HIGH THOUGHPUT EXPOSOMIC STUDIES FOR NEW INSIGHTS INTO SMOKE EXPOSURES IN OCCUPATIONAL AND POPULATION HEALTH

HIGH THOUGHPUT EXPOSOMIC STUDIES FOR NEW INSIGHTS INTO SMOKE EXPOSURES IN OCCUPATIONAL AND POPULATION HEALTH

By Biban Gill, B.Sc.

A thesis submitted to the School of Graduate Studies in Partial Fulfilment of the Requirements for the Degree
Doctor of Philosophy

July 2022

DOCTOR OF PHILOSOPHY (2022)
(Chemistry and Chemical Biology)

McMaster University
Hamilton, ON

TITLE:	High Throughput Exposomic Studies for New Insights into Smoke Exposures in Occupational and Population Health
AUTHOR:	Biban Gill, B.Sc. (University of Toronto)
SUPERVISOR:	Professor Philip Britz-McKibbin
PAGES	xix, 259

Abstract

Exposomics aims to characterize the totality of exposures over the lifespan, and their impact on human health. Currently, chronic exposure to harmful chemicals from air pollution and/or tobacco smoke, along with a suboptimal diet, remain leading causes for preventable mortality and morbidity worldwide. As a result, new analytical methods are needed to measure robust biomarkers of smoke exposure and food intake for improved risk assessment of clinical events. This thesis aims to develop high throughput methods to rapidly quantify urinary biomarkers of environmental smoke in high-risk occupations, and diverse global populations using multisegment injection-capillary electrophoresis-mass spectrometry (MSI-CE-MS) technology. *Chapter II* outlines an inter-laboratory method comparison for the targeted analysis of urinary 1-hydroxypyrene (HP) when using gas chromatography-high resolution mass spectrometry (GC-HRMS) and liquid chromatography-tandem mass spectrometry (LC-MS/MS) on urine samples collected from firefighters. This work revealed the critical role of incomplete enzymatic deconjugation on method bias and underreporting of true smoke exposures. *Chapter III* introduces a high throughput MSI-CE-MS/MS method (< 3 min/sample) to directly analyze the intact glucuronide conjugate of HP (HP-G) in urine without complex pre-column enzyme deconjugation and derivatization procedures. Importantly, firefighters deployed under emergency conditions at the 2016 Fort McMurray wildfire had creatinine normalized HP-G concentrations below the biological exposure index, likely caused by delays in urine collection under emergency conditions, at early stages of firefighting. *Chapter IV* extends from targeted biomonitoring of occupational smoke exposure, towards elucidating the relative risk of tobacco smoking in an international cohort of participants (n=1000) from the Prospective Urban and Rural Epidemiological (PURE) study. Comprehensive analysis of nicotine metabolites in urine by

MSI-CE-MS allowed for reliable determination of the total nicotine equivalent and nicotine metabolic ratio as robust indicators of recent tobacco smoke exposure and nicotine dependence, respectively. This method also offers a more accurate approach for biochemical verification of smoking status in large-scale epidemiological studies that are prone to social desirability and gender bias when relying on standardized questionnaires. Lastly, *Chapter V* employs a nontargeted metabolomics workflow using MSI-CE-MS to identify urinary metabolites that may serve as objective dietary biomarkers of food intake in participants across 14 countries from the PURE cohort. A panel of robust and generalizable metabolites were validated for biomonitoring of complex dietary exposures, that may further exacerbate the hazards of tobacco smoking. In summary, this thesis contributes high throughput analytical tools for characterizing the human urine exposome to better decipher the roles of smoke exposure, and suboptimal diet on chronic disease burden among diverse populations and regions worldwide.

Acknowledgements

First and foremost, I would like to thank my supervisor, Dr. Philip Britz-McKibbin for welcoming me into his group and providing me with the opportunity to expand my knowledge and learn in an environment with continuous optimism. I have gained invaluable skills and experiences, with the help of his guidance and encouragement. I would also like to thank my committee members Dr. Karl Jobst, and Dr. Greg Slater for their feedback, and unanticipated questions, which have inspired me to broaden my thinking and grow as a well-rounded scientist. I would also like to extend my appreciation to all our collaborators including Dr. Salim Yusuf, Dr. Guillaume Paré, Dr. Teo Koon, Dr. Sathish Thirunavukkarasu, Dr. Nicola Cherry, and Dr. David W. Kinniburgh for their time and advice along the way.

My sincerest gratitude goes out to the Britz lab, including a special thank you to Dr. Meera Shanmuganathan, who first introduced me to the lab and has been a good friend and mentor over the years. I would like to thank all the current lab members including, Zach Kroezen, Stellena Mathiaparanam, Ritchie Ly, Megan Magee, and Vanessa, your kindness, charisma, and ambition have made my experience here truly enjoyable. To all the past lab members; Dr. Michelle Saoi, Dr. Sandi Azab, Dr. Mai Yamamoto, Dr. Nadine Wellington, Dr. Adriana N. de. Macedo, Dr. Alicia DiBattista, and Dr. Karen Lam thank you for the knowledge and much needed wisdom that you passed on.

To my parents, sister, and grandparents, I am sincerely grateful for your kind words, comfort, and enthusiasm. Your work ethic and perseverance have given me the motivation to pursue my PhD. Finally, I would like to thank my partner Erick for his encouragement and unwavering support. Your positivity has always been a much-needed light during the most difficult times.

Table of Contents

List of Figures

Supporting Figures

List of Tables

Supporting Tables

List of Abbreviations and Symbols

$[M+H]^+$	Protonated molecular ion
$[M-H]^-$	Deprotonated molecular ion
1-OH-Pyr	1-Hydroxypyrene
2-OH-Nap	2-Hydroxynapthalene
ACN	Acetonitrile
AHEI	Alternative healthy eating index
APCI	Atmospheric pressure chemical ionization
ASK	Acesulfame K
BEI	Biological exposure index
BGE	Background electrolyte
BMI	Body mass index
CAA	Clean Air Act
CDC	Centers for Disease Control and Prevention
CE	Capillary electrophoresis
CHD	Coronary heart disease
CID	Collision induced dissociation
Cl-Tyr	3-Chloro-l-tyrosine
CPD	Cigarettes per day
CRA	Cumulative risk assessment
CV%	Coefficient of variation
CVD	Cardiovascular disease
CYP450	Cytochrome P450
DBS	Dry blood spot
DDT	Dichlorodiphenyltrichloroethane
EIE	Extracted ion electropherogram
EMV	Electron multiplier voltage
EPA	Environmental Protection Agency
ESI	Electrospray ionization
EWAS	Exposure wide association studies
FDR	False discovery rate
FFQ	Food frequency questionnaires
FMO	Flavin monooxygenase
FOD	Frequency of detection
GATS	Global Adult Tobacco Survey
GC	Gas chromatography
GC-HRMS	Gas chromatography-high resolution mass spectrometry
GC-MS	Gas chromatography-mass spectrometry
GWAS	Genome wide association studies
HCA	Hierarchical correlation analysis

HCL	Hydrochloric acid
HELIX	Human Early Life Exposome Study
HIC	High-income countries
HMDB	Human metabolome database
HP	1-Hydroxypyrene
HP-G	Hydroxypyrene glucuronide
HRMS	High resolution mass spectrometry
IBD	Inflammatory bowel disease
ICC	Interclass coefficients
IM	Ion mobility
KNN	k-nearest neighbors
LC	Liquid chromatography
LC-FD	Liquid chromatography with fluorescence detection
LC-FLD	Liquid chromatography with fluorescence detection
LC-MS/MS	Liquid chromatography-tandem mass spectrometry
LIC	Low-income country
LLE	Liquid-liquid extraction
LOD	Limit of detection
LOQ	Limit of quantification
m/z	Mass-to-charge ratio
MEOH	Methanol
MIC	Middle-income country
MICE	Multiple imputation by chained equations
MPs	Methoxyphenols
MSI-CE-MS	Multisegment injection-capillary electrophoresis-mass spectrometry
MSI-CE-MS/MS	Multisegment injection-capillary electrophoresis-tandem mass spectrometry
MSI-NACE-MS	Multisegment injection-nonaqueous capillary electrophoresis-mass spectrometry
MTBE	Methyl-tert-butyl-ether
MUFA	Monounsaturated fats
NACE	Non-aqueous capillary electrophoresis
NCD	Non-communicable diseases
NHANES	National Health and Nutrition Examination Survey
NIH	National Institute of Health
NIOSH	National Institute for Occupational Safety and Health
NMR	Nuclear magnetic resonance
NMR	Nicotine metabolic ratio
NMS	2-Napthalenesulfonic acid
OH-PAH	Hydroxylated polycyclic aromatic hydrocarbons
PAH	Polycyclic aromatic hydrocarbons

PATH	Population Assessment of Tobacco and Health
PCA	Principal component analysis
PFAS	Per- and polyfluoroalkyl substances
PLS-DA	Partial least squares-discriminant analysis
PQN	Probabilistic quotient normalization
PUFA	Polyunsaturated fats
PURE	Prospective Urban and Rural Epidemiology
QA	Quality assurance
QC	Quality control
r	Pearson correlation coefficient
RMT	Relative migration time
ROC	Receiver operating curve
RPA	Relative peak area
SD	Standard deviation
SHS	Second-hand smoke
SNPs	Single nucleotide polymorphisms
SPE	Solid phase extraction
TNE7	Total nicotine equivalents
TOF-MS	Time of flight mass spectrometer
u_{ep}	Electrophoretic mobility
UGT	UDP glucuronosyltransferase
Vcap	Capillary voltage
WHO	World Health Organization

Declaration of Academic Achievement

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

Chapter II: Reprinted and adapted from Gill B, Mell A, Shanmuganathan M, Jobst K, Zhang X, Kinniburgh D, Cherry N, Britz-McKibbin P. Urinary hydroxypyrene determination for biomonitoring of firefighters deployed at the Fort McMurray wildfire: An inter-laboratory method comparison. *Anal. Bioanal. Chem.* **2019** 411, 1397-1407

Chapter III: Reprinted and adapted from Gill B, Jobst K, Britz-McKibbin P. Rapid screening of urinary 1-hydroxypyrene glucuronide by multisegment injection–capillary electrophoresis–tandem mass spectrometry: A high-throughput method for biomonitoring of recent smoke exposures. *Anal. Chem.* **2020**, 92, 13558-13564

Chapter I:

Introduction to Exposure Assessment:

Where the Exposome Meets the Metabolome

Chapter I: Introduction to Exposure Assessment: Where the Exposome Meets the Metabolome

1.1 Origins of Chronic Disease

Prior to the early 1800's mortality was driven by pathogens from acute bacterial, viral and parasitic infections.[1] However, with the onset of industrialization came a better understanding of hygiene and sanitation practices, minimizing infectious disease rates prior to the development of vaccines and antibiotics. In addition, the industrial revolution brought forth improved water quality, food availability and nutritional sufficiency resulting in the doubling of human lifespans in most developed countries.[1–3] The causes of mortality subsequently shifted from pathogen-driven infections to chronic diseases of aging, accelerated by long-term exposure to environmental toxins and tobacco usage.[1]

Elucidating the potential environmental causes of chronic diseases is best exemplified by Percivall Pott, who made an account of a rare form of scrotum cancer prevalent among young chimney sweepers in 1775, which is now known as the first association between occupational exposure and cancer.[4,5] However, it was not until the early 1950's when Sir Richard Doll, first reported a causal link between tobacco smoking and lung cancer, which subsequently led him to quit smoking.[6] The past 40 years has brought on major changes in dietary habits in contemporary society from a traditional plant-based diet rich in fiber, fruits and vegetables to a high-fat, energy-dense diet containing refined grains, alcohol, and ultra-processed foods.[7,8] Collectively with a more sedentary lifestyle, tobacco use and air pollution from urbanization, chronic diseases including cardiovascular disease, respiratory illnesses, and various cancers have become predominant causes of disability and premature death worldwide.[8] Moreover, socioeconomic differences between developing and developed countries have resulted in a

substantial divide in chronic disease burden.[7] While biological risk factors are fixed (age, sex and genetics), behavioral (e.g. diet, tobacco use and physical activity), societal (e.g. socioeconomic, gender, cultural) and environmental (e.g. air and water quality) variables are modifiable.[8] Consequently, a deeper understanding of the complex interaction between genetic and environmental exposures may help more accurately predict disease incidence and progression on an individual level.[9,10]

1.2 Biomarkers in the Context of Environmental Health

The National Institute of Health (NIH) defines a biomarker as a characteristic that is objectively measured and evaluated as an indicator of biological, pathogenic or pharmacological processes and/or responses to external exposures or stimuli.[11] Although the term "biomarker" was first coined in the late 1960s, the concept itself has dated back centuries.[12] In their earliest form biomarkers were objective physical findings and observations such as heart rate, temperature and/or physical patient features, which later evolved into the qualitative analysis of biological fluids such as urine (e.g., colour, surface foam).[12,13] However, the 20th century brought on technical advances in analytical techniques (e.g., chromatography, mass spectrometry) which led to the quantitative assessment of biological samples, vastly increasing the number of detectable biomarkers.[11] Accordingly, a transition from anatomical and/or physiological measurements to discrete molecular biomarkers (e.g., genes, proteins, carbohydrates, lipids, hormones) has led to the routine monitoring of biological responses in both a clinical and environmental context.[11]

Biomarkers used in environmental health studies can be divided to three main classes; biomarkers of exposure, effect and susceptibility.[14,15] While biomarkers of exposure can be determined through the measurement of contaminants in various environmental samples, (e.g., air, water, food) interindividual variability in absorption, distribution and excretion are not

captured. A more accurate representation of internal dose and subsequent risk to an exposure can be determined by measuring the compound of interest, or its transformed biproducts in a biological sample (e.g., urine, saliva, sweat, blood).[16] For example, tobacco biomarkers such as nicotine and its respective metabolites can indicate both exposure and dosage when using a simple non-invasive urine collection.[17] Overall, an ideal biomarker of exposure should be both specific and quantitative, while also remaining in the body long enough (i.e., pharmacokinetics) to obtain a reliable measurement with a straightforward sample collection.[16,18] Conversely, biomarkers of effect are often less specific, and measure biochemical, physiological or behavioral alterations which may be associated with potential health outcomes or clinical events.[19] In particular, oxidative stress is a common underlying consequence, indicative of cumulative exposure effects over time. For instance, urinary 8-hydroxy-2-deoxyguanosine (8-OHdG) and 8-isoprostane function as biomarkers of DNA damage and lipid peroxidation, respectively which may result from various acute or chronic exposure to harmful xenobiotics, such as polycyclic aromatic hydrocarbons (PAHs), phthalates, and flame retardants.[20,21] Lastly, biomarkers of susceptibility may suggest the intensity and form of response, as certain individuals may have an inherent or acquired ability to mitigate the biological consequences associated with an exposure.[19] Genetic biomarkers (e.g., variants) are widely accepted indicators of susceptibility and chronic disease risk, independent of the original exposure. As a result they can denote differences in uptake, metabolism, and bioactivation of chemical carcinogens.[15,22] In humans, genetic variants for tumor suppressor proteins, DNA repair enzymes, and cytochrome P450 (CYP P450) enzymes involved in phase I metabolism, have all been implicated with increased cancer risk.[23] For example, a greater risk for lung cancer among smokers has been linked to variants of CYP 2A6, given its key role in the metabolic activation of several tobacco

carcinogens.[24] However, genetic biomarkers alone have limited applicability given the phenotypic heterogeneity as a result of complex environmental interactions.[9,25] Thus, additional external factors such as age, sex, diet, metabolic health, occupational exposures and lifestyle are important considerations in accurately predicting disease onset susceptibility.

1.3 Human Biomonitoring for Exposure Assessment

As early as the 18th century, the link between environmental exposures and increased risk for chronic disease was observed.[26] However, the turning point in exposure assessment was the publishing of "Silent Spring" by Rachel Carson (1962), who raised concerns on the adverse health consequences linked to synthetic organic chemical pesticides (e.g. DDT), which ultimately led to the establishment of the US Environmental Protection Agency (EPA).[26] Despite the introduction of regulatory agencies, and efforts toward replacing harmful chemicals, the rapid expansion of industry, agriculture and population growth have brought forth the continued emergence of new substances yet to be fully characterized, commonly referred to as "regrettable substitutions".[27] With a current list of 83000 chemicals in the US Toxic Substances Control Act (TSCA), only 3% have available data, often favoring such substitutions.[27] A well-known example is the substitution of Bisphenol A (BPA), with Bisphenol S (BPS), where a push towards "BPA-Free" products lead to a rapid and simple solution, largely driven by consumer demand. [27] As a result, human exposure studies in monitoring the presence, severity and health impact of these substances has been defined as human biomonitoring.[28] Overall, this is considered the most direct way of identifying the quantity of the exposure, it's associated risk and the underlying respective mechanism, in an attempt to provide early warning signs, both at the individual and population level.[29] Thus solutions towards providing early warning signs, and evidence-based policies are critical. As a result, human biomonitoring, described as exposure

studies for the purpose of monitoring the presence, severity and health impact of a contaminant, is considered the most direct way of identifying the quantity, risk and underlying mechanisms of an exposure.[28]

Conventional methods for biomonitoring of exposures (e.g., diet, smoking, air pollution) have relied on self-reports using standardized questionnaires, given their low-cost and convenience.[30] While, still commonly implemented in environmental epidemiological analyses (e.g., Global Adult Tobacco Survey), survey methods are often prone to misreporting, error and bias.[31] Consequently, biomonitoring transitioned toward targeted chemical analysis in various biological matrices and has been since used by a number of regulatory agencies for emerging chemicals of concern, with new additions continuously being added to public databases (>250 chemicals).[32] The majority of biomonitoring programs are currently conducted by high-income countries (HICs), such as those lead by the Center for Disease Control and Prevention (CDC). However, they remain in their infancy among most low-income countries (LICs).[28] Given the prevalence of contaminants in our environment, we are subject to cumulative exposure to thousands of chemicals over time. As a result, a paradigm shift from traditional hypothesis driven biomonitoring techniques to multi-omics investigations of exposure are critical.

1.4 Characterizing the Human Exposome

In the last 20 years exposure analysis has evolved from the concept of "one exposure one-outcome" to examining a myriad of environmental exposures over time and space.[33] In 2005, Dr. Christopher Wild first described this idea as the "exposome", to capture the totality of exposures from conception to death, and their impact on human health.[9] The three primary domains encompassing the exposome include our internal exposome (e.g., metabolism), specific external (e.g., environmental pollutants) and general external (e.g., socioeconomic status) exposomes

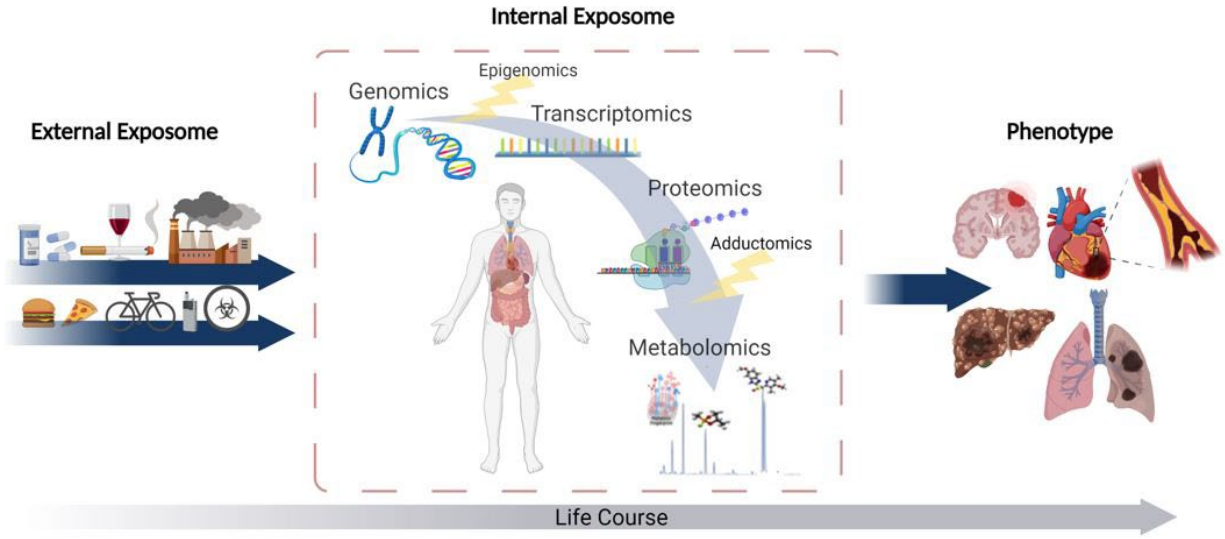


Figure 1.1. Schematic of the human exposome including both general and specific individual exposures (external exposome), which subsequently impact the complex internal exposome comprised of the -omics cascade. Overall, exposomics aims to capture the totality of exposures over one's entire lifetime which results in various phenotypic responses (e.g., chronic disease burden). *Created with BioRender.com*

(**Figure 1.1**).[30] Wild[34] recognized that although mapping the human genome revolutionized our ability to determine one's genetic susceptibility to disease, it also revealed its limitations in predictive power. As such, similar to genome wide association studies (GWAS) which generated extensive data on genetic variants, exposure wide association studies (EWAS) can provide information on the effects of multiple environmental exposures and their relationship to health risks.[10,33] Thus, integration of both genome and exposure data can help better characterize the human exposome, including susceptibility to chronic disease risk.

Although exposome approaches do include traditional targeted biomonitoring strategies, they also incorporate nontargeted discovery, and theoretically aim to encompass all detectable chemical exposures of significance.[30] As a result, current objectives in exposome research include; scaling up the coverage of exposures, while also expanding the diversity and size of study cohorts to better understand the genetic and environmental impacts on disease risk.[35,36] To

date, multi-omics technologies have been deployed in exposome research as outlined in **Table 1.1**.[30] Overall, genome sequencing (genomics) has been complemented by analogous tools that characterize downstream biological events in the form of genetic modifications (epigenomics), RNA expression (transcriptomics), adduct formation (adductomics), protein characterization (proteomics) and metabolite characterization (metabolomics) in biological matrices.[10] Consequently, each of these -omics methodologies provides information on thousands of molecular components that can be correlated with specific clinical events, similar to GWAS.[35]

While whole-genome sequencing is available, routine implementation on a large-scale is difficult given the price, storage and computational efforts that are required (**Table 1.1**).[30] Alternatively, single nucleotide polymorphisms (SNPs)-based arrays are widely used to impute genotypes for a majority of genes with a known function. However, gene-environmental interactions are dependent on epigenetic mechanisms that are beyond the scope of conventional gene sequencing approaches.[28] This can include alterations in chromatin structure, DNA methylation and histone modifications.[28] Research suggests that epigenomics can provide insight on early life exposures (**Table 1.1**), where offspring of undernourished mothers have demonstrated unique epigenetic profiles compared to their siblings.[37,38] Similarly, chemical exposures can change RNA expression by activation of signaling pathways, thus requiring monitoring via transcriptomic technologies.[39] For instance, a recent study by Pintarelli *et al*.[40] investigated the differences in transcript levels of lung tissue from current and never smokers to identify a gene panel that reflects recent exposure to cigarette smoke. While informative, these technologies often required more invasive sampling and can be difficult to implement in large-scale population studies.[46]

Table 1.1 Summary of Omics approaches used in exposomic research[41–45]

-Omics Approach	Description	Techniques	Temporal Variability	Influenced by Disease or Environment	Throughput
Genomics	Study of genes and their function.	• PCR based methods • DNA microarrays • Parallel-seq	None	No	Medium to High
Epigenomics	The study of all heritable changes in gene expression which are independent of the DNA sequence.	• DNA methylation profiling • ChIP-seq • RNA-seq	Low	Yes	Low to High
Transcriptomics	The study of all RNA transcripts produced by the genome.	• Hybridization-based technologies • RNA-seq • RNA splicing • Microarrays	Moderate	Yes	High
Proteomics	The large-scale study of proteins encoded by the genome, including their structure, expression and function.	• Soft ionization mass spectrometry • Antibody microarrays	High	Yes	Low to Medium
Adductomics	The characterization and detection of macromolecular adducts formed with DNA or proteins.	• High resolution mass spectrometry	High	Yes	Medium to High
Metabolomics	The study of small molecules (1.5<kDa) within cells, biofluids, tissues or organisms.	• NMR • Mass Spectrometry	High	Yes	High

Less invasive monitoring can be performed in biological fluids such as serum using proteomic and adductomic technologies (**Table 1.1**).[46] Environmental exposures eliciting changes in the post-translational state of proteins (e.g., methylation) have been previously evaluated in serum, saliva, and feces.[29] Similarly, adduct formation of macromolecules with reactive chemicals can lead to an altered biological state of a cell, tissue or organism.[47] Given the longer half-lives of many adducts in serum, adductomics has been previously reported to identify individuals experiencing high levels of exposure long after the initial exposure event.[48] Recently, adducts in human serum albumin have been correlated with several human diseases including diabetes, cancer, and cardiovascular disease.[28,49,50] However, both proteomics and adductomics are limited in specificity, and are best used as biomarkers of effect with few candidates having been successfully applied for routine clinical use.[51] Additionally, while adducts and mutated proteins may confirm greater risk of developing disease, little information regarding the initial source of exposure, rate of disease progression, or severity is determined.[28]

While numerous -omics technologies have proven useful, many of those described require multi-omics collaborative efforts to provide insight on susceptibility, exposure, and biological response simultaneously. This is often costly, and impractical for large-scale analysis.[52] In contrast, the metabolome consists of small molecules (< 1.5 kDa) involved in chemical reactions that maintain cell and organ functions (**Table 1.1**).[53] Alterations made in the human genome, proteome, and transcriptome are readily detected in the metabolome, with close associations between clinical outcomes and aberrant metabolism having been previously reported.[54] Overall, current analytical technologies are capable of detecting several thousands of small molecules in a variety of biological matrices.[55] Currently, databases such as the Human Metabolome Database (HMDB) provide information on over 200,000 endogenous and

exogenous metabolites.[56] Given the extensive amount of information the human metabolome has to offer, it is a very useful tool for exposome characterization, and can provide high resolution multifactorial phenotypic signatures for complex disease etiologies.[55] However metabolomic changes are very dynamic, interdependent, and often subject to numerous external factors.[57] Such -omics strategies have also struggled with inadequate study power, poor method validation and lack of reproducibility.[22]

1.5 Pre-analytical Considerations

While no standardized protocol can resolve, detect and identify the totality of endogenous and exogenous chemicals that comprise the human exposome, a general workflow can be outlined consisting of study design, sample preparation/storage, data acquisition, post analytical processing, unknown identification, biological interpretation and translation (**Figure 1.2**).[57] Overall, exposomic studies, no matter the approach, should be assumption-driven, where any possible confounding exposures is considered in the study design.[58] Although it is not feasible to quantitatively measure every exposure, the variables most relevant should be incorporated, while attempting to expand coverage by other means (e.g., questionnaires, metadata).[58] Study designs in exposomic analysis can vary between observational, cross-sectional, longitudinal, cohort-based, or case-controlled.[36,43] However, the majority of observational and cross-sectional studies are limited by the temporal discrepancy between exposure and effect.[33,58] As a result, exposomic approaches benefit from longitudinal and temporal variability in sampling.[22,30,59] For instance, the human early life exposome study (HELIX) combines six longitudinal population-based birth cohorts from six European countries.[60] In addition, the selection of sample size is crucial during study design, as it can later impact the statistical power of the analysis. This is a common challenge in exposomics as it can be difficult to estimate the sample size required given the

analytical variability and wide concentration range of exposures.[57,61] Moreover, during cohort selection, several factors such as age, sex, ethnicity, overall health status, comorbidities, diet and lifestyle, are all important considerations to record and adjust for, as they can influence both the metabolome and consequently the exposome.[57] As such, mitigation strategies to minimize bias can include stratified randomization to ensure balance of experimental groups in respect to confounding variables, and the collection of clinical meta-data to enable post-processing corrections.[57] Lastly, given that complex study designs can lead to challenging experimental models, *a priori* considerations of controls, reference standards, and data fidelity (e.g., quality assurance or QA, and quality control or QC) need to be carefully considered to minimize false discoveries.[41,62]

When conducting metabolomic based exposome analysis, various biological matrices can be used, with the most common being blood and urine.[63,64] Additional matrices of interest that have been previously reported include, feces, hair, saliva, sweat, and teeth.[58] Given the wide dynamic range of metabolite concentrations reported in exposomic studies, the choice of biofluid must be selected based on the research question and context of analysis.[65] Importantly, the detection of xenobiotics in biological matrices typically requires larger sample volumes, thus the simplicity of collection and availability of the biofluid should be considered.[66] For example, less persistent environmental compounds such as phthalates and bisphenols are easily detected in urine, while more persistent compounds are commonly detected in blood.[57,67] Irrespective of the choice of specific biofluid, the chemical stability of metabolites, time of sample collection, handling, pretreatment and storage are all critical variables to be evaluated.[67] Overall, biological samples should be aliquoted into small volumes to minimize effects from repeated freeze-thaw cycles.[68] In addition, it is recommended that biological samples be stored at -80°C immediately

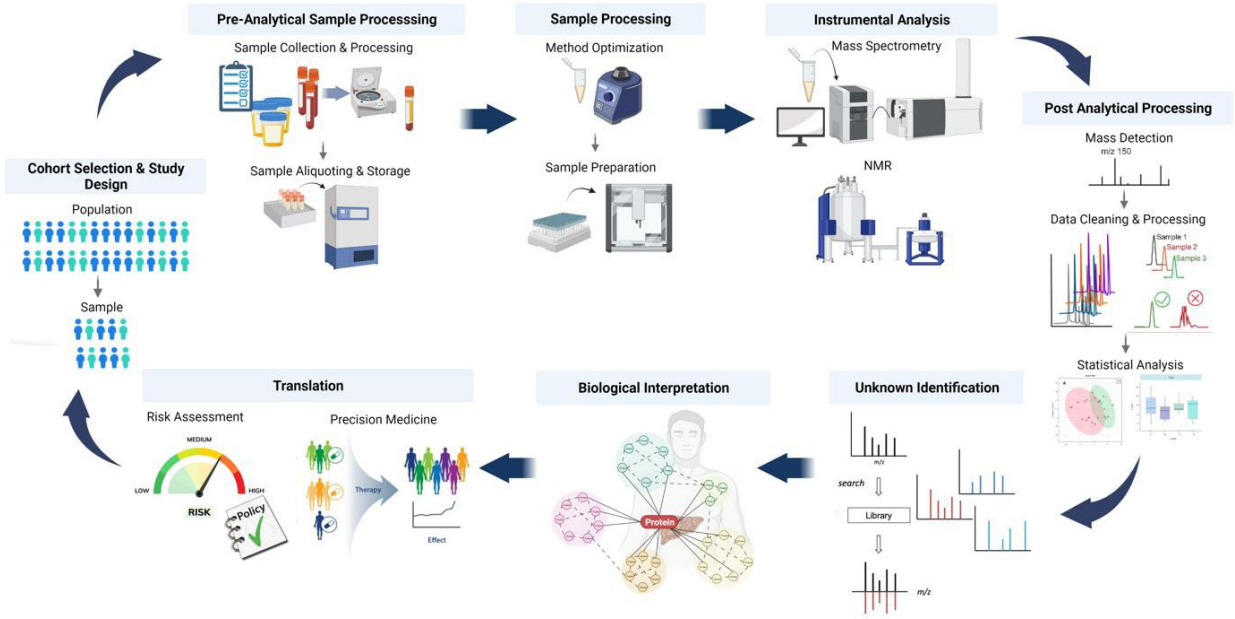


Figure 1.2. Overview of major steps in the exposomic workflow for biomarker discovery from cohort selection and study design to pre-analytical considerations, including proper sampling and storage techniques, followed by optimized sample processing techniques for targeted or nontargeted analysis. Subsequent instrumental analysis with validated analytical platforms are then performed with appropriate post-analytical processing techniques, including mass detection, peak picking, peak filtering and processing, allowing for statistical analysis of filtered data. Lastly, unknown identification and biological interpretation can be conducted on top-ranked features prior to validation and translation in the context of risk assessment, public policy and precision medicine. *Created with BioRender.com*

after collection whenever possible.[69] For long-term storage Hebels *et al.* [70] has reported that as long as consistent procedures have been implemented, biobank samples stored at -80°C are suitable for analysis for up to 15 years. While long-lived persistent contaminants (e.g., pesticides) are stable and can remain in sample for prolonged periods of time, more polar and labile contaminants (e.g., nicotine) are subject to greater variability as a result of poor analytical practices.[63,71] Thus, the number of repeat freeze-thaw cycles must be recorded, and stability studies for metabolites of significance are recommended to be evaluated as sources of potential pre-analytical bias.

1.6 Sample Processing and Analytical Method Considerations

The biggest challenge in translating the exposome from concept to practice remains the diversity and scale of chemical burden an individual may experience, with current estimates upward of a million chemicals over a lifetime.[72] With recent advances in analytical chemistry metabolomics has been identified as a key readout of the exposome.[69] Due to the extensive diversity in physiochemical properties and abundances of metabolites, exacerbated by the dynamic range of xenobiotics, a single analytical platform to monitor the entirety of the metabolome does not currently exist.[73] Differences in polarity, stability and volatility challenge metabolomic analyses and necessitate the need for complementary analytical platforms for detection, identification and quantification of both known and unknown metabolites.[73] However, implementation of cross-platform studies is costly, inefficient and not always feasible in large-scale analysis.[74]

When using metabolomics for characterization of the human exposome, both nontargeted and targeted acquisition strategies have been commonly employed.[74–78] A nontargeted approach enables broad scope chemical profiling for unbiased and comprehensive monitoring of metabolites without any *a priori* information.[67,73] However, the main challenge remains the identification and assignment of molecular features with unambiguous structural identification. Conversely, targeted approaches rely on *a priori* knowledge to detect chemicals expected in a sample.[67] An advantage of targeted analysis is the ability to optimize analytical strategies for improved sensitivity and quantification of low abundance metabolites without interferences. Overall, nontargeted acquisition can generate a data driven hypothesis which can facilitate subsequent biomonitoring of novel compounds of significance.[75,79] As a result, the choice of analytical platform can dictate the window through which the metabolome is viewed and should be carefully considered based on study objectives and practical constraints (e.g., sample matrix,

sample volume, budget, time).[69]

The most widely used analytical tools are nuclear magnetic resonance (NMR) spectroscopy and mass spectrometry (MS) coupled along with various separation platforms, including liquid chromatography (LC), gas chromatography (GC), capillary electrophoresis (CE), and ion mobility (IM) **(Table 1.2)**.[67,74,80] Although NMR is frequently used for metabolomics analysis due to its high selectivity, reproducibility, simple sample preparation, and non-destructive nature, the poor concentration sensitivity and large sample volume requirements it has limited utility in exposomics**.**[74] In contrast, high resolution MS (HRMS) is widely used in exposomic studies due to its improved sensitivity, better selectivity, and wide dynamic range that is amenable to measuring a broader range of metabolites in complex biological samples.[67,81] Additionally, current strategies of full scan and parallel reaction monitoring workflows allows for simultaneous targeted and nontargeted data acquisition.[36] However, as described by Rappaport *et al*.[47] critical challenges remain, including the vast number of metabolites to be characterized and identified, as well as the difference in concentrations profiles of xenobiotics as compared to endogenous metabolites. Consequently, sample extraction and/or enrichment are often required for comprehensive detection for low abundance intoxicants.[67] Thus, low throughput methodologies and complicated sample work up procedures must be balanced with achieving highly reproducible yet robust workflows applicable for large-scale analysis.[57,67]

While the analytical platform, biological matrix, and metabolites of interest do result in an abundance of sample preparation possibilities, all protocols aim to maximize metabolite recovery and minimize interferences.[73] Generally, more concentrated endogenous analytes (e.g., creatinine in urine) require a dilution, whereas lower concentration analytes (e.g., hydroxylated PAHs or OH-PAHs) may require of-line sample preconcentration to improve sensitivity.[73,84,85]

Table 1.2. Summary of analytical platforms commonly used for exposome analysis[55,57,67,82,83]

Analytical Platform	Analysis Time	Advantages	Limitations	Applications
Liquid Chromatography Mass Spectrometry (LC-MS)	Minutes to Hours	• Large selection of column chemistry • Structural elucidation • Highly sensitive and selective • Retention time reproducibility • Broad exposome coverage	• Large solvent volumes/waste (Costly) • Longer analysis times • Prone to suppression and sample carry over • Complex elution programming (e.g HILIC)	Captures polar and non-polar chemical exposures in biofluids, tissues and environmental samples with applications in food, pharmaceutical and clinical analysis.
Gas Chromatography Mass Spectrometry (GC-MS)	Minutes to Hours	• Extensive spectral library (EI-MS) • Structural elucidation • Excellent separation efficiency and sensitivity • Ideal for volatile stable compounds • Retention time reproducibility	• Destructive (EI-MS) • Suitable for only sufficiently volatile compounds • Complex sample workup • Longer analysis times	Captures polar to volatile chemicals, with applications in environmental, food, pharmaceutical, forensic, waste, petrochemical and industrial analysis.
Inductively Coupled Plasma Mass Spectrometry (ICP-MS)	Seconds to Minutes	• Low detection limits • Multi-element analysis • Robust against matrix effects (e.g salts) • High throughput	• Expensive infrastructure and operational costs • Large sample volumes • Not ideal for solid samples • Limited analyte compatibility	Quantification of isotopes in various environmental, biological and geological samples.
Ion Mobility Mass Spectrometry (IM-MS)	Milliseconds to Seconds	• Fast analysis times • Highly reproducible • Portability and miniaturization • Isobaric/Isomeric separation	• Prone to suppression • Lower mass resolution • Standards required • Limited linear range	Quantification of organic compounds with isomers, applicable in clinical, environmental and pharmaceutical settings.

Analytical Platform	Analysis Time	Advantages	Limitations	Applications
Direct Infusion Mass Spectrometry (DI-MS)	Minutes	• High throughput (<3min/sample) • Lower cost (low solvent consumption) • Minimal sample volumes • Sensitive	• Not recommended for complex sample matrices • No isomeric/isobaric resolution • Minimal selectivity • Requires stable isotope-internal standards for quantification (costly)	Aids in metabolite fingerprinting with common application in the analysis of small proteins, macromolecular complexes and rapid drug screening.
Nuclear Magnetic Resonance (NMR)	Minutes	• Fast analysis times • Non-destructive • Highly reproducible • Structural elucidation • Minimal sample preparation	• Expensive infrastructure and high operational costs • Large sample volumes (< 0.5 mL) • Low sensitivity • Lower resolution in complex samples	Liquid state analysis of wide range of chemicals with applications in imaging, forensics, and food technology.
Capillary Electrophoresis Mass Spectrometry (CE-MS)	Minutes	• Low samples injection volumes (<20 nL) • Minimal solvent consumption (low cost) • Minimal sample preparation • Multiplexing capability (injection versatility) • High separation efficiency • Ideal for polar/ionic compounds • Effective desalting abilities	• Lower sensitivity • Less exposome coverage • Poor migration time reproducibility • Few large-scale validation studies	Quantification of polar/non-polar ionic endogenous and exogenous compounds in biofluids and tissues with applications in food, environmental, pharmaceutical and clinical analysis.

Additionally, for more heterogenous matrices (e.g., tissues, fecal samples and cell cultures) sample homogenization is needed to obtain consistent extraction efficiencies, while precipitation of solutes and proteins is essential for other sample types (e.g., urine, serum).[64,67] However, an ongoing challenge in sample preparation remains the balance between selectivity and sensitivity.

