A) Lead Plasma Biomarkers of Contrasting Diets from DIGEST



(B) Lead Urinary Biomarkers of Contrasting Diets from DIGEST

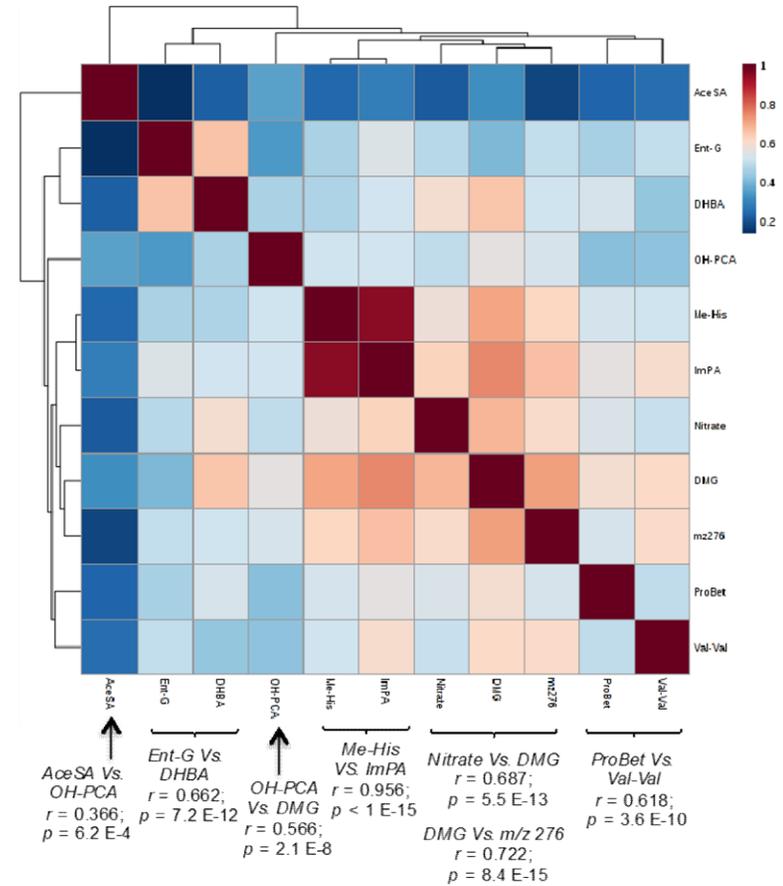


Figure 4S.14: 2D heat maps and correlation matrices for top-ranked (A) plasma and (B) urinary metabolites associated with contrasting diets from food provisions when using a Pearson correlation analysis on glog-transformed data. Distinctive clusters of metabolite classes suggest common dietary sources and/or biochemical pathways for their regulation, such as circulating ketone bodies (*kLeu*, *kVal*, 3-OH-BA), fatty acids (C14:0, C15:0, C18:2, C18:3) and amino acids/carnitines (C0, *Pro*, *Ala*) in plasma, as well as plant-derived biotransformed phenols (*Ent-G*, *DHBA*) and imidazole metabolites (*Me-His*, *ImPA*) in urine. In many cases, urinary metabolites reflective of recent dietary patterns were broadly co-linear with other compounds (e.g., *OH-PCA*, *Me-His*, *ProBet* and unknown ion, *m/z* 276) or had modest correlations to other compounds overall.

4.7 Supplemental References

1. Zulyniak, M. A.; de Souza, R. J.; Mente, A.; Kandasamy, S.; Nundy, M.; Desai, D.; Raman, K.; Hasso, R.; Paré, G.; Beyene, J., Anand, S. S. *BMC Nutr.* **2016**, 2, 34.
2. Kuehnbaum, N. L.; Kormendi, A.; Britz-McKibbin, P. *Anal. Chem.* **2013**, 85, 10664-10669.
3. Yamamoto, M.; Ly, R.; Gill, B.; Zhu, Y.; Moran-Mirabal, J.; Britz-McKibbin, P. *Anal. Chem.* **2016**, 88, 10710–10719.
4. Nori de Macedo, A.; Mathiaramanam, S.; Brick, L.; Keenan, K.; Gonska, T.; Pedder, L.; Hill, S.; Britz-McKibbin, P. *ACS Cent. Sci.* **2017**, 3, 904–913.
5. DiBattista, A.; McIntosh, N.; Lamoureux, M.; Al-Dirbashi, O. Y.; Chakraborty, P.; Britz-McKibbin, P. *J. Proteome Res.* **2019**, 18, 841–854.
6. Saoi, M.; Percival, M.; Nemr, C.; Li, A.; Gibala, M. J.; Britz-McKibbin, P. *Anal. Chem.* **2019**, 91, 4709–4718.
7. Wehrens, R.; Hageman, J. A.; van Eeuwijk, F.; Kooke, R.; Flood, P. J.; Wijnker, E.; Keurentjes, J. J. B.; Lommen, A.; van Eekelen, H. D. L. M.; Hall, R. D.; Mumm, R.; de Vos, R. C. H. *Metabolomics* **2016**, 12, 88.
8. Mahieu, N. G.; Patti, G. J. *Anal. Chem.* **2017**, 89, 10397–10406.
9. Dunn, W. B., Erban, A., Weber, R. J. M., Creek, D. J., Brown, M., Breitling, R., Hankemeier, T.; Goodacre, R.; Neumann, S.; Kopka, J.; Viant, M. R. *Metabolomics* **2013**, 9, 44–66.
10. Allen, F.; Pon, A.; Wilson, M.; Greiner, R.; Wishart, D. *Nucleic Acids Res.* **2014**, 42, W94–W99.
11. Jafari, N.; Ahmed, R.; Gloyd, M.; Bloomfield, J.; Britz-McKibbin, P.; Melacini G. *J. Med. Chem.* **2016**, 59, 7457-7465.

12. Nori de Macedo, A. N.; Teo, K.; Mente, A.; McQueen, M. J.; Zeidler, J.; Poirier, P.; Lear, S. A.; Wielgosz, A.; Britz-McKibbin, P. *Anal. Chem.* **2014**, *86*, 10010–10015.
13. Chong, J.; Soufan, O.; Li, C.; Caraus, I.; Li, S.; Bourque, G.; Wishart, D. S.; Xia, J. *Nucleic Acids Res.* **2018**, *46*, W486-W494.
14. Tai, B. Y.; Speed, T. P. *Ann. Statistics* **2006**, *34*, 2387–2412.

Chapter 5: Towards Accessing the Exposome for Precision Interventions

Contributions of the Thesis and Future Work

Chapter 5: Revealing the Exposome for Evidence-based Public Health Policies

5.1 Thesis Discoveries and Contributions

Work in this thesis is primarily focused on 1) method development to identify molecular targets important to classifying the exposures of stratified populations, and 2) multi-platform molecular phenotyping to determine patterns of personal environmental exposures that can be used to clinically identify individuals with increased risk of chronic illness based on their exposome. Advancing towards these holistic multi-factor methods of clinical decision-making and away from traditional generalized approaches is fundamental to advancing a new paradigm of precision risk reduction. Molecular biomarkers are the key to gaining greater insight into these complex gene-environment interactions, thus sensitive, selective, and robust methods to identify these molecules are of fundamental importance.

Chapter 1 of this thesis provides an abridged history of important biomolecular discoveries, subsequently employed in pre-clinical health assessments and their relation to modern analytical applications to the human metabolome to identify physiological changes related to the etiology of chronic disease. The analytical methods and instruments fundamental to metabolomics applications are detailed to provide a thorough background on instrumental technologies and methodologies used in this thesis. The work described in *Chapter 2* details the development of a highly sensitive self-reporting optical assay for the clinically relevant biomarker *N*-acetylneuraminic acid (Neu5Ac) and other acidic metabolites that attains sub-micromolar detection limits without relying on expensive analytical infrastructure or derivatization commonly used in metabolomics studies. A previously unexplored binding mode between an arylboronic acid probe, 4-isoquinoline boronic acid, and Neu5Ac in acidic media (pH 3) was found to involve the formation of a zwitterionic complex via a Lewis