A number of sample preparation techniques involving varying degrees of selectivity span from protein precipitation (PP) or dilution, to solid-phase extraction (SPE) and liquid-liquid extraction (LLE).[86] The development of complementary LLE protocols, allows for fractionation of different classes of analytes based on their hydrophobicity and polarity, which can in turn be processed separately for expanded metabolome coverage.[67] Additionally, certain analytes may require chemical or enzymatic pretreatment to promote their release from the matrix or to enhance senstivity.[67] For instance, enzyme deconjugation is commonly employed in several targeted assays (e.g. bisphenols, OH-PAHs, nicotine metabolites) to measure the total concentration of parent compounds.[85,87,88] However, varying reaction rates make it difficult to deconjugate a wide range of analytes under a single operating condition, unless rigorously validated. Also, challenges in standardization of incubation protocols has contributed to substantial intra- and inter-lab variability.[77,85,89] As described by Vitale *et al*.[67] future chemical exposomic studies should advance towards a more comprehensive screening of phase II metabolism, as understanding underlying mechanisms of detoxification can provide insight on potential biological response to environmental stressors. Comparatively, chemical derivatization is often needed in GC-MS to generate more volatile compounds or improve retention in reverse-phase LC-MS, with the overall intention of enhancing sensitivity and resolution of analytes.[67] Although enzyme deconjugation and chemical derivatization protocols may enhance overall analytical performance, the slow and variable reaction speeds greatly limit sample throughput.

For example, GC-HRMS methods for the analysis of OH-PAHs in urine require enzyme deconjugation, precolumn derivatization and SPE making them both costly and time consuming.[90] Similarly, previous LC-MS/MS methods have predominantly used enzyme deconjugation followed by either SPE or LLE extraction with a preconcentration for the detection of low abundance tobacco specific metabolites (e.g., nitrosamines).[91,92] Due to the extensive sample processing, risk for analyte loss, compound degradation and poor technical precision, the need for recovery standards is essential.[73] Accordingly, simpler single-step protocols are favored to minimize inaccuracies. However, regardless of the assay, randomization of sample order, and batch number should be conducted in addition to the inclusion of internal or external measures to objectively evaluate method performance. The use of pooled representative samples (e.g., QCs) and/or a National Institute of Standards and Technology (NIST) reference standards with repeat measurements allows for day-to-day monitoring and subsequent post-processing correction of batch effects and long-term instrumental signal drift.[93–95]

1.7 Instrumental Platforms in Exposomics

Following sample processing protocols, several instrumentation platforms are used in contemporary exposomic studies (**Table 1.2**). GC-MS is ideal for more volatile, low molecular weight and thermally stable compounds. While it is highly sensitive, selective, and reproducible it is restricted by the range of compounds amenable for separation. In addition, GC-MS often relies on more extensive sample work-up procedures for gas-phase partitioning, which can be costly and time consuming, and unideal for high polarity compound resolution.[57,67,80] In contrast, LC-MS methods are highly versatile given the variety of mobile and stationary phases available.[80] For example, although hydrophilic interaction liquid chromatography (HILIC) is preferential for polar metabolites and reversed-phase columns are more suitable for non-polar

metabolites, a HILIC based method was proposed by Cai *et al.*[96] for the concurrent profiling of both polar metabolites and lipids in plasma. Overall, given the different column phases, modifiers and chromatographic temperatures available, LC-MS methods can be fairly adaptable in regards to sensitivity and resolution. However, methods typically require larger sample volumes, while also producing more solvent waste, making them less cost-effective.[57,73,74] As highlighted by Vitale *et al.*[67] novel developments should be directed toward the combination of different stationary phases and mobile phases to capture a larger diversity of chemicals having varying polarity within a single analysis.

Capillary electrophoresis (CE) is a microseparation technique which remains an underexploited tool in exposome studies despite its high resolving power compared to LC and GC separations. Coupled to electrospray ionization (ESI)-MS, CE separations have demonstrated excellent resolution of polar ionic compounds spanning amino acids, nucleotides and organic acids in addition to exogenous compounds such as drugs of abuse, plasticizers, and OH-PAH glucuronide conjugates.[95,97–99] With extensive tunability, CE has also been applied to low polarity ionic compounds using previously reported nonaqueous capillary electrophoresis (NACE) techniques.[100,101] Due to the nanoliter (nL) sample volumes which are hydrodynamically introduced on-capillary, CE is particularly well suited for the analysis of volume-restricted samples such as tissue biopsies, infant sweat and dry blood spot (DBS) punch.[102–104] Overall, minimal solvent consumption and waste enables lower daily maintenance costs when using low-cost fused-silica capillaries. Also, when working with highly saline samples such as sweat and urine, CE permits the separation of involatile electrolytes (e.g., sodium) prior to ESI-MS, minimizing matrix effects which facilitates simple sample work up procedures.[74] Conversely, limitations in concentration sensitivity and metabolome coverage for non-ionic/neutral

compounds, with few validated protocols for large-scale analysis makes applications in exposomics a challenge. However, recent studies by Harada *et al*.[105] and Ishibashi *et al*.[106] have shown the validity and reproducibility of CE-MS for large-scale population-based cohorts (n > 6000 samples), in both plasma and urine demonstrating its suitability for application in epidemiology.

As compared to LC and GC methods, CE is performed under isocratic and isothermal conditions within an aqueous buffer (BGE) filled capillary which facilitates a consistent environment for electromigration and ionization.[107] Accordingly, this presents a unique advantage of CE for mutliplexed separations in order to increase sample throughput, as observed by multisegment injection-capillary electrophoresis mass spectrometry (MSI-CE-MS).[108] **Figure 1.3** outlines recent advances in MSI-CE-MS where following a hydrodynamically injected sample, an electrokinetic BGE spacer is utilized, and migration of analytes begins, taking advantage of a longer effective capillary length for separations. Along with rapid analysis times (< 3 min/sample), improved peak capacity and resolution are also obtained relative to a single injection.[109] Overall this helps facilitate high throughput screening, where recently a large-scale analysis of the maternal serum metabolome demonstrated excellent long-term reproducibility.[95] MSI-CE-MS also offers improved data fidelity and unique data workflows when taking advantage of temporal signal pattern recognition which can be utilized in nontargeted analysis protocols to rapidly filter spurious features.[110] Nontargeted applications have recently included high throughput screening of drugs of abuse, serum metabolome analysis in physically inactive older persons, and nontargeted nutritional epidemiology.[98,109,111] Thus, MSI-CE-MS offers a valuable analytical platform in the field of exposomics that will be the major focus of this thesis.

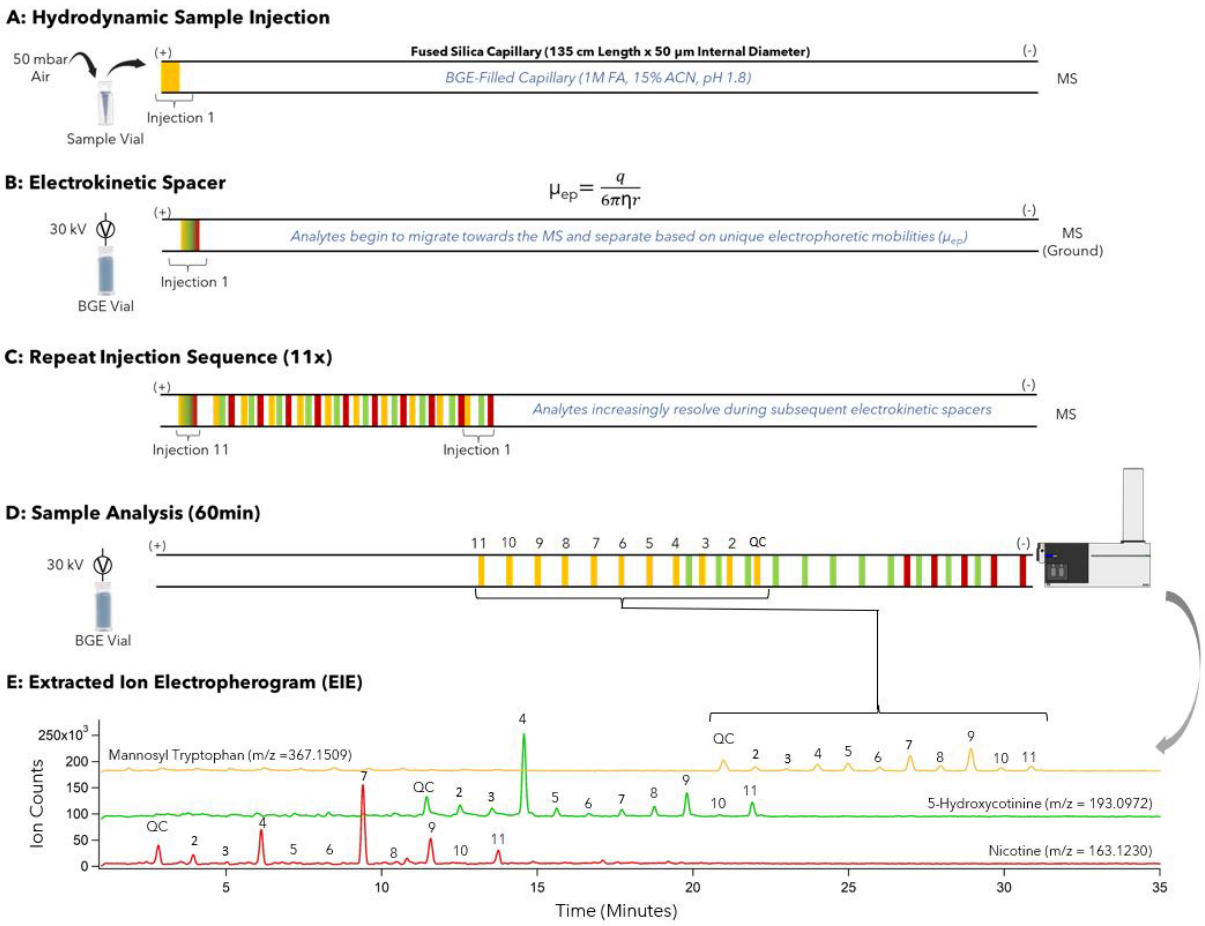


Figure 1.3 Schematic of MSI-CE-MS operation. (A) sample is hydrodynamically injected into a narrow-fused silica capillary using a positive pressure (50mbar/10seconds). (B) the sample vial at the inlet of the capillary to swapped with a BGE vial and a voltage (30 kV) is applied. During this time cationic analytes begin to migrate forward (backwards for anions) at rates according to their unique electrophoretic mobilities. (C) steps (A) and (B) are repeated until 11 injections have been completed, where separation occurs during each subsequent injection. (E) EIEs showing 11 peaks corresponding to the 11 injections. In some instances, as shown with nicotine, fewer than 11 peaks are detected indicating nicotine <LOD.

1.8 Post-Analytical Considerations

The processing of post-acquisition exposomic data can be complicated and time-consuming, especially in nontargeted analysis, which includes the conversion of raw and vendor-specific data files to a curated tabulated file.[77] General data pre-processing involves a series of steps including filtering, peak picking, spectral deconvolution, and time alignment.[73] Following which data normalization, transformation and scaling can be performed prior to multivariate statistical

analysis. Regardless of targeted or nontargeted analysis, MS-based exposomic studies require the conversion of raw data (signal intensities over mass and time dimensions) into a series of ion features characterized by an integrated peak area, accurate mass (*m/z*) and retention or migration time (RT or MT).[73] For nontargeted analysis, data must then be filtered to identify authentic peaks while rejecting background signals, chemical noise, and spurious ions.[57,73] Often, MS-based methods requiring ESI (e.g. CE, LC) can generate in-source fragments contributing to artifact signals.[76,81,95] Moreover, isotopic and adduct signals from the same metabolite also result in chemical redundancy in the data.[76] As such, with careful peak filtering, data overfitting and false discoveries can be largely avoided. Consequently, this is considered a major bottleneck in metabolomics and often involves various software tools for storage, processing and filtering of large amounts of data (e.g., XCMS, MzMine, MetAlign, MassHunter).[93,112] Although these programs can help eliminate redundancies, manual peak filtering is often still required. Subsequent time alignment procedures involving peak clustering between runs can be utilized to correct for retention time RT or migration time MT drifts.[80,95] Alignment to an internal standard may also be applied to authentic peaks to correct for RT/MT and injection variation ensuring consistent selection of a peak across multiple runs. This offers a second orthogonal parameter to accurate mass (*m/z*:RT) which can aid in molecular feature annotation.[113]

Metabolomic and exposomic studies often include QC samples to monitor and correct for variance between samples and batches.[95,114,115] In large-scale analysis, data collected intermittently over time is subject to long-term signal drift in intensity and mass accuracy.[115] Numerous approaches for batch correction have been employed based on the consistent inclusion of QCs, either using empirical Bayes methods or regression models.[116,117] Processed QC data also allows for a secondary filtering approach, where features can be removed according to

parameters such technical precision (e.g., < 30%) and frequency of detection (e.g., > 75%). Notably, parameters are subject to variation based on study objectives, where lower abundance xenobiotics may allow for higher precision cut-offs or intermittently detected drugs of abuse can be satisfied with lower detection frequency. In the case of missing values, care must be taken to avoid data skewing and overfitting. For example, consistently detected features (> 75%), where instrumental limitations (e.g., < limit of detection or LOD, ion suppression) are the likely cause of missing data, can employ the imputation of a small value (e.g., minimum value, LOD/2).[118–120] Conversely, features with extensive missing values as a result of biological response, or lack of exposure (e.g., drugs of abuse) should be more carefully considered. Customizable computational algorithms such as weighted k-nearest neighbors (KNN) or multiple imputation by chained equations (MICE) can be applied in such instances.[119] Next, normalization can be performed depending on the specific sample type. It is well known that hydration status of an individual can cause variation in effective metabolite concentrations. As a result, urinary metabolites can be normalized to creatinine, total solute content or by means of a probabilistic quotient normalization (PQN).[121] Similarly, heterogenous tissue and fecal samples are normalized to wet or dried weight, while hemodynamically controlled samples such as serum do not require normalization.

1.9 Statistical Analysis of Complex Exposomic Data Sets

Due to biological variation, dynamic range, and heteroscedasticity typically associated with complex exposome data, transformation and scaling techniques are often essential.[122,123] Correctly employed transformation methods can prevent misleading statistical outcomes as a result of artifacts, while correcting for skewed and non-normalized data distributions. Most commonly log_{10} and generalized *log* transformations are employed. For example, negative and

zero values are undefined when using log_{10} transformation, while *glog* transformation can accept positive, negative and zero values while stabilizing variance from high and low intensity signals.[122] Subsequently, scaling techniques such as autoscaling, Pareto scaling, range scaling can help account for differences in metabolite abundances.[123]

To determine the appropriate statistical model (e.g., parametric vs non-parametric) to be applied in follow up analysis, data normality can be determined using a Shapiro-Wilk or Kolmogrov-Smirnov tests.[124] Univariate or multivariate analysis can then be used to uncover significant metabolite features in a study. Single variable statistical analysis using univariate techniques can be applied by ANOVA/students t-test or Kruskall-Wallis/Mann-Whiney U tests depending on the structure of the data.[124] This offers an efficient solution to determine biological modulations as a result of toxicants by comparing predefined groups (e.g., smokers vs never smokers). As a result, true differences can be delineated from random variation. In exposome analysis, often hundreds of simultaneous univariate tests are required, and if done one by one, risk of false discovery is drastically increased. For large high-dimensional data sets, multivariate approaches can better elucidate subsets of metabolites which may be biologically linked and reduce dimensionality.[55,125,126] Unsupervised multivariate techniques aim to identify subgroups within a data set irrespective of known phenotypes or clinical data.[41] For example, principle component analysis (PCA) can help reduce data dimensionality through the projection of individual data points on to principle components (PCs), where the first few PCs explain the greatest degree of variance, ultimately aiming to preserve a high degree of variance, while reducing data dimensionality.[114]Thus reducing noise, identifying outliers, and allowing for the recognition of patterns in an efficient manner. Similarly, cluster analysis can identify groups based on similarity between features, determined according to a distance metric.[41] Additional,

network or association analysis (e.g., Heatmap) offers another unsupervised technique to determine relationships between molecular features to interpret the influence of xenobiotics on specific biological pathway components.[127,128] This offers a more exploratory data analysis approach, and is effective for visualization of data structure (e.g., group clustering). In contrast, supervised approaches such as partial least squares-discriminant analysis (PLS-DA), where *a priori* information is incorporated into the data matrix (e.g., pre- vs post-), to provide insight on top-ranked features contributing to the variation between groups, such as variable importance in projection (VIP) score in the first principle component.[126] In addition, receiver operating characteristic (ROC) curves can be applied to identify the performance of putative biomarkers for binary classification (e.g., smokers vs never smokers) once select features of significance have been identified.[129] Regardless of the form of analysis, exposome data involves a multitude of comparisons, thus risk for type 1 error (false positive) is a large concern.[73] However, overly conservative correction methods can also lead to type 2 errors (false negatives). For instance, Bonferroni may produce exceedingly conservative thresholds, when exploratory approaches in exposomics typically requires a less conservative approach, such as a Benjamini-Hochberg or false discovery rate (FDR) adjustment.[43,55]

1.10 MS Strategies for Unknown Identification

Characterizing unknown compounds of significance remains an ongoing challenge given the limited number of exposome databases.[30] Nontargeted approaches in exposomic studies may identify novel and unreported molecular features for hypothesis generation, yet biochemical interpretation is hampered without their structural elucidation. Even when using the most comprehensive mass spectral libraries, the number of identified peaks rarely exceeds 30% in contemporary metabolomic publications.[75] As a first step monoisotopic mass or most likely

molecular formula can be used to acquire a list of potential chemical candidates from various databases (e.g. PubChem, ChemSpider, HMDB, Exposome Explorer).[22,93] However, given the number of unknowns and the risk of redundant signals, further exacerbated in ESI-MS (e.g. in-source fragments, adducts), this process can be time consuming, therefore statistical significance is often first determined prior to devoting efforts towards structural identification.

MS-based approaches are largely employed for elucidation of select compounds of significance, where fragmentation patterns obtained from collision induced dissociation (CID) experiments can be utilized to aid in identification of unknowns.[28] However, given the complexities of biological samples, compounded with the risk of redundant signals, spectral elucidation is often a major challenge in exposomics, even when spectral libraries (e.g METLIN, HMDB) are available.[28,76] Importantly, the confidence level in compound identification can be determined according to the quality and orthogonality of data to support a unique chemical structure (**Table 1.3**).[130] For instance *level 1* metabolites are unambiguously identified with a minimum of two orthogonal properties such as accurate mass and RT/MT, along with matching MS/MS spectra with an authentic chemical standard under identical operating conditions. [81,113] **Figure 1.4** outlines a recent example by DiBattista *et al.*[103] of an unambiguous *level 1* identification of an unknown, putatively identified as ophthalmic acid (OPA) prior to confirmation using a representative DBS and authentic OPA chemical standard where CID spectra are obtained under matched experimental conditions with low mass error (< 5ppm). *Level 2* identification refers to putatively identified compounds with annotated fragmentation spectra matched to a publicly available spectral library or database in conjunction with a matching accurate mass and RT/MT.[76] Often, this can be difficult given different conditions for spectra with various platforms and limited spectral database availability. Additionally, matching

Table 1.3 Recommended standards for unknown metabolite identification confidence levels.[81,130]

Confidence Level	Minimum Requirements
Level 1	*"Validated Structure"*; Matched identification using authentic standards under identical analytical conditions with minimum two orthogonal properties (e.g m/z, retention/migration time)
Level 2	*"Putative ID"*; Probable structure based on physiochemical characteristics and a matched library MS/MS structure.
Level 3	*"Tentative Structure";* Spectral and/or physiochemical properties indicating compounds class.
Level 4	*"Unknown";* Reproducible and quantifiable signal with an assigned accurate mass, retention/migration time and most likely molecular formula.

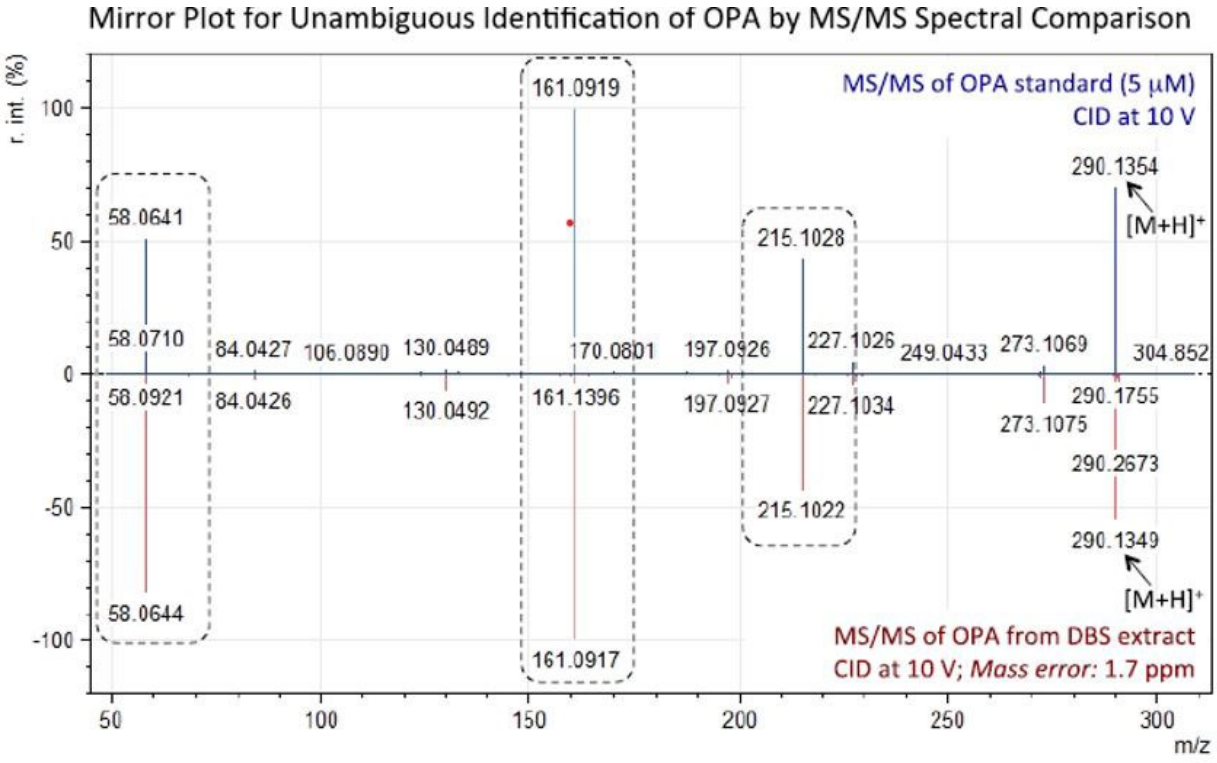


Figure 1.4. A mirror plot comparing MS/MS spectra of an unknown putatively identified as ophthalmic acid (OPA, *m/z* 290.1349), measured from a pooled DBS extract (red trace), with an authentic OPA standard (blue trace) at a CID of 10V. Unambiguous *Level 1* identification is confirmed with a low mass error (1.7 ppm) and excellent match in both abundance and accurate mass for three diagnostic product ions (*m/z* 58.0644, *m/z* 161.0917, *m/z* 215.1022) Reproduced with permission from DiBattista *et al.* 2019.

RT and MTs are not currently available for LC and CE based separations making it difficult to obtain a second level of identification. *Level 3* features include compounds with a tentative structure relying on a comparison of measured physiochemical properties and/or obtained MS/MS spectra to other compounds with similar chemical characteristics (e.g., same chemical class).[76] Lastly, *level 4* features are classified as unknowns with a reproducible signal and assigned accurate mass, RT/MT and most likely molecular formula.[76] Consequently, tools for predicting spectral fragmentation and RT have also been developed (e.g., MS FINDER, CFM ID) when databases fall short.[75] In addition, bioinformatics tools can be used with biological pathway information to filter and rank lists of candidates for feature identification (e.g., XCMS, xMSannotator).[72,81]. However, despite the improvements in high resolution mass spectrometry, identification without a commercially available standard presents a major challenge, as one accurate mass may be common with multiple structures (e.g., isobar, stereoisomer),[30,130] Moreover, acquisition of MS/MS spectra for reliable interpretation may be difficult for low abundance precursor ions. Consequently, filtering based on chemical reactivity, enzymatic reactivity, in silico modeling can also be applied to help remove isobaric and isomeric features.[130,131]

1.11 Risk Assessment for Occupational and Population Health

Risk assessment can be defined as the likelihood of an environmental hazard causing adverse effects to the health and safety of individuals and populations.[132] The process involves integrating information on the causally linked health effects of an exposure (hazard identification) with the route, duration and magnitude of that exposure (dose-response). Risk assessment is used in a variety of contexts including occupational settings, food safety and public health policy.[133,134] As a result, occupational health programs such as National Institute for

Occupational Safety and Health (NIOSH) have been implemented to evaluate risk and establish biological exposure index (BEI) thresholds.[135] For instance occupational exposure to smoke (e.g. firefighters) further increases the incidence of cardiorespiratory illnesses (e.g. chronic obstructive pulmonary disease, cardiac infarction) and is associated with a higher risk for cancer.[84,136] Despite mitigation strategies such as self-contained breathing apparatuses', hygienic practices and personal protective equipment (e.g., bunker gear) **(Figure 1.5A)** to reduce inhalation, ingestion and dermal absorption of intoxicants, higher cancer rates are still prevalent among firefighters as compared with the general public.[137] Similarly, risk assessment is applied in the context of public health policy. For instance, air pollution remains one of the largest fields of environmental risk assessment worldwide.[138] Consequently, public health policies such as the Clean Air Act (CAA) have been employed to limit emissions of air pollutants from various industrial sources.[139]

Various methods of risk assessment have been employed, including cumulative risk assessment (CRA), toxicity testing, biomonitoring and most recently exposomics.[28,140] CRA is a broader approach which addresses multiple aggregate exposures and combines risks for common health outcomes.[140] While CRAs emphasize multiple exposures, it is often challenging to identify groups and develop common metrics to evaluate dissimilar risks.[140] CRAs are also yet to be applied in an occupational context.[140] Conversely toxicity testing is a targeted testing system to determine underlying pathways and adverse health effects of acute exposures.[28] Additionally, biomonitoring can be used as a tool to observe the efficacy of public health policies and investigate work protection factors to better mitigate chronic disease risk and improve long-term health outcomes.[141] Accordingly, many population-based biomonitoring efforts such as the national survey conducted by the CDC, such as the National Health and Nutrition Examination

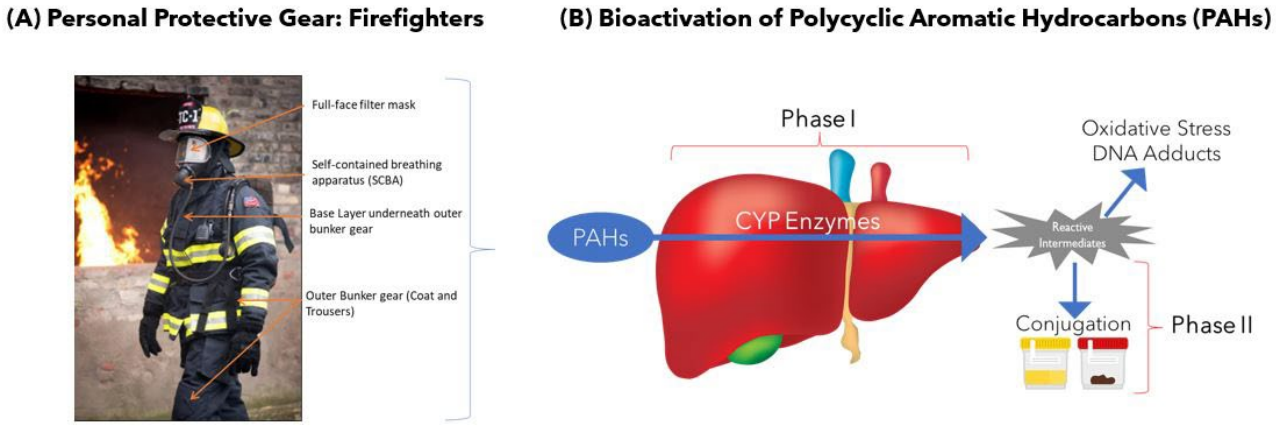


Figure 1.5. (A) Personal protective gear commonly worn by firefighters, where the use of a self-contained breather apparatus (SCBA), and heavy outer bunker gear may vary depending on fire event (ex. Wildland vs structural fire). **(B)** Subsequent intake of PAHs either due to inhalation, ingestion or dermal exposure results in bioactivation by phase I metabolism (CYP P450 isoforms: CYP 1A1, CYP 1B1, CYP 1A2) leading to the formation of reactive intermediates which can have mutagenic and carcinogenic effects. Reactive intermediates are subsequently conjugated to form a glucuronide or sulfate conjugate prior to excretion into urine and/or feces

Survey (NHANES) are ongoing to establish reference ranges of chemicals in the United States.[32]

Understanding the composition of an initial exposure (e.g., tobacco, wood smoke), can provide insight on the potential biomarkers which may be most amenable for biomonitoring for reliable risk assessment. In the context of smoke exposure, biomarkers reflecting recent exposure, either from air pollution, wood smoke or cigarettes should demonstrate adequate sensitivity, and specificity, while also being able to delineate from background sources of exposure are critical.[90] For instance, OH-PAHs have been widely reported for biomonitoring smoke exposure in both wildland and structural firefighters, however their ubiquitous nature (e.g. barbecuing, fossil fuels, second-hand-smoke), makes them less specific and more difficult to interpret in population-based studies.[90] Importantly, PAHs such as benzo[a]pyrene require metabolic activation by specific CYP isoforms (e.g. CYP 1A1, CYP 1A2, CYP 1B1), in order to induce carcinogenic effects.[142] As a result cancer risk is strongly impacted by interindividual

differences in bioactivation and subsequent detoxification (e.g., glucuronidation) (**Figure 1.5B.**). Notably, cigarette smoke is an inducer of PAH bioactivation and its subsequent carcinogenicity, thus it may provide evidence for current smokers who are at greater risk for cancer.[142] Accordingly, PAH-DNA adducts, formed from the bioactivation of PAHs by CYP P450 enzymes can provide a more accurate measure of the biological effect associated with the smoke exposure, thus serving as both biomarkers of exposure and physiological response (**Figure 1.5B**).[90] In contrast, nicotine and its respective metabolites (**Figure 1.6**) serve as more specific biomarkers for smoke derived from tobacco use.[143] For example, cotinine is a widely measured biomarker for tobacco exposure, which has shown a correlation with self-reported cigarettes per day (CPD), and has also demonstrated a dose-dependent association with lung cancer risk.[144] However, excretion rates of individual nicotine metabolites are often largely impacted by several factors including genetics, diet and medication use, which has resulted in demographic variations of cut-points used to distinguish current and never smokers.[145] In contrast, the sum of nicotine metabolites (e.g., total nicotine equivalents or TNE) mitigates the impact of metabolism differences and serves as a more reliable measure of exposure, especially in more diverse populations. [146] Additionally, phenotypic biomarkers based on metabolite ratios can be employed to better evaluate the impact of nicotine metabolism.[147] Overall, careful consideration in the selection of biomarkers for biomonitoring either in the context of occupational exposure or population health should be taken.

In both occupational and population health, risk assessment, often a linear relationship between the initial exposure and biomarkers of exposure/effect is assumed. However, diet and lifestyle factors remain major sources of variation which may limit the utility of many recognized that nutrition plays a significant role metabolism, oxidation and inflammation,

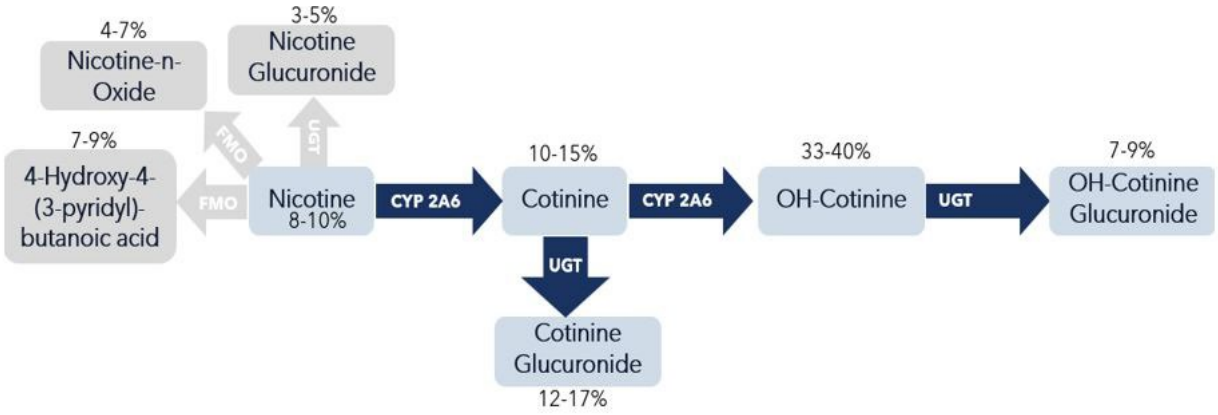


Figure 1.6. Overview of nicotine metabolism and its subsequent excretion profile in urine, where individual metabolites range in their excretion as a result of genetic and environmental factors

biomarkers as specific indicators of exposure.[36] Consequently, it has become increasingly where a suboptimal diet is currently one of the leading causes of preventable deaths worldwide.[148] Importantly, the relationship between diet and host physiology is considered bidirectional, and further influenced by cultural and lifestyle factors.[148] Consumption of nutrients such as vitamins, antioxidants, amino acids and bioactive compounds (e.g., chlorogenic acids), can largely influence chronic disease risk. For instance, chlorogenic acids found in coffee have demonstrated beneficial health effects against cardiovascular disease and overall mortality.[149] Conversely, ultra-processed foods containing artificial sweeteners have been associated with increased risk of irritable bowel syndrome, diabetes and certain cancers.[150] Many of these dietary exposures may also impact metabolism and bioactivation of contaminants.[151] For instance, cruciferous vegetables are known to modulate CYP P450 activity, specifically the induction of CYP 1A.[151] However, often environmental biomonitoring and occupational risk assessment either do not consider the contribution of diet, or it is recorded based on self-reported questionnaires, which are subject to various forms of bias and misreporting, which are further exacerbated in certain regions of the world.[152] As result robust biomarkers reflective of recent dietary intake can

provide further insight on the complex relationships between exposure and disease. Given the complexity in exposures and the influence of diet, lifestyle and genetics, a more comprehensive approach towards exposome characterization can provide more accurate and risk assessment.[66] As outlined by Martyn *et al.*[153] populations with the highest "totality of exposures" (e.g., firefighters, current smokers) should be the focus of future exposome studies as they are at increased risk for chronic disease. Overall, the utilization of exposomics for risk assessment in diverse populations can contribute to the mapping the genetic vs environmental risk factors of chronic disease burden.

1.12 Thesis Motivation: Exposome Characterization for Risk Assessment in Occupational and Population Health

In the last century, with the rapid expansion of industry, humans have been exposed to an increasing number of man-made and naturally occurring chemicals, in the workplace, at home and the wider envrionment.[28] The impact of such complex exposures from sources such as air pollution and tobacco smoke, compounded with poor diet quality are among the leading causes of preventable disability and mortality worldwide.[7] However, the underlying mechanisms of exposure-disease relationships remain poorly understood, and emphasizes the need for advancements in human exposome characterization for improved strategies in risk assessment. With the ability of metabolomics to depict a diverse series of endogenous and exogenous metabolites reflective of genetics, environment, and phenotype, it offers a promising approach toward understanding the developmental origins of health and disease.[154] Consequently, identification and validation of metabolites serving as reliable biomarkers, along with reliable analytical methods amenable to comprehensive biomonitoring are urgently needed. Thus, the work presented in this thesis centers around the development of high throughput technologies to

quantify robust biomarkers of environmental smoke exposure in high-risk occupations, with additional applications towards global health initiatives. In this context, the validation and inter-laboratory method comparison of urinary 1-hydroxypyrene (HP) for PAH exposure assessment was conducted and applied to firefighters deployed under emergency conditions to the 2016 Fort McMurray wildfire, where minimal access to protective equipment was available (*Chapter II*). This work sheds light on limitations in current analytical techniques for smoke exposure analysis, including reporting bias and the need for standardization in sample pre-processing protocols (e.g., enzyme deconjugation). As a result, an MSI-CE-MS/MS method was developed and rigorously validated for the direct analysis of urinary HP as its intact glucuronide conjugate (HP-G) in firefighters, minimizing bias and greatly improving sample throughput, while avoiding complex sample workup procedures often used for GC-MS. (*Chapter III*) From biomonitoring smoke exposure in an occupational health context, this thesis expanded to elucidating health risks among tobacco users on an international scale. In this case, MSI-CE-MS was first validated for analysis of a comprehensive panel of nicotine metabolites in urine to biochemically verify smoking status and smoking behavior in a large cohort (n=1000) from the Prospective Urban and Rural Epidemiological (PURE) study (*Chapter IV*). Lastly, nontargeted profiling of the urine metabolome was employed in the PURE cohort for a wide range of polar/ionic metabolites using MSI-CE-MS. This enabled the identification of a panel of robust and generalizable dietary biomarkers for recent food intake, which may also reflect regional variation in nutrition, and complex diet-disease relationships among smokers (*Chapter V*).

1.13 Biomonitoring Smoke Exposure in Firefighters

PAHs are a class of contaminants formed from the incomplete combustion of organic materials known for their carcinogenic and mutagenic properties.[155] The ubiquitous nature of PAHs in

numerous forms of smoke, charred meats and ambient air pollution has made them a global health concern for chronic disease risk.[90] Occupational exposure to PAHs has demonstrated increased rates of cancer and cardiovascular disease among industrial workers and firefighters.[136] Mitigation strategies such as personal protective gear and self-contained breathing apparatus are not consistent among urban and wildland firefighters, where wildland firefighters often have limited access to protective equipment.[84,156] For instance, a recent collaboration with Cherry *et al*.[157] demonstrated dermal absorption as an important route for PAH exposure, outlining a reduction in exposure among individuals with better access to proper hygiene practices. However, under emergency circumstances such as those observed during the 2016 Fort McMurray wildfire, firefighters are often subject to heavy smoke exposure with limited access to protective equipment, making biomonitoring in these contexts critical for risk assessment.[157] Despite the widespread use of urinary HP as a biomarker of recent smoke exposure[90] few studies have reported a rigorous validation of HP, with no inter-laboratory method studies having been done to ensure reliable quantification. *Chapter II* describes an inter-laboratory method comparison for urinary HP when using GC-HRMS and LC-MS/MS on urine specimens collected from firefighters ($n = 42$) deployed to the Fort McMurray wildfire. Overall, despite acceptable accuracy and precision for both methods, a modest bias was observed, likely attributed to variations in enzyme hydrolysis protocols between both laboratories. In addition, given the complex nature of wood smoke, distribution of HP and its correlation to several other PAH and methoxyphenols (MP) metabolites was also investigated. However, only two OH-PAH isomers exhibited a significant correlation, demonstrating distinct excretion profiles among PAH metabolites likely outlining variations in PAH metabolism and bioactivation by CYP isoforms. Accordingly, the work presented in *Chapter II* highlights the need for standardization of complex

analytical protocols (e.g., enzyme hydrolysis) and the potential importance of biomarker selection based on differences in urinary metabolite excretion.

1.14 Rapid Screening of HP-G in Urine

Previously reported methods for urinary HP analysis, including LC with native fluorescence detection, LC-MS/MS and GC-MS have demonstrated excellent sensitivity and selectivity, but are limited by low sample throughput and complicated sample workup protocols.[85,90] *Chapter III* introduces a high throughput MSI-CE-MS/MS method to directly analyze the intact glucuronide conjugate of HP (HP-G) following a simple acidified ether extraction on urine specimens obtained from firefighters following recent smoke exposure during the Fort McMurray wildfire. The multiplexed injection approach enables the analysis of 13 independent urine extracts within a single run (< 3 min/sample) with stringent QC, while avoiding enzyme deconjugation and precolumn chemical derivatization. Rigorous method optimization to achieve adequate sensitivity, comparable to that of previous LC-MS/MS methods was performed to reach low ng/mL detection limits. Subsequent method validation demonstrated good linearity, technical precision, and recoveries over a wide concentration range. A cross-platform inter-laboratory method comparison between the direct analysis of HP-G by MSI-CE-MS/MS and total hydrolyzed urinary HP by GC–MS further establishes excellent mutual agreement with minimal bias. Subsequent quantification of HP-G among firefighters revealed creatinine corrected concentrations below the biological exposure index, likely a consequence of delays in sample collection under emergency conditions. Overall, MSI-CE-MS/MS offers a multiplexed separation platform, ideal for large-scale biomonitoring studies and risk assessment among occupationally exposed firefighters.

1.15 Biomonitoring Tobacco Exposure and Smoking Patterns in the PURE Cohort

Tobacco use is the leading cause of preventable death and disease worldwide, with over one billion current smokers.[158] However, tobacco related disease risk is not uniform across populations, where recent work by Sathish *et al*.[159] has demonstrated discrepancies in the hazards associated with smoking between countries of varying income status. Common practice for epidemiological studies investigating tobacco use patterns relies on self-reported questionnaires, which are often prone to error and reporting bias.[160] Robust biomarkers for quantitative assessment of recent tobacco smoke exposure include analysis of the total nicotine equivalent (TNE-7), which comprises the sum of seven major nicotine metabolites excreted in urine.[161] Specific nicotine metabolite ratios have also exhibited excellent utility as phenotypic biomarkers of enzyme activity (e.g. CYP 2A6) to assess smoking behavior, such as the nicotine metabolic ratio (NMR).[88] However, current methods for comprehensive analysis of nicotine and its metabolites in urine rely on complex sample preparation techniques or two-step enzyme deconjugation protocols, which are limited in sample throughput and suffer from high operating costs.[88] *Chapter IV* introduces a simple and cost-effective method for direct analysis of up to seven nicotine metabolites by MSI-CE-MS that was applied in a large-scale PURE cohort with acceptable technical precision, good stability and adequate limits of detection. Due to the influence of smoking behavior on tobacco related disease risk, several urinary metabolites and their ratios were explored to monitor enzyme activity among current smokers to better evaluate the impact of metabolism on smoking habits (e.g., puff volume) and nicotine dependence. Overall, TNE-7 demonstrated a moderate correlation with self-reported CPD with the highest rates of misreporting occurring among LICs. Additionally, tobacco smoke exposure was highest in HICs among heavy smokers based on biochemical verification. Subsequent stratification of

NMR enabled the identification of fast metabolizers, where those with an elevated ratio, are more likely to exhibit increased CYP 2A6 activity. High rates of nicotine metabolism have been previously associated with larger puff volumes per cigarette, and greater nicotine dependence.[162] Similarly, this work demonstrated elevated tobacco exposure based on increased TNE-7 concentrations among individuals classified as fast metabolizers. Consequently, these select participants were predicted to have the highest risk for toxicant exposure because of increased tobacco intake. *Chapter IV* outlines a robust and reliable method for the biochemical verification of tobacco smoke exposure to distinguish smoking status, while providing new insights into smoking behaviors and regional differences in tobacco-related disease risk worldwide.

1.16 High-Throughput Metabolomics for Dietary Biomarker Classification

In addition to tobacco smoking, a suboptimal diet poses a considerable burden on human health, where several chronic non-communicable diseases (NCDs) including cancer, cardiovascular disease and diabetes have demonstrated strong associations with select lifestyle factors (e.g., smoking, diet, physical activity).[163] Specifically, poor diet quality defined as the habitual intake of high caloric, nutrient-poor, ultra-processed foods, is a leading modifiable risk factor for preventable morbidity and mortality worldwide.[164] Current methods for monitoring dietary exposures primarily rely on self-reports either through 24 h diet records or food frequency questionnaires (FFQs).[165] Accordingly, scoring methods such as the Alternative Healthy Eating Index (AHEI) have been developed, where adherence to an AHEI based diet have indicated reduced risk for mortality.[166] However, diet recall methods are prone to random and systematic error, exacerbated in low socioeconomic groups.[167] Thus, *Chapter V* utilizes MSI-CE-MS to demonstrate a nontargeted, high-throughput metabolomics data workflow for the characterization of dietary exposures in an international cohort of participants from 14 different countries in the

PURE study. Overall, a total of 116 polar/ionic urinary metabolites were reliably measured (CV<30%) and compared against semi-quantitative food records from FFQs (*n*=60), with subsequent evaluation of dose response. This work identified a panel of eight top-ranked candidate dietary biomarkers in urine, indicative of five distinct food classes, while also delineating regional trends in dietary habits, their association with smoking and overall diet quality. Importantly, *Chapter V* outlines the identification of objective dietary biomarkers, as a means of providing more comprehensive and reliable assessments of nutrition across diverse populations, which may provide further insight on the complex relationships between environmental exposures, diet, and chronic disease risk.

1.17 References

(1) Trumble, B. C.; Finch, C. E. The Quarterly Review of Biology The Exposome in Human Evolution: From Dust to Diesel. *Q Rev Biol.* **2019**, Vol. 94.

(2) Fogel, R. W.; Costa, D. L. A Theory of Technophysio Evolution, with Some Implications for Forecasting Population, Health Care Costs, and Pension Costs. *Demography* **1997**, *34* (1), 49–66.

(3) Drevenstedt, G. L.; Crimmins, E. M.; Vasunilashorn, S.; Finch, C. E. The Rise and Fall of Excess Male Infant Mortality. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105* (13), 5016–5021.

(4) Brian, C.; Clarke, B.; Blackadar, B.; Blackadar, C. B.; Researcher, V. Historical Review of the Causes of Cancer.*World Journal of Clinical Oncology*. **2016**, 7(1), 54.

(5) BROWN, J. R.; THORNTON, J. L. Percivall Pott (1714-1788) and Chimney Sweepers' Cancer of the Scrotum. *Br. J. Ind. Med.* **1957**, *14* (1), 68.

(6) Doll, R.; Hill, A. B. Smoking and Carcinoma of the Lung. *Br. Med. J.* **1950**, *2* (4682), 739.

(7) Furman, D.; Campisi, J.; Verdin, E.; Carrera-Bastos, P.; Targ, S.; Franceschi, C.; Ferrucci, L.; Gilroy, D. W.; Fasano, A.; Miller, G. W.; Miller, A. H.; Mantovani, A.; Weyand, C. M.; Barzilai, N.; Goronzy, J. J.; Rando, T. A.; Effros, R. B.; Lucia, A.; Kleinstreuer, N.; Slavich, G. M. Chronic Inflammation in the Etiology of Disease across the Life Span. *Nat. Med.* **2019**, *25* (12), 1822–1832.

(8) World Health Organization. Diet, Nutrition, and the Prevention of Chronic Diseases : Report of a Joint WHO/FAO Expert Consultation.; World Health Organization, **2003**.

(9) Wild, C. P. Complementing the Genome with an "Exposome": The Outstanding Challenge of Environmental Exposure Measurement in Molecular Epidemiology. *Cancer Epidemiology Biomarkers and Prevention*. **2005**, *14*(8), 1847–1850.

(10) Wild, C. P. The Exposome: From Concept to Utility. *International Journal of Epidemiology*. **2012**, *41*(1), 24–32.

(11) Strimbu, K.; Tavel, J. A. What Are Biomarkers? *Curr Opin HIV AIDS*. **2010** Nov; 5(6): 463–466.

(12) Vaidya, V. S., & Bonventre, J. V. Biomarkers: In Medicine, Drug Discovery, and Environmental Health. John Wiley & Sons, **2010**.

(13) Haber, M. H. Pisse Prophecy: A Brief History of Urinalysis. *Clin. Lab. Med.* **1988**, *8* (3), 415–430.

(14) Silins, I.; Högberg, J. Combined Toxic Exposures and Human Health: Biomarkers of Exposure and Effect. *International Journal of Environmental Research and Public Health*. MDPI **2011**, 8(3), 629–647.

(15) Au, W. W. Usefulness of Biomarkers in Population Studies: From Exposure to Susceptibility and to Prediction of Cancer. *Int. J. Hyg. Environ. Health* **2007**, *210* (3–4), 239–246.

(16) Timbrell, J. A. Biomarkers in Toxicology. *Toxicology* **1998**, *129* (1), 1–12.

(17) Murphy, S. E. Nicotine Metabolism and Smoking: Ethnic Differences in the Role of P450 2A6. *Chemical Research in Toxicology*. **2017**, 30(1), 410–419.

(18) Biomarkers of Exposure - Canada.ca https://www.canada.ca/en/health-canada/services/publications/healthy-living/biomarkers-exposure.html (accessed 2022 -04 -28).

(19) World Health Organization. Biomarkers In Risk Assessment: Validity And Validation – Environmental Health Criteria 222. **2001**.

(20) Barros, B.; Oliveira, M.; Morais, S. Urinary Biohazard Markers in Firefighters. *Adv. Clin. Chem.* **2021**, *105*, 243–319.

(21) Lu, S. Y.; Li, Y. X.; Zhang, T.; Cai, D.; Ruan, J. J.; Huang, M. Z.; Wang, L.; Zhang, J. Q.; Qiu, R. L. Effect of E-Waste Recycling on Urinary Metabolites of Organophosphate Flame Retardants and Plasticizers and Their Association with Oxidative Stress. *Environ. Sci. Technol.* **2017**, *51* (4), 2427–2437.

(22) Vineis, P.; Robinson, O.; Chadeau-Hyam, M.; Dehghan, A.; Mudway, I.; Dagnino, S. What Is New in the Exposome? *Environment International*. **2020**.143,105887.

(23) Basu, A. K. DNA Damage, Mutagenesis and Cancer. *Int. J. Mol. Sci.* **2018**, *19* (4).

(24) Yuan, J. M.; Nelson, H. H.; Carmella, S. G.; Wang, R.; Kuriger-Laber, J.; Jin, A.; Adams-Haduch, J.; Hecht, S. S.; Koh, W. P.; Murphy, S. E. CYP2A6 Genetic Polymorphisms and Biomarkers of Tobacco Smoke Constituents in Relation to Risk of Lung Cancer in the Singapore Chinese Health Study. *Carcinogenesis.* **2017**, *38* (4), 411–418.

(25) Novelli, G.; Biancolella, M.; Latini, A.; Spallone, A.; Borgiani, P.; Papaluca, M. Precision Medicine in Non-Communicable Diseases. *High-throughput.* **2020**, *9*(1), 3.

(26) Ragas, A. M. J. Trends and Challenges in Risk Assessment of Environmental Contaminants. *Journal of Integrative Environmental Sciences*. **2011**, *8*(3), 195–218.

(27) Maertens, A.; Golden, E.; Hartung, T. Avoiding Regrettable Substitutions: Green Toxicology for Sustainable Chemistry. *ACS Sustain. Chem. Eng.* **2021**, *9* (23), 7749–7758.

(28) Bocato, M. Z.; Bianchi Ximenez, J. P.; Hoffmann, C.; Barbosa, F. An Overview of the Current Progress, Challenges, and Prospects of Human Biomonitoring and Exposome Studies. *J. Toxicol. Environ. Heal. - Part B Crit. Rev.* **2019**, *22* (5–6), 131–156.

(29) Paustenbach, D. J.; Galbraith, D. Biomonitoring and Biomarkers: Exposure Assessment Will Never Be the Same. *Environ. Health Perspect.* **2006**, *114* (8), 1143–1149.

(30) Dennis, K. K.; Marder, E.; Balshaw, D. M.; Cui, Y.; Lynes, M. A.; Patti, G. J.; Rappaport, S. M.; Shaughnessy, D. T.; Vrijheid, M.; Barr, D. B. Biomonitoring in the Era of the Exposome. *Environmental Health Perspectives.* **2017**, *25*(4), 502–510.

(31) Palipudi, K. M.; Morton, J.; Hsia, J.; Andes, L.; Asma, S.; Talley, B.; Caixeta, R. D.; Fouad, H.; Khoury, R. N.; Ramanandraibe, N. Methodology of the Global Adult Tobacco Survey—2008–2010. *Glob. Health Promot.* **2016**, *23* (2), 3–23.

(32) Sobus, J. R.; DeWoskin, R. S.; Tan, Y. M.; Pleil, J. D.; Phillips, M. B.; George, B. J.; Christensen, K.; Schreinemachers, D. M.; Williams, M. A.; Hubal, E. A. C.; Edwards, S. W. Uses of NHANES Biomarker Data for Chemical Risk Assessment: Trends, Challenges, and Opportunities. *Environ. Health Perspect.* **2015**, *123* (10), 919–927.

(33) Haddad, N.; Andrianou, X. D.; Makris, K. C. A Scoping Review on the Characteristics of Human Exposome Studies. *Current Pollution Reports*. **2019**, *5*(4), 378–393.

(34) Vermeulen, R.; Schymanski, E. L.; Barabási, A.-L.; Miller, G. W. The Exposome and Health: Where Chemistry Meets Biology. *Science*. **2020**, *367*(6476), 392-396.

(35) Wild, C. P.; Scalbert, A.; Herceg, Z. Measuring the Exposome: A Powerful Basis for Evaluating Environmental Exposures and Cancer Risk. *Environmental and Molecular Mutagenesis*. **2013**, *54*(7), 480–499.

(36) Zhang, P.; Carlsten, C.; Chaleckis, R.; Hanhineva, K.; Huang, M.; Isobe, T.; Koistinen, V. M.; Meister, I.; Papazian, S.; Sdougkou, K.; Xie, H.; Martin, J. W.; Rappaport, S. M.; Tsugawa, H.; Walker, D. I.; Woodruff, T. J.; Wright, R. O.; Wheelock, C. E. Defining the Scope of Exposome Studies and Research Needs from a Multidisciplinary Perspective. *Environmental Science and Technology Letters*. **2021**, *8*(10), 839–852.

(37) Heijmans, B. T.; Tobi, E. W.; Stein, A. D.; Putter, H.; Blauw, G. J.; Susser, E. S.; Slagboom, P. E.; Lumey, L. H. Persistent Epigenetic Differences Associated with Prenatal Exposure to Famine in Humans. *Proc. Natl. Acad. Sci.*. **2008**, *105* (44), 17046–17049.

(38) Holland, N. Future of Environmental Research in the Age of Epigenomics and Exposomics. *Reviews on Environmental Health*. **2017**, *32*(1-2), 45–54.

(39) O'Donnell, S. T.; Ross, R. P.; Stanton, C. The Progress of Multi-Omics Technologies: Determining Function in Lactic Acid Bacteria Using a Systems Level Approach. *Front. Microbiol.* **2020**, *10*, 3084.

(40) Pintarelli, G.; Noci, S.; Maspero, D.; Pettinicchio, A.; Dugo, M.; De Cecco, L.; Incarbone, M.; Tosi, D.; Santambrogio, L.; Dragani, T. A.; Colombo, F. Cigarette Smoke Alters the Transcriptome of Non-Involved Lung Tissue in Lung Adenocarcinoma Patients. *Sci. Reports* **2019**, *9* (1), 1–10.

(41) Santiago-Rodriguez, T. M.; Hollister, E. B. Multi 'omic Data Integration: A Review of Concepts, Considerations, and Approaches. *Semin. Perinatol.* **2021**, *45* (6).

(42) Jain, R. B. Use of Urinary Thiocyanate as a Biomarker of Tobacco Smoke. *Epidemiol. Open Access* **2016**, *6* (5), 151456.

(43) DeBord, D. G.; Carreón, T.; Lentz, T. J.; Middendorf, P. J.; Hoover, M. D.; Schulte, P. A. Use of the "Exposome" in the Practice of Epidemiology: A Primer on -Omic Technologies. *Am. J. Epidemiol.* **2016**, *184* (4), 302–314.

(44) Vlaanderen, J.; Portengen, L.; Rappaport, S. M.; Glass, D. C.; Kromhout, H.; Vermeulen, R. The Impact of Saturable Metabolism on Exposure-Response Relations in 2 Studies of Benzene-Induced Leukemia. *Am. J. Epidemiol.* **2011**, *174* (5), 621–629.

(45) Sadeghi, L.; Jenzer, H. Omics Technologies: The Most Convincing Tools for a (Fully) Individualized Diet? *J. Mol. Biomark. Diagn.* **2016**, *8*(311),2.

(46) Sawyers, C. L. The Cancer Biomarker Problem. *Nature* **2008**, *452* (7187), 548–552.

(47) Rappaport, S. M. Discovering Environmental Causes of Disease. *J. Epidemiol. Community Health* **2012**, *66* (2), 99–102.

(48) Beland, F. A.; Poirier, M. C. Significance of DNA Adduct Studies in Animal Models for Cancer Molecular Dosimetry and Risk Assessment. In *Environmental Health Perspectives*, **1993,** 99, 5–10.

(49) Pereira, S. A.; Antunes, A. M. M. Special Issue "Adductomics: Elucidating the Environmental Causes of Disease." *High-Throughput.* **2019**, *8* (3), 2–3.

(50) Preston, G. W.; Phillips, D. H. Protein Adductomics: Analytical Developments and Applications in Human Biomonitoring. *Toxics* **2019**, *7* (2), 29.

(51) Hanash, S. M.; Pitteri, S. J.; Faca, V. M. Mining the Plasma Proteome for Cancer Biomarkers. *Nature* **2008**, *452* (7187), 571–579.

(52) Olivier, M.; Asmis, R.; Hawkins, G. A.; Howard, T. D.; Cox, L. A. The Need for Multi-Omics Biomarker Signatures in Precision Medicine. *Int. J. Mol. Sci.* **2019**, *20* (19), 4781.

(53) Idle, J. R.; Gonzalez, F. J. Metabolomics. *Cell Metab.* **2007**, *6* (5), 348–351.

(54) Ryan, D.; Robards, K. Metabolomics: The Greatest Omics of Them All? *Anal. Chem.* **2006**, *78* (23), 7954–7958.

(55) Tzoulaki, I.; Ebbels, T. M. D.; Valdes, A.; Elliott, P.; Ioannidis, J. P. A. Design and Analysis of Metabolomics Studies in Epidemiologic Research: A Primer on-Omic Technologies. *Am. J. Epidemiol.* **2014**, *180* (2), 129–139.

(56) Wishart, D. S.; Guo, A. C.; Oler, E.; Wang, F.; Anjum, A.; Peters, H.; Dizon, R.; Sayeeda, Z.; Tian, S.; Lee, B. L.; Berjanskii, M.; Mah, R.; Yamamoto, M.; Jovel, J.; Torres-Calzada, C.; Hiebert-Giesbrecht, M.; Lui, V. W.; Varshavi, D. D.; Varshavi, D. D.; Allen, D.; Arndt, D.; Khetarpal, N.; Sivakumaran, A.; Harford, K.; Sanford, S.; Yee, K.; Cao, X.;

Budinski, Z.; Liigand, J.; Zhang, L.; Zheng, J.; Mandal, R.; Karu, N.; Dambrova, M.; Schiöth, H. B.; Greiner, R.; Gautam, V. HMDB 5.0: The Human Metabolome Database for 2022. *Nucleic Acids Res.* **2022**, *50* (D1), D622–D631.

(57) Hyötyläinen, T. Analytical Challenges in Human Exposome Analysis with Focus on Environmental Analysis Combined with Metabolomics. *Journal of Separation Science*. John Wiley and Sons Inc April 1, **2021***, 44*(8), 1769–1787.

(58) Zhang, P.; Arora, M.; Chaleckis, R.; Isobe, T.; Jain, M.; Meister, I.; Melén, E.; Perzanowski, M.; Torta, F.; Wenk, M. R.; Wheelock, C. E. Tackling the Complexity of the Exposome: Considerations from the Gunma University Initiative for Advanced Research (GIAR) Exposome Symposium. *Metabolites* **2019**, *9*(6), 106.

(59) Oliveira, M.; Slezakova, K.; Delerue-Matos, C.; Pereira, M. C.; Morais, S. Children Environmental Exposure to Particulate Matter and Polycyclic Aromatic Hydrocarbons and Biomonitoring in School Environments: A Review on Indoor and Outdoor Exposure Levels, Major Sources and Health Impacts. *Environment International*. **2019**.124,180-204

(60) Maitre, L.; De Bont, J.; Casas, M.; Robinson, O.; Aasvang, G. M.; Agier, L.; Andrušaitytė, S.; Ballester, F.; Basagaña, X.; Borràs, E.; Brochot, C.; Bustamante, M.; Carracedo, A.; De Castro, M.; Dedele, A.; Donaire-Gonzalez, D.; Estivill, X.; Evandt, J.; Fossati, S.; Giorgis-Allemand, L.; Gonzalez, J. R.; Granum, B.; Grazuleviciene, R.; Gützkow, K. B.; Haug, L. S.; Hernandez-Ferrer, C.; Heude, B.; Ibarluzea, J.; Julvez, J.; Karachaliou, M.; Keun, H. C.; Krog, N. H.; Lau, C. H. E.; Leventakou, V.; Lyon-Caen, S.; Manzano, C.; Mason, D.; McEachan, R.; Meltzer, H. M.; Petraviciene, I.; Quentin, J.; Roumeliotaki, T.; Sabido, E.; Saulnier, P. J.; Siskos, A. P.; Siroux, V.; Sunyer, J.; Tamayo, I.; Urquiza, J.; Vafeiadi, M.; Van Gent, D.; Vives-Usano, M.; Waiblinger, D.; Warembourg, C.; Chatzi, L.; Coen, M.; Van Den Hazel, P.; Nieuwenhuijsen, M. J.; Slama, R.; Thomsen, C.; Wright, J.; Vrijheid, M. Human Early Life Exposome (HELIX) Study: A European Population-Based Exposome Cohort. *BMJ Open* **2018**, *8* (9), 1–17.

(61) Yuxia Cui, 1 David M. Balshaw, 1 Richard K. Kwok, 2 Claudia L. Thompson, 3 Gwen W. Collman, 4 and Linda S. Birnbaum5. Perspectives | Brief Communication The Exposome : Embracing the Complexity for Discovery in Environmental Health. *Environ. Health Perspect.* **2007**, 137–140.

(62) Plebani, M. Quality Indicators to Detect Pre-Analytical Errors in Laboratory Testing. *Clin. Biochem. Rev.* **2012**, *33* (3), 85–88.

(63) Athersuch, T. Metabolome Analyses in Exposome Studies: Profiling Methods for a Vast Chemical Space. *Archives of Biochemistry and Biophysics*. **2016**, 589, 177–186.

(64) Smith, L.; Villaret-Cazadamont, J.; Claus, S. P.; Canlet, C.; Guillou, H.; Cabaton, N. J.; Ellero-Simatos, S. Important Considerations for Sample Collection in Metabolomics Studies with a Special Focus on Applications to Liver Functions. *Metabolites*. **2020**. *10*(3),104.

(65) Orešič, M.; McGlinchey, A.; Wheelock, C. E.; Hyötyläinen, T. Metabolic Signatures of the Exposome—Quantifying the Impact of Exposure to Environmental Chemicals on Human Health. *Metabolites* **2020**, *10* (11), 1–31.

(66) Sillé, F. C. M.; Karakitsios, S.; Kleensang, A.; Koehler, K.; Maertens, A.; Miller, G. W.;

Prasse, C.; Quiros-Alcala, L.; Ramachandran, G.; Rappaport, S. M.; Rule, A. M.; Sarigiannis, D.; Smirnova, L.; Hartung, T. The Exposome - A New Approach for Risk Assessment. *ALTEX* **2020**, *37* (1), 3–23.

(67) Maria Vitale, C.; Price, E. J.; Miller, G. W.; David, A.; Antignac, J.-P.; Barouki, R.; Klá nová, J. Analytical Strategies for Chemical Exposomics: Exploring Limits and Feasibility. *Exposome* **2021**, *1* (1) osab003.

(68) Teahan, O.; Gamble, S.; Holmes, E.; Waxman, J.; Nicholson, J. K.; Bevan, C.; Keun, H. C. Impact of Analytical Bias in Metabonomic Studies of Human Blood Serum and Plasma. *Anal. Chem.* **2006**, *78* (13), 4307–4318.

(69) Athersuch, T. J.; Keun, H. C. Metabolic Profiling in Human Exposome Studies. *Mutagenesis*. Oxford University Press **2015**, *30*(6), 755–762.

(70) Hebels, D. G. A. J.; Georgiadis, P.; Keun, H. C.; Athersuch, T. J.; Vineis, P.; Vermeulen, R.; Portengen, Ü.; Bergdahl, I. A.; Hallmans, G.; Palli, D.; Bendinelli, B.; Krogh, V.; Tumino, R.; Sacerdote, C.; Panico, S.; Kleinjans, J. C. S.; de Kok, T. M. C. M.; Smith, M. T.; Kyrtopoulos, S. A. Performance in Omics Analyses of Blood Samples in Long-Term Storage: Opportunities for the Exploitation of Existing Biobanks in Environmental Health Research. *Environ. Health Perspect.* **2013**, *121* (4), 480–487.

(71) Shen, H.; Xu, W.; Peng, S.; Scherb, H.; She, J.; Voigt, K.; Alamdar, A.; Schramm, K. W. Pooling Samples for "Top-down" Molecular Exposomics Research: The Methodology. *Environ. Heal. A Glob. Access Sci. Source* **2014**, *13* (1),1-8.

(72) Walker, D. I.; Valvi, D.; Rothman, N.; Lan, Q.; Miller, G. W.; Jones, D. P. The Metabolome: A Key Measure for Exposome Research in Epidemiology. *Curr. Epidemiol. Reports* **2019**, *6* (2), 93–103.

(73) Olesti, E.; González-Ruiz, V.; Wilks, M. F.; Boccard, J.; Rudaz, S. Approaches in Metabolomics for Regulatory Toxicology Applications. *Analyst* **2021**, *146* (6), 1820–1834.

(74) David, A.; Chaker, J.; Price, E. J.; Bessonneau, V.; Chetwynd, A. J.; Vitale, C. M.; Klánová, J.; Walker, D. I.; Antignac, J. P.; Barouki, R.; Miller, G. W. Towards a Comprehensive Characterisation of the Human Internal Chemical Exposome: Challenges and Perspectives. *Environ. Int.* **2021**, *156*, 106630.

(75) Bloszies, C. S.; Fiehn, O. Using Untargeted Metabolomics for Detecting Exposome Compounds; **2018**; 8, 87–92.

(76) Ljoncheva, M.; Stepišnik, T.; Džeroski, S.; Kosjek, T. Cheminformatics in MS-Based Environmental Exposomics: Current Achievements and Future Directions. *Trends in Environmental Analytical Chemistry*., **2020**, 28.

(77) Pourchet, M.; Debrauwer, L.; Klanova, J.; Price, E. J.; Covaci, A.; Caballero-Casero, N.; Oberacher, H.; Lamoree, M.; Damont, A.; Fenaille, F.; Vlaanderen, J.; Meijer, J.; Krauss, M.; Sarigiannis, D.; Barouki, R.; Le Bizec, B.; Antignac, J. P. Suspect and Non-Targeted Screening of Chemicals of Emerging Concern for Human Biomonitoring, Environmental Health Studies and Support to Risk Assessment: From Promises to Challenges and Harmonisation Issues. *Environ. Int.* **2020**, *139*, 105545.

(78) Sun, J.; Fang, R.; Wang, H.; Xu, D. X.; Yang, J.; Huang, X.; Cozzolino, D.; Fang, M.; Huang, Y. A Review of Environmental Metabolism Disrupting Chemicals and Effect Biomarkers Associating Disease Risks: Where Exposomics Meets Metabolomics. *Environment International*. **2022**,158.

(79) Chen, L.; Zhong, F.; Zhu, J. Bridging Targeted and Untargeted Mass Spectrometry-Based Metabolomics via Hybrid Approaches. *Metabolites* (**2020**) 10.9, 348

(80) Misra, B. B. Metabolomics Tools to Study Links Between Pollution and Human Health: An Exposomics Perspective. *Current Pollution Reports*. Springer September 15, **2019**, pp 93–111.

(81) Xue, J.; Lai, Y.; Liu, C. W.; Ru, H. Towards Mass Spectrometry-Based Chemical Exposome: Current Approaches, Challenges, and Future Directions; *Toxics*. **2019**; *7*(3),41.

(82) Zheng, X.; Wojcik, R.; Zhang, X.; Ibrahim, Y. M.; Burnum-Johnson, K. E.; Orton, D. J.; Monroe, M. E.; Moore, R. J.; Smith, R. D.; Baker, E. S. Coupling Front-End Separations, Ion Mobility Spectrometry, and Mass Spectrometry for Enhanced Multidimensional Biological and Environmental Analyses. *Annual Review of Analytical Chemistry*. **2017**, 10(1), 71–92.

(83) Fernández Maestre, R. Ion Mobility Spectrometry: History, Characteristics and Applications. *Rev. U.D.C.A Actual. Divulg. Científica.* **2012**, *15* (2).467-479.

(84) Adetona, O.; Simpson, C. D.; Li, Z.; Sjodin, A.; Calafat, A. M.; Naeher, L. P. Hydroxylated Polycyclic Aromatic Hydrocarbons as Biomarkers of Exposure to Wood Smoke in Wildland Firefighters. *J. Expo. Sci. Environ. Epidemiol.* **2017**, *27* (1), 78–83.

(85) Gill, B.; Mell, A.; Shanmuganathan, M.; Jobst, K.; Zhang, X.; Kinniburgh, D.; Cherry, N.; Britz-McKibbin, P. Urinary Hydroxypyrene Determination for Biomonitoring of Firefighters Deployed at the Fort McMurray Wildfire: An Inter-Laboratory Method Comparison. *Anal. Bioanal. Chem.* **2019**, *411* (7), 1397–1407.

(86) Nováková, L.; Vlčková, H. A Review of Current Trends and Advances in Modern Bio-Analytical Methods: Chromatography and Sample Preparation. *Anal. Chim. Acta* **2009**, *656* (1–2), 8–35.

(87) Gys, C.; Kovačič, A.; Huber, C.; Yin Lai, F.; Heath, E.; Covaci, A. Suspect and Untargeted Screening of Bisphenol S Metabolites Produced by in Vitro Human Liver Metabolism. **2018,** 295,115-123.

(88) Taghavi, T.; Novalen, M.; Lerman, C.; George, T. P.; Tyndale, R. F. A Comparison of Direct and Indirect Analytical Approaches to Measuring Total Nicotine Equivalents in Urine. *Cancer Epidemiol. Biomarkers Prev.* **2018**, *27* (8), 882–891.

(89) Dwivedi, P.; Zhou, X.; Powell, T. G.; Calafat, A. M.; Ye, X. Impact of Enzymatic Hydrolysis on the Quantification of Total Urinary Concentrations of Chemical Biomarkers. *Chemosphere* **2018**, 199, 256-262.

(90) Gill, B.; Britz-McKibbin, P. Biomonitoring of Smoke Exposure in Firefighters: A Review. *Curr. Opin. Environ. Sci. Heal.* **2020**, 15, 57-65.

(91) Xia, B.; Xia, Y.; Wong, J.; Nicodemus, K. J.; Xu, M.; Lee, J.; Guillot, T.; Li, J. Quantitative Analysis of Five Tobacco-Specific N-Nitrosamines in Urine by Liquid

Chromatography–Atmospheric Pressure Ionization Tandem Mass Spectrometry. *Biomed. Chromatogr.* **2014**, *28* (3), 375–384.

(92) Lai, F. Y.; Been, F.; Covaci, A.; Van Nuijs, A. L. N. Novel Wastewater-Based Epidemiology Approach Based on Liquid Chromatography-Tandem Mass Spectrometry for Assessing Population Exposure to Tobacco-Specific Toxicants and Carcinogens. *Anal. Chem.* **2017**, *89* (17), 9268–9278.

(93) Monteiro Bastos da Silva, J.; Chaker, J.; Martail, A.; Costa Moreira, J.; David, A.; Le Bot, B. Improving Exposure Assessment Using Non-Targeted and Suspect Screening: The ISO/IEC 17025: 2017 Quality Standard as a Guideline. *J. Xenobiotics* **2021**, *11* (1), 1–15.

(94) Godzien, J.; Alonso-Herranz, V.; Coral, B.; Armitage, G. Controlling the Quality of Metabolomics Data: New Strategies to Get the Best out of the QC Sample. *Metabolomics.* **2015**, 11(3), 518-528.

(95) Shanmuganathan, M.; Kroezen, Z.; Gill, B.; Azab, S.; de Souza, R. J.; Teo, K. K.; Atkinson, S.; Subbarao, P.; Desai, D.; Anand, S. S.; Britz-McKibbin, P. The Maternal Serum Metabolome by Multisegment Injection-Capillary Electrophoresis-Mass Spectrometry: A High-Throughput Platform and Standardized Data Workflow for Large-Scale Epidemiological Studies. *Nat. Protoc.* **2021**, *16* (4), 1966–1994.

(96) Cai, X.; Li, R. Concurrent Profiling of Polar Metabolites and Lipids in Human Plasma Using HILIC-FTMS. *Sci. Reports 2016 61* **2016**, *6* (1), 1–10.

(97) Yamamoto, M.; Pinto-Sanchez, M. I.; Bercik, P.; Britz-McKibbin, P. Metabolomics Reveals Elevated Urinary Excretion of Collagen Degradation and Epithelial Cell Turnover Products in Irritable Bowel Syndrome Patients. *Metabolomics* **2019**, *15* (6),1-18.

(98) Dibattista, A.; Rampersaud, D.; Lee, H.; Kim, M.; Britz-Mckibbin, P. High Throughput Screening Method for Systematic Surveillance of Drugs of Abuse by Multisegment Injection-Capillary Electrophoresis-Mass Spectrometry. *Anal. Chem.* **2017**, *89* (21), 11853–11861.

(99) Gill, B.; Jobst, K.; Britz-Mckibbin, P. Rapid Screening of Urinary 1-Hydroxypyrene Glucuronide by Multisegment Injection-Capillary Electrophoresis-Tandem Mass Spectrometry: A High-Throughput Method for Biomonitoring of Recent Smoke Exposures. *Anal. Chem.* **2020**, *92* (19), 13558–13564.

(100) Azab, S.; Ly, R.; Britz-Mckibbin, P. Robust Method for High-Throughput Screening of Fatty Acids by Multisegment Injection-Nonaqueous Capillary Electrophoresis-Mass Spectrometry with Stringent Quality Control. *Anal. Chem.* **2019**, *91* (3), 2329–2336.

(101) Azab, S. M.; De Souza, R. J.; Teo, K. K.; Anand, S. S.; Williams, N. C.; Holzschuher, J.; McGlory, C.; Philips, S. M.; Britz-McKibbin, P. Serum Non-Esterified Fatty Acids Have Utility as Dietary Biomarkers of Fat Intake from Fish, Fish Oil, and Dairy in Women. *J. Lipid Res.* **2020**, *61* (6), 933–944.

(102) Saoi, M.; Kennedy, K. M.; Gohir, W.; Sloboda, D. M.; Britz-McKibbin, P. Placental Metabolomics for Assessment of Sex-Specific Differences in Fetal Development During Normal Gestation. **2020**, *10* (1), 1–10.

(103) Dibattista, A.; McIntosh, N.; Lamoureux, M.; Al-Dirbashi, O. Y.; Chakraborty, P.; Britz-

Mckibbin, P. Metabolic Signatures of Cystic Fibrosis Identified in Dried Blood Spots for Newborn Screening Without Carrier Identification. *J. Proteome Res.* **2019**, *18* (3), 841–854.

(104) Macedo, A. N.; Mathiaparanam, S.; Brick, L.; Keenan, K.; Gonska, T.; Pedder, L.; Hill, S.; Britz-Mckibbin, P. The Sweat Metabolome of Screen-Positive Cystic Fibrosis Infants: Revealing Mechanisms beyond Impaired Chloride Transport. **2017**, *3*(8), 904-913.

(105) Harada, S.; Hirayama, A.; Chan, Q.; Kurihara, A.; Fukai, K.; Iida, M.; Kato, S.; Sugiyama, D.; Kuwabara, K.; Takeuchi, A.; Akiyama, M.; Okamura, T.; Ebbels, T. M. D.; Elliott, P.; Tomita, M.; Sato, A.; Suzuki, C.; Sugimoto, M.; Soga, T.; Takebayashi, T. Reliability of Plasma Polar Metabolite Concentrations in a Large-Scale Cohort Study Using Capillary Electrophoresis-Mass Spectrometry. *PLoS One* **2018**, *13* (1).e0191230.

(106) Ishibashi, Y.; Harada, S.; Takeuchi, A.; Iida, M.; Kurihara, A.; Kato, S.; Kuwabara, K.; Hirata, A.; Shibuki, T.; Okamura, T.; Sugiyama, D.; Sato, A.; Amano, K.; Hirayama, A.; Sugimoto, M.; Soga, T.; Tomita, M.; Takebayashi, T. Reliability of Urinary Charged Metabolite Concentrations in a Large-Scale Cohort Study Using Capillary Electrophoresis-Mass Spectrometry. *Sci. Reports.* **2021**, *11*(1), 1-9.

(107) Chiari, M.; Ceriotti, L. Capillary Electrophoresis Capillary Electrophoresis. *Food Toxicants Anal. Tech. Strateg. Dev.* **2014**, *1997* (2), 561–597.

(108) Kuehnbaum, N. L.; Kormendi, A.; Britz-Mckibbin, P. Multisegment Injection-Capillary Electrophoresis-Mass Spectrometry: A High-Throughput Platform for Metabolomics with High Data Fidelity. *Anal. Chem.* **2013**, *85* (22), 10664–10669.

(109) Saoi, M.; Li, A.; McGlory, C.; Stokes, T.; von Allmen, M. T.; Phillips, S. M.; Britz-Mckibbin, P. Metabolic Perturbations from Step Reduction in Older Persons at Risk for Sarcopenia: Plasma Biomarkers of Abrupt Changes in Physical Activity. *Metabolites* **2019**, *9* (7), 134.

(110) Ramautar, R.; Somsen, G. W.; de Jong, G. J. CE-MS for Metabolomics: Developments and Applications in the Period 2012–2014. *Electrophoresis* **2015**, *36* (1), 212–224.

(111) Wellington, N.; Shanmuganathan, M.; De Souza, R. J.; Zulyniak, M. A.; Azab, S.; Bloomfield, J.; Mell, A.; Ly, R.; Desai, D.; Anand, S. S.; Britz-McKibbin, P. Metabolic Trajectories Following Contrasting Prudent and Western Diets from Food Provisions: Identifying Robust Biomarkers of Short-Term Changes in Habitual Diet. *Nutrients* **2019**, *11* (10), 2407.

(112) Yi, L.; Dong, N.; Yun, Y.; Deng, B.; Ren, D.; Liu, S.; Liang, Y. Chemometric Methods in Data Processing of Mass Spectrometry-Based Metabolomics: A Review. *Anal. Chim. Acta* **2016**, *914*, 17–34.

(113) Blaženović, I.; Kind, T.; Sa, M. R.; Ji, J.; Vaniya, A.; Wancewicz, B.; Roberts, B. S.; Torbašinović, H.; Lee, T.; Mehta, S. S.; Showalter, M. R.; Song, H.; Kwok, J.; Jahn, D.; Kim, J.; Fiehn, O. Structure Annotation of All Mass Spectra in Untargeted Metabolomics. *Anal. Chem.* **2019**, *91* (3), 2155–2162.

(114) Espín-Pérez, A.; Portier, C.; Chadeau-Hyam, M.; van Veldhoven, K.; Kleinjans, J. C. S.; de Kok, T. M. C. M. Comparison of Statistical Methods and the Use of Quality Control Samples for Batch Effect Correction in Human Transcriptome Data. *PLoS One* **2018**, *13*

(8), e0202947.

(115) Han, W.; Li, L. Evaluating and Minimizing Batch Effects in Metabolomics. *Mass Spectrom. Rev.* **2022**, *41* (3), 421–442.

(116) Kirwan, J. A.; Broadhurst, D. I.; Davidson, R. L.; Viant, M. R. Characterising and Correcting Batch Variation in an Automated Direct Infusion Mass Spectrometry (DIMS) Metabolomics Workflow. *Anal. Bioanal. Chem.* **2013**, *405* (15), 5147–5157.

(117) Johnson, W. E.; Li, C.; Rabinovic, A. Adjusting Batch Effects in Microarray Expression Data Using Empirical Bayes Methods. *Biostatistics* **2007**, *8* (1), 118–127.

(118) Armitage, E. G.; Godzien, J.; Alonso-Herranz, V.; López-Gonzálvez, Á.; Barbas, C. Missing Value Imputation Strategies for Metabolomics Data. *Electrophoresis* **2015**, *36* (24), 3050–3060.

(119) Do, K. T.; Wahl, S.; Raffler, J.; Molnos, S.; Laimighofer, M.; Adamski, J.; Suhre, K.; Strauch, K.; Peters, A.; Gieger, C.; Langenberg, C.; Stewart, I. D.; Theis, F. J.; Grallert, H.; Kastenmüller, G.; Krumsiek, J. Characterization of Missing Values in Untargeted MS-Based Metabolomics Data and Evaluation of Missing Data Handling Strategies. *Metabolomics* **2018**, *14* (10), 1–18.

(120) Wei, R.; Wang, J.; Su, M.; Jia, E.; Chen, S.; Chen, T.; Ni, Y. Missing Value Imputation Approach for Mass Spectrometry-Based Metabolomics Data. *Sci. Reports 2018 81* **2018**, *8* (1), 1–10.

(121) Vollmar, A. K. R.; Rattray, N. J. W.; Cai, Y.; Santos-Neto, Á. J.; Deziel, N. C.; Jukic, A. M. Z.; Johnson, C. H. Normalizing Untargeted Periconceptional Urinary Metabolomics Data: A Comparison of Approaches. *Metabolites* **2019**, *9* (10),198.

(122) Karaman, I. Preprocessing and Pretreatment of Metabolomics Data for Statistical Analysis. *Adv. Exp. Med. Biol.* **2017**, *965*, 145–161.

(123) van den Berg, R. A.; Hoefsloot, H. C. J.; Westerhuis, J. A.; Smilde, A. K.; van der Werf, M. J. Centering, Scaling, and Transformations: Improving the Biological Information Content of Metabolomics Data. *BMC Genomics* **2006**, *7* (1), 1–15.

(124) Hu, X.; Walker, D. I.; Liang, Y.; Smith, M. R.; Orr, M. L.; Juran, B. D.; Ma, C.; Uppal, K.; Koval, M.; Martin, G. S.; Neujahr, D. C.; Marsit, C. J.; Go, Y. M.; Pennell, K. D.; Miller, G. W.; Lazaridis, K. N.; Jones, D. P. A Scalable Workflow to Characterize the Human Exposome. *Nat. Commun. 2021 121* **2021**, *12* (1), 1–12.

(125) Cai, Y.; Rosen Vollmar, A. K.; Johnson, C. H. Analyzing Metabolomics Data for Environmental Health and Exposome Research. *Methods Mol. Biol.* **2020**, *2104*, 447–467.

(126) Jain, P.; Vineis, P.; Liquet, B.; Vlaanderen, J.; Bodinier, B.; Van Veldhoven, K.; Kogevinas, M.; Athersuch, T. J.; Font-Ribera, L.; Villanueva, C. M.; Vermeulen, R.; Chadeau-Hyam, M. A Multivariate Approach to Investigate the Combined Biological Effects of Multiple Exposures. *J. Epidemiol. Community Health* **2018**, *72* (7), 564–571.

(127) Martino, D.; Ben-Othman, R.; Harbeson, D.; Bosco, A. Multiomics and Systems Biology Are Needed to Unravel the Complex Origins of Chronic Disease. *Challenges* **2019**, *10* (1), 23.

(128) Van Der Hooft, J. J. J.; Wandy, J.; Young, F.; Padmanabhan, S.; Gerasimidis, K.; Burgess, K. E. V.; Barrett, M. P.; Rogers, S. Unsupervised Discovery and Comparison of Structural Families Across Multiple Samples in Untargeted Metabolomics. *Anal. Chem.* **2017**, *89* (14), 7569–7577.

(129) Mortimer, M., Fang, W., Zhou, X., Vodovnik, M., & Guo, L. H. (2022). Omics Approaches in Toxicological Studies. In Advances in Toxicology and Risk Assessment of Nanomaterials and Emerging Contaminants. *Springer, Singapore*. **2022**, 61-94.

(130) 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.; Dunn, W. B.; Brown, Á. M.; Erban, A.; Kopka, Á. J.; Weber, R. J. M.; Viant, Á. M. R.; Creek, D. J.; Breitling, R.; Hankemeier, T.; Goodacre, R. Mass Appeal: Metabolite Identification in Mass Spectrometry-Focused Untargeted Metabolomics. *Metabolomics* **2013**, *9(1)*, 44–66.

(131) D'Agostino, L. A.; Lam, K. P.; Lee, R.; Britz-McKibbin, P. Comprehensive Plasma Thiol Redox Status Determination for Metabolomics. *J. Proteome Res.* **2011**, *10* (2), 592–603.

(132) Selection, P. Risk Assessment Risk Assessment Risk Assessment. *Risk Manag.* **2008**, *24* (4), 1–7.

(133) Hallenbeck, W. H. Quantitative Risk Assessment for Environmental and Occupational Health; *CRC Press*., **1993**.

(134) Brunekreef, B. Environmental Epidemiology and Risk Assessment. *Toxicol. Lett.* **2008**, *180* (2), 118–122.

(135) Pinto, A., Nunes, I. L., & Ribeiro, R. A. Occupational risk assessment in construction industry–Overview and reflection. *Safety science.*, **2011**, 49(5), 616-624.

(136) Laitinen, J.; Mäkelä, M.; Mikkola, J.; Huttu, I. Firefighters' Multiple Exposure Assessments in Practice. *Toxicol. Lett.* **2012**, *213* (1), 129–133.

(137) Guidotti, T. I. Evaluating Causality for Occupational Cancers: The Example of Firefighters. *Occup. Med.* **2007**, 57.7, 466-471.

(138) Manisalidis, I.; Stavropoulou, E.; Stavropoulos, A.; Bezirtzoglou, E. Environmental and Health Impacts of Air Pollution: A Review. *Front. Public Heal.* **2020**, *8(1)*, 14.

(139) McCarthy, J. E., Copeland, C., Parker, L., & Schierow, L. J., Clean Air Act: A summary of the act and its major requirements. Congressional Research Service, Library of Congress. **2007.**

(140) Williams, P. R., Dotson, G. S., & Maier, A. Risk Assessment's New Era: part 2: Evolving Methods and Future Directions. *Synergist*, **2012**; *25*(5), 46.

(141) Barr, D. B.; Wang, R. Y.; Needham, L. L. Biologic Monitoring of Exposure to Environmental Chemicals throughout the Life Stages: Requirements and Issues for Consideration for the National Children's Study. *Environ. Health Perspect.* **2005**, *113*(8),1083-1091.

(142) Luo, K.; Luo, X.; Cao, W.; Hochalter, J. B.; Paiano, V.; Sipe, C. J.; Carmella, S. G.; Murphy, S. E.; Jensen, J.; Lam, S.; Golin, A. P.; Bergstrom, L.; Midthun, D.; Fujioka, N.; Hatsukami, D.; Hecht, S. S. Cigarette Smoking Enhances the Metabolic Activation of the

Polycyclic Aromatic Hydrocarbon Phenanthrene in Humans. *Carcinogenesis* **2021**, *42* (4), 570–577.

(143) Chang, C. M., Edwards, S. H., Arab, A., Del Valle-Pinero, A. Y., Yang, L., & Hatsukami, D. K. Biomarkers of tobacco exposure: summary of an FDA-sponsored public workshop. *Cancer Epidemiology and Prevention Biomarkers*. **2017**, 26(3), 291-302.

(144) Yuan, J. M.; Gao, Y. T.; Murphy, S. E.; Carmella, S. G.; Wang, R.; Zhong, Y.; Moy, K. A.; Davis, A. B.; Tao, L.; Chen, M.; Han, S.; Nelson, H. H.; Yu, M. C.; Hecht, S. S. Urinary Levels of Cigarette Smoke Constituent Metabolites Are Prospectively Associated with Lung Cancer Development in Smokers. *Cancer Res.* **2011**, *71* (21), 6749–6757.

(145) Benowitz, N. L.; Bernert, J. T.; Caraballo, R. S.; Holiday, D. B.; Wang, J. Optimal Serum Cotinine Levels for Distinguishing Cigarette Smokers and Nonsmokers Within Different Racial/Ethnic Groups in the United States Between 1999 and 2004. *Am. J. Epidemiol.* **2009**, *169* (2), 236–248.

(146) Mazumder, S.; Shia, W.; Bendik, P. B.; Achilihu, H.; Sosnoff, C. S.; Alexander, J. R.; Luo, Z.; Zhu, W.; Pine, B. N.; Feng, J.; Blount, B. C.; Wang, L. Nicotine Exposure in the U.S. Population: Total Urinary Nicotine Biomarkers in NHANES 2015–2016. *Int. J. Environ. Res. Public Health.* **2022**, *19* (6), 3660.

(147) Allenby, C. E.; Boylan, K. A.; Lerman, C.; Falcone, M. Precision Medicine for Tobacco Dependence: Development and Validation of the Nicotine Metabolite Ratio. *Journal of Neuroimmune Pharmacology*. Springer New York LLC September 1, **2016**, 11(3), 471–483.

(148) Maruvada, P.; Lampe, J. W.; Wishart, D. S.; Barupal, D.; Chester, D. N.; Dodd, D.; Djoumbou-Feunang, Y.; Dorrestein, P. C.; Dragsted, L. O.; Draper, J.; Duffy, L. C.; Dwyer, J. T.; Emenaker, N. J.; Fiehn, O.; Gerszten, R. E.; Hu, F. B.; Karp, R. W.; Klurfeld, D. M.; Laughlin, M. R.; Little, A. R.; Lynch, C. J.; Moore, S. C.; Nicastro, H. L.; O'brien, D. M.; Ordovás, J. M.; Osganian, S. K.; Playdon, M.; Prentice, R.; Raftery, D.; Reisdorph, N.; Roche, H. M.; Ross, S. A.; Sang, S.; Scalbert, A.; Srinivas, P. R.; Zeisel, S. H. Perspective: Dietary Biomarkers of Intake and Exposure-Exploration with Omics Approaches. *Adv Nutr* **2019**, *11(2)*, 200-215.

(149) Poole, R.; Kennedy, O. J.; Roderick, P.; Fallowfield, J. A.; Hayes, P. C.; Parkes, J. Coffee Consumption and Health: Umbrella Review of Meta-Analyses of Multiple Health Outcomes. *BMJ.* **2017**, *359*, 5024.

(150) Narula, N.; Wong, E. C. L.; Dehghan, M.; Mente, A.; Rangarajan, S.; Lanas, F.; Lopez-Jaramillo, P.; Rohatgi, P.; Lakshmi, M.; Varma, R. P.; Orlandini, A.; Avezum, A.; Wielgosz, A.; Poirier, P.; Almadi, M. A.; Altuntas, Y.; Ng, K. K.; Chifamba, J.; Yeates, K.; Puoane, T.; Khatib, R.; Yusuf, R.; Boström, K. B.; Zatonska, K.; Iqbal, R.; Weida, L.; Yibing, Z.; Sidong, L.; Dans, A.; Yusufali, A.; Mohammadifard, N.; Marshall, J. K.; Moayyedi, P.; Reinisch, W.; Yusuf, S. Association of Ultra-Processed Food Intake with Risk of Inflammatory Bowel Disease: Prospective Cohort Study. *BMJ* **2021**, *374*, 1554.

(151) Hodges, R. E.; Minich, D. M. Modulation of Metabolic Detoxification Pathways Using Foods and Food-Derived Components: A Scientific Review with Clinical Application. *J. Nutr. Metab.* **2015**, *2015*.

(152) Turner, C.; Kalamatianou, S.; Drewnowski, A.; Kulkarni, B.; Kinra, S.; Kadiyala, S. Food Environment Research in Low- and Middle-Income Countries: A Systematic Scoping Review. *Adv. Nutr.* **2020**, *11* (2), 387–397.

(153) Smith, M. T.; de la Rosa, R.; Daniels, S. I. Using Exposomics to Assess Cumulative Risks and Promote Health; *Environmental and molecular mutagenesis*, **2015**, 56(9), 715–723.

(154) van Roekel, E. H.; Loftfield, E.; Kelly, R. S.; Zeleznik, O. A.; Zanetti, K. A. Metabolomics in Epidemiologic Research: Challenges and Opportunities for Early-Career Epidemiologists. *Metabolomics* **2019**, *15* (1), 1–7.

(155) Gao, P.; da Silva, E.; Hou, L.; Denslow, N. D.; Xiang, P.; Ma, L. Q. Human Exposure to Polycyclic Aromatic Hydrocarbons: Metabolomics Perspective. *Environment International*. **2018**, *119*, 466-477.

(156) Wingfors, H.; Nyholm, J. R.; Magnusson, R.; Wijkmark, C. H. Impact of Fire Suit Ensembles on Firefighter PAH Exposures as Assessed by Skin Deposition and Urinary Biomarkers. *Ann. Work Expo. Heal.* **2018**, *62* (2), 221–231.

(157) Cherry, N.; Aklilu, Y. A.; Beach, J.; Britz-Mckibbin, P.; Elbourne, R.; Galarneau, J. M.; Gill, B.; Kinniburgh, D.; Zhang, X. Urinary 1-Hydroxypyrene and Skin Contamination in Firefighters Deployed to the Fort McMurray Fire. *Ann. Work Expo. Heal.* **2019**, 63(4), 448-458.

(158) Perez-Paramo, Y. X.; Lazarus, P. Pharmacogenetics Factors Influencing Smoking Cessation Success; the Importance of Nicotine Metabolism. *Expert Opinion on Drug Metabolism and Toxicology*. **2021**, 17(3), 333–349.

(159) Sathish, T.; Teo, K. K.; Britz-McKibbin, P.; Gill, B.; Islam, S.; Paré, G.; Rangarajan, S.; Duong, M. L.; Lanas, F.; Lopez-Jaramillo, P.; Mony, P. K.; Pinnaka, L.; Kutty, V. R.; Orlandini, A.; Avezum, A.; Wielgosz, A.; Poirier, P.; Alhabib, K. F.; Temizhan, A.; Chifamba, J.; Yeates, K.; Kruger, I. M.; Khatib, R.; Yusuf, R.; Rosengren, A.; Zatonska, K.; Iqbal, R.; Lui, W.; Lang, X.; Li, S.; Hu, B.; Dans, A. L.; Yusufali, A. H.; Bahonar, A.; O'Donnell, M. J.; McKee, M.; Yusuf, S. Variations in Risks from Smoking between High-Income, Middle-Income, and Low-Income Countries: An Analysis of Data from 179 000 Participants from 63 Countries. *Lancet Glob. Heal.* **2022**, *10* (2), e216–e226.

(160) Jena, P. K.; Kishore, J.; Jahnavi, G. Correlates of Digit Bias in Self-Reporting of Cigarette per Day (CPD) Frequency: Results from Global Adult Tobacco Survey (GATS), India and Its Implications. *Asian Pacific J. cancer Prev.* **2013**, *14* (6), 3865–3869.

(161) Schick, S. F.; Blount, B. C.; Jacob, P.; Saliba, N. A.; Bernert, J. T.; El Hellani, A.; Jatlow, P.; Steven Pappas, R.; Wang, L.; Foulds, J.; Ghosh, A.; Hecht, S. S.; Gomez, J. C.; Martin, J. R.; Mesaros, C.; Srivastava, S.; St Helen, G.; Tarran, R.; Lorkiewicz, P. K.; Blair, I. A.; Kimmel, H. L.; Doerschuk, C. M.; Benowitz, N. L.; Bhatnagar, A.; Hellani, E. A.; Helen, S. G.; Schick, S. F. Biomarkers of Exposure to New and Emerging Tobacco Delivery Products. *Am J Physiol Lung Cell Mol Physiol.* **2017**, *313*, 425–452.

(162) Chen, A.; Krebs, N. M.; Zhu, J.; Muscat, J. E. Nicotine Metabolite Ratio Predicts Smoking Topography: The Pennsylvania Adult Smoking Study. *Drug Alcohol Depend.* **2018**, *190*, 89–93.

(163) Ng, R.; Sutradhar, R.; Yao, Z.; Wodchis, W. P.; Rosella, L. C. Smoking, Drinking, Diet

and Physical Activity—Modifiable Lifestyle Risk Factors and Their Associations with Age to First Chronic Disease. *Int. J. Epidemiol.* **2020**, *49* (1), 113–130.

(164) Guillermo, C.; Boushey, C. J.; Franke, A. A.; Monroe, K. R.; Lim, U.; Wilkens, L. R.; Le Marchand, L.; Maskarinec, G. Diet Quality and Biomarker Profiles Related to Chronic Disease Prevention: The Multiethnic Cohort Study. *J. Am. Coll. Nutr.* **2020**, *39* (3), 216–223.

(165) Rafiq, T.; Azab, S. M.; Teo, K. K.; Thabane, L.; Anand, S. S.; Morrison, K. M.; De Souza, R. J.; Britz-Mckibbin, P. Nutritional Metabolomics and the Classification of Dietary Biomarker Candidates: A Critical Review. *Adv. Nutr.* **2021**, *12* (6), 2333–2357.

(166) Onvani, S.; Haghighatdoost, F.; Surkan, P. J.; Larijani, B.; Azadbakht, L. Adherence to the Healthy Eating Index and Alternative Healthy Eating Index Dietary Patterns and Mortality from All Causes, Cardiovascular Disease and Cancer: A Meta-Analysis of Observational Studies. *J. Hum. Nutr. Diet.* **2017**, *30* (2), 216–226.

(167) Archer, E.; Marlow, M. L.; Lavie, C. J. Controversy and Debate: Memory-Based Methods Paper 1: The Fatal Flaws of Food Frequency Questionnaires and Other Memory-Based Dietary Assessment Methods. *J. Clin. Epidemiol.* **2018**, *104*, 113–124.

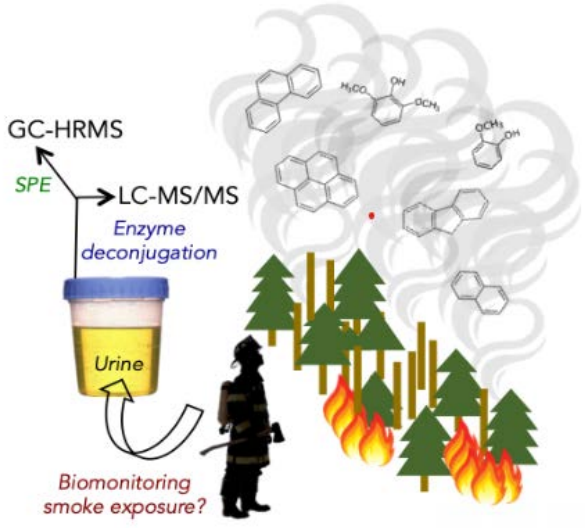
Chapter II:

An Inter-laboratory Method Comparison of Urinary 1-Hydroxypyrene Determination for Biomonitoring of Firefighters Deployed at the Fort McMurray Wildfire

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

Biban Gill,[1] Alicia Mell,[1] Meera Shanmuganathan,[1] Karl Jobst,[2] Xu Zhang,[3] David Kinniburgh,[3] Nicola Cherry,[4] Philip Britz-McKibbin.[1] Urinary hydroxypyrene determination for biomonitoring of firefighters deployed at the Fort McMurray wildfire: An inter-laboratory method comparison. *Anal. Bioanal. Chem.* **2019** 411(7), 1397-1407.

[1]*Department of Chemistry and Chemical Biology, McMaster University, Hamilton, Ontario, L8S 4M1, Canada*
[2]*Laboratory Branch, Ontario Ministry of the Environment, Conservation and Parks, Toronto, Ontario, M9P 3V6, Canada*
[3]*Alberta Centre for Toxicology, University of Calgary, Calgary, Alberta, T2N 1N4, Canada*
[4]*Division of Preventative Medicine, University of Alberta, Edmonton, Alberta, T6G 2R3, Canada*



B.G. conducted the enzyme hydrolysis time-point studies, as well as the statistical analysis and data interpretation. K.J, A.M., and M.S., were involved optimization, sample analysis and data processing for the GC-HRMS data included in the study. D.K. and X.Z. conducted experiments, validation and sample analysis by LC-MS/MS for the inter-laboratory comparison. B.G wrote the initial draft for publication and P.B.M provided critical feedback and revisions of the final manuscript for publication.

Chapter II: An Inter-laboratory Method Comparison of Urinary 1-Hydroxypyrene Determination for Biomonitoring of Firefighters Deployed at the Fort McMurray Wildfire

2.1 Abstract

Urinary 1-hydroxypyrene (1-OH-Pyr) is a widely used for biomonitoring human exposures to polycyclic aromatic hydrocarbons (PAHs) from air pollution and tobacco smoke. However, there have been few rigorous method validation studies reported to ensure reliable 1-OH-Pyr determination for risk assessment. Herein, we report an inter-laboratory method comparison for urinary 1-OH-Pyr when using gas chromatography-high resolution mass spectrometry (GC-HRMS) and liquid chromatography-tandem mass spectrometry (LC-MS/MS) on urine specimens collected from firefighters (n=42) deployed at the 2016 Fort McMurray wildfire. Overall, there was good mutual agreement in urinary OH-Pyr quantification following enzyme deconjugation with an average bias of 39% with no significant deviation from linearity (slope = 1.36; $p > 0.05$), whereas technical precision (< 12%) and average recovery (> 85%) were acceptable when using a stable-isotope internal standard. Faster analysis times (4 min) were achieved by LC-MS/MS without chemical derivatization, whereas lower detection limits (0.64 ng/L, S/N =3) was realized with solid-phase extraction prior to GC-HRMS. A median creatinine normalized 1-OH-Pyr concentration of 128 ng/g was measured for firefighters that was below the recommended biological exposure index due to extended delays between early stages of firefighting and urine sample collection. Similar outcomes were measured for 3-hydroxyphenanthrene and 9-hydroxyfluorene that were positively correlated with urinary 1-OH-Pyr ($p < 0.05$), implying similar uptake, distribution and liver biotransformation processes. Optimal specimen collection strategies post-deployment together with standardized protocols for OH-PAH analysis are critical to accurately evaluate smoke exposure in firefighters.

2.2 Introduction

The Fort McMurray wildfire made international headlines in May 2016 over a three month period as it engulfed 590,000 ha of land in the province of Alberta with over 88,000 people being forced from their homes, which resulted in the most expensive natural disaster in Canadian history [1,2]. Firefighters from a number of Canadian provinces along with personnel from the Canadian Armed Forces, and numerous rescue volunteers provided support in both fire suppression and evacuation. During the initial days of the fire, deployed firefighters experienced heavy smoke exposure during work shifts over 24 h with limited access to respiratory equipment or hygienic practices [3]. Polycyclic aromatic hydrocarbons (PAHs) are major chemical constituents of smoke resulting from incomplete combustion of organic materials [4], which have been measured in air samples, skin wipes and urine specimens collected from wildland and structural firefighters [5–7]. The International Agency for Research on Cancer has evaluated exposures in firefighting as possibly carcinogenic [8], with many jurisdictions adopting presumptive disability benefits for certain cancers in structural (urban) firefighters [9]. In this context, new strategies are needed for better risk assessment and mitigation of chemical exposures to firefighters for chronic disease prevention.

PAHs are excreted in human urine as their hydroxylated PAH (OH-PAH) metabolites due to biotransformation processes involving xenobiotic detoxification and elimination, including formation of their water-soluble *O*-glucuronide and sulfate conjugates [10]. OH-PAH metabolites derive primarily from enzyme-mediated reactions by various cytochrome P450 (CYP450) isoforms in the liver, which catalyze PAH oxidization with concomitant formation of reactive epoxide intermediates responsible for the mutagenic capacity of certain bioactivated PAH metabolites [11]. Numerous studies have investigated smoke exposure using human urine

as a convenient and non-invasive biofluid for sample collection in occupational health and population-based epidemiological studies [12–14]. In this case, 1-hydroxypyrene (1-OH-Pyr) is a widely used indicator of smoke exposure since it can be readily measured in the urine of occupationally unexposed populations. Additionally, it generates a single OH-PAH isomer with an average half-life within 24 h resulting in peak urinary excretion following exposure [4,15]. Most carcinogenic PAHs and their DNA or protein adducts are present at extremely low concentrations in urine, and are preferentially eliminated in feces, such as benzo[a]pyrene and its various hydroxylated metabolites and positional isomers [16]. Thus, measurement of 1-OH-Pyr from random/single-spot urine samples provides a convenient way for biomonitoring of recent smoke exposure that is derived from various sources, including habitual diet (charred meats), lifestyle (tobacco smoke), urban environments (vehicle exhaust) and notably specific occupational exposures (coke oven/asphalt workers) [17]. Due to the complex chemical composition of wood smoke, other classes of organic compounds may also serve as smoke exposure biomarkers for firefighters besides OH-PAH metabolites, including sterols, alcohols, alkylbenzenes, levoglucosan and methoxyphenols (MPs) [5]. Indeed, MPs are less widely studied despite being a major combustion by-product of lignin which constitutes up to 35% of wood by mass [18], which may represent more sensitive indicators of recent smoke exposure notably under suboptimal sampling and storage conditions [19].

Current analytical methods for urinary 1-OH-Pyr determination have relied on instrumental methods needed for detection of low exposure levels in the population, including liquid chromatography with fluorescence detection (LC-FD), gas chromatography-mass spectrometry (GC-MS) and LC with tandem mass spectrometry (LC-MS/MS). Direct analysis of 1-OH-Pyr and its intact glucuronide conjugate can be achieved when using immunoaffinity [20]

or reverse-phase LC-FD [21–23], whereas LC-MS/MS offers greater selectivity with fewer chemical interferences while analyzing a wider range of OH-PAHs in complex urine samples [24], including 6-hydroxynitropyrene [25]. Alternatively, GC-MS offers higher separation efficiency with better resolution of OH-PAH isomers and lower detection limits [26, 27], which is also optimal for the analysis of airborne PAHs from particular matter when using adsorbent filters/collection devices [5–7]. Various GC-MS configurations, including complementary ion sources and mass analyzers have been used for accurate quantification of parent PAH or their biotransformed OH-PAH metabolites, however most protocols require extensive sample workup, including enzyme hydrolysis, solid-phase extraction (SPE) and pre-column chemical derivatization [28]. To the best of our knowledge, no previous study has performed a cross-platform validation study to ensure reliable urinary 1-OH-Pyr determination despite its widespread application as a biomarker of PAH exposure relevant to global health initiatives, including indoor air pollution from wood smoke in household settings [29]. Herein, we report a rigorous inter-laboratory method comparison for 1-OH-Pyr quantification in human urine following enzyme deconjugation on a cohort of firefighters recently deployed at the Fort McMurray wildfire when using reverse-phase LC coupled to electrospray ionization-tandem mass spectrometry (LC-MS/MS) and GC with atmospheric pressure chemical ionization-high resolution mass spectrometry (GC-HRMS). Method performance and analytical figures of merit for urinary 1-OH-Pyr determination are presented with good mutual agreement demonstrated between the two instrumental platforms. Assessment of the distribution of urinary 1-OH-Pyr concentrations measured among recently deployed firefighters and its correlation to other OH-PAHs and MPs are also discussed.

2.3 Materials and Methods

2.3.1 Chemicals and Reagents

HPLC grade reagents including acetonitrile, toluene, and glacial acetic acid in addition to sodium acetate (reagent grade, > 99%), L-Ascorbic acid (> 99%), potassium dihydrogen phosphate (ACS reagent, ≥ 99%), Surine® negative urine, β-glucuronidase (from *Helix pomatia*, Type HP-2, aqueous solution, > 100,000 units/mL) and Methyl-N-(trimethylsilyl)trifluoroacetamide (MSTFA, > 98.5%) were purchased from Sigma Aldrich (Mississauga, ON, Canada). Dichloromethane (distilled in glass) and HPLC grade methanol were purchased from Caledon (Canada). OH-PAH standards, including 1-hydroxypyrene (OH-Pyr), deuterated 1-hydroxypyrene (OH-Pyr-d_9, as recovery standard), 2-hydroxynapthalene (OH-Nap), 3-hydroxyphenanthrene (OH-Phe), 9-hydroxyfluorene (OH-Flu), and other OH-PAH standards were obtained from Toronto Research Chemicals Inc. (Toronto, ON, Canada), whereas pyrene-d_{10} (as internal standard for GC-HRMS) was supplied by Cambridge Isotope Laboratories (Tewksbury, MA, USA). Methoxyphenols (MPs), including, guaiacol, methylguaiacol, ethylguaiacol, syringol and methylsyringol were purchased from Sigma Aldrich Inc. All stock solutions were prepared by dissolving in acetonitrile to obtain a 1 mg/mL concentration following further serial dilutions in acetonitrile.

2.3.2 Sample Collection

This study was granted ethics approval (#Pro00065284) by the Health Ethics Review Board at the University of Alberta prior to collection of urine specimens from firefighters recently deployed to the Fort McMurray wildfire. All firefighters completed questionnaires to determine health status, smoking status and overall lifestyle information in addition to time and duration of

deployment. For this study, we focused on a cohort of 42 firefighters from amongst those deployed from one structural fire service in the Edmonton area. These firefighters provided a random, spot, mid-stream urine sample in a 100 mL sterile container. This was immediately aliquoted and stored at -70°C in the mobile clinical laboratory stationed at the fire hall. De-identified urine samples were transported on dry ice for analysis by GC-HRMS (McMaster University/Ministry of Environment) and LC-MS/MS (Alberta Centre for Toxicology) for an independent inter-laboratory method comparison. Of the 42 individuals, 20 participants also provided a second repeat single-spot urine specimen about three months following the initial sampling period, which was included in the inter-laboratory method comparison. All 42 firefighters were male with an average age of 25 years (range of 22-53 years) and average BMI range 29 kg/m^2 (range of 24-37 kg/m^2) with one self-reported current smoker. The deployment pattern for these firefighters was a rapid rotation of 2-3 days prior to return to base. Sample collection was carried out between 1-13 days following the return from deployment with an average collection time of 6 days after exposure, which was a major limitation in this study.

2.3.3 Sample Preparation

Sample preparation prior to LC-MS/MS analysis involved an enzyme deconjugation step followed by protein precipitation and dilution. To a 1.0 mL aliquot of urine, 5 µL of ascorbic acid (1.5 M), 200 µL of 1.0 M acetate pH 5.2 acetate buffer and 10 µL of diluted β-glucuronidase/arylsulfatase were added in addition to 10 µL of recovery standard (1-OH-Pyr-d_9, 100 ng/mL). The diluted mixture was subsequently incubated on a shaking water bath for 2 h at 60°*C* and then cooled to room temperature. A 100 µL aliquot of the hydrolyzed sample was further diluted with 5 µL of ascorbic acid and 900 µL of a dilution solvent (acetonitrile/0.1 M

phosphate buffer, pH 6 buffer, 2:1) representing an overall dilution of 12-fold. The mixture was then vortexed and centrifuged at 2,630 *g* for 10 min. The supernatant was transferred to a 2 mL auto sample vial for LC-MS/MS analysis. Similarly, the GC-HRMS procedure involved enzyme deconjugation of urine samples following solid-phase extraction (SPE) prior to pre-column chemical derivatization using MSTFA as described with minor modifications by Fernando *et al.* [5]. Briefly, urine samples were first centrifuged to sediment particulates for 5 min at 10 000 *g* from which a 1.0 mL urine aliquot was transferred into an 8 mL glass vial consisting of 5 mL of sodium acetate buffer (0.1 M, pH 5.5) and 10 µL of recovery standard, 1-OH-Pyr-d_9. Subsequently, each urine sample was incubated with 10 µL glucuronidase/arylsulfatase at 37°*C* for 17-18 h to ensure complete deconjugation of OH-PAH metabolites. SPE extraction was then carried out manually on a vacuum manifold using Varian Focus (50 mg, 6 mL) cartridges (Agilent Technologies Inc., Santa Clara, CA, USA). The SPE cartridges were conditioned with 1 mL of methanol, followed by 1 mL of water each at a flow rate of 10 mL/min. The sample was introduced at a flow rate of 1 mL/min with a subsequent addition of 1 mL nanopure water and 3 mL of 20% methanol in sodium acetate buffer (0.1 M, pH 5.5). Sorbents in the cartridges were dried by aspirating air through the cartridge for 5 min followed by nitrogen for an additional 5 min. The deconjugated OH-PAH metabolites and MPs retained on the stationary phase were eluted with dichloromethane (DCM) at a rate of 0.5 min/mL. The DCM extract was then dried with 10-20 mg of magnesium sulfate anhydrous. Next, the contents were passed through a glass pipette containing glass wool into an 8 mL vial to prevent salt carryover during sample pretreatment. The final extract was then blown down to 100 µL under a gentle stream of nitrogen prior to transfer to a GC vial insert in which the sample was blown down further to dryness and reconstituted to 10 µL in HPLC grade toluene. An internal standard, 1 µL (pyrene-d_{10}, 100 pg/

µL), was also spiked into the sample. Finally, the sample was derivatized with 15 µL of MSTFA following incubation at 60°*C* for 20 min, which resulted in a 25-fold overall enrichment of the original urine sample as required for measuring low concentrations of urinary 1-OH-Pyr in non-smokers/non-occupationally exposed subjects. Additionally, creatinine was measured in all single-spot urine specimens to correct for hydration status between subjects using an AU 480 Olympus Chemistry Analyzer (Beckman Coulter Inc., Mississauga, ON, Canada). This assay is based upon a kinetic method using the Jaffé reaction, where the creatinine concentration was determined colorimetrically via formation of a red complex from reaction of picrate and creatinine under alkaline conditions. DRI® creatinine-detect calibrator set including DRI® creatinine-detect 7.5 mg/dL and DRI® creatinine-detect 23.0 mg/dL were purchased from Thermo Scientific Inc. (Mississauga, ON, Canada) and used for calibration and quality control. All urinary OH-PAH concentrations in this study were reported as their normalized concentrations to creatinine (ng/g creatinine).

2.3.4 Instrumental Analysis

Samples were analyzed on both GC-HRMS and LC-MS/MS platforms. LC-MS/MS analysis was performed using Agilent 1200 high performance liquid chromatograph system (Agilent Technologies, Santa Clara, CA, USA) coupled with a Sciex 5500 Q-trap mass spectrometer (AB Sciex, Concord, Ontario, Canada). Separation of OH-Pyr was achieved on an Agilent SB-C18 column (Poreshell 120, 3.0 X 100 mm, 2.7 µm) using isocratic elution of water (40%) and acetonitrile (60%) at a flow rate of 0.4 mL/min, and an injection volume of 10 µL with the column kept at room temperature. The mass spectrometer was operated in negative ion mode with multiple reaction monitoring (MRM). The source temperature was 550°C with ion spray

potential of -4.5 kV, curtain gas at 20 psi, gas 1 at 45, gas 2 at 55. MRM transition for the detection of 1-OH-Pyr was 217.0/188.9 with a collision energy (CE) at -45 V and declustering potential (DP) at -140 V. While MRM transitions for 1-OH-Pyrene-d_9 was 226.1/198.1 with CE at -48 V and DP at -150 V. Data processing was conducted with Multiquant 3.0.1 (AB Sciex, Canada). In addition, GC-HRMS analysis was performed on an Agilent 7890B gas chromatograph (Agilent Technologies, Santa Clara, CA, USA) coupled to a Water Xevo G2-XS quadrupole time-of-flight mass spectrometer (Waters Corporation, Wilmslow, UK) operated with an atmospheric pressure chemical ionization (APCI) source. The primary column (DB-17 30 m 0.25 mm x 0.15 μm film) was connected to a Custom MXT tubing (sulfinert treated, Restek Corporation, Bellefonte, PA) 0.8 m x 0.18 mm, inserted into the heat transfer line (340°C). Helium was used as the carrier gas, and the optimal flow was set at 1 mL/min. An injection volume of 1 μL of sample extract was used with a splitless injector set to 280°C. The initial oven temperature was set to 70°C, and then ramped at a rate of 8°C/min to 300°C, and held for 5 min. The corona current was set to 3 μA and the ion source temperature was 150°C, whereas the cone and auxiliary gas flows were 100 L/hr and 175 L/hr. Full scan spectra was acquired over a mass range from *m/z* 50 to 1200 at a mass resolving power > 20 000 (FWHM). Trimethylsilylated (TMS)-1-OH-Pyr and its 1-OH-Pyr-d_9 deuterated analog were monitored as MRM transitions (290.1/275.1305 and 299.1/284.1305, respectively) recorded in parallel with the full-scan measurements. These transitions correspond to a loss of a methyl radical as base peak in the mass spectrum at an optimum collision energy of 30 V. Data processing was conducted using MassLynx 4.2 (Waters Corporation, USA) with quantification by relative response factor using a recovery standard (1-OH-Pyr-d_9).

2.3.5 Method Validation and Quality Control

For LC-MS/MS analysis a set of eight matrix matched calibrators were prepared by spiking known amount of 1-OH-Pyr into Surine® negative urine from 20 ng/L to 10,000 ng/L. 10 μL was injected on-column, which is equivalent to 0.016 pg to 8 pg of 1-OH-Pyr on column for calibrators. Two in-house quality controls (QCs) were prepared at 0.1 ng/mL and 0.5 ng/mL by a different analyst. A standard reference material (SRM 3673, organic contaminants in non-smoker's urine) which contains 29.3 ± 1.8 ng/L of 1-OH-Pyr in urine was purchased from National Institute of Standards & Technology (NIST) and used as an external QC. Each batch consisted of 40 Fort McMurray samples, a set of calibrators, a solvent blank, a synthetic urine blank, two in-house QCs and SRM 3673. All blanks, calibrators and QCs were hydrolyzed and cleaned up along with the samples. The blanks and QCs were run after calibrators, in the middle of the sequence and at the end of sequence to insure all samples were bracketed with blanks and QCs. QCs were considered acceptable within ± 20% of targeted concentration. The quantification of 1-OH-Pyr was based on the measured ion response ratio relative to the recovery standard (1-OH-Pyr-d_9) through the use of an external calibration curve with $R^2 > 0.999$. In the case for GC-HRMS, a QC was first prepared by pooling together a mixture of residual urine samples from the firefighter cohort, which was subsequently aliquoted and processed together as a batch of samples comprising individual urine specimens (seven), and a synthetic urine blank (one). All samples were spiked with a recovery standard (1-OH-Pyr-d_9) prior to performing the standardized sample workup protocol, including enzyme deconjugation, SPE, solvent reconstitution and chemical derivatization. Overall, a total of ten pooled QC specimens were analyzed intermittently by GC-HRMS in this study in order to evaluate overall technical precision, which was also used in spike/recovery studies. Additionally, prior to each set of

analyses on the GC-HRMS system, a 1 μL injection of 1-OH-Pyr standard at 50 ng/L was injected to assess the instrument performance. Throughout the study, at the end of each batch sequence, a toluene blank was also injected to assess sample carry-over throughout the sequence, for which no significant carry over was noted. Individual urine extracts from firefighters, synthetic urine blanks and pooled QC were run in a randomized order within a batch sequence when using GC-HRMS. A total of two external calibration curves were acquired for 1-OH-Pyr, as well as other OH-PAHs and MPs, which were derived from nine different calibrant concentrations over a linear dynamic range of three orders of magnitude with correlation coefficient greater than $R^2 > 0.999$. All stock calibration solutions were prepared in acetonitrile and derivatized using MSTFA resulting in formation of their *O*-trimethylsilyl derivatives. Statistical analyses were carried out using SPSS 18.0 (IBM Corporation, NY, USA), Minitab 17 (Minitab Inc., PA, USA), MedCalc (MedCalc Software, Ostend, Belgium), and MetaboAnalyst 4.0 [30].

2.4 Results and Discussion

2.4.1 Urinary 1-OH-Pyr Determination by GC-HRMS

1-OH-Pyr is a widely measured indicator of recent PAH exposure in human urine, which shows consistently higher levels of excretion in occupationally exposed relative to non-exposed workers [4, 31], as well as smokers, including secondhand smoke exposure at home as compared to non-smokers [32]. However, other factors also contribute to total PAH exposures, including diet, lifestyle, and genotype that impact OH-PAH metabolic activity even with low exposures due to polymorphic genes encoding CYP enzymes [33]. As a result, careful interpretation of findings is needed when analyzing urinary 1-OH-Pyr concentrations in small cohorts given large

between-subject variations in pharmacokinetics that is also dependent on exposure mechanisms [34]. Optimally this would be applied to biomonitoring studies for PAH exposure on a population level. Data interpretation is also dependent on method accuracy notably when comparing reference ranges for 1-OH-Pyr acquired by different instrumental methods. In this work, two complementary analytical strategies were performed independently at different laboratory sites in order to quantify 1-OH-Pyr from de-identified urine specimens collected from firefighters recently deployed at the Fort McMurray wildfire. In both cases, a matching deuterated recovery standard (1-OH-Pyr-d_9) was spiked in all urine specimens prior to enzyme hydrolysis in order to analyze total 1-OH-Pyr content given differences in secondary metabolism as its glucuronide conjugate is a major excreted metabolite [22]. Extensive in-source fragmentation often arises from electron ionization (EI), a hard ionization technique commonly used in GC/MS. In this study, APCI was employed to significantly reduce in-source fragmentation of the incipient molecular ions by stabilizing collisions with ambient nitrogen. This enabled the sensitive detection of trace levels of OH-PAH metabolites in conjunction with the selectivity afforded by high mass resolution, accurate mass measurements of the Q-TOF-MS [35]. The APCI mass spectra of 1-OH-Pyr and 1-OH-Pyr-d_9 (not shown) are dominated by their molecular ions ($M^{\bullet+}$) at *m/z* 290.113 and *m/z* 299.169, respectively. Upon collision-induced dissociation (CID), the molecular ions undergo loss of a methyl radical (15.024 Da) (**Supplementary Figure S2.1**). These dissociations were monitored by MRM experiments performed in parallel with the full-scan HRMS measurements in order to achieve greater sensitivity. Extracted ion chromatograms (EICs) for representative urine samples analyzed by GC-HRMS are depicted for 1-OH-Pyr-d_9 (**Figure 2.1A**) and 1-OH-Pyr (**Figure 2.1B**) as their trimethylsilyated (TMS) derivatives that elute within 24 min. Note that the native and deuterated

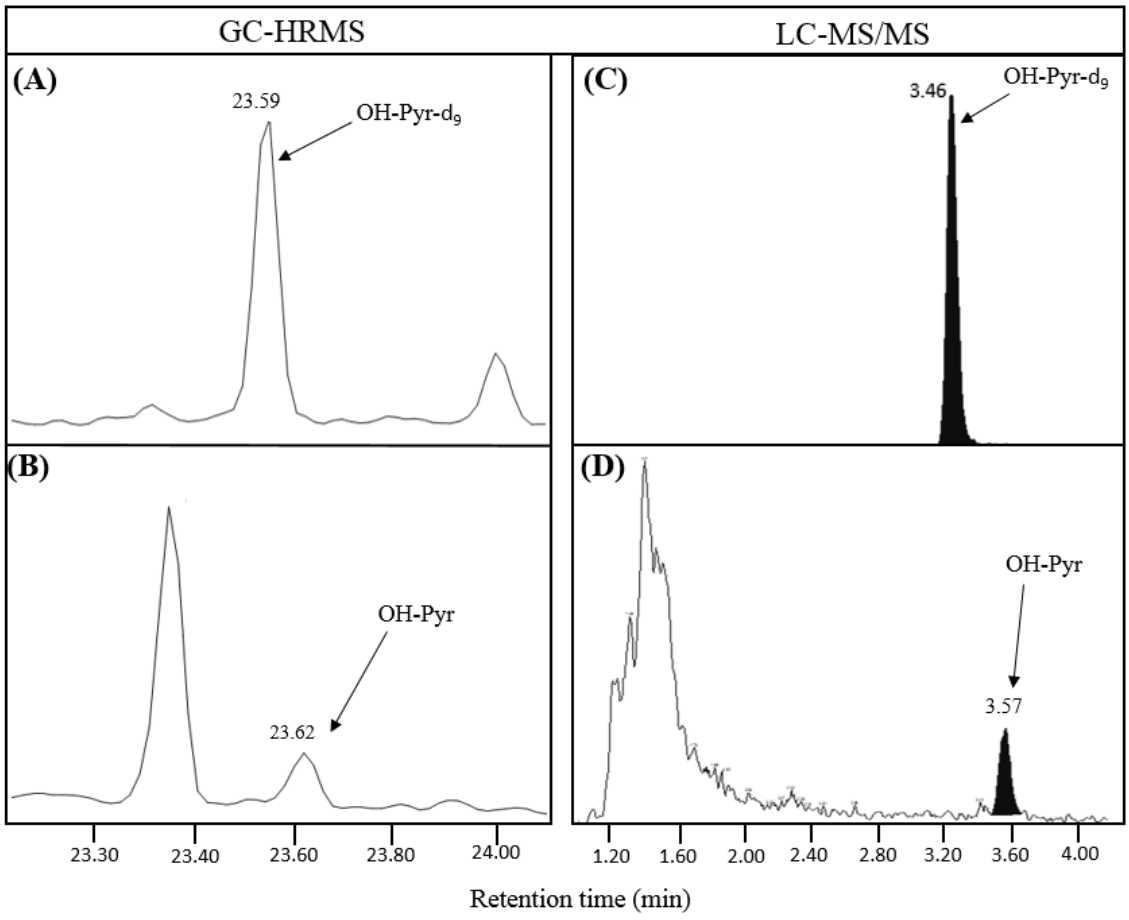


Figure 2.1. Representative extracted ion chromatograms (EIC) for urinary 1-OH-Pyr with its stable-isotope internal standard, 1-OH-Pyr-d9 following enzyme deconjugation when using GC-HRMS and LC-MS/MS. The GC-HRMS method included pre-column SPE to lower detection limits together with atmospheric pressure chemical ionization for improved detection of 1-OH-Pyr in urine specimens, whereas the LC-MS/MS method using an isocratic elution analyzed samples rapidly without chemical derivatization when using electrospray ionization and multiple reaction monitoring.

analogues of TMS-1-OH-Pyr differ in mass by 9 Da while having a minor retention time deviation of only 0.03 min. Method validation of GC-HRMS included assessment of long-term technical (inter-day) precision based on intermittent analysis of pooled urine specimens as QCs over the duration of the study with an average coefficient of variance (*CV*) of 11.5% (*n*=10). This is acceptable given the extensive sample workup that is required for processing urine samples prior to analysis. Method bias was evaluated by recovery experiments in representative urine specimens after spiking at two concentration levels for 1-OH-Pyr in triplicate relevant to

occupationally unexposed smokers (500 ng/L) and non-smokers (200 ng/L) [31] that resulted in an overall accuracy of 108% ranging from 85-122%. Additionally, an external calibration curve for 1-OH-Pyr normalized to 1-OH-Pyr-d_9 has a wide linear dynamic range (R^2 > 0.999) over three orders of magnitude from 2.0 to 2000 ng/L (**Supplementary Figure S2.2**). The limit of detection (LOD, *S/N* = 3) and limit of quantification (LOQ, *S/N* = 10) for the GC-HRMS method when using calibrators prepared in acetonitrile was 0.64 ng/L and 2.16 ng/L, respectively. However, average blank-limited concentrations measured for 1-OH-Pyr in processed synthetic urine extracts (*n*=7) was about 30 ng/L reflecting higher background levels present in real-world samples. As summarized in **Table 2.1**, GC-HRMS offers higher sensitivity (*i.e.,* sharper slope) when using APCI-MS and lower detection limits (LOD) since SPE was used as an off-line sample preconcentration step (*i.e.,* 25-fold sample enrichment) prior to chemical derivatization. As a result, reliable quantification was realized for 1-OH-Pyr in 98.4% of urine samples collected from firefighters (61 of 62) above the blank-limited background (30 ng/L). Overall, ten OH-PAH metabolites/isomers and MPs were measured consistently in the majority (> 80%) of urine specimens, including 1- and 3-hydroxyphenanthrene (1-, 3-OH-Phe), 2- and 9-hydroxyfluorene (2-, 9-OH-Flu), 2-hydroxynapthalene (2-OH-Nap), as well as a series of MPs, including guaiacol, methylguaiacol, ethylguaicol, syringol and methylsyringol.

2.4.2 Urinary 1-OH-Pyr Determination by LC-MS/MS

Total 1-OH-Pyr analysis in human urine samples following enzyme deconjugation was also performed by reverse-phase LC-MS/MS when using a fast isocratic elution program within 4 min as shown in **Figure 2.1C** and **Figure 2.1D** for 1-OH-Pyr-d_9 and 1-OH-Pyr, respectively. This method offers higher throughput for 1-OH-Pyr biomonitoring in human urine when using

Table 2.1. Summary of various figures of merit for the determination of urinary OH-Pyr as a marker of smoke exposure when using GC-HRMS and LC-MS/MS platforms.

Figures of Merit	GC-HRMS	LC-MS/MS
Technical precision[a]	11.5%	5.1%
Recovery/bias (range)[b]	108% (85-122%)	93% (88-98%)
LOD (*S/N* = 3)	0.64 ng/L	10 ng/L
LOQ (*S/N* =10)	2.16 ng/L	20 ng/L
Linearity (slope; R^2)[c]	0.997; 0.9999	0.728; 0.9991
% Non-detected in urine[d]	1.6% (1/62)	29% (18/62)

[a] Inter-day precision was determined based on CV for intermittent analysis of 10 pooled urine samples as QC performed over the study when using GC-HRMS, whereas LC-MS/MS precision was determined using a NIST SRM (29.3 ng/L) in triplicate?.
[b] Spike/recovery experiments to estimate average bias and range were performed in triplicate two concentration levels in pooled urine samples for GC-HRMS, and three concentration levels in synthetic urine matrix for LC-MS/MS.
[c] External calibration curves were based on normalization to OH-Pyr-d_9 over a 1000 and 500-fold linear dynamic range for OH-Pyr using seven calibrators when using GC-HRMS and LC-MS/MS, respectively, where slope units are $(ng/mL)^{-1}$.
[d] Analysis of 42 single-spot urine specimens from recently deployed firefighters, as well as 20 repeat urine specimens > 3 mo.

ESI under negative ion mode with multiple reaction monitoring, which avoids the need for additional sample pretreatment steps, such as SPE or pre-column chemical derivatization. However, there was a small yet significant retention time difference of about 0.11 min between 1-OH-Pyr and 1-OH-Pyr-d_9 that was used as a stable-isotope internal standard to correct for ion suppression or enhancement effects. In this case, a deuterium isotope effect results in earlier elution of 1-OH-Pyr-d_9, which may contribute to ion suppression when analyzing diverse urine specimens that vary widely in matrix composition/hydration status by LC-MS/MS without extensive sample cleanup [36]. However, recovery studies for 1-OH-Pyr demonstrated minimal bias with an overall accuracy of 93% when spiked at three different concentration levels in representative urine samples. In addition, replicate analysis of SRM 3673 by LC-MS/MS had an accuracy of 88% with excellent precision (*CV* = 5.5%, *n*=6), as compared to the certified value (29.3 ± 1.8 ng/L) , determined from three governmental agencies (NIST, CDC and INSPQ) using different protocols based on GC-MS/MS with SPE or GC-HRMS with liquid-liquid extraction

and enzyme deconjugation [37]. Similar to GC-HRMS, excellent linearity (R^2 > 0.999) was demonstrated for 1-OH-Pyr quantification over a 500-fold dynamic range when normalizing ion responses relative to its stable-isotope internal standard. In this case, selective reaction monitoring was used to monitor the product ion (*m/z* 188.9) for OH-Pyr corresponding to $[M\text{-}H\text{-}CO]^-$, which is a common fragmentation pathway following collisional-induced dissociation of OH-PAHs resulting in a neutral loss of carbon monoxide [38]. Although LC-MS/MS provides excellent recovery and precision with faster analysis times, lower sensitivity was the major limitation as compared to the GC-HRMS protocol, where 29% (18 of 62) of diluted urine samples had 1-OH-Pyr concentrations measured below the method detection limit. Further improvements in concentration sensitivity in LC-MS/MS is realized by SPE for sample enrichment of OH-PAHs in urine [24] and/or pre-column chemical derivatization of OH-PAHs as their pentaflurobenzyl ethers to increase solute ionization efficiency when using APCI in negative ion mode [39].

2.4.3 Inter-laboratory Method Comparison and Mutual Agreement

Following method validation, an inter-laboratory method comparison was performed based on analysis of paired urine samples using GC-HRMS and LC-MS/MS protocols to ensure consistent 1-OH-Pyr quantification. Due to the skewed data distribution for urinary 1-OH-Pyr concentrations measured in the firefighter cohort (Shapiro-Wilks test, p > 0.05), a Passing-Bablok regression analysis, and a Bland-Altman % difference plot were performed to assess mutual agreement [40] as shown in **Figure 2.2**. Of the total 62 urine samples analyzed from the firefighters, 44 were reliably quantified by both methods with 1-OH-Pyr concentrations ranging from 47 ng/L to 940 ng/L with 18 non-detected samples among diluted/processed urine samples

when using LC-MS/MS. Elevated urinary 1-OH-Pyr concentrations due to occupational PAH exposures have been reported to range from a median of 560 ng/L for chimney sweeps [41] to over 7,000 ng/L for coke oven workers [42]. Importantly, all measured concentrations from deployed firefighters were found to be well below the recommended biological exposure index (BEI) [43] threshold of 2,500 ng/L based on hydrolyzed (total) urinary 1-OH-Pyr for end of shift at end of work week sampling [4]. **Figure 2.2A** demonstrates that there was a random data distribution independent of 1-OH-Pyr concentration with a mean bias of 39% for the GC-HRMS method relative to LC-MS/MS. Despite this positive bias, there were few outliers ($p < 0.05$; or 2 of 44) that exceeded agreement limits at a 95% confidence interval. Additionally, **Figure 2.2B** reveals that there was no statistically significant difference from the line of unity ($p > 0.05$) despite a positive slope of 1.36. Thus, this work confirms good mutual agreement in measured urinary 1-OH-Pyr concentrations between two laboratory sites with a modest bias likely attributed to matrix-induced ion suppression and different enzyme hydrolysis efficiencies. The LC-MS/MS deconjugation protocol involved a two hour incubation time while GC-HRMS carried out an 18 hour deconjugation contributing to incomplete release of the major *O*-glucuronide of 1-OH-Pyr with the LC-MS/MS deconjugation method (**Supplementary Figure S2.3**). A bias of 24% was evaluated with the LC-MS/MS protocol from the two hour to 18 hour deconjugation time points (CV = 8.2%, n=3) largely explaining the 39% bias observed by the GC-HRMS method. Moreover, urine drug testing and biomarker measurements have reported differences in analyte recoveries and systematic error due to batch variations in commercial enzymes, including recombinant and wild-type β-glucuronidases/arylsulfatases that possess variable substrate affinities [44, 45]. Thus, standardized pre-analytical protocols for urine

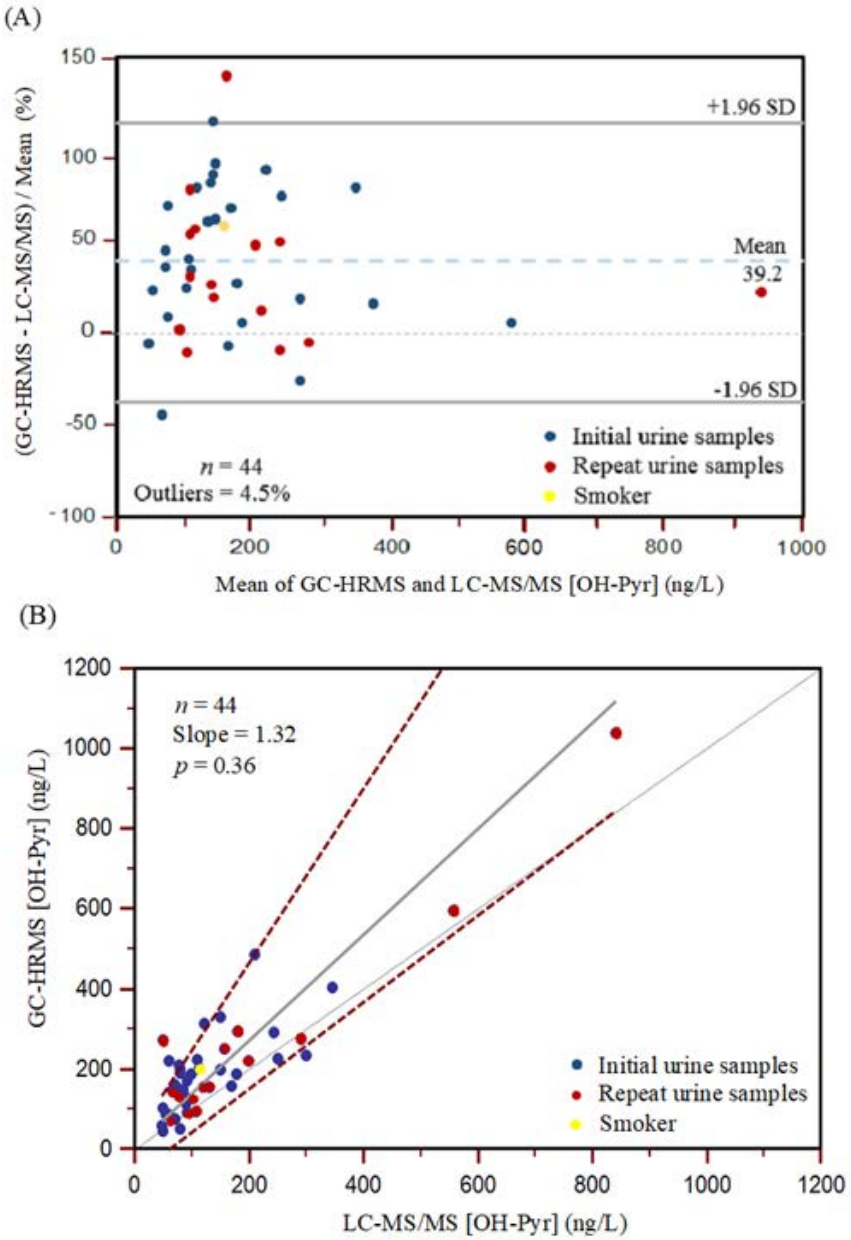


Figure 2.2. (A) A Bland-Altman percent difference plot for comparing the mutual agreement between urinary 1-OH-Pyr concentrations measured independently by GC-HRMS and LC-MS/MS methods at two different laboratories from firefighters deployed in the 2016 Fort McMurray wildfire. Overall, the data is randomly distributed with a modest average positive bias of 39% with two outliers outside agreement limits (B) A Passing-Bablok regression analysis demonstrates no significant deviation (dotted lines; 95% confidence interval) from the line of equality ($p > 0.05$) with a slope of 1.36 (grey line; regression line). There was one self-reported tobacco smoker (n=1; yellow symbol) among recently deployed firefighters (n=28; blue symbol), whereas repeat urine specimens (red symbol) collected from a sub-set of firefighters (n=15) 3 months following their initial deployment. Overall, 18 urine samples were not detected by LC-MS/MS due to its higher detection limits from the total number of urine sample collected from study (n=62).

processing using enzyme reagents are critical for achieving consistent measurements between laboratories.

2.4.4. Correlation Analysis of Other Urinary Smoke Markers to 1-OH-Pyr

Since wood smoke is comprised of a complex mixture of PAHs that can modulate the metabolism of other carcinogens [11], a wider range of urinary OH-PAHs can provide more accurate assessment of total chemical exposures than 1-OH-Pyr alone. For instance, a panel of eight urinary OH-PAH metabolites have been shown to have modest positive associations with cardiovascular disease risk among US adults that is independent of smoking status and other confounders [46]. Similarly, eight urinary OH-PAHs were also associated with significant decreases in lung function in a Canadian population with larger effects reported for the sum of compound classes [47]. Due to variations in hydration status when relying on random single-spot urine collections, creatinine is frequently used to reduce biological variances in measured analyte concentrations provided subjects have normal kidney function and protein intake [48]. Additionally, since firefighters in this study were all male adults, creatinine was deemed appropriate for data normalization as it was also found to show linear correlations ($R^2 > 0.800$) to both urinary specific gravity (data not shown) and osmolality (**Supplementary Figure S2.4**) – two independent bulk solution properties also used to correct for urine hydration status. **Figure 2.3A** depicts a correlation matrix/heat map with hierarchical cluster analysis (HCA) in order to explore the relationship among ten urinary smoke markers (including 1-OH-Pyr) measured consistently among firefighters by GC-HRMS based on their creatinine normalized concentrations (ng/g creatinine). As expected, both syringol (syringol/methylsyringol; $\rho = 0.462$, $p = 0.0002$) and guaicol (methylguaicol/ethylguaicol; $\rho = 0.667$, $p = 3 \times 10^{-9}$) analogs as major

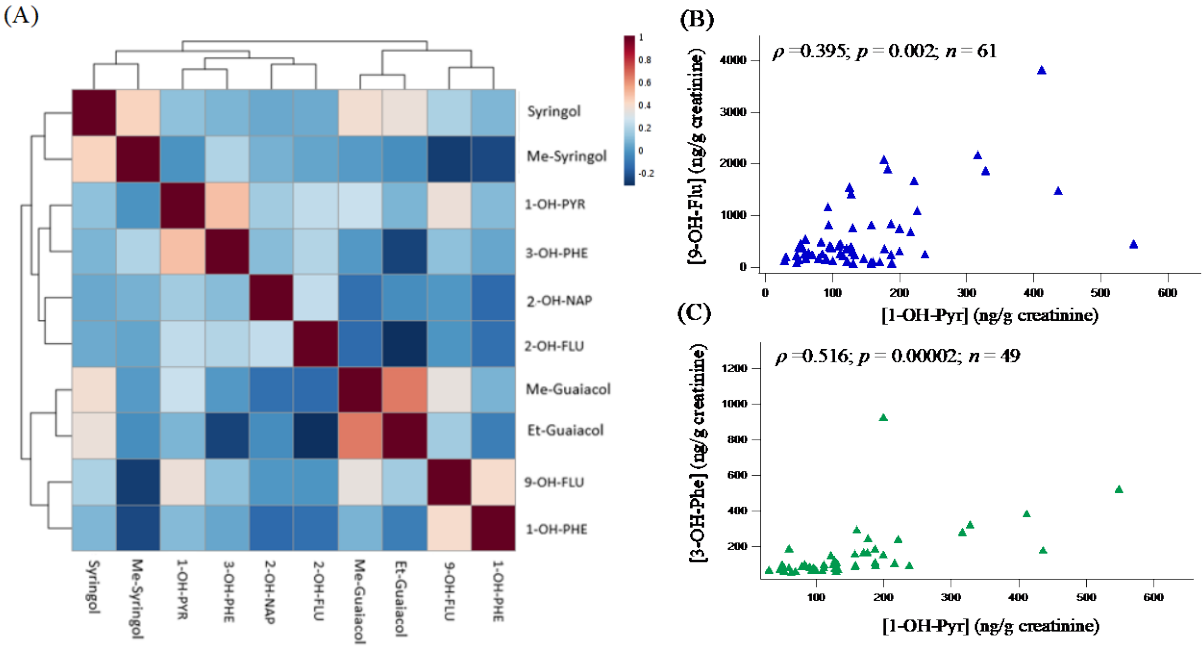


Figure 2.3. (A) A correlation matrix heat map with hierarchical cluster analysis (HCA) for classifying the relationship of ten urinary smoke markers measured by GC-HRMS comprising different OH-PAHs and MPs from firefighters recently deployed at the Fort McMurray wildfire. Correlation plots based on a Spearman rank test confirm that **(B)** 9-OH-Flu, and **(C)** 3-OH-Phe were the two most significant compounds associated with urinary 1-OH-Pyr excretion following a Bonferroni adjustment ($p < 0.005$) implying similar uptake, distribution and metabolic pathways.

MPs were strongly correlated with each other when using a Spearman rank test with a high degree of similarity in their urinary excretion patterns as shown in the dendrogram.

Additionally, 9-OH-Flu ($\rho = 0.395$, $p = 0.002$) and notably 3-OH-Phe ($\rho = 0.516$, $p = 0.0002$) were two specific isomers of OH-PAHs directly correlated with urinary 1-OH-Pyr as shown in **Figure 2.3B.** This data is consistent with previous studies that have demonstrated a dose-response relationship of urinary 1-OH-Pyr excretion which is elevated among smokers, and also correlated with 3-OH-Phe, but not for 1-OH-Phe [49]. In this study, urinary 1-OH-Phe was only correlated to 9-OH-Flu excretion ($\rho = 0.434$, $p = 0.0004$) unlike 1-OH-Pyr ($\rho = 0.146$, $p = 0.258$). Thus, smoke exposure among firefighters contribute to distinctive excretion patterns among OH-PAHs and their isomers since they are dependent on the expression, induction and

inhibition of specific PAH-metabolizing CYP isoforms (1A1, 1B1, 1A2 and 3A4) that are further modulated by diet, lifestyle, and other environmental exposures [34].

2.4.5 Urinary 1-OH-Pyr Concentrations of Firefighters

As there were no significant differences in measured urinary 1-OH-Pyr concentrations among a subset of repeat urine specimens ($p > 0.05$, paired t-test, n=20) as shown in **Figure 2.2A**, original urine specimens collected from recently deployed firefighters at the Fort McMurray wildfire (n=42) were evaluated. Indeed, 1-OH-Pyr is reported to be excreted within 12 h following controlled feeding experiments on non-occupationally exposed/non-smokers [34], whereas much longer yet more variable median lifetimes (up to 35 h) [50] have been reported following inhalation and/or dermal exposure in coke oven workers [51], as well as controlled smoke exposure studies involving non-smoking volunteers [19]. However, the average delay to urine collection of 6 days (ranging from 1 to 13 days) from firefighters in this study exceeded all reported half-life times for 1-OH-Pyr. For instance, **Figure 2.4A** depicts the creatinine adjusted concentration distribution in urinary 1-OH-Pyr for firefighters as measured by GC-HRMS, which ranged from 30 to 485 ng/g creatinine with a median/interquartile range (IQR) of (128 ± 100) ng/g creatinine, and a geometric mean of 118 ng/g creatinine. In this non-random sampling of 42 firefighters, from the best equipped and organized fire service [3], there were 18 non-smokers who provided a urine sample within 24 hr of exposure. Even among this sub-set of recently deployed firefighters, 1-OH-Pyr has a median concentration of 129 ng/g creatinine that is similar to overall data trends in **Figure 2.4**. These findings are lower than reported from a recent study of structural firefighters with a geometric mean for 1-OH-Pyr of 270 ng/g creatinine [7], including post-shift wildland firefighters with a geometric mean of 750 ng/g creatinine [52].

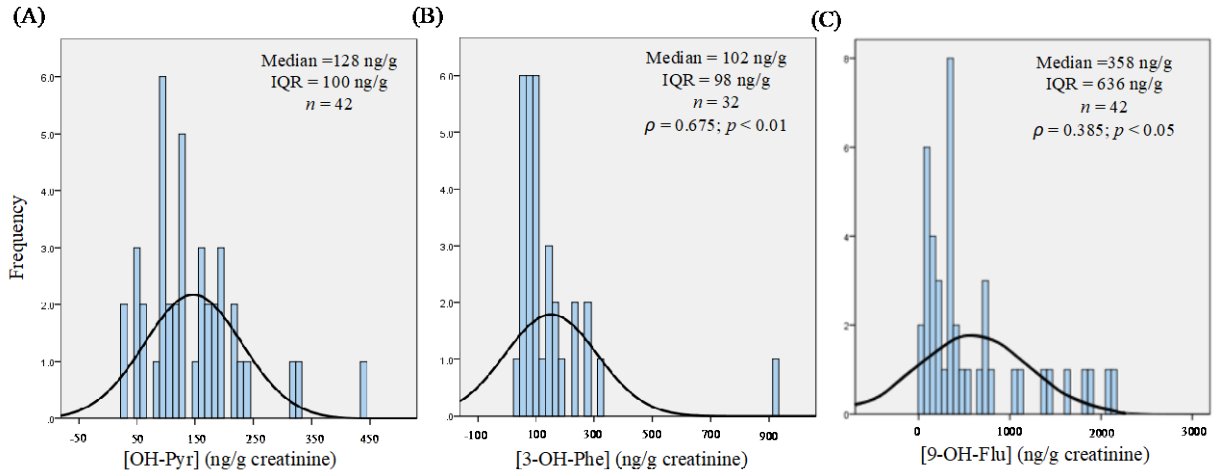


Figure 2.4. Histograms depicting creatinine-normalized concentrations measured using GC-HRMS for (A) 1-OH-Pyr, (B) 3-OH-Phe, and (C) 9-OH-Flu in urine for biomonitoring of 42 firefighters deployed at the Fort McMurray wildfire. Overall, firefighters had lower OH-PAH concentrations than anticipated due to prolonged delays in urine collection that does not accurately reflect recent smoke exposure during early stages of intense firefighting.

Urinary concentrations for firefighters are also depicted for the two OH-PAH isomers that were directly correlated with 1-OH-Pyr, namely 3-OH-Phe and 9-OH-Flu as shown in **Figure 2.4B** and **2.4C**, including the total urinary concentration of all three OH-PAHs (**Supplementary Figure S2.5**). However, it must be recognized that even the same day samples were collected from firefighters exposed only after the second week of the fire, and that their OH-PAH concentrations can in no way be taken as indicative of smoke exposure during the earlier days of intense firefighting under very difficult conditions [3].

2.5 Conclusion

In summary, two complementary instrumental methods were compared for urinary 1-OH-Pyr determination as required for biomonitoring of recent PAH exposure among a cohort of firefighters deployed at the Fort McMurray wildfire. Direct analysis by isotope-dilution LC-MS/MS following enzyme deconjugation allows for rapid screening of 1-OH-Pyr without

complicated sample pretreatment. In contrast, GC-HRMS together with SPE allows for consistent detection of low concentration levels of 1-OH-Pyr and the resolution of a wider range of low abundance OH-PAHs and their isomers. Overall, there was good mutual agreement between methods for urinary 1-OH-Pyr determination performed independently at two different laboratories with excellent technical precision while incorporating intermittent quality controls. Good accuracy was demonstrated following method validation, and a modest extent of bias was evident between methods, which was likely attributed to incomplete enzyme deconjugation when using shorter incubation times. Only two OH-PAH isomers were found to be significantly correlated with urinary 1-OH-Pyr excretion patterns among firefighters, including 3-OH-Phe and 9-OH-Flu. Overall, firefighters were found to have urinary 1-OH-Pyr concentrations well below the recommended BEI for health hazard, which was a study limitation due to the practical logistic challenges in coordinating urine collection following the intense early stages of firefighting under emergency conditions. Future studies that allow for facile collection/storage of multiple spot urine samples in the field (*e.g.,* filter paper cards) and the analysis of alternative biospecimens for assessing PAH exposures of firefighters over a longer time window post-deployment (*e.g.,* stool) or continuously during deployment (*e.g.,* sweat pads) are recommended for new advances in risk assessment and personalized health.

2.6 References

1. Cherry N, Haynes W. Effects of the Fort McMurray wildfires on the health of evacuated workers: follow-up of 2 cohorts. CMAJ Open. 2017;5:E638–45.

2. Landis MS, Edgerton ES, White EM, Wentworth GR, Sullivan AP, Dillner AM. The impact of the 2016 Fort McMurray Horse River Wildfire on ambient air pollution levels in the Athabasca Oil Sands Region, Alberta, Canada. Sci Total Environ. 2018;618:1665–76.

3. Cherry, N.; Akililu, Y.-A.; Beach, J.; Britz-McKibbin, P.; Elbourne, R.; Galarneau, J.-M.; Gill, B.; Kinniburgh, D. J. Ann. Urinary 1-hydroxypyrene and skin contaminantion in firefighters deployed to the Fort McMurray fire. Work Expo. Health 2018, 63:448-458.

4. Polycyclic Aromatic Hydrocarbons (PAHs): BEI(R), 7th Edition Documentation, American Conference of Governmental Industrial Hygienists (ACGIH), Cinanatti, OH, USA, 2017;1-25.

5. Fernando S, Shaw L, Shaw D, Gallea M, Vandenenden L, House R, et al. Evaluation of Firefighter Exposure to Wood Smoke during Training Exercises at Burn Houses. Environ Sci Technol. 2016;50:1536–43.

6. Navarro KM, Cisneros R, Noth EM, Balmes JR, Hammond SK. Occupational Exposure to Polycyclic Aromatic Hydrocarbon of Wildland Firefighters at Prescribed and Wildland Fires. Environ Sci Technol. 2017;51:6461-69.

7. Keir JLA, Akhtar US, Matschke DMJ, Kirkham TL, Chan HM, Ayotte P, et al. Elevated Exposures to Polycyclic Aromatic Hydrocarbons and Other Organic Mutagens in Ottawa Firefighters Participating in Emergency, On-Shift Fire Suppression. Environ Sci Technol. 2017;51:12745–55.

8. Reports C. Fatalities among volunteer and career firefighters--United States, 1994-2004. MMWR Morb Mortal Wkly Rep. 2006;55:453–5.

9. Guidotti TI. Evaluating causality for occupational cancers: The example of firefighters. Occup Med (Chic Ill). 2007;57:466-71.

10. Strickland P, Kang D. Urinary 1-hydroxypyrene and other PAH metabolites as biomarkers of exposure to environmental PAH in air particulate matter. Toxicol. Lett. 1999;108:191-9.

11. Shimada T, Guengerich FP. Inhibition of human cytochrome P450 1A1-, 1A2-, and 1B1-mediated activation of procarcinogens to genotoxic metabolites by polycyclic aromatic hydrocarbons. Chem Res Toxicol. 2006;19:288-94.

12. Jongeneelen FJ, Anzion RBM, Scheepers PTJ, Bos RP, Henderson PT, Nijenhuis EH, et al. 1-hydroxypyrene in urine as a biological indicator of exposure to polycyclic aromatic hydrocarbons in several work environments. Ann Occup Hyg. 1988;32:35-43

13. Li Z, Sandau CD, Romanoff LC, Caudill SP, Sjodin A, Needham LL, et al. Concentration and profile of 22 urinary polycyclic aromatic hydrocarbon metabolites in the US population. Environ Res. 2008;107:320–31.

14. Hansen ÅM, Mathiesen L, Pedersen M, Knudsen LE. Urinary 1-hydroxypyrene (1-HP) in environmental and occupational studies-A review. Int. J. Hyg. Environ. Health. 2008. p. 471–

503.

15. Lu PL, Chen ML, Mao IF. Urinary 1-hydroxypyrene levels in workers exposed to coke oven emissions at various locations in a coke oven plant. Arch Environ Health. 2002;57:255-61.

16. Bouchard M, Viau C. Urinary excretion kinetics of pyrene and benzo(a)pyrene metabolites following intravenous administration of the parent compounds or the metabolites. Toxicol Appl Pharmacol. 1996;139:301-9.

17. Chang CM, Edwards SH, Arab A, Del Valle-Pinero AY, Yang L, Hatsukami DK. Biomarkers of tobacco exposure: Summary of an FDA-sponsored public workshop. Cancer Epidemiol. Biomarkers Prev. 2017;26:291-302.

18. Nolte CG, Schauer JJ, Cass GR, Simoneit BR. Highly polar organic compounds present in wood smoke and in the ambient atmosphere. Env Sci Technol. 2001;35:1912–9.

19. Li Z, Trinidad D, Pittman EN, Riley EA, Sjodin A, Dills RL, et al. Urinary polycyclic aromatic hydrocarbon metabolites as biomarkers to woodsmoke exposure-results from a controlled exposure study. J Expo Sci Environ Epidemiol. 2016;26:241-8.

20. Fagundes RB, Abnet CC, Strickland PT, Kamangar F, Roth MJ, Taylor PR, et al. Higher urine 1-hydroxy pyrene glucuronide (1-OHPG) is associated with tobacco smoke exposure and drinking maté in healthy subjects from Rio Grande do Sul, Brazil. BMC Cancer. 2006;6:139-46.

21. Jongeneelen FJ, Anzion RBM, Henderson PT. Determination of hydroxylated metabolites of polycyclic aromatic hydrocarbons in urine. J Chromatogr B Biomed Sci Appl. Elsevier; 1987;413:227–32.

22. Strickland PT, Kang D, Bowman ED, Fitzwilliam A, Downing TE, Rothman N, et al. Identification of 1-hydroxypyrene glucuronide as a major pyrene metabolite in human urine by synchronous fluorescence spectroscopy and gas chromatography-mass spectrometry. Carcinogenesis. 1994;15:483–7.

23. Campo L, Fustinoni S, Buratti M, Cirla PE, Martinotti I, Foà V. Unmetabolized polycyclic aromatic hydrocarbons in urine as biomarkers of low exposure in asphalt workers. J Occup Environ Hyg. 2007;4:100-10.

24. Ramsauer B, Sterz K, Hagedorn H-W, Engl J, Scherer G, McEwan M, et al. A liquid chromatography/tandem mass spectrometry (LC-MS/MS) method for the determination of phenolic polycyclic aromatic hydrocarbons (OH-PAH) in urine of non-smokers and smokers. Anal Bioanal Chem. 2011;399:877–89.

25. Raponi F, Bauleo L, Ancona C, Forastiere F, Paci E, Pigini D, et al. Quantification of 1-hydroxypyrene, 1- and 2-hydroxynaphthalene, 3-hydroxybenzo[a]pyrene and 6-hydroxynitropyrene by HPLC-MS/MS in human urine as exposure biomarkers for environmental and occupational surveys. Biomarkers. Taylor & Francis; 2017;22:575–83.

26. Campo L, Rossella F, Fustinoni S. Development of a gas chromatography/mass spectrometry method to quantify several urinary monohydroxy metabolites of polycyclic aromatic hydrocarbons in occupationally exposed subjects. J Chromatogr B Anal Technol Biomed Life Sci. 2008;875:531–40.

27. Rossella F, Campo L, Pavanello S, Kapka L, Siwinska E, Fustinoni S. Urinary polycyclic aromatic hydrocarbons and monohydroxy metabolites as biomarkers of exposure in coke oven workers. Occup Environ Med. 2009;66:509–16.

28. Poster DL, Schantz MM, Sander LC, Wise SA. Analysis of polycyclic aromatic hydrocarbons (PAHs) in environmental samples: A critical review of gas chromatographic (GC) methods. Anal. Bioanal. Chem. 2006;386:859-81.

29. Rylance J, Gordon SB, Naeher LP, Patel A, Balmes JR, Adetona O, et al. Household air pollution: a call for studies into biomarkers of exposure and predictors of respiratory disease. AJP Lung Cell Mol Physiol. 2013;304:L571-8.

30. Chong J, Soufan O, Li C, Caraus I, Li S, Bourque G, et al. MetaboAnalyst 4.0: Towards more transparent and integrative metabolomics analysis. Nucleic Acids Res. 2018;46:W486-94.

31. Jongeneelen FJ. Benchmark guideline for urinary 1-hydroxypyrene as biomarker of occupational exposure to polycyclic aromatic hydrocarbons. Ann Occup Hyg. 2001;45:3–13.

32. Ram B. Jain. Trends and concentrations of selected polycyclic aromatic hydrocarbons in general US population: Data from NHANES 2003-2008. Cogent Environ Sci. 2015;1:1031508.

33. Ciarrocca M, Rosati MV, Tomei F, Capozzella A, Andreozzi G, Tomei G, et al. Is urinary 1-hydroxypyrene a valid biomarker for exposure to air pollution in outdoor workers? A meta-analysis. J Expo Sci Environ Epidemiol. 2014;24:17–26.

34. Li Z, Romanoff L, Bartell S, Pittman EN, Trinidad DA, McClean M, et al. Excretion Profiles and half-lives of ten urinary polycyclic aromatic hydrocarbon metabolites after dietary exposure. Chem Res Toxicol. 2012;25:1452-61.

35. Bristow T, Harrison M, Sims M. The application of gas chromatography/atmospheric pressure chemical ionisation time-of-flight mass spectrometry to impurity identification in pharmaceutical development. Rapid Commun Mass Spectrom. 2010;24:1673–81.

36. Wang S, Cyronak M, Yang E. Does a stable isotopically labeled internal standard always correct analyte response?: A matrix effect study on a LC/MS/MS method for the determination of carvedilol enantiomers in human plasma. J Pharm Biomed Anal. 2007;43:701–7.

37. Li Z, Romanoff LC, Trinidad DA, Hussain N, Jones RS, Porter EN, et al. Measurement of urinary monohydroxy polycyclic aromatic hydrocarbons using automated liquid-liquid extraction and gas chromatography/isotope dilution high-resolution mass spectrometry. Anal Chem. 2006;78:5744–51.

38. Xu X, Zhang J, Zhang L, Liu W, Weisel CP. Selective detection of monohydroxy metabolites of polycyclic aromatic hydrocarbons in urine using liquid chromatography/triple quadrupole tandem mass spectrometry. Rapid Commun Mass Spectrom. John Wiley & Sons, Ltd; 2004;18:2299–308.

39. Jacob P, Wilson M, Benowitz NL. Determination of phenolic metabolites of polycyclic aromatic hydrocarbons in human urine as their pentafluorobenzyl ether derivatives using liquid chromatography-tandem mass spectrometry. Anal Chem. 2007;79:587–98.

40. Macedo A, Macri J, Hudecki P, Saoi M, McQueen MJ, Britz-McKibbin P. Validation of a

capillary electrophoresis assay for monitoring iodine nutrition in populations for prevention of iodine deficiency: An interlaboratory method comparison. J Appl Lab Med. 2017;3:649-60.

41. Alhamdow A, Lindh C, Albin M, Gustavsson P, Tinnerberg H, Broberg K. Early markers of cardiovascular disease are associated with occupational exposure to polycyclic aromatic hydrocarbons. Sci Rep. 2017;7:9426-37.

42. Strickland P, Kang D, Sithisarankul P. Polycyclic aromatic hydrocarbon metabolites in urine as biomarkers of exposure and effect. Environ Health Perspect. 1996;104:927-32.

43. Morgan MS. The biological exposure indices: A key component in protecting workers from toxic chemicals. Environ. Health Perspect. 1997;105:105-15.

44. Mastrianni KR, Andrew Lee L, Brewer WE, Dongari N, Barna M, Morgan SL. Variations in enzymatic hydrolysis efficiencies for amitriptyline and cyclobenzaprine in urine. J Anal Toxicol. 2016;40:732-7.

45. Dwivedi P, Zhou X, Powell TG, Calafat AM, Ye X. Impact of enzymatic hydrolysis on the quantification of total urinary concentrations of chemical biomarkers. Chemosphere. 2018;199:256-62.

46. Alshaarawy O, Elbaz HA, Andrew ME. The association of urinary polycyclic aromatic hydrocarbon biomarkers and cardiovascular disease in the US population. Environ Int. 2016;89:174-8.

47. Cakmak S, Hebbern C, Cakmak JD, Dales RE. The influence of polycyclic aromatic hydrocarbons on lung function in a representative sample of the Canadian population. Environ Pollut. 2017;228:1-7.

48. Barr DB, Wang RY, Needham LL. Biologic monitoring of exposure to environmental chemicals throughout the life stages: Requirements and issues for consideration for the National Children's Study. Environ Health Perspect. 2005;113:1083-91.

49. Heudorf U, Angerer J. Urinary monohydroxylated phenanthrenes and hydroxypyrene - The effects of smoking habits and changes induced by smoking on monooxygenase-mediated metabolism. Int Arch Occup Environ Health. 2001;74:177–83.

50. Li Z, Romanoff LC, Lewin MD, Porter EN, Trinidad DA, Needham LL, et al. Variability of urinary concentrations of polycyclic aromatic hydrocarbon metabolite in general population and comparison of spot, first-morning, and 24-h void sampling. J Expo Sci Environ Epidemiol. 2010;20:526-35.

51. Jongeneelen FJ, Leeuwen FE Van, Oosterink S, Anzion RBM. Ambient and biological monitoring of cokeoven workers : determinants of the internal dose of polycyclic aromatic hydrocarbons. Br J Ind Med. 1990;47:454-61.

52. Adetona O, Simpson CD, Li Z, Sjodin A, Calafat AM, Naeher LP. Hydroxylated polycyclic aromatic hydrocarbons as biomarkers of exposure to wood smoke in wildland firefighters. J Expo Sci Environ Epidemiol. 2017;27:78–83.

2.7 Supporting Information

2.7.1 Chemicals and Reagents.

Reagents used for HPLC analysis were ≥ 98.5% purity, acetonitrile, L-Ascorbic acid (> 99%), Surine® negative urine, β-glucuronidase (from *Helix pomatia*, Type HP-2, aqueous solution, > 100,000 units/mL) and Methyl-N-(trimethylsilyl) trifluoroacetamide (MSTFA, > 98.5%) were purchased from Sigma Aldrich (Mississauga, ON, Canada). Dichloromethane (distilled in glass) was purchased from Caledon (Canada). OH-PAH standards, including 1-hydroxypyrene (1-OH-Pyr), isotope labelled 1-hydroxypyrene ($^{13}C_6$ 1-OH-Pyr, as recovery standard), 1-OH-Pyr glucuronide were obtained from Toronto Research Chemicals Inc. (Toronto, ON, Canada), whereas pyrene-d_{10} (as internal standard) was supplied by Cambridge Isotope Laboratories (Tewksbury, MA, USA). All stock solutions and serial dilutions were prepared in acetonitrile.

2.7.2 Sample Preparation

Sample preparation prior to GC-MS analysis involved an enzyme deconjugation with multiple time points over 18 hours. Three 200 µL aliquots of Surine® negative urine containing 10 µg/mL 1-OH-Pyr-glucuronide were prepared with 1 µL of ascorbic acid (1.5 M), 40 µL of 1.0 M acetate pH 5.2 buffer and 2 µL of β-glucuronidase/ arylsulfatase with the addition of 20 µL of recovery standard ($^{13}C_6$-1-OH-Pyr, 10 µg/mL). The mixture was subsequently incubated *at 60*°C for 18 hours. At time points *0,* 0.5, 1, 2, 4, 8, 12, 18 hours, the reaction was stopped with the addition of 300 µL of ice-cold dichloromethane (DCM), followed by centrifugation at 2750 g for 5 min at 4°*C*; after which samples were stored on ice and extracted by shaking for 30 min. A 10 µL aliquot of the DCM extract was collected in a GC-vial insert to which 1 µL of internal standard (Pyrene-d_{10}, 10 µg/mL), was spiked into the sample and derivatized with 15 µL of MSTFA following an incubation at 60°*C* for 20 min

2.7.3 Instrument Conditions

Samples were analyzed by gas chromatography mass spectrometry (GC-MS) using an Agilent 6890 gas chromatograph equipped with a DB-17 column (30 m x 0.25 mm x 0.15 μm film) coupled to an Agilent 5973 single quadrupole mass spectrometer with electron impact ionization (EI-MS) operating in selected ion monitoring (SIM) mode. Helium was used as the carrier gas with an optimal flow set to 1 mL/min. An injection volume of 1 μL was used with a splitless injector set to 280°*C.* The initial oven temperature was set to 70°C and held for 1 min, and then ramped at a rate of 20°C/min to 300°C and held for 5 min. Trimethylsilylated (TMS)-1-OH-Pyr and its isotope labelled $^{13}C_6$-1-OH-Pyr analog were monitored at *m/z* 290 and 296 respectively. Data processing was conducted using ChemStation software, version D.03.00 (Agilent Technologies, USA) with quantitation based on relative ion response using the isotope labelled 1-OH-Pyr as the recovery standard. All were depicted using Igor Pro 5.0 software (Wavemetric Inc., Lake Oswego, OR, USA).

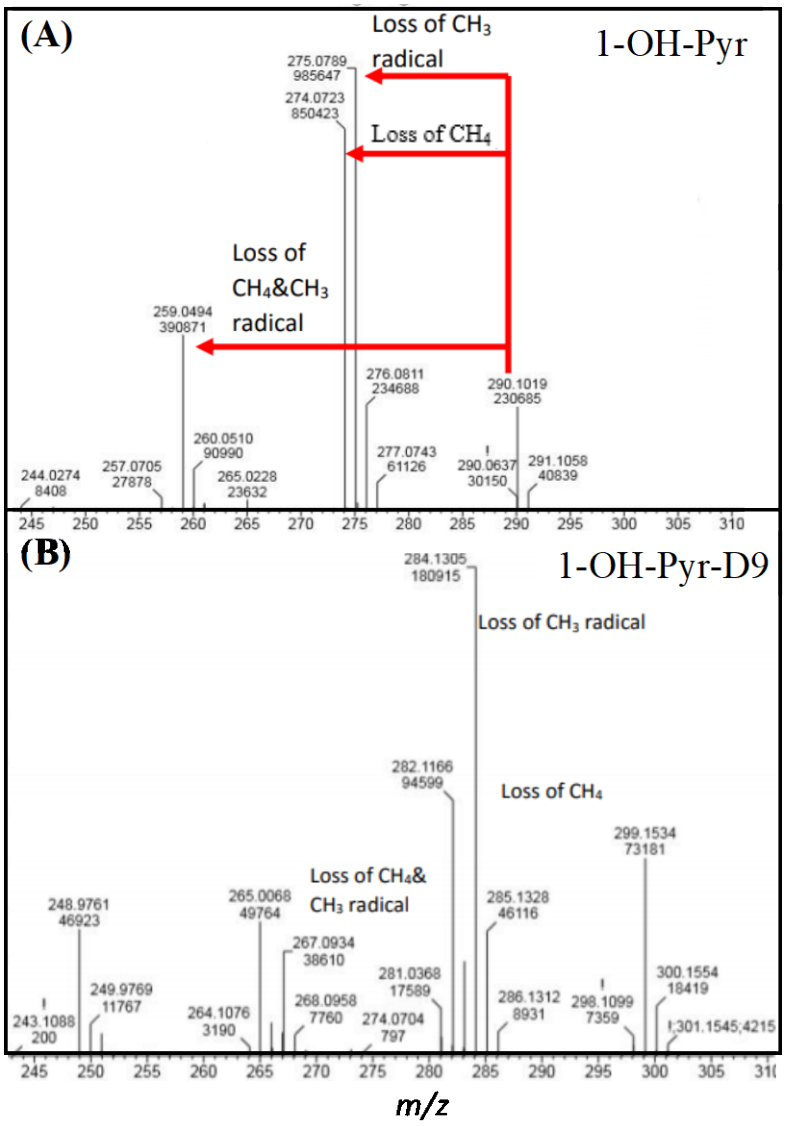


Figure S2.1. (A) Mass spectrum of 1-OH-Pyr as its TMS derivative with the molecular ion indicated at a *m/z* of 290.1019 along with the ^{13}C isotope at *m/z* 291.1058 when using GC-HRMS with APCI (positive ion mode). The loss of a methyl radical from TMS, and the loss of neutral methane are demonstrated by the fragments ions, *m/z* 275.0789 and *m/z* 274.0730, respectively. **(B)** Similarly, the mass spectrum of 1-OH-Pyr-d9 as its TMS derivative shows an analogous fragmentation pattern with a mass shift of 9 Da.

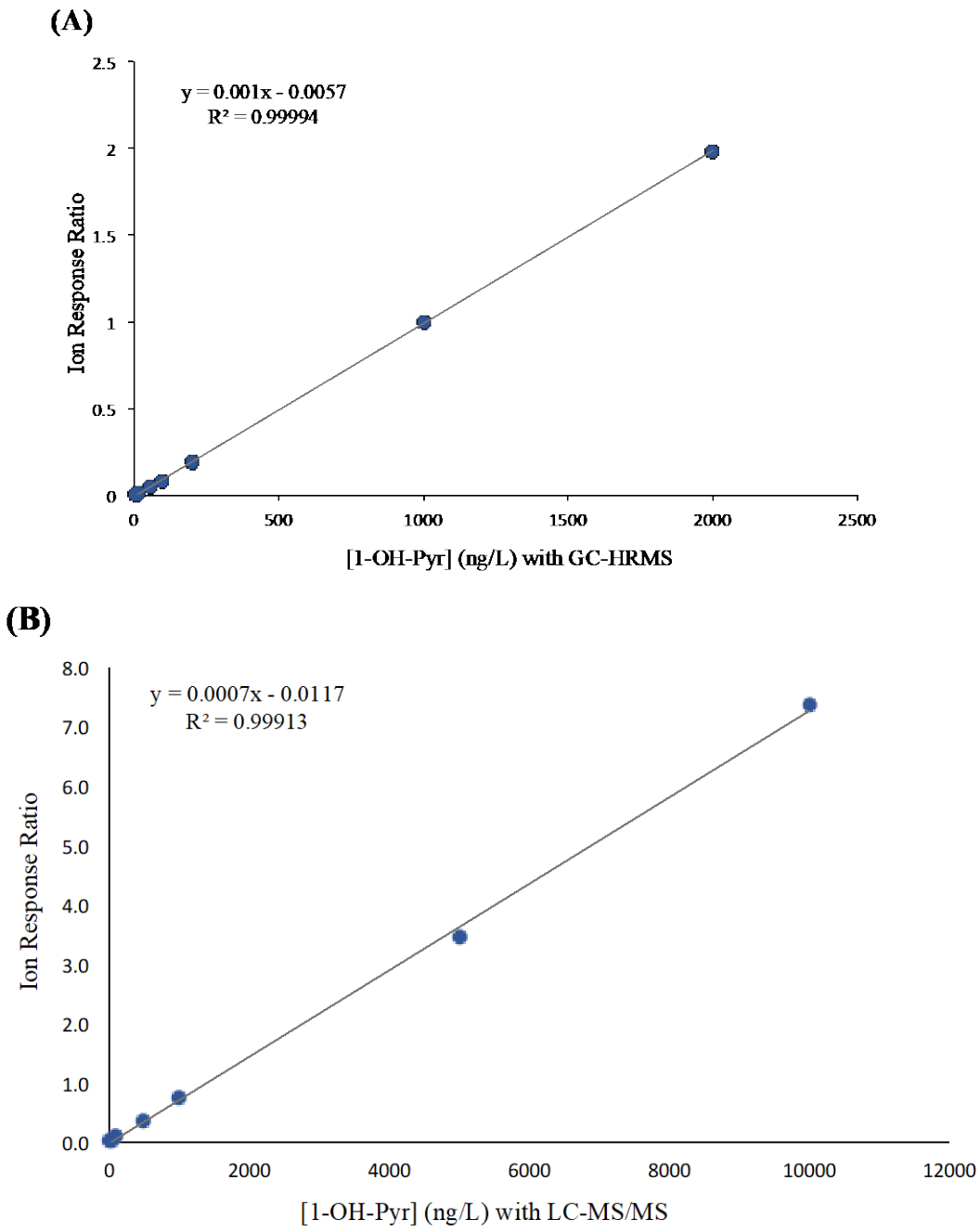


Figure S2.2. External calibration curves for 1-OH-Pyr when using **(A)** GC-HRMS (that includes pre-column SPE in the protocol for a 25-fold enrichment of urine and chemical derivatization), and **(B)** LC-MS/MS (direct analysis after enzyme hydrolysis), where ion responses were normalized to 1-OH-Pyr-d9 for all calibrators prepared in toluene. Both methods demonstrate a wide linear dynamic range over three orders of magnitude with good linearity (R^2 = 0.999). In all cases, enzyme deconjugation was performed for all urine specimens prior to analysis by GC-HRMS and LC-MS/MS for determination of total 1-OH-Pyr.

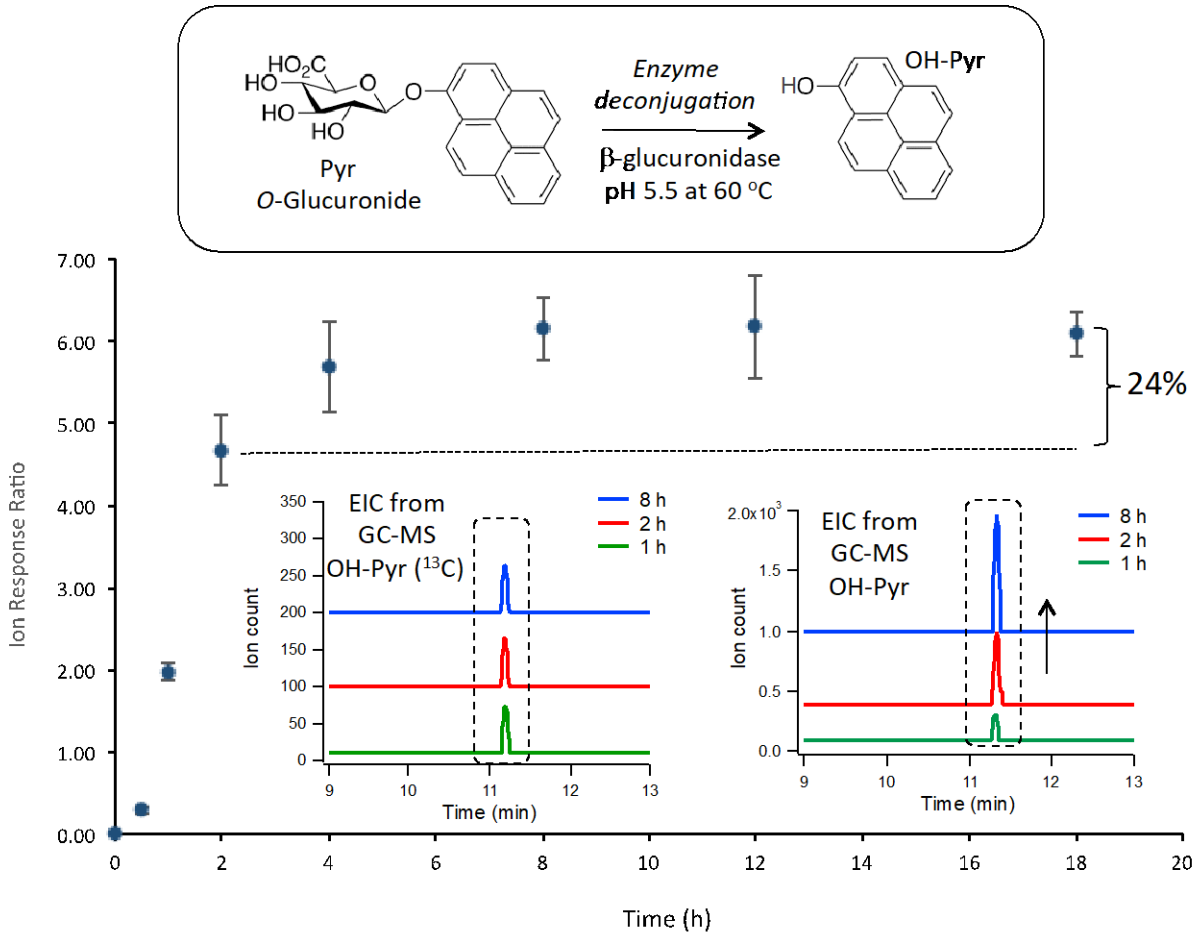


Figure S2.3. The enzymatic reaction under LC-MS/MS conditions using β-glucuronidase (from *Helix pomatia*, Type HP-2, aqueous solution, > 100,000 units/mL) is depicted with 1-OH-Pyr spiked Surine® negative urine (10 µg/mL). A time course of 18 hours at a temperature of 60°C with multiple time points (0, 1, 1.5, 2, 4, 8, 12, 18 hours) was used, where the reaction was stopped with the addition of ice-cold dichloromethane (DCM), followed by centrifugation at 4°*C* at 2750 g for 5 min prior to liquid-liquid extraction with DCM (1:1 ratio). The product, 1-OH-Pyr, ion response ratio normalized to $^{13}C_6$-1-OH-Pyr, as a function of time is indicated with a 24% increase observed between the 2 hour and 18 hour time points with triplicate measurements (CV = 8.2%). Error bars indicating one standard deviation are highlighted for all time points. Extracted ion chromatograms (EICs) demonstrated in the green trace indicates the 1 hour enzyme reaction time point, followed by the 2 hour reaction product in red and 8 hour reaction product indicated in the blue trace following which a plateau is observed. The recovery standard $^{13}C_6$-1-OH-Pyr trace indicated in brown and the internal standard (Pyr-d_{10}) spiked prior to injection depicted in the black trace. A clear increase in product is observed as a result of increase incubation time and likely contributes to the 39% bias observed by the GC-HRMS method.

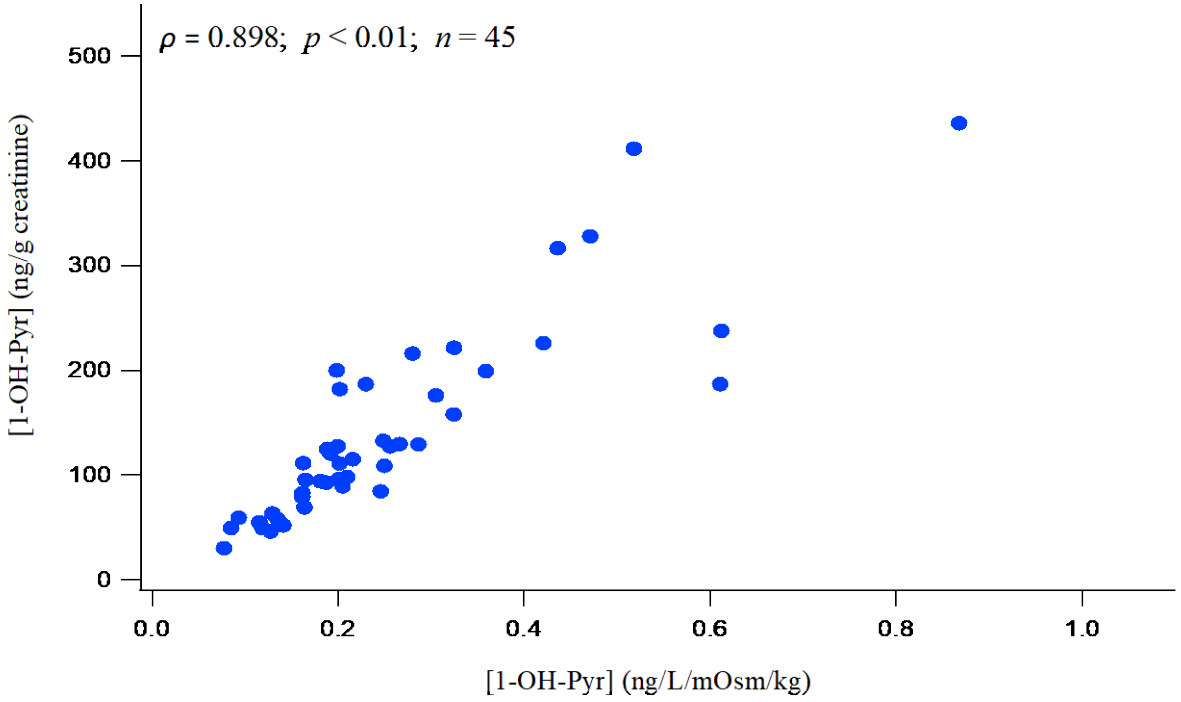


Figure S2.4. Creatinine normalized 1-OH-Pyr concentrations was used for correction of hydration status when evaluating smoke exposure from firefighters based on collection of random/single-spot urine specimens after deployment to the Fort McMurray wildfire. Also, osmolality normalized urine concentrations for 1-OH-Pyr was also measured for the same subset of 45 urine samples. A Spearman rank correlation plot between measured creatinine and osmolality adjusted 1-OH-Pyr concentrations demonstrate a strong positive correlation (ρ = 0.898, p < 0.001) between these two orthogonal normalization methods.

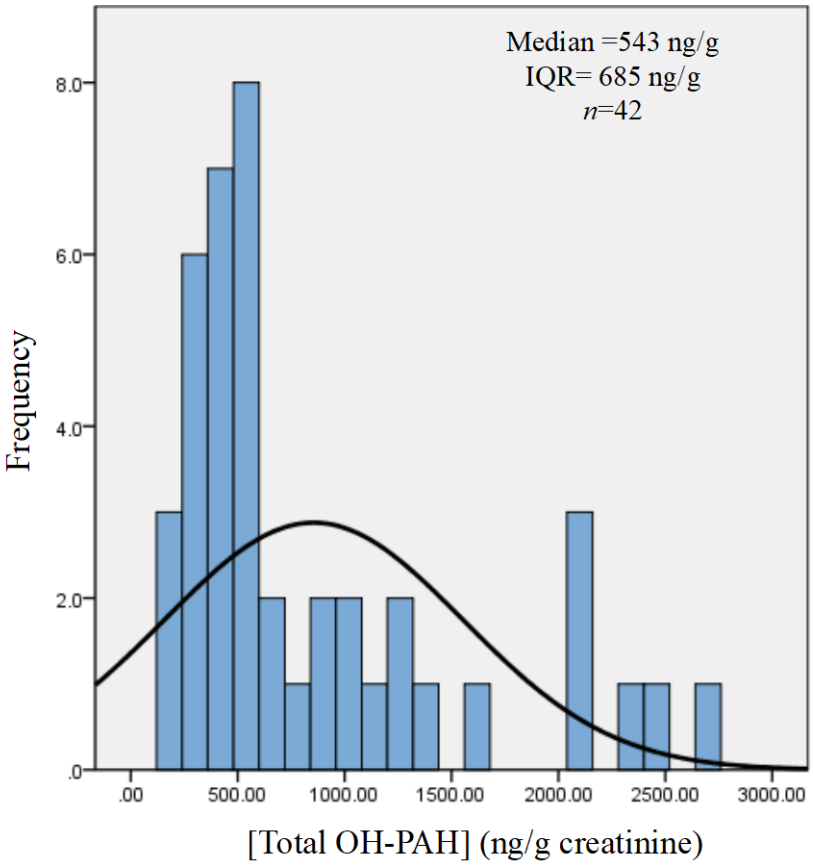


Figure S2.5. Histogram depicting the urinary excretion profile based on total excretion of the three major OH-PAH metabolites strong correlated with each other, including 1-OH-Pyr, 3-OH-Phe and 9-OH-Flu. This distribution is displayed for 42 firefighters (excluding repeat urine samples) with total concentrations lower than anticipated as a result of delays in urine collection following their return from deployment (average 6 days) after the second week of the Fort McMurray wildfire that does not accurately reflect their occupational exposure during early stages of intense firefighting.

Table S2.1- Table of values of urinary 1-OH-Pyr as a marker for exposure when using GC-HRMS and LC-MS/MS platforms.

Sample Number	LC-MS/MS (ng/L)	GC-HRMS (ng/L)
1	112	203
2	199	222
3	77.1	210
4	n.d.	83.7
5	92.3	93.3
6	92.0	173
7	110	225
8	47.4	59.1
9	121	315
10	n.d.	67.6
11	n.d.	62.1
12	178	189
13	n.d.	43.6
14	n.d.	72.3
15	84.4	150
16	n.d.	48.7
17	n.d.	49.1
18	108	96.5
19	54.9	86.9
20	84.1	127
21	66.1	148
22	244	294
23	92.1	124
24	345	403
25	91.4	129
26	99.3	187
27	557	597
28	88.5	115
29	156	251
30	69.5	161
31	n.d.	53.1
32	66.0	76.9
33	207	486
34	170	159
35	249	227
36	47.1	104
37	124	156
38	84.6	50.9
39	126	156
40	n.d.	114

Sample Number	LC-MS/MS (ng/L)	GC-HRMS (ng/L)
41	n.d.	78.6
42	149	199
43	62.6	86.2
44	n.d.	45.2
45	n.d.	30.2
46	n.d.	72.3
47	n.d.	n.d.
48	83.8	138
49	56.7	221
50	51.3	274
51	n.d.	62.6
52	45.8	47.8
53	838	1040
54	78.2	191
55	151	332
56	179	295
57	n.d.	77.7
58	84.7	203
59	n.d.	57.3
60	295	234
61	293	275
62	n.d.	136

n.d. indicates non-detects determined by the LOQ.

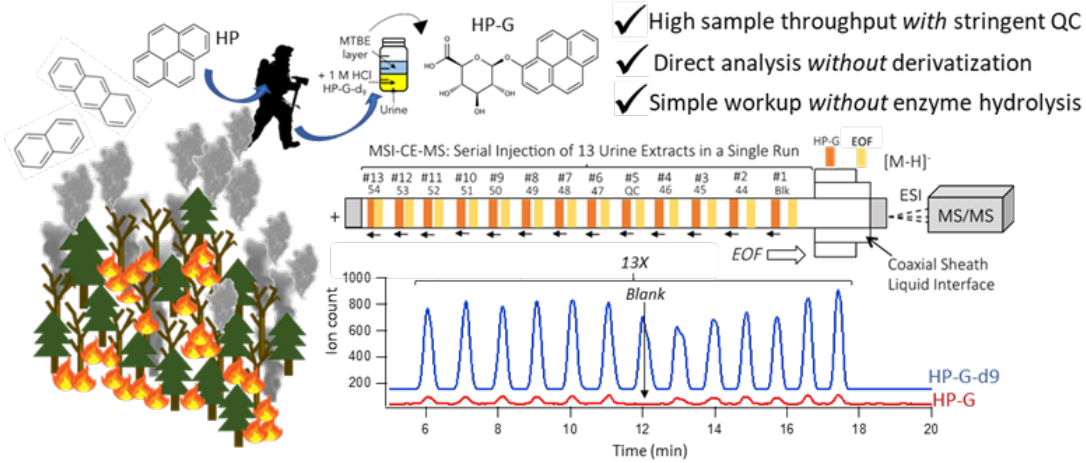
Chapter III:

Rapid Screening of Urinary 1-Hydroxypyrene Glucuronide by Multisegment Injection-Capillary Electrophoresis-Tandem Mass Spectrometry: A High Throughput Method for Biomonitoring of Recent Smoke Exposures

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

Biban Gill,[1] Karl Jobst,[2] Philip Britz-McKibbin[1] Rapid Screening of Urinary 1-Hydroxypyrene Glucuronide by Multisegment Injection–Capillary Electrophoresis–Tandem Mass Spectrometry: A High-Throughput Method for Biomonitoring of Recent Smoke Exposures. *Anal. Chem.* **2020**, 92 (19), 13558-13564.

[1]*Department of Chemistry and Chemical Biology, McMaster University, Hamilton, Ontario, L8S 4M1, Canada.*
[2]*Department of Chemistry, Memorial University of Newfoundland, St. John's, Newfoundland and Labrador, A1B 3X7, Canada.*



B.G. conducted all method optimization and validation experiments by MSI-CE-MS/MS. B.G. performed the statistical analysis and wrote the initial draft for publication. K.J. provided the GC-HRMS data as well as valuable feedback and revisions for the manuscript draft. P.B.M. provided critical feedback for experiments and data analysis, as well as final revisions of the manuscript for publication.

Chapter III: Rapid Screening of Urinary 1-Hydroxypyrene Glucuronide by Multisegment Injection-Capillary Electrophoresis-Tandem Mass Spectrometry: A High Throughput Method for Biomonitoring of Recent Smoke Exposures

3.1 Abstract

Urinary 1-hydroxypyrene (HP) is a widely used biomarker of polycyclic aromatic hydrocarbon exposures relevant for biomonitoring the deleterious health impacts from tobacco smoke, ambient air pollution, and the hazards of certain occupations. Conventional methods for urinary HP analysis based on liquid chromatography with native fluorescence detection or tandem mass spectrometry (MS/MS), as well as gas chromatography-mass spectrometry (GC-MS) are limited by low sample throughput and complicated sample workup protocols that are prone to bias. Herein, we introduce a high throughput method to directly analyze the intact glucuronide conjugate of HP (HP-G) in human urine after a simple acidified ether extraction procedure when using multisegment injection-capillary electrophoresis-tandem mass spectrometry (MSI-CE-MS/MS). Multiplexed analysis of thirteen independent urine extracts was achieved in a single run (< 3 min/sample) with stringent quality control while avoiding enzyme deconjugation and pre-column chemical derivatization. Method validation demonstrated good technical precision (CV = 7.7%, n = 45) and accuracy with a mean recovery of (93 ± 3%) for urinary HP-G at three concentration levels with adequate detection limits (7 ng/L, S/N = 3). An inter-laboratory method comparison of urine samples collected from firefighters deployed in the 2016 Fort McMurray wildfire also confirmed good mutual agreement with an acceptable negative bias (mean bias = 15%, n = 55) when measuring urinary HP-G by MSI-CE-MS/MS as compared to total hydrolyzed urinary HP by GC-MS due to low residual levels of free HP and its sulfate conjugate. This multiplexed separation platform is optimal for large-scale biomonitoring studies of air pollution relevant to global health, as well as occupational smoke exposures in firefighters susceptible to dermal PAH absorption when using personal protective equipment.

3.2 Introduction

Polycyclic aromatic hydrocarbons (PAHs) are a complex group of environmental toxicants formed by incomplete combustion of organic materials, some of which possess carcinogenic, mutagenic, inflammatory and endocrine disrupting properties.[1] The ubiquitous nature of PAHs in the environment contributes to their deleterious impacts on global health from chronic exposures to urban air pollution, indoor smoke and/or tobacco.[2,3] However, occupational exposure to PAHs is most pronounced among industrial workers and notably firefighters, corresponding to their greater risk for cardiorespiratory diseases[4,5] and certain cancers[6,7] relative to the general population. Although inhalation and ingestion are common mechanisms of PAH absorption, whole body dermal exposure is predominant in firefighters using personal protective equipment, including a self-contained breathing apparatus.[8-11] Monohydroxylated PAH (OH-PAH) metabolites and their isomers derived from naphthalene, fluorene, phenanthrene and pyrene, are widely measured as biomarkers of recent PAH exposure given the convenience of non-invasive urine sampling.[12,13] Overall, urinary 1-hydroxypyrene (HP) is frequently reported in human biomonitoring studies given that pyrene is abundant in most smoke mixtures, and its biotransformation results in formation of a single isomer with an average urinary half-life of about 18 h.[14] Ciarocca *et al.*[15] performed a meta-analysis confirming urinary HP as a sensitive yet specific biomarker of air pollution after considering other potential sources and confounders, including environmental, genetic and lifestyle factors. For these reasons, the American Conference of Governmental Hygienists recommends HP as the biological exposure index for human exposure to PAHs,[16] and occupational guidelines propose a minimum no-observed genotoxic effect for HP of 1000 nmol/mol creatinine in end-of-shift urine samples.[17]

The biotransformation of pyrene is mediated by specific monooxygenase cytochrome P450 enzyme isoforms resulting in the formation of HP in the liver.[18] This is followed by phase II metabolism generating a water-soluble glucuronide conjugate of HP (HP-G) that comprises the majority (> 85%) of total pyrene excreted in human urine,[19] with about 10% attributed to its sulfate conjugate and less than 1% as free HP.[20] Conventional methods for urinary HP determination includes immunoaffinity chromatography with synchronous fluorescence spectroscopy,[21,22] liquid chromatography (LC) with native fluorescence detection (LC-FLD),[23] and increasingly LC with tandem mass spectrometry (LC-MS/MS)[24] as it offers better selectivity with fewer chemical interferences.[25] However, stable isotope dilution GC-MS/MS remains the gold standard since it is a robust separation platform used in large-scale population health studies,[12,26] while GC with high resolution mass spectrometry (GC-HRMS) allows for more comprehensive surveillance of urinary OH-PAHs and other environmental contaminants.[27] Most methods to date have been constrained by low sample throughput, high operating costs, and complicated sample workup protocols, including solid-phase extraction (SPE), enzyme deconjugation and/or pre-column chemical derivatization.[28] Importantly, enzyme hydrolysis can contribute to greater inter-batch variations with different substrate efficiencies and unexpected by-products depending on the purity and activity of the enzyme source.[29]

Herein, we introduce multisegment injection-capillary electrophoresis-tandem mass spectrometry (MSI-CE-MS/MS) as a high throughput platform for direct analysis of HP-G following a simple acidified ether extraction step of human urine. MSI-CE-MS/MS takes advantage of a serial sample injection format with stringent quality control (QC) that is ideal for the analysis of mass-limited tissue[30] and volume-restricted bio-banked specimens.[31] Our recent inter-laboratory method comparison for urinary HP determination by LC-MS/MS and GC-

HRMS revealed that enzyme hydrolysis is a major source of bias when using suboptimal incubation conditions,[28] which is relevant when assessing other urinary exposure biomarkers excreted mainly as their phase II conjugates.[29] Urinary OH-PAH conjugates are also more chemically stable when refrigerated than their corresponding free OH-PAHs while being resistant to multiple freeze-thaw cycles thereby reducing pre-analytical variances.[32] This motivated us to develop a rapid yet direct method for urinary HP-G determination that forgoes the need for time-consuming sample workup procedures and slow data acquisition when relying on single sample injection separation methods.

3.3 Experimental Section

3.3.1 Chemicals and Reagents

HPLC-grade methanol and water, were used to prepare sheath liquid and background electrolyte (BGE) solutions, respectively (Caledon Inc., Georgetown, ON, Canada). HPLC grade toluene, and acetonitrile were used for stock solutions and were purchased from Caledon. Methyl-*tert*-butyl-ether (MTBE) was purchased from Caledon for urine liquid extraction protocol using 1.0 M hydrochloric acid (HCl) for acidification. Ammonium bicarbonate and ammonium hydroxide were purchased from Sigma-Aldrich Inc. (Oakville, ON, Canada), whereas standards for HP-G and HP-G-d_9 were obtained from Toronto Research Chemicals (Toronto, ON, Canada). All stock solutions were prepared by dissolving in toluene, followed by serial dilutions of calibrant solutions in acetonitrile and subsequently HPLC grade water.

3.3.2 Instrumental Method for Direct Determination of Urinary HP-G by MSI-CE-MS/MS

An Agilent 6370 triple quadrupole mass spectrometer with a coaxial sheath liquid electrospray (ESI) ionization source coupled to an Agilent G7100A CE unit was used for all experiments

(Agilent Technologies Inc., Mississauga, ON, Canada). An Agilent 1260 Infinity isocratic pump and 1260 Infinity degasser were used to deliver 70:30 MeOH:water at a flow rate of 10 µL/min using a CE-MS coaxial sheath liquid interface. The nebulizer spray was turned off during serial sample injection to avoid suctioning effects and current instabilities.[33] The spray was then turned on at 4 psi upon initiation of CE separation with the source temperature set to 300 °C, and the drying gas delivered at 4.0 L/min. The instrument was operated under negative ion mode detection with a Vcap set at 3500 V while MS/MS was acquired via multiple reaction monitoring (MRM). Quantifier and qualifier transitions for the detection of HP-G were 393/217 and 393/198, whereas HP-G-d_9 used 402/226 and 402/198, respectively. Both sets of transitions required a collision energy (CE) of -60 V and -80 V respectively, a fragmentor voltage of 120 V and a delta electron multiplier voltage (EMV) of -400 V as summarized in Table S1 of the Supporting Information. CE separations used a 50 µm internal diameter fused-silica capillary with an outer diameter of 360 µm and a total length of 135 cm (Polymicro Technologies Inc., AZ). A capillary window maker (microSolv, Leland, NC) was used to remove 7.0 mm of polyimide coating on both ends of the capillary since it is prone to swelling with organic solvent and aminolysis with strongly alkaline (pH > 9) solutions containing ammonia.[34] A 50 mM ammonium bicarbonate buffer, pH 8.5 was used as an alkaline BGE that was adjusted with ammonium hydroxide. A serial sample injection sequence was performed by MSI-CE-MS/MS with 13 urine extracts introduced hydrodynamically for 20 s at 100 mbar that alternated with a transient electrokinetic separation using the BGE at 30 kV for 45 s. The process was repeated to introduce all 13 urine extracts in series within the capillary (temperature set to 25 °C) once applying a constant voltage of 30 kV with a pressure gradient of 2.0 mbar/min to reduce total analysis times.[35]

3.3.3 Urine Collection and Sample Workup for Urinary HP-G Determination from Firefighters

Single spot mid-stream urine samples (n = 62) were collected and aliquoted prior to their storage at -80°C from a cohort of firefighters (fire station A) who were deployed to the 2016 Fort McMurray wildfire.[11] This study design was approved by the Health Ethics Review Board of the University of Alberta (Pro00065284). Aliquots of these urine samples were originally used to develop and validate two complementary analytical procedures for urinary HP determination based on LC-MS/MS and GC-HRMS following enzyme deconjugation.[28] In this work, a single-step liquid extraction procedure was optimized for direct analysis of urinary HP-G using a modified MTBE extraction protocol[36] for the rapid determination of serum fatty acids under nonaqueous buffer conditions.[33] Overall, 42 of the 62 urine samples were collected from individual male firefighters (mean age = 36 years; mean BMI = 29 kg/m^2) following their deployment; however, there were unexpected logistic delays in urine collection (median = 4.5 days) ranging from one day (~ 36% of samples collected < 24 h) to up to 13 days.[11] The remaining urine samples were repeat samples collected from a subset of the same firefighters (n = 20) at a later time interval as a control, which did not show differences in recent PAH exposures.[11,28] All 62 urine samples were thawed on ice and then centrifuged to remove particulates. A 0.5 mL aliquot was then transferred for extraction and enrichment in a pre-washed 2.0 mL glass GC vial. Next, 1.0 µL of 100 ng/mL HP-G-d$_9$ recovery standard and 125 µL of 1.0 M HCl were added to urine prior to vortex mixing for 5 min. Following urine acidification (pH ~ 2-3), 1.0 mL of MTBE was then added and mixed by vortex for 20 min at room temperature. The upper MTBE layer was transferred in several 100 µL aliquots to 100 µL GC vial inserts to improve overall recovery, and then dried under a gentle stream of nitrogen.

The urine extracts were pooled together and then reconstituted to a final volume of 10 µL with 50 mM ammonium bicarbonate buffer, pH 8.5 as the alkaline BGE that corresponded to an overall 50-fold sample enrichment from 0.5 mL urine. Urine extracts were stored at -80 °C prior to MSI-CE-MS/MS analysis. Sample preparation for total urinary HP determination using GC-HRMS with atmospheric pressure chemical ionization[28] included an overnight enzyme hydrolysis, solid-phase extraction (SPE), and pre-column chemical derivatization. Further details on urine sample workup procedures and GC-HRMS instrumental configuration are described in the Experimental of the Supporting Information.

3.3.4 Method Validation, Quality Control and Interlaboratory Method Comparison

Stock solutions for calibrants were initially prepared in toluene and then diluted in acetonitrile (1000 mg/L) followed by serial dilution in HPLC grade water (10 mg/L), which were used to prepare calibrant solutions ranging from 20 ng/L to 2000 ng/L in triplicate. A pooled urine sample derived from all firefighters was also used as a quality control (QC) specimen. All integrated peak areas for the deprotonated molecular ion $[M\text{-}H]^-$ for quantifier and qualifier ions for HP-G were normalized to a HP-G-d_9 that was added to all calibrant solutions (10 µg/L). The limit of detection and limit of quantification for HP-G was determined based on serial dilutions of calibrant solutions that generated a signal-to-noise ratio of 3 and 10, respectively. Technical blank extracts (water) were also analyzed to ensure a lack of sample carry-over effects when using multiplexed separations in MSI-CE-MS/MS. Reproducibility was assessed based on intra-day precision (n = 45) and inter-day (n = 108) using spiked standard and quality control (QC) runs conducted at the start, middle and end of each day performed over the course of three days. Method accuracy for HP-G determination by MSI-CE-MS/MS was evaluated based on spike-

recovery studies at three different concentrations (40 ng/L, 200 ng/L, 20 µg/L) that were spiked into pooled human urine samples collected from non-smoking volunteers without occupational smoke exposure. The percent recovery for HP-G determination was calculated based on percentage difference between spiked and known concentration divided by known spiked concentration. Furthermore, an inter-laboratory method comparison was also conducted to evaluate method accuracy relative to a previously validated GC-HRMS method using matching urine samples (n = 62) from firefighters.[28] All urinary HP-G and HP concentrations were reported as their normalized concentrations values to creatinine (ng/g creatinine) to correct for hydration status when collecting random single-spot urine samples from firefighters post-deployment.

3.3.5 Data Processing and Statistical Analysis

All data acquired by MSI-CE-MS/MS was analyzed with Agilent Mass Hunter Workstation Software (Qualitative Analysis, version B.06.00, Agilent Technologies Inc.). HP-G and HP-G-d_9 MRM transitions were extracted in profile mode using 10 ppm mass window. Extracted ion electropherograms (EIEs) were integrated after smoothing using the quadratic/cubic Savitzky-Golay function (7 points) while peak areas, migration times and signal to noise were transferred to Excel (Microsoft office, Edmond, WA, USA) for calculation of relative peak areas (RPAs) that corrects for differences in injection volume between samples, as well as matrix-induced ion suppression/enhancement effects when using a co-migrating deuterated internal standard (HP-G-d_9) that is introduced to all calibrants and acidified urine samples prior to MTBE extraction. All electropherograms were depicted using Igor Pro 5.0 software (Wavemetric Inc., Lake Oswego, OR, USA). Analysis for external calibration data, calculation of figures of merit, and control

charts for HP-G measured in QC samples were performed using Excel. Bland-Altman % difference plots and Passing-Bablok regression were performed using MedCalc version 12.5 (Ostend, Belgium) for the inter-laboratory method comparison of urinary HP-G and total (hydrolyzed) HP measured by MSI-CE-MS/MS and GC-HRMS, respectively.

3.4 Results and Discussion

3.4.1 High Throughput Biomonitoring of Urinary HP-G by MSI-CE-MS/MS

In this work, we implemented a serial sample injection program in MSI-CE-MS/MS that enables the analysis of 13 urine extracts within an effective duty cycle ~ 2.7 min/sample, including a 20 min analytical run following a 10 min automated injection program and 5 min capillary flush/equilibration time. This is achieved when an alternating series of samples are introduced hydrodynamically (20 s at 100 mbar) followed by a transient electrokinetic separation (30 kV for 45 s) that is then repeated on-capillary when coupled to a single MS/MS instrument.[35] Unlike complicated column switching or elution hopping methods in gradient elution LC-MS/MS that offer limited throughput gains,[37] multiplexed electrophoretic separations relies on an isocratic alkaline BGE (50 mM ammonium acetate, pH 8.5) together with a coaxial sheath liquid solution (70:30 MeOH:water) to enable steady-state ionization of migrating HP-G zones under negative ion mode detection **(Figure 3.1A**). Robust separations of acidic urinary metabolites is achieved in MSI-CE-MS/MS by preventing polyimide aminolysis when using strongly alkaline ammonia-based buffers (pH > 10)[34] while turning off the nebulizer spray during serial sample introduction to avoid suctioning of air within capillary and frequent current failures.[33] Importantly, this hybrid serial injection format retains the total capillary length to further boost sample throughput and separation resolution unlike hydrodynamic injections used to spatially displace sample and BGE segments prior to voltage application in MSI-CE-MS/MS.[38] Multiple reaction monitoring was

performed using quantifier (*m/z* 217) and qualifier (*m/z* 198) ion transitions optimized for HP-G (*m/z* 393 as intact quasimolecular ion, $[M-H]^-$) that co-migrates with its matching stable-isotope internal standard, HP-G-d_9 (**Supporting Information Table S3.1; Figure 3.1B**) without sample carry-over in process/extraction blanks (**Figures 3.1C**). In this case, no deuterium effect is evident since electrophoretic separations occur in free solution, where HP-G and HP-G-d_9 from 13 independent samples co-migrate into the ion source, but are offset in time over a 12 min period (~ 6-18 min). This in turn enables better correction of potential sample matrix effects given retention time differences for deuterated OH-Pyr analogs in chromatographic separations[28] especially when using gradient elution conditions in LC-MS/MS.

3.4.2 Optimization of a Quantitative Urinary HP-G Extraction Protocol

We also optimized a simple yet efficient liquid extraction protocol for sample enrichment of low levels of urinary HP-G (< 20 ng/L) in non-smoking populations not exposed to occupational smoke. A modified methyl-*tert*-butyl ether (MTBE) extraction protocol[33] used an initial acidification step of urine (pH ~ 2-3, 12.5 μL of 1.0 M HCl) prior to liquid extraction to improve partitioning of urinary HP-G (pK_a ~ 3.8) to the top ether layer as reflected by a 7.6-fold greater recovery as compared to unmodified urine (pH ~ 6.5) as a control (**Supporting Information Figure S3.1**). Further optimization of the MTBE:urine volume ratio was conducted to simplify extractions using standard 2.0 mL GC vials. The original MTBE extraction protocol by Matyash *et al.*[36] for lipid analysis used a 5:1 volume ratio of MTBE:serum or plasma, however HP-G recovery studies in pooled urine demonstrated that a 2:1 MTBE:urine volume ratio was adequate to achieve high solute recoveries (**Supporting Information Figure S3.1**). Importantly, a single aliquot of MTBE solvent is sufficient to yield quantitative recovery of HP-G (~ 90%, 10 μg/mL)

with no residual amounts detected in second or third extract fractions applied to the same urine sample (**Supporting Information Figure S3.1**). Thus, a single-step acidified MTBE extraction protocol using 0.5 mL of urine provided an overall 50-fold enrichment of HP-G after a reconstitution step to 10 μL prior to MSI-CE-MS/MS analysis.

3.4.3 Versatile Data Workflows and Method Validation for Reliable Urinary HP-G Determination

Various serial injection configurations can be designed in MSI-CE-MS/MS to reflect different experimental designs, where mass spectral information is encoded temporally in the separation[35] allowing for accelerated data workflows for metabolite authentication and biomarker discovery together with stringent QC and batch correction.[30,31] For example, extraction efficiency is evaluated using HP-G spiked into a pooled urine sample (0.5 mL) at three different concentration levels (40 ng/L, 200 ng/L, 20 μg/L) in triplicate within a single run (**Figure 3.1D**), including blank extracts as controls. Overall, an average recovery of (93 ± 3)% demonstrates good method accuracy consistent with exhaustive extraction of urinary HP-G (**Figure 3.1G**). Moreover, a six-point calibration curve for HP-G normalized to HP-G-d_9 is acquired in duplicate in a single run over a 100-fold linear dynamic range when using MSI-CE-MS/MS (**Figure 3.1E**) with excellent linearity (R^2 = 0.999) and good repeatability with a mean CV of 6.8% at all calibrant levels (**Figure 3.1H**).

Also, the limit of detection (LOD, *S/N* ~ 3) and limit of quantification (LOQ, *S/N* ~ 10) for HP-G were about 7 ng/L and 20 ng/L, respectively following processing of all calibrants using the optimized acidified MTBE extraction protocol; this is comparable to recent methods developed for direct urinary HP-G determination by LC-FLD with SPE using 1.0 mL of urine,[25]

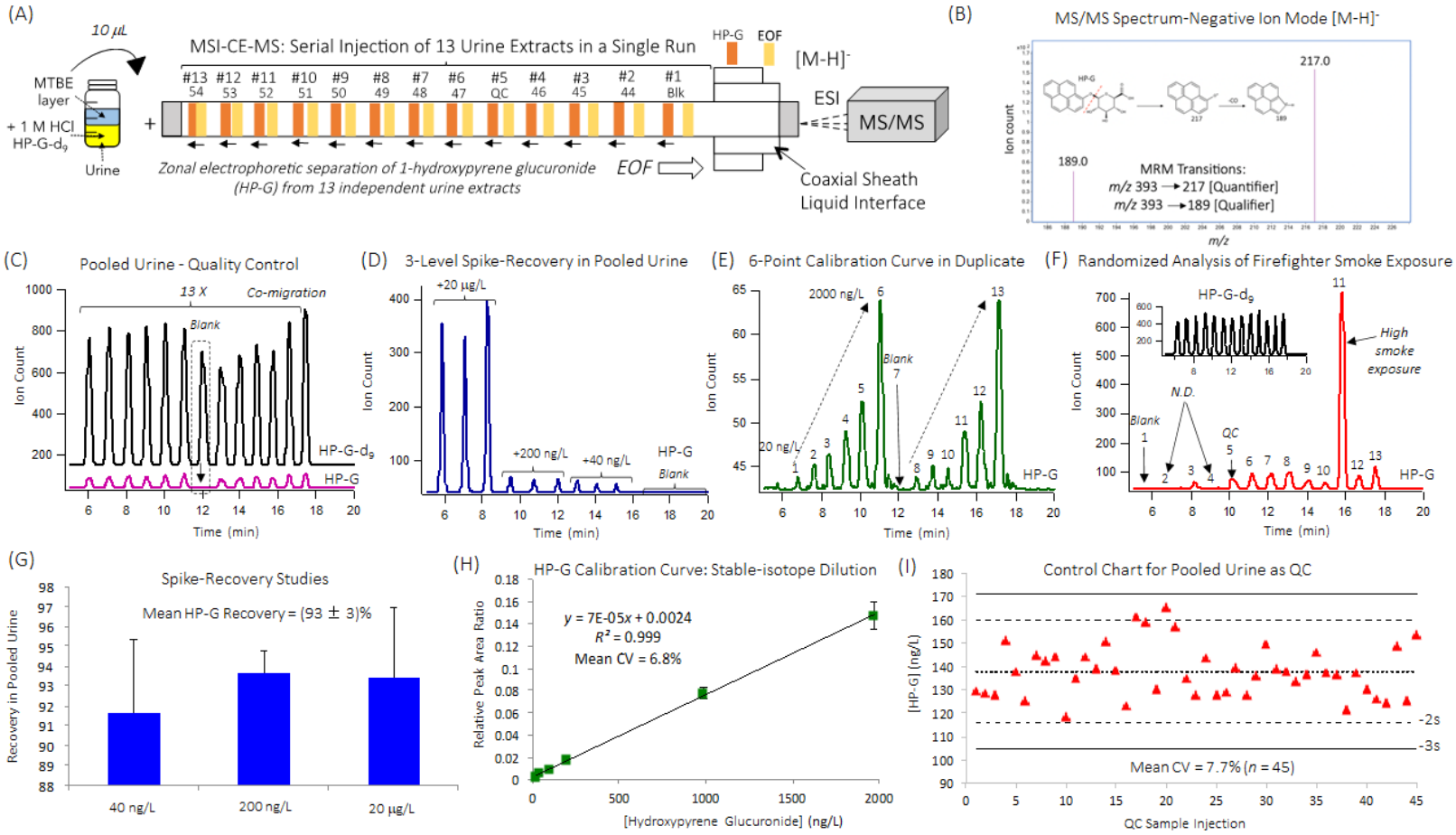


Figure 3.1. **(A)** High throughput biomonitoring of smoke exposures using MSI-CE-MS/MS based on serial injections of 13 discrete samples and zonal electrophoretic separation of HP-G in an alkaline buffer system following an acidified MTBE urine extraction protocol with a 50-fold sample enrichment. **(B)** A MS/MS spectrum of precursor ion (*m/z* 393, $[M-H]^-$) under negative ion mode detection showing two product ions used as quantifier (*m/z* 217) and qualifier (*m/z* 198). **(C)** An extracted ion electropherogram (EIE) overlay of HP-G and HP-G-d_9 as a co-migrating stable-isotope internal standard analyzed from 12 pooled firefighter urine extracts used as a quality control (QC) together with a technical blank extract within a single run. **(D)** An EIE depicting a spike-recovery study performed in pooled urine from non-occupational exposed volunteers, where HP-G was spiked at three different concentration levels in triplicate together with non-spiked urine extracts. **(E)** An EIE showing a 6-point calibration curve performed in duplicate by MSI-CE-MS/MS together with a blank extract based on the mean ion response ratio of HP-G to HP-G-d_9. **(F)** Randomized analysis of urine extracts from individual firefighters deployed at the Fort McMurray wildfire together with a QC and a blank extract in each run highlighting the large between-subject variations in smoke exposures based on differences in urinary HP-G. **(G)** Excellent mean recoveries (> 90%) for urinary HP-G are achieved using a simple ether extraction protocol **(H)** with good linearity over a 100-fold range (20 ng/L-2000 ng/L), and **(I)** acceptable technical precision (CV < 8%) as shown in the control chart for QC samples ($n = 45$) analyzed in this study.

as well as UPLC-MS/MS with liquid-phase extraction.[39] Importantly, MSI-CE-MS/MS offers more than an order of magnitude greater sample throughput than GC-MS/MS for urinary HP determination in large-scale epidemiological studies to assess smoke exposures by the Centers for Disease Control and Prevention (~ 40 samples per day per analyst).[26] Additional cost and time savings are anticipated with implementation of an automated liquid handling system as compared to multi-step sample workup procedures for processing urine prior to GC-MS/MS. The figures of merit achieved for urinary HP-G by MSI-CE-MS/MS are summarized relative to previously validated LC-FLD, LC-MS/MS, and GC-MS/MS protocols for urinary HP or HP-G determination (**Supporting Information Table S3.2**).

3.4.4 Inter-laboratory Method Comparison for Biomonitoring of Smoke Exposures in Firefighters

Our method was further validated by independently quantifying HP-G in matching urine samples collected from firefighters (n = 62) deployed at the Fort McMurray wildfire previously measured by GC-HRMS after an extended enzyme hydrolysis incubation (~ 17 h) to ensure quantitative urinary HP deconjugation.[28] In our case, all urine extracts were analyzed by MSI-CE-MS/MS in randomized sample positions (*i.e.,* 11 samples per run) together with a blank extract and a pooled urine extract derived from all firefighters that served as internal QC (**Figure 3.1F**). Overall, non-detectable levels of HP-G was measured in 7 out of 62 urine samples (11%), which was better than a fast isocratic LC-MS/MS method also developed for urinary HP quantification that had higher detection limits (29% of all urine samples were non-detected) since it did not include liquid extraction or SPE for sample enrichment.[28] Overall, the technical precision for MSI-CE-MS/MS was acceptable (mean CV = 7.7%, n = 45) as depicted in the control chart for repeated

analysis of QC urine extracts in every run (**Figure 3.1I**) as compared to intermittent QC samples (~ 8-10 runs) analyzed in sample batches when using single injection chromatographic methods.[28] An inter-laboratory method comparison between (total) urinary HP by GC-HRMS as compared to direct determination of urinary HP-G by MSI-CE-MS/MS using a Passing-Bablok regression (**Figure 3.2A**) confirms good mutual agreement as reflected by a slope of 1.10 with no significant difference from the line of unity (p = 0.72). Moreover, a Bland-Altman %difference plot (**Figure 2B**) highlights that there is a mean bias of 15% between both methods with few outliers (3.3%; 2 of 55 urine samples) exceeding the agreement limits (± 2 s). A modest positive bias for GC-HRMS as compared to MSI-CE-MS/MS is reasonable given that up to 10% of total HP measured in urine is derived from the hydrolysis of the sulfate conjugate of HP[19] when using a mixed enzyme solution with glucuronidase/sulfatase activity. Thus, the exclusion of other minor urinary HP metabolites apart from HP-G only contributes to a small extent of negative bias relative to total HP that does not impact decision making in terms of smoke exposures. A box plot of urinary HP-G concentrations in non-repeat samples collected from deployed emergency firefighters (n = 36) in a single fire station (mean age = 36 years; mean BMI = 29 kg/m^2) depicts median (range) concentrations of 125 (42-444) ng/L or 108 (25-342) ng/g creatinine (**Figure 3.2C**). As expected, there is considerable between-subject variability, yet with low overall PAH exposures below the biological exposure index (< 1000 ng/g creatinine). This result is likely attributed to delays in urine collection during emergency conditions (median = 4.5 days; only 14 of 36 samples collected < 24 h following deployment) when estimated PAH exposures were lower than peak levels at earlier stages of the wildfire as discussed by Cherry *et al.*[11] This cohort of non-smoking male firefighters also had lower dermal PAH loadings as they were able to maintain good hygiene practices (*e.g.,* change of clothing, access to showers) during

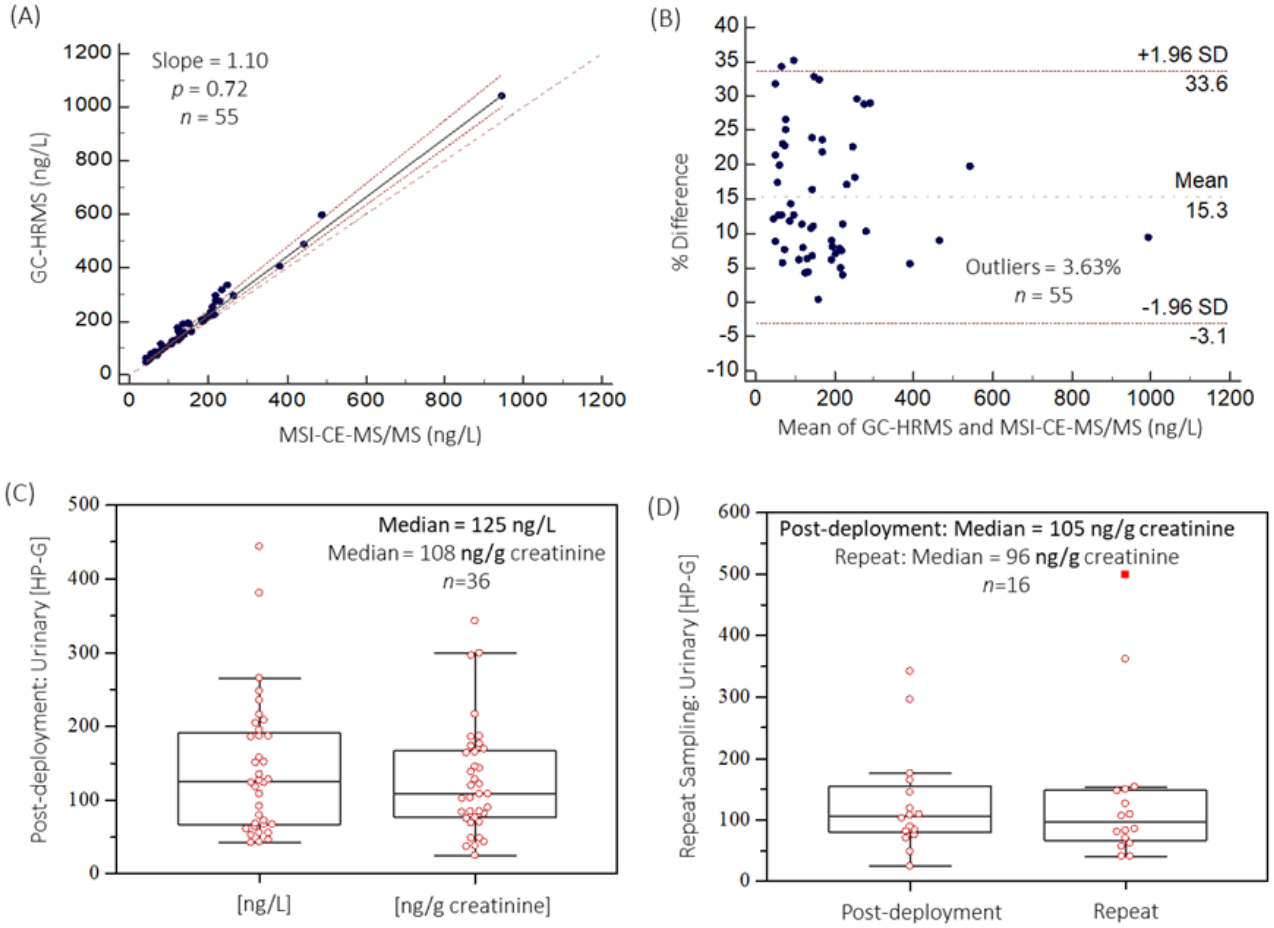


Figure 3.2. Inter-laboratory method comparison between urinary HP-G by MSI-CE-MS/MS relative to total HP by GC-HRMS measured consistently from firefighters (n = 55) following their deployment to the 2016 Fort McMurray wildfire. **(A)** A Passing-Bablok regression analysis and **(B)** Bland-Altman %difference plot confirm good mutual agreement in smoke exposure assessment between both methods based on total urinary HP following overnight enzyme hydrolysis as compared to the direct analysis of HP-G using an acidified ether extraction procedure. **(C)** Box plot comparing urinary HP-G measured in non-repeat urine samples from firefighters (n = 36) based on their absolute concentrations (ng/L) and creatinine-normalized concentrations (ng/g) indicative of normal PAH exposures due to delays in urine collection, as well as firefighters practicing good hygiene practices with use of personalized protective equipment at later stages of the wildfire. **(D)** Comparison of urinary HP-G profiles measured from a subset of firefighters (n = 16) post-deployment relative to matching repeat urine samples collected at a later time period with no significant change in overall PAH exposures ($p > 0.05$).

less intense stages of firefighting while also using respiratory protective equipment.[11] In fact, there were no differences ($p > 0.05$) in overall PAH exposures in a subset of firefighters (n = 16) based on their creatinine-normalized urinary HP-G concentrations measured post-deployment relative to repeat samples collected at later periods while off-shift (**Figure 3.2D**). The extent of

smoke exposure assessed in this study is comparable to normal levels for the US population (urinary HP ~ 119 ng/L in 2009-2010),[40] but far below populations sampled from inland regions of China with poor air quality (urinary HP-G ~ 510 ng/g creatinine),[39] as well as urine samples collected post-shift from structural (urinary HP ~ 270 ng/g creatinine)[9] or wildland firefighters (urinary HP ~ 576 ng/g creatinine)[27] in North America. Future studies that enable practical collection and storage of repeat urine specimens in the field are needed for more accurate risk assessment of smoke exposures following active firefighting activities. Robust nanospray ion source designs are also needed to further boost concentration sensitivity in CE-MS for ultra-trace environmental analyses[41] as required for new advances in exposomics research.

3.5 Conclusion

MSI-CE-MS/MS offers a cost-effective approach for rapid screening of recent smoke exposures based on direct analysis of urinary HP-G following a simple acidified ether extraction procedure using 0.50 mL of urine. This multiplexed separation method offers rapid analysis times (< 3 min/sample), as well as acceptable reproducibility (CV < 8%) and recovery (~ 93%) with good mutual agreement (mean bias ~ 15%) as compared to GC-HRMS that avoids complicated sample workup procedures, including enzyme hydrolysis, solid-phase extraction and pre-column chemical derivatization. To the best of our knowledge, this is the first report of urinary HP-G as a convenient surrogate for total HP when biomonitoring PAH exposures in firefighters, where delays in urine collection can underestimate true smoke exposures during prolonged emergency fire suppression activities in the field. Also, good hygiene practices and proper use of personal protective equipment by firefighters can prevent dermal PAH absorption. MSI-CE-MS/MS offers a high throughput platform for reliable determination of urinary OH-PAHs and other

classes of anionic environmental contaminants[42] that is optimal for large-scale biomonitoring of populations in support of evidence-based policies to mitigate smoke exposures and chronic disease burden.

3.6 Acknowledgments

PBM acknowledges funding from the Natural Sciences and Engineering Research Council of Canada, and Genome Canada. The authors thank Nicola Cherry from the Division of Preventative Medicine at the University of Alberta for access to firefighter urine samples, as well as Karl Jobst from Memorial University, and David Kinniburgh from the Alberta Centre for Toxicology. The authors also acknowledge Marcus Kim and John Saucen from Agilent Technologies Inc.

3.7 References

(1) Idowu, O.; Semple, K. T.; Ramadass, K.; O'Connor, W.; Hansbro, P.; Thavamani, P. Beyond the obvious: environmental health implications of polar polycyclic aromatic hydrocarbons. *Environ. Int.* **2019**, 123, 543–557.
(2) Tong, S. Air pollution and disease burden. *Lancet Planet. Heal.* **2019**, 3, e49–e50.
(3) West, J. J.; Cohen, A.; Dentener, F.; Brunekreef, B.; Zhu, T.; Armstrong, B.; Bell, M. L.; Brauer, M.; Carmichael, G.; Costa, D. L.; Dockery, D. W.; Kleeman, M.; Krzyzanowski, M.; Künzli, N.; Liousse, C.; Lung, S. C. C.; Martin, R. V.; Pöschl, U.; Pope, C. A.; Roberts, J. M.; Russell, A. G.; Wiedinmyer, C. What we breathe impacts our health: improving understanding of the link between air pollution and health. *Environ. Sci. Technol.* **2016**, 50, 4895–4904.
(4) Alshaarawy, O.; Elbaz, H. A.; Andrew, M. E. The association of urinary polycyclic aromatic hydrocarbon biomarkers and cardiovascular disease in the US population. *Environ. Int.* **2016**, 89, 174–178.
(5) Soteriades E.S.; Smith D.L.; Tsismenakis A.J.; Baur D.M.; Kales S.N. Cardiovascular disease in US firefighters: a systematic review. *Cardiol. Rev.* **2011**, 19, 202–215.
(6) Daniels R. D.; Bertke S.; Dahm M. M.; Yiin J. H.; Kubale T. L.; Hales T.R.; Baris D.; Zahm S. H.; Beaumont J. J.; Waters K. M.;, Pinkerton L. E.. Exposure-response relationships for selectcancer and non-cancer health outcomes in a cohort of US firefighters from San Francisco, Chicago and Philadelphia (1950–2009). *Occup. Environ. Med.* **2015**, 72, 699–706.
(7) Soteriades E.S.; Kim J.; Christophi C.A.; Kales S.N. Cancer incidence and mortality in

firefighters: A state-of-the-art review and meta-́analysis. *Asian Pac. J. Cancer Prev.* **2019**, 20, 3221–3231.

(8) Fernando, S.; Shaw, L.; Shaw, D.; Gallea, M.; Vandenenden, L.; House, R.; Verma, D. K.; Britz-McKibbin, P.; McCarry, B. E. Evaluation of firefighter exposure to wood smoke during training exercises at burn houses. *Environ. Sci. Technol.* **2016**, 50, 1536–1543.

(9) Keir, J.L.A.; Akhtar, U.S.; Matschke, D.M.J.; Kirkham, T.L.; Chan, H.M.; Ayotte, P.; White, P.A.; Blais, J.M. Elevated exposures to polycyclic aromatic hydrocarbons and other organic mutagens in Ottawa firefighters participating in emergency, on-shift fire suppression. *Environ. Sci. Technol.* **2017,** 51, 12745–12755.

(10) Wingfors, H.; Nyholm, J. R.; Magnusson, R.; Wijkmark, C. H. Impact of fire suit ensembles on firefighter PAH exposures as assessed by skin deposition and urinary biomarkers. *Ann. Work Expo. Heal.* **2018**, 62, 221–231.

(11) Cherry, N.; Aklilu, Y.A.; Beach, J.; Britz-McKibbin, P.; Elbourne, R.; Galarneau, J.M.; Gill, B.; Kinniburgh, D.; Zhang X. Urinary 1-hydroxypyrene and skin contamination in firefighters deployed to the Fort McMurray fire. *Ann. Work Expo. Health.* **2019**, 63, 448–458.

(12) Li Z.; Sandau C. D.; Romanoff L. C.; Caudill S. P.; Sjodin A.; Needham L. L.;, Patterson D. G. Jr. Concentration and profile of 22 urinary polycyclic aromatic hydrocarbon metabolites in the US population. *Environ. Res.* **2008**, 107, 320–331.

(13) Wang Y.; Wong L. Y.; Meng L.*;* Pittman E.N.; Trinidad, D.A.; Hubbard, K.L.; Etheredge, A.; Del Valle-Pinero, E.Y.; Zamoiski, R.; Van Bemmel D.M.; Borek N. Urinary concentrations of monohydroxylated polycyclic aromatic hydrocarbons in adults from the U.S. Population Assessment of Tobacco and Health (PATH) Study Wave 1 (2013-2014). *Environ. Int.* **2019**, 123, 201–208.

(14) Gill, B.; Britz-McKibbin, P. Biomonitoring of smoke exposure in firefighters: a review. *Curr. Opin. Environ. Sci. Health* **2020**, 15, 57–65.

(15) Ciarrocca, M.; Rosati, M. V.; Tomei, F.; Capozzella, A.; Andreozzi, G.; Tomei, G.; Bacaloni, A.; Casale, T.; Andrè, J. C.; Fioravanti, M.; Cuartas, M. F.; Caciari, T. Is urinary 1-hydroxypyrene a valid biomarker for exposure to air pollution in outdoor workers? A meta-analysis. *J. Expo. Sci. Environ. Epidemiol.* **2014**, 24, 17–26.

(16) ACGIH. (2017) Polycyclic aromatic hydrocarbons (PAHS): BEI(R), 7th Edition Documentation, American Conference of Governmental Industrial Hygienists (ACGIH), Cincinnati, OH, USA, **2017**; 1–15.

(17) Jongeneelen F. J. A guidance value of 1-hydroxypyrene in urine in view of acceptable occupational exposure to polycyclic aromatic hydrocarbons. *Toxicol. Lett.* **2014**, 231, 239–248.

(18) Henkler F., Stolpmann K., Luch A. Exposure to polycyclic aromatic hydrocarbons: bulky DNA adducts and cellular responses. *Exp. Suppl.* **2012**, 101, 107–131.

(19) Strickland, P. T.; Kang, D.; Bowman, E. D.; Fitzwilliam, A.; Downing, T. E.; Rothman, N.; Groopman, J. D.; Weston, A. Identification of 1-hydroxypyrene glucuronide as a major pyrene metabolite in human urine by synchronous fluorescence spectroscopy and gas chromatography-mass spectrometry. *Carcinogenesis* **1994**, 15, 483–487.

(20) Singh, R.; Tuček, M.; Maxa, K.; Jana, T.; Weyand, E. H. A rapid and simple method for the analysis of 1-hydroxypyrene glucuronide: A potential biomarker for polycyclic aromatic hydrocarbon exposure. *Carcinogenesis.* **1995,** 16, 2909–2915.

(21) Lee, C. K.; Cho, S. H.; Kang, J. W.; Lee, S. J.; Ju, Y. S.; Sung, J.; Strickland, P. T.; Kang,

D. Comparison of three analytical methods for 1-hydroxypyrene glucuronide in urine after non-occupational exposure to polycyclic aromatic hydrocarbons. *Toxicol. Lett.* **1999**, 108, 209–215.

(22) Hofmann, J. N.; Liao, L. M.; Strickland, P. T.; Shu, X.-O.; Yang, G.; Ji, B.-T.; Li, H.-L.; Rothman, N.; Kamangar, F.; Gao, Y.-T. Polycyclic aromatic hydrocarbons: determinants of urinary 1-hydroxypyrene glucuronide concentration and risk of colorectal cancer in the Shanghai women's health study. *BMC Cancer* **2013**, 13, 282.

(23) Yang, M., Wang, Y., Ren, J. Li; M.; Wang, Q.; Li, N.; Zhu, J.; Zou, Z. A rapid and sensitive method of determination of 1-hydroxypyrene glucuronide in urine by UPLC–FLD. *Chromatographia* **2019**, 82, 835–842.

(24) Ramsauer, B.; Sterz, K.; Hagedorn, H.-W.; Engl, J.; Scherer, G.; McEwan, M.; Errington, G.; Shepperd, J.; Cheung, F. A liquid chromatography/tandem mass spectrometry (LC-MS/MS) method for the determination of phenolic polycyclic aromatic hydrocarbons (OH-PAH) in urine of non-smokers and smokers. *Anal. Bioanal. Chem.* **2011**, 399, 877–889.

(25) Kakimoto, K.; Toriba, A.; Ohno, T.; Ueno, M.; Kameda, T.; Tang, N.; Hayakawa, K. Direct measurement of the glucuronide conjugate of 1-hydroxypyrene in human urine by using liquid chromatography with tandem mass spectrometry. *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.* **2008**, 867, 259–263.

(26) Centers for Disease Control and Prevention. Laboratory procedure manual: Monohydroxy-polycyclic aromatic hydrocarbons (OH-PAHs), **2013**.

(27) Adetona, O.; Simpson, C. D.; Li, Z.; Sjodin, A.; Calafat, A. M.; Naeher, L. P. Hydroxylated polycyclic aromatic hydrocarbons as biomarkers of exposure to wood smoke in wildland firefighters. *J. Expo. Sci. Environ. Epidemiol.* **2017**, 27, 78–83

(28) Gill, B.; Mell, A.; Shanmuganathan, M.; Jobst, K.; Zhang, X.; Kinniburgh, D.; Cherry, N.; Britz-McKibbin, P. Urinary hydroxypyrene determination for biomonitoring of firefighters deployed at the Fort McMurray wildfire: An inter-laboratory method comparison. *Anal. Bioanal. Chem.* **2019**, 411, 1397–1407.

(29) Dwivedi, P.; Zhou, X.; Powell, T.G.; Calafat, A.M.; Ye, X. Impact of enzymatic hydrolysis on the quantification of total urinary concentrations of chemical biomarkers. *Chemosphere* **2018**, 199, 256–262.

(30) Saoi, M.; Kennedy, K.M.; Gohir, W.; Sloboda, D.M.; Britz-McKibbin, P. Placental metabolomics for assessment of sex-specific differences in fetal development during normal gestation. *Sci. Rep.* **2020**, 10, 9399.

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

(32) Gaudreau, É.; Bérubé, R.; Bienvenu, J.F.; Fleury, N. Stability issues in the determination of 19 urinary (free and conjugated) monohydroxy polycyclic aromatic hydrocarbons. *Anal. Bioanal. Chem.* **2016**, 408, 4021–4033.

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

(34) Yamamoto, M.; Ly, R.; Gill, B.; Zhu, Y.; Moran-Mirabal, J.; Britz-McKibbin, P. Robust and high-throughput method for anionic metabolite profiling: preventing polyimide aminolysis and capillary breakages under alkaline conditions in capillary electrophoresis-

mass spectrometry. *Anal. Chem.* **2016**, 88, 10710–10719.

(35) Saoi, M.; Li, A.; McGlory, C.; Stokes, T.; von Allmen, M. T.; Phillips, S. M.; Britz-McKibbin, P. Metabolic perturbations from step reduction in older persons at risk for sarcopenia: Plasma biomarkers of abrupt changes in physical activity. *Metabolites* **2019**, 9, 134–153.

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

(37) Livesay, E.A.; Tang, K.; Taylor, B.K.; Buschbach, M.A.; Hopkins, D.F.; LaMarche, B.L.; Zhao, R.; Shen, Y.; Orton, D.J.; Moore, R.J.; Kelly, R.T.; Udseth, H.R.; Smith, R.D. Fully automated four-column capillary LC-MS system for maximizing throughput in proteomic analyses. *Anal. Chem.* **2008**, 80, 294–302.

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

(39) Li, M.; Wang, Q.; Zhu, J.; Li, N.; Zou, X. A simple analytical method of determining 1-hydroxypyrene glucuronide in human urine by isotope dilution with ultra performance liquid chromatography-tandem mass spectrometry. *Anal. Bioanal. Chem.* **2017**, 409, 1513–1518.

(40) CDC. National Report on Human Exposure to Environmental Chemicals. Centers for Disease Control and Prevention, Volume 1, Atlanta, GA, **March 2018**.

(41) Höcker, O; Bader, T.; Schmidt, T. C.; Schulz, W.; Neusüß, C. Enrichment-free analysis of anionic micropollutants in the sub-ppb range in drinking water by capillary electrophoresis-high resolution mass spectrometry. *Anal. Bioanal. Chem*. **2020,** 412, 4857–4865.

(42) Azab, S. M.; Hum, R.; Britz-McKibbin P. Rapid biomonitoring of perfluoroalkyl substance exposures in serum by multisegment injection-nonaqueous capillary electrophoresis-tandem mass spectrometry. *Anal. Sci. Advances* **2020**, 1, 173-182

3.8 Supporting Information

3.8.1 Instrumental Method for Total Hydrolyzed Urinary HP Determination by GC-HRMS

Sample preparation for total urinary HP determination using gas chromatography with high resolution mass spectrometry (GC-HRMS) included enzyme hydrolysis, solid-phase extraction (SPE), and pre-column chemical derivatization.[1,2] Briefly, a 1.0 mL urine aliquot was incubated with 5.0 mL sodium acetate buffer (0.10 M, pH 5.5), 10 µL of recovery standard (OH-Pyr-d_9) and 10 µL glucuronidase/arylsulfatase (*Helix pomatia*, Sigma-Aldrich Inc.) for 17-18 h at 37 °C to ensure complete hydrolysis.[2] Next, solid-phase extraction (SPE) was conducted on a vacuum manifold using Varian Focus (50 mg, 6.0 mL) cartridges (Agilent Technologies, Santa, Clara, CA, USA). Deconjugated (total) HP were eluted with dichloromethane at a rate of 0.5 min/mL. This was followed by drying the urine extract with 10-20 mg of magnesium sulfate anhydrous that was filtered in a glass pipette containing glass wool. The final extract was dried under a gentle stream of nitrogen to 100 µL, transferred to a GC vial insert, blown down to dryness and reconstituted to 10 µL in toluene. An internal standard of Pyr-d_{10} (1.0 µL, 9100 µg/L) was spiked into all reconstituted urine extracts. The samples were then derivatized with 15 µL of *N*-methyl-*N*-(trimethylsilyl)trifluoroacetamide (MSTFA), using an incubation step of 20 min at 60 °C, which corresponded to an 25-fold sample enrichment from 1.0 mL of urine. The GC-HRMS analysis was performed on an Agilent 7890B gas chromatograph (Agilent Technologies, Santa, Clara, CA, USA) coupled to a Waters Xevo G1-XS quadrupole time of flight (qTOF) mass spectrometer with an atmospheric pressure chemical ionization source for urinary HP determination.[2] Also, creatinine was measured in all urine samples using an AU 480 Olympus Chemistry Analyzer (Beckman Coulter Inc., Mississauga, ON, Canada). All urinary HP-G and HP concentrations were reported as their normalized concentrations values to creatinine (ng/g

creatinine) to correct for hydration status when collecting random single-spot urine samples from firefighters post-deployment.

Table S3.1. MRM transitions for HP-G and HP-G-d_9 in urine using MSI-CE-MS/MS

	Compound	**Transition**	**CE(eV)**	**Fragmentor (V)**	**Delta EMV(V)**
Quantifier	1-Hydroxypyrene Glucuronide	393 →217	-60	120	-400
Qualifier	1-Hydroxypyrene Glucuronide	393 →189	-80	120	-400
Quantifier	1-Hydroxypyrene Glucuronide- d_9	402 →226	-60	120	-400
Qualifier	1-Hydroxypyrene Glucuronide- d_9	402 →189	-80	120	-400

Table S3.2. Comparison of major figures of merit based on different validated methods used for urinary HP-G determination.

Analytical Method	Extraction Method	Precision (Inter-day)	Recovery	LOD	Analysis Time (per sample)	Reference
HPLC-FLD	SPE (Sep-Pak C_{18})	-	95.0%	8-20 ng/L	30 min	Singh *et al.* (1995)[3]
IAC-SFS	SPE (Sep-Pak C_{18})	8.0-10.0%	85.0-95.0%	17.8 ng/mL	-	Sithisarankul *et al.* (1997)[4]
IAC-SFS	SPE (Sep-Pak C_{18})	8.0-10.0%	-	13.4 ng/mL	-	Roth *et al.* (2001)[5]
IAC-SFS	SPE (Sep-Pak C_{18})	8.0-10.0%	80.0%	26.7 ng/L	-	Lai *et al.* (2004)[6]
LC-MS/MS	SPE (Oasis MAX)	4.40-5.70%	98.0-99.0%	0.13 fmol/inj	20 min	Kakimoto *et al.* (2008)[7]
IAC-SFS	SPE (Sep-Pak C_{18})	9.0%	82.0%	10.0 ng/L	-	Lee *et al.* (2009)[8]
UPLC-MS/MS	LLE (Ethyl Acetate)	6.70%	79.4-106%	15 ng/L	10 min	Li *et al.* (2017)[9]
UPLC-FLD	SPE (PRiME HLB)	1.81-5.00%	91.9-107%	3.2 ng/L	8 min	Yang *et al.* (2019)[10]
MSI-CE-MS/MS	LLE (MTBE)	7.73%	87.0-102%	7 ng/L	2.3 min	*Current work*

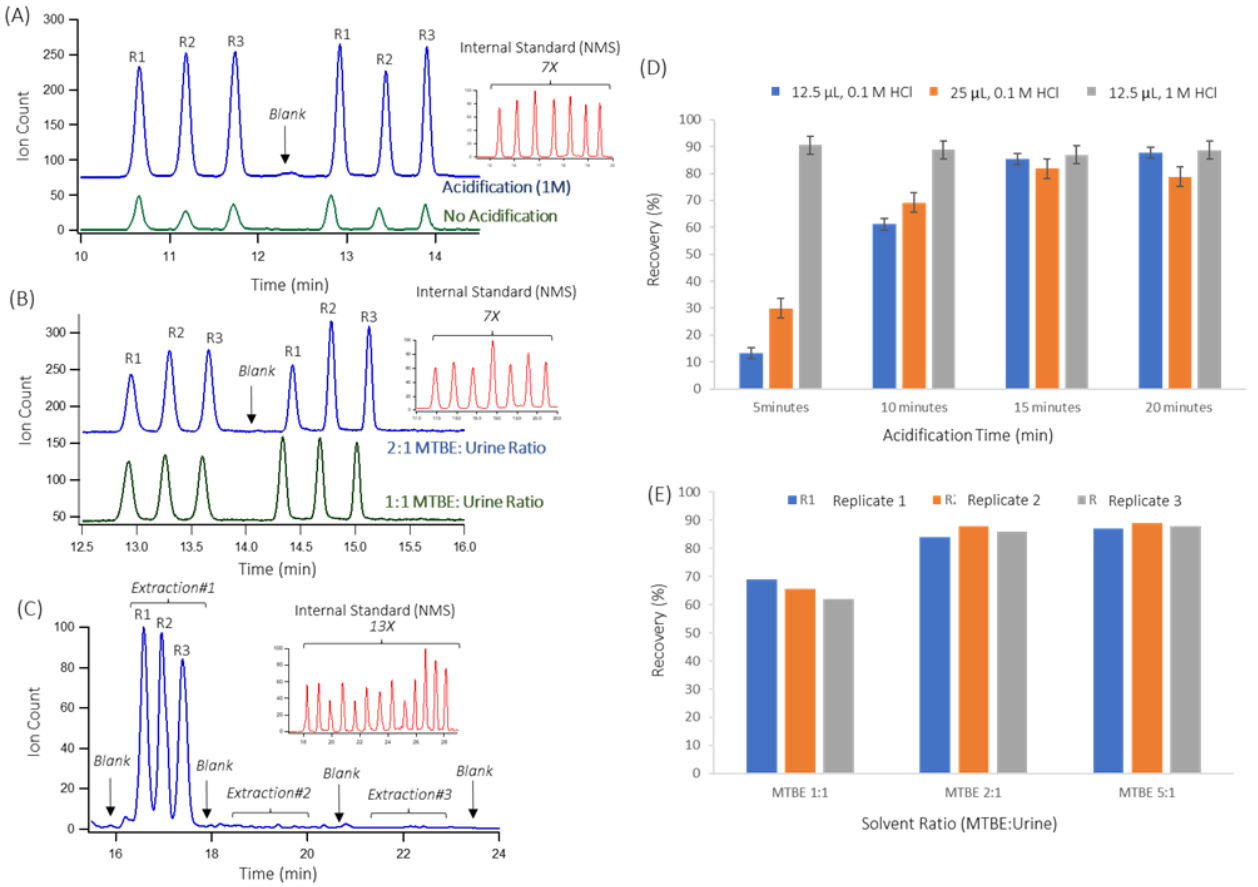


Figure S3.1 Optimization of the urine extraction protocol for sample enrichment of HP-G as a biomarker of recent PAH exposures in smoke by MSI-CE-MS/MS when using a 7 or 13 sample plug serial injection format. **(A)** Extracted ion electropherogram (EIE) confirming an average 7.6-fold increase in the recovery of HP-G upon the acidification of pooled urine using HCl (pH ~ 3) as compared to unmodified urine as control in replicate (R) samples, where naphthalene monosulfonic acid (NMS) was used as an internal standard in all reconstituted urine extracts. **(B)** Good recovery (~ 90%) was achieved for spiked HP-G in non-occupational exposed urine samples with no evidence of sample carry over and background contamination in the blank extract when using an optimal solvent ratio of 2:1 MTBE-urine that allows for use of standard 2.0 mL GC vials. **(C)** Quantitative extraction of urinary HP-G is achieved in a single aliquot of MTBE with no HP-G detected in repeat extractions of urine samples using a second or third aliquot of MTBE that was performed in triplicate with suitable blanks by MSI-CE-MS/MS. **(D)** Bar graph highlights that optimal urine acidification not only improved recovery of weakly acidic HP-G in human urine samples, but also allowed for faster equilibration times (< 5 min) prior to MTBE extraction. **(E)** Bar graph confirms good mean recoveries as required for accurate urinary HP-G quantification that is achieved using a minimum volume ratio of 2:1 for MTBE:urine with a single aliquot of ether solvent.

3.8.2 Supporting References

(1) Cherry, N.; Aklilu, Y.A.; Beach, J.; Britz-McKibbin, P.; Elbourne, R.; Galarneau, J.M.; Gill, B.; Kinniburgh, D.; Zhang X. Urinary 1-hydroxypyrene and skin contamination in firefighters deployed to the Fort McMurray fire. *Ann. Work Expo. Health.* **2019**, 63, 448–458.

(2) Gill, B.; Mell, A.; Shanmuganathan, M.; Jobst, K.; Zhang, X.; Kinniburgh, D.; Cherry, N.; Britz-McKibbin, P. Urinary hydroxypyrene determination for biomonitoring of firefighters deployed at the Fort McMurray wildfire: An inter-laboratory method comparison. *Anal. Bioanal. Chem.* **2019**, 411, 1397-1407.

(3) Singh, R.; Tuček, M.; Maxa, K.; Jana, T.; Weyand, E. H. A rapid and simple method for the analysis of 1-hydroxypyrene glucuronide: A potential biomarker for polycyclic aromatic hydrocarbon exposure. *Carcinogenesis.* **1995,** 16, 2909–2915.

(4) Sithisarankul, P.; Vineis, P.; Kang, D.; Rothman, N.; Caporaso, N.; Strickland, P. Association of 1-hydroxypyrene glucuronide in human urine with cigarette smoking and broiled or roasted meat consumption. *Biomarkers.* **1997**, 2, 217-221.

(5) Roth, M. J.; Qiao, Y.-L.; Rothman, N.; Tangrea, J. A.; Dawsey, S. M.; Wang, G.-Q.; Cho, S.-H.; Kang, D.; Taylor, P. R.; Strickland, P. T. High urine 1-hydroxypyrene glucuronide concentrations in Linxian, China, an area of high risk for squamous oesophageal cancer. *Biomarkers.* **2001**, 6, 381–386.

(6) Lai, C. H.; Liou, S. H.; Shih, T. S.; Tsai, P. J.; Chen, H. L.; Buckley, T. J.; Strickland, P. T.; Jaakkola, J. J. Urinary 1-hydroxypyrene glucuronide as a biomarker of exposure to various vehicle exhausts among highway toll-station workers in Taipei, Taiwan. *Arch. Environ. Health.* **2004**, 59, 61-69.

(7) Kakimoto, K.; Toriba, A.; Ohno, T.; Ueno, M.; Kameda, T.; Tang, N.; Hayakawa, K. Direct measurement of the glucuronide conjugate of 1-hydroxypyrene in human urine by using liquid chromatography with tandem mass spectrometry. *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.* **2008**, 867, 259–263.

(8) Lee, K. H.; Vermeulen, R.; Lenters, V.; Cho, S. H.; Strickland, P. T.; Kang, D. Determinants of urinary 1-hydroxypyrene glucuronide in South Korean children. *Int. Arch. Occup. Health.* **2009**, 82, 961-968.

(9) Li, M.; Wang, Q.; Zhu, J.; Li, N.; Zou, X. A simple analytical method of determing 1-hydroxypyrene glucuronide in human urine by isotope dilution with ultra performance liquid chromatography-tandem mass spectrometry. *Anal. Bioanal. Chem*. **2017**, 409, 1513-1518.

(10) Yang, M., Wang, Y., Ren, J. Li; M.; Wang, Q.; Li, N.; Zhu, J.; Zou, Z. A rapid and sensitive method of determination of 1-hydroxypyrene glucuronide in urine by UPLC–FLD. *Chromatographia* **2019**, 82, 835–842.

Chapter IV:

Urine Biomonitoring of Tobacco Smoke Exposure in the Prospective Urban and Rural Epidemiological (PURE) Study: Regional Variations in Nicotine Dependence, Smoking Topography and Misreporting

Biban Gill,[1] Stellena Mathiaparanam,[1] Sathish Thirunavukkarasu,[2] Guillaume Paré,[2,3]

Koon Teo,[2] Salim Yusuf,[2]Philip Britz-McKibbin*[1]

[1]*Department of Chemistry & Chemical Biology, McMaster University, Hamilton, ON, Canada*

[2]*Population Health Research Institute, McMaster University, Hamilton, ON, Canada*

[3] *Department of Pathology and Molecular Medicine, McMaster University, Canada*

* *Manuscript in preparation for submission*

B.G. performed all experiments, including method optimization, validation, sample preparation and data acquisition by MSI-CE-MS. B.G. also carried out the data processing, statistical analysis, and interpretation. S.M. was involved in sample processing and provided data for thiocyanate. S.Y. is the principal investigator for the PURE study, K.T. was involved in the study design for PURE, and G.P., aided in the design and supervision of the PURE urine sub study. S.T. was involved in the selection of samples for the PURE urine sub study and provided participant information for the statistical analysis. B.G. wrote the initial draft for publication. P.B.M. was involved in the design of the PURE urine sub study and provided feedback for experiments as well as revisions for the manuscript draft.

Chapter IV: Urine Biomonitoring of Tobacco Smoke Exposure in the Prospective Urban and Rural Epidemiological (PURE) Study: Regional Variations in Nicotine Dependence, Smoking Topography and Misreporting

4.1 Abstract

4.1.1 Background

Tobacco use remains a leading preventable cause of death and chronic disease burden globally. Total nicotine equivalent (TNE) and nicotine metabolic ratio (NMR) represent phenotypic biomarkers to objectively assess recent tobacco smoke exposures and smoking behavior as compared to self-reports that are prone to bias. However, current analytical methods for determining urinary TNE and/or NMR are low throughput and require complicated sample workup procedures limiting their applicability for large-scale epidemiological studies.

4.1.2 Methods

We introduce a rapid screening method for urinary nicotine metabolites using multisegment injection-capillary electrophoresis-mass spectrometry (MSI-CE-MS) with full-scan data acquisition under positive ion mode. Morning spot urine samples were directly analyzed for up to seven major nicotine metabolites and their intact conjugates from participants (n=1000) in the Prospective Urban and Rural Epidemiological (PURE) study from fourteen countries at different income status. Chemical stability and sensitivity studies were performed to confirm reliable analysis of a panel of seven urinary nicotine metabolites with few non-detects in current smokers as compared to single nicotine species. Complementary correlation and univariate statistical analyses with covariate adjustments were applied to compare TNE-7 with self-reported smoking intensity (i.e., cigarettes per day, CPD), as well as stratify urinary NMR into quartiles to classify fast (Q4) from slow (Q1-Q4) nicotine metabolizers in current smokers from high-, medium- and

low-income countries (HICs, MICs, and LICs). A similar approach was also performed for assessment of the UDP-glucuronosyltransferase activity for cotinine as related to tobacco detoxification capacity.

4.1.3 Results

Method validation of MSI-CE-MS demonstrated acceptable inter-day reproducibility for major nicotine metabolites (median CV=18%, n=156), good chemical stability that tolerates repeat freeze-thaw cycles (<15%), with adequate limits of detection (0.05-0.11 μM). Overall, TNE-7 had a moderate correlation with self-reported CPD (r = 0.391, $p = 4.91\times10^{-22}$, n = 564) that was elevated in heavy smokers from HICs (F = 10.2, $p = 2.51\times10^{-10}$, n = 322) as compared to MICs and LICs. Fast nicotine metabolizers were also identified in the PURE study, who had greater tobacco smoke exposure (F = 38.2, $p = 1.0\times10^{-33}$, n = 568) after adjustment for smoking intensity and other covariates.

4.1.4 Conclusions

This work outlines a high throughput method to biochemically verify smoking status and assess smoking topography, which is urgently needed to elucidate regional variations in tobacco-related disease risk in diverse populations not feasible by standardized questionnaires.

4.1.5 Impact

A high throughput platform for comprehensive nicotine analyses enables accurate assessment of the hazards of smoking for global health that is confounded by varying smoking habits, genetic variations in the population, and a myriad of tobacco products consumed in different countries.

4.2 Introduction

The tobacco epidemic remains the leading cause of preventable, premature death and chronic disease burden worldwide (1), despite its well-established health risks (2). Smoking causes diminished overall health, with current and former smokers demonstrating an increased risk over never smokers for several chronic diseases, including heart disease, stroke, and lung cancer (3). While intensive public health efforts (e.g., pack warning, ad bans, public space policies) have reduced overall smoking prevalence (4), global disparities in policy performance remain an ongoing concern, including regional variations within developed countries with higher smoking rates impacting vulnerable populations (5). These challenges are further exacerbated by a dismal 3-5% (6) success rate in smoking cessation (7), with relapse and smoking rates reported to have increased during the COVID-19 pandemic (8).

Globally, there is an estimated 1.14 billion current tobacco users with > 80% of current smokers residing in developing countries (9). Smoking habits and tobacco-related disease risk differ widely between countries with varying cultural and socioeconomic conditions (10,11). As a result, large-scale longitudinal surveillance studies have been conducted using standardized surveys (12). Although questionnaires related to smoking history remain an inexpensive tool for large-scale epidemiological studies, they are prone to error and misreporting (13,14), including gender and social desirability bias (15). Also, self-reports are unable to account for variations in commercial cigarette product designs or regional hand rolled varieties (e.g., bidis), nicotine content and tar chemical composition, as well as smoking topography (e.g., puff volume and frequency) that impacts carcinogen exposure (16). For instance, a recent study comparing tobacco related all-cause mortality and clinical events globally, the hazard ratios were reported to be elevated in current smokers from high-income countries (HICs) as compared to middle-

(MICs) and low-income countries (LICs), which could not be explained by self-reported smoking history (e.g., age at initiation, quantity, duration) (17). Surprisingly, few epidemiological studies studying smoking behaviors across multi-ethnic populations or different countries, such as the National Health and Nutrition Examination Survey (NHANES), Population Assessment of Tobacco and Health (PATH), and Global Adult Tobacco Survey (GATS), have made use of tobacco-specific biomarkers for biochemical verification of smoking status and smoking behavior (18–21).

Regardless of the form of tobacco usage (e.g., chewing tobacco, bidi, cigarette), nicotine and its metabolites are considered to be the most reliable measures of recent tobacco smoke exposure (22). Although various biospecimens (e.g., serum, urine, saliva, hair) have been used for surveillance (23), urine is best suited for non-invasive sampling whilst providing the most concentrated and comprehensive assessment of recent tobacco exposure within ~ 2 days depending on smoking intensity and frequency (24). Nicotine, inhaled from tobacco smoke or ingested from smokeless tobacco, is rapidly metabolized in circulation into cotinine by cytochrome P450 (CYP) 2A6, which is then then biotransformed into 3-hydroxycotinine (25). Subsequent phase II metabolism generates the glucuronide conjugates of nicotine, cotinine, and 3-hydroxycotinine via UDP glucuronosyltransferase (UGT) to facilitate their renal elimination (25). A minor pathway for nicotine detoxification occurs via flavin monooxygenase (FMO) mediated *N*-oxidation to form nicotine-*N*-oxide. Other low abundance urinary metabolites of nicotine are also generated, including nornicotine, norcotinine, and cotinine-*N*-oxide, however all these species comprise less than 10% of total excreted nicotine (26).

Measurement of a single major nicotine metabolite in urine, such as cotinine is frequently performed in routine drug screening to confirm smoking status (27,28), however it suffers from

reduced sensitivity and extensive interindividual and ethnic variability that does not reflect true levels of nicotine exposure (29–32). As a result, the current gold standard for recent tobacco exposure is the molar sum of nicotine and its major metabolites in urine, referred to as total nicotine equivalents (TNE) (22,26,33–35). While accurate and reliable TNE measurements are critical, assessment of the rate of nicotine clearance has also been shown to provide insight on nicotine dependence, smoking relapse, and toxicant exposure (16,35–38). Consequently, the nicotine metabolite ratio (NMR) is a validated phenotypic biomarker of CYP 2A6 activity, which accounts for ~ 90% of nicotine clearance (16,39–41). The highest concordance with plasma NMR can be achieved when measuring the sum of 3-hydroxycotinine and its glucuronide to cotinine in urine (42). Moreover, ratiometric indicators of UGT activity have also been used as measures of phase II metabolism, which may be indicative of differential inactivation of tobacco smoke toxicants, such as carcinogenic tobacco-specific nitrosamines (35,43,44).

Conventional approaches for determination of nicotine and its metabolites in urine include commercial immunoassay kits, immunoaffinity chromatography, gas chromatography-mass spectrometry and increasingly liquid chromatography-tandem mass spectrometry (LC-MS/MS) (45). LC-MS/MS is the gold standard for the analysis nicotine metabolites in various biological samples (35,45,46), which often relies on indirect methods following enzyme deconjugation (42). While LC-MS/MS methods are sensitive, selective and enable comprehensive surveillance of nicotine metabolites in urine, most methods to date are constrained by high costs, long analysis times (> 15 min/sample) and complicated sample-pretreatment protocols, including solid-phase extraction, liquid-liquid extraction, and enzyme deconjugation (47).

Herein, we introduce a high-throughput platform for the direct analysis of seven urinary nicotine metabolites and their intact glucuronide conjugates by multisegment injection-capillary electrophoresis-mass spectrometry (MSI-CE-MS) with full-scan data acquisition under positive ion mode conditions (48). This multiplexed separation platform takes advantage of serial injection of 11 samples in a single run (< 5 min/sample) with stringent quality control that is ideal for large-scale epidemiological studies (49) and comprehensive drug surveillance applications (50). Following method validation, MSI-CE-MS was applied for biochemical verification of smoking status in 1000 participants from the Prospective Urban and Rural Epidemiological (PURE) cohort, including self-identified never smokers and current smokers. For the first time, we characterize recent tobacco smoke exposures and nicotine dependence in verified current smokers from fourteen different countries based on their urinary TNE, NMR, and UGT distributions. This work offers new insights into regional differences in nicotine metabolism, smoking topography, and tobacco-related disease risk across diverse populations when compared to self-reports (17).

4.3 Materials and Methods

4.3.1 Study Design

The PURE study is an international prospective cohort that enrolled 166,762 participants (aged 35-70 years) from 21 countries between January 2001 and October 2016 (17). A sub-study of 8073 participants from fourteen countries were selected based on a nested case-cohort design (17), from which urine samples of 1000 participants with one or less freeze-thaw cycles were randomly selected in about equal proportions of never smokers (n=335), as well as light (< 10 cigarettes per day, n=324), and heavy (≥ 10 cigarettes per day, n=341) current smokers (17). All

samples were morning void urine collected from consenting adults and subsequently aliquoted prior to storage at -80°C using a standardized procedure across all clinical sites. Standardized questionnaires were completed by all participants in the PURE study, including age, sex, education, household income, location, smoking status, alcohol use, diet, physical activity, psychosocial stress, and use of medications, as described in detail elsewhere (51).

4.3.2 Chemicals and Reagents

Ultra HPLC grade LC-MS solvents (water, methanol, acetonitrile) were obtained from Caledon Laboratories Ltd (Georgetown, ON, Canada) for preparation of buffer, sheath liquid and stock solutions. nicotine, cotinine, nicotine-*N*-oxide, in addition to their matched deuterated internal standards (nicotine-d_3, cotinine-d_3, nicotine-*N*-oxide-d_3) were purchased from Cayman Chemicals (Ann Arbor, MICH, USA). Glucuronide conjugates including nicotine glucuronide, nicotine-d_3-glucuronide, Cotinine-glucuronide, and 3-hydroxycotinine glucuronide were purchased from Toronto Research Chemicals Inc. (Toronto, ON, Canada) while all other chemicals were obtained from Sigma-Aldrich Inc. (St. Louis, MO, USA).

4.3.3 Multisegment Injection-Capillary Electrophoresis-Mass Spectrometry (MSI-CE-MS)

Urine aliquots were thawed slowly on ice, vortexed for 30 seconds and then centrifuged to sediment particulates (14000 *g* for 5 min). Samples were subsequently diluted five-fold with HPLC-grade water and an internal standard mixture containing nicotine-d_3, cotinine-d_3, nicotine-*N*-oxide-d_3, 3-hydroxycotinine-d_3, nicotine-d_3-glucuronide, 3-chlorotyrosine (Cl-Tyr), D-glucose-$^{13}C_6$, choline-d_9, gaba-d_6, and 2-napthalenesulfonic acid (NMS) with a final concentration of 20 μM. All diluted urine samples were analyzed directly on an Agilent G1700A

capillary electrophoresis (CE) unit (Agilent Technologies Inc., Santa Clara, CA, USA) coupled to an Agilent 6230 time-of-flight mass spectrometer (TOF-MS) with a coaxial sheath liquid electrospray ionization interface. An Agilent 1260 infinity isocratic pump and a 1260 infinity degasser were used to deliver 60 % vol MeOH with 0.1 % vol formic acid sheath liquid for positive ion mode at a flow rate of 10 μL/min. For real-time mass correction, reference ions purine and hexakis (2,2,3,3-tetrafluoropropoxy)phosphazine (HP-921) were spiked into the sheath liquid at 0.02 % vol. The nebulizer spray was turned off during serial sample injections to avoid suctioning effects and current instabilities (52). It was then programmed to turn on at 4 psi following voltage application with the source temperature set to 300 °C, and the drying gas delivered at 8 L/min. The Vcap was set to 2000 V, and the nebulizer gas was delivered at 10 psi, while the sheath gas was set to 195 °C and delivered at 3.5 L/min. Additionally, MS voltage settings were set to a fragmentor voltage of 120 V, the skimmer at 65 V, and Oct1 RF at 750 V. Data acquisition was performed in full-scan mode on the TOF-MS system set to a mass range of 50-1700 *m/z* at an acquisition rate of 500 ms/spectrum.

Separations were performed on bare fused-silica capillaries with a 50 μm internal diameter, 360 μm outer diameter and 135 cm total length (Polymicro Technologies Inc., AZ). A capillary window maker (MicroSolv Leland, NC, USA) was used to remove ~7 mm of the polyamide coating on both ends of the capillary. The background electrolyte (BGE) used for electrophoretic separation consisted of 1.0 M formic acid with 15 % vol acetonitrile (pH 1.80) for cationic metabolite analysis. The applied voltage during CE separations was set to 30 kV at 25 °C. A serial sample injection sequence was also used, which consisted of 11 discrete hydrodynamic sample injections (5 seconds at 100 mbar) interspaced with a BGE spacer injected electrokinetically at 30 kV for 75 seconds (53) (**Figure S4.1A**). Between all runs the capillary

was flushed for five minutes with BGE at 950 mbar. In addition, the TOF-MS system was calibrated daily prior to analysis using an Agilent tune mix within a mass error of ± 0.2 ppm. Preventative maintenance was also performed each morning including cleaning of the ion source with 50 % vol isopropanol using a lint-free cloth and cleaning of the CE electrode with MeOH and HPLC-grade water. Subsequently, the capillary was flushed with BGE for 15 min prior to running a pooled quality control (QC) sample (prepared from all 1000 PURE participants) and a blank to assess system stability before beginning a randomized analysis of batches of diluted urine samples from PURE. Overall, the individual urine samples were analyzed over six consecutive days by MSI-CE-MS under positive ion mode conditions. High resolution tandem mass spectrometry (MS/MS) spectra were acquired for verification of nicotine metabolites from a pooled heavy smoker QC urine sample on an Agilent G7100A CE system with a coaxial sheath liquid Jetstream electrospray ion source connected to an Agilent 6550 iFunnel QTOF instrument. A single urine sample was injected hydrodynamically at 100 mbar for 90 seconds, followed by a BGE spacer at 100 mbar for 5 seconds. Precursor ions were then selected for collision-induced dissociation experiments at varying collision energies (10-40 V). The ESI conditions were set to a Vcap of 3500 V, with a nozzle voltage at 2000 V, nebulizer gas set to 8 psi, drying gas delivered at 14 L/min at 225 °C, while the MS voltage settings included a fragmentor of 380 V and Oct1 RF set to 750 V. Subsequently, acquired MS/MS spectra were annotated based on characteristic product ions, neutral mass losses, and given an unambiguous identification (level 1) if confirmed with a matching MS/MS spectra and co-migration with an authentic standard obtained under the same experimental conditions. While a level 2 identification was determined based on the comparison of acquired MS/MS spectra to an open-access public database (e.g., HMDB) (54).

4.3.4 Method Validation and Quality Control

Stock solutions for calibrants were initially prepared in MeOH (1.0 mg/mL) and then diluted in HPLC-grade water (0.1 mg/mL) followed by serial dilution to prepare calibrant solutions in triplicate ranging from 40-3000 ng/mL for nicotine (0.25-18.5 μM) and cotinine (0.23-17.0 μM), 40-600 ng/mL (0.22-3.36 μM) for nicotine-*N*-oxide, 40-5000 ng/mL for both 3-hydroxycotinine (0.52-26.0 μM), 3-hydroxycotinine glucuronide (0.27-13.6 μM), 40-1000 ng/mL (0.12-2.95 μM) for nicotine glucuronide and 40-6000 ng/mL (0.28-17.0 μM) for cotinine glucuronide. A pooled QC sample was run in every MSI-CE-MS to monitor for technical precision together with 10 individual urine samples from PURE participants that were analyzed in a randomized injection order. All integrated peak areas were normalized to their respective deuterated internal standard, apart from 3-hydroxycotinine glucuronide and cotinine glucuronide which were normalized to Cl-Tyr. The limit of detection (LOD) and limit of quantification (LOQ) for each nicotine metabolite was determined based on the serial dilutions of calibrant solutions that generated a signal-to-noise ratio (*S/N*) of ~ 3 and ~ 10, respectively. Method accuracy was also evaluated through spike-recovery studies at three distinct concentrations for each nicotine metabolite. All spikes were performed in pooled never smoker urine samples, and percent recovery was determined based on the percentage difference between the spiked and known concentrations. Additionally, inter-day reproducibility (CV%) was evaluated over the entire course of data acquisition by MSI-CE-MS based on repeated analysis of all seven nicotine metabolites in pooled urine QCs (n = 156). The chemical stability of all seven nicotine metabolites was evaluated through analysis of spiked pooled never smoker urine following three repeat freeze-thaw cycles together. Also, a delay to storage study using freshly collected urine from a never

smoker analysis that was repeatedly analyzed at room temperature (23 °C) prior to freezing (0, 2, 4, 8, 12, 24 h).

4.3.5 Data Processing and Statistical Analysis

All data acquired by MSI-CE-MS was analyzed with the Agilent Mass Hunter Workstation Software (Qualitative Analysis version B.06.00, Agilent Technologies Inc.). Nicotine metabolites were extracted using a 10 ppm mass window and integrated after smoothing using the quadratic/cubic Savitzky-Golay function (15 points), whereas peak areas, migration times and *S/N* were transferred to Microsoft Excel for calculation of relative peak areas (RPAs) and relative migration times (RMT) to correct for differences in injection volume, as well as matrix induced ion suppression. Analysis of external calibration data, and calculation of TNE, enzyme ratios and of figures of merit were all performed in Excel, while box plots, histograms, scatter plots, receiver operating curves (ROC), and control charts were all made in MedCalc version 12.5 (Ostend, Belgium). Repeated measures ANCOVA and partial Pearson correlation were performed with covariant adjustment (age, sex, BMI, and CPD) on *log*-transformed data, in addition to a Spearman rank correlation analysis on untransformed data using SPSS (IBM SPSS statistics for windows, version 20.0 NY, USA).

4.4 Results

4.4.1 Study Design and Metabolomics Workflow

This nested case-cohort study included a subset of PURE participants (n=1000) from fourteen countries (3 HICs,7 MICs, 4 LICs; **Table S4.1**) with completed tobacco history and available spot morning urine samples (17). More women (65%) self-identified as never smokers, while a

majority of heavy smokers (CPD ≥ 10, 70%) were men as highlighted in **Table 4.1**. Also, heavy smokers had disproportionally a longer smoking history (i.e., pack years of smoking) than light smokers. Amongst current smokers, cigarettes were the most common form of tobacco smoke exposure (91%, *n*=607 of 665), with less than 5% using exclusively smokeless tobacco products (e.g., snuff, chewing tobacco). However, hand-rolled cigarettes (e.g bidi), were found exclusively in LICs in 30% (*n*=40 of 132) of current smokers. Overall, 33% of never smokers reported second-hand smoke exposure (SHS) with 19% reporting daily SHS exposure (≥ 1 time/day). Otherwise, no significant difference in age (mean of 53 years, *p*=0.846) or BMI (mean of 26 kg/m^2, *p*=0.850) was found between current and never smokers from PURE (**Table 4.1**).

This multiplexed separation format allowed for unique serial sample injection configurations designed to encode mass spectral information temporally (50) depending on experimental design. For example, a pooled sub-group analysis of urine samples from PURE participants (i.e., never *versus* light *versus* heavy smokers) were analyzed in triplicate together with a blank sample by MSI-CE-MS, which showed a progressive increase in urinary cotinine levels as a function smoking intensity. In contrast, urinary cotinine was not detected in the pooled urine from never smokers as well as blank sample (**Figure S4.1B**). Subsequent confirmation of urinary cotinine was achieved by high-resolution MS/MS based on two diagnostic products ions after comparison to an authentic standard under the same collision energy conditions (20 V) as shown in a mirror plot (**Figure S4.1C**). Overall, seven nicotine metabolites were analyzed in urine samples from most habitual smokers in the PURE study, including nicotine-*N*-oxide, as well as nicotine, cotinine, 3-hydroxycotinine and their respective glucuronide conjugates, whose sum is defined as TNE-7. Moreover, migrations time shifts were evident when comparing different singly charged nicotine species in MSI-CE-MS (e.g., cotinine

versus 3-hydroxycotinine; cotinine *versus* cotinine glucuronide) based on their characteristic electrophoretic mobilities (i.e., hydrodynamic size) (**Figure S4.1D**).

4.4.2 Method Validation for Reliable Nicotine Metabolite Determination

Next, method validation was performed for all nicotine metabolites based on several figures of merit summarized in **Table S4.2**. Matching deuterated internal standards available for most nicotine metabolites allowed for correction for potential matrix-induced ion suppression from involatile salts and abundant urinary creatinine due to their co-migration (i.e., no deuterium effect). Recovery studies using authentic nicotine standards spiked into pooled never-smoker urine samples (in triplicate) at three concentration levels (3.9 μM, 13 μM, 130 μM), were used to evaluate method accuracy as shown by 3-hydroxycotinine (**Figure S4.2A**), with acceptable recoveries observed for all nicotine metabolites (> 80%; **Table 4.1**). Overall, excellent linearity (R^2 > 0.980) was achieved