acid coordination of an anionic carboxylate coupled with cyclization via condensation with a vicinal α -hydroxy group. This interaction has generally been overlooked due to the instability of boronic esters formed with neutral sugars at low pH. The additional electrostatic interaction stabilizes this complex toward hydrolysis at low pH which usually impairs the formation of these cyclic boronates with neutral sugar analogs, yielding an anomalously high affinity constant of $K = 5390 \pm 190 \text{ M}^{-1}$, approximately 40% larger than similar probes examined near physiological pH. Of immediate interest is the examination of the effect of Neu5Ac titration on the 276 nm peak observed to decrease linearly between 0-40 μM of sialic acid titrated to determine the linear dynamic range and linearity of this response. Furthermore, it is important to identify if other competing ligands, such as lactate or gluconic acid, elicit the same effect. By monitoring two wavelengths in a diagnostic array, greater selectivity for Neu5Ac may be possible and help to overcome the challenges of quantifying sialic acid in biological fluids. Biological specimens for Neu5Ac analysis is most commonly collected via skin, blood, or urine sampling, which are inherently invasive and potentially traumatic procedures, where sample handling requires trained personnel, specialized storage, and time-consuming processing and workup steps. To avoid these pitfalls, an assay that targets free Neu5Ac in saliva would be minimally invasive and facile to prepare. These benefits could rapid turnaround times to a salivary Neu assay with greater compatibility for use with larger populations. Easy access to saliva is particularly beneficial for juvenile patients in whom early and rapid postnatal identification of aberrant sialic acid metabolism (*e.g.* sialic acid storage disease, Salla disease) is fundamental to the mitigation of deleterious physiological and cognitive effects. There are also promising oncological applications for the detection of hypersialylated tumor cells in low-pH environments of pH 3 to 4 caused by tumor acidosis using fluorescent the 4-iqba probe. Indeed, it has been shown that *in vivo* optical pattern recognition of over-sialylated cancer cells is possible.¹

Binding between IQBA and Neu was only slightly reduced by a unit drop in pH, thus one approach to fine tuning the selectivity of this interaction for Neu5Ac could be to modulate pH to reduce ionization of potential competitive ligands, For instance, at the sub-optimal pH of 2.8 ionization of lactic acid is less than 10% which could allow for the calibration of Neu5Ac recovery in biological samples. Another approach would use strong ions and other chelating agents to reduce interferences: for instance, calcium pretreatment for the selective trapping of lactate would be an immediate first step towards enhancing the response of IQBA to Neu5Ac in biological samples.

Chapter 3 used high resolution separation methods based on a GC-MS platform to target a suite of wood smoke markers, including resin acids as novel urinary markers of smoke exposure. Firefighters are a subset of the population who endure extreme personal risk to ensure the health and safety of others, and suffer from chronic exposures to hazardous toxicants that are strongly linked to cancer, heart disease and several other chronic illnesses. This study used a targeted analytical multi-platform method to analyze 10 samples per firefighter among 3 cohorts ($n = 18$) to identify markers of wood smoke exposure in ambient air samples, skin wipes, and subsequently excreted in urine despite the use of bunker gear and a self-contained breathing apparatus (SCBA). Widespread contamination of the cheeks and arms was discovered in 3 groups of firefighters after short-term (30 min) exposure to wood smoke as part of their normal training exercises, indicating inadequate protection from regulatory protective gear against harmful chemical intrusions from pro-carcinogenic and proinflammatory chemicals, such as naphthalene and fluorene. Highly abundant smoke markers syringol, methylsyringol, propylsyringol, guaiacol, methylguaiacol, ethylguaiacol, and naphthalene (as 1-hydroxynaphthalene, a biotransformed metabolite of naphthalene) reached peak excretion within 6 h following acute smoke exposure. Concentrations of several chemicals in the urine had moderate to strong correlations

to amounts found in soot deposited under the bunker coat on the arm or under the SCBA on the cheek. In fact, current standards for equipment integrity testing were identified as areas of concern in this study as the SCBA collected significant amounts of soot inside the device during the exposure challenge, allowing some chemicals to deposit at levels up to 330-times amounts found on the smoke-exposed outer lens of the mask. Coincidentally, the inadequacy of current non-standardized cleaning protocols of fire stations and equipment was found through elevated concentrations of chemical contaminants found on the skin and bunker gear prior in baseline control samples. Overall, immediate intervention to mitigate direct and inadvertent chemical exposures to firefighters is necessary to reduce the long-term risk for chronic diseases that can be prevented by revised standards for SCBA fit testing, improved hygiene practices, and novel fire-resistant materials to mitigate chemical deposition onto skin surface.

Applications of this work could have immediate benefits in a longitudinal health and safety surveillance initiative to measure MP, PAH, and RA in firefighter workspaces and residences to develop and monitor more effective hygiene protocols. For instance, guaiacols and syringols were among the relatively concentrated wood smoke markers associated with high background levels in this study that would help to track exposures to volatile low molecular weight compounds. Away from the fireground these chemicals may appear in the environment from soiled and/or off-gassing equipment and clothing, and residue adsorbed to the skin and hair.² Ideally, a longitudinal surveillance program would monitor molecular targets alongside the implementation of new standards of turnout gear handling, fire station cleaning, and personal hygiene to quantitatively prove mitigation of local accumulation and personal risk. In tandem with local monitoring, a standardized definition of “clean” regarding firefighting equipment is badly needed to effectively reduce second- and thirdhand exposures to chemical toxicants that settle or embed within the material of the turnout coat and SCBA.

Currently, there are no regulatory protocols in place for proper cleaning to remove diverse classes of chemical contaminants firefighters encounter in the field. In fact, there exists no definition for the word “clean,” as NFPA and province of Ontario guidelines rely on each firefighter to make a judicious assessment about the integrity and cleanliness of their suit based on their recent exposures.^{3,4}

While not targeted in this study, of critical importance to firefighter safety is characterizing the toxic compounds present in the field due to burning structural materials such as plastics and solvents (*e.g.*, dioxins and heavy PAH) that are likely to accumulate on gear and skin by similar mechanisms observed in this study. An large scale follow-up study would focus on sampling important areas of the turnout gear of firefighters after fire suppression activities like the jacket collar, hood, jacket cuff, waistline, neck, cheeks/lips, and hands, with efficient samplers that are easily handled and stored quickly without risk to sample integrity. For instance, the isopropanol-soaked sample wipes used in *Chapter 3* were effective at recovering trace amounts of soot on the skin contaminants and could be applied in a large-scale occupational biomonitoring program. “Self-sampling” kits would allow participating firefighters to easily prepare and collect sample wipes from target areas, seal, and ship them for analysis. To elucidate the effects of these exposures, concurrent intermittent biofluid sampling would demonstrate how the firefighter exposome changes due to these exposures, including any metabolites related to metabolic dysfunction. Similar large-scale biomonitoring programs are already underway to determine the links between firefighting and the breast cancer in female firefighters to identify unique impacts to their health.⁵ While firefighting comes with inherent risk, the current degree of hazard firefighters experience is dangerous and can be remediated. The findings detailed in this these provide critical findings for important next steps by industry and governments to address it.

The discovery of novel dietary biomarkers in plasma and urine in *Chapter 4* is an important contribution towards the development of more objective diet

classification tools for nutritional epidemiology, as suboptimal diet plays an important role in population chronic disease risk. The parallel structure of this study allowed for direct comparisons of metabolite trajectories, rather than single-spot measurements, to define two highly distinct eating patterns by metabolite profiles and the direction of flux after exposure. Proline-betaine and 3-methylhistidine were two biochemical markers strongly associated with Prudent eating during the dietary intervention, and 3-methylhistidine in particular may provide further insight into the upstream physiological processes linked to the beneficial exposures in this eating pattern as it is strongly linked to the regeneration of skeletal muscle in humans. Furthermore, a series of amino acids and other organic acids were highly correlated to self-reported classes of food intakes, such as alanine and proline revealing recent consumption of fats, while enterolactone glucuronide and dihydroxybenzoic acid associated with healthier foods containing insoluble fibre, protein, essential nutrients and bioactive phytochemicals. Importantly, our work was able to evaluate the urine metabolome of to identify individuals in each cohort whose dietary records indicated food consumption was inconsistent with their provided meal plan. This is an important step towards reducing reliance on error-prone nutritional assessment tools (*i.e.*, FFQ and food diaries) currently used in clinical contexts. The development of a ratiometric diet quality index showed good discrimination among samples in the study, however an exciting next step would involve a larger-scale longitudinal study to test the efficacy of this classification tool, while also seeking new metabolites for inclusion into a dietary index panel, as characterizing recent dietary exposures is fundamental to realizing precision healthcare due to the frequency and diversity of exposures that comprise the dietary exposome. Additionally, a large-scale, longitudinal study presents the opportunity to determine strong links between the dietary exposome and long-term clinical outcomes that will then be applied to medical decision making on an individual level.

5.2 References

1. A. Matsumoto, A. J. Stephenson-Brown, T. Khan, T. Miyazawa, H. Cabral, K. Kataoka and Y. Miyahara, *Chem. Sci.*, 2017, **8**, 6165–6170.
2. K. W. Fent, B. Alexander, J. Roberts, S. Robertson, C. Toennis, D. Sammons, S. Bertke, S. Kerber, D. Smith and G. Horn, *J. Occup. Environ. Hyg.*, 2017, **14**, 801–814.
3. A. C. Mayer, K. W. Fent, S. Bertke, G. P. Horn, D. L. Smith, S. Kerber and M. J. La Guardia, *J. Occup. Environ. Hyg.*, 2019, **16**, 129–140.
4. Government of Ontario, *Ontario Regulation 253/07: FIREFIGHTERS*, 2018.
5. WFBC, Women Firefighters Biomonitoring Collaborative, <http://womenfirefighterstudy.com/> (accessed 2 August 2019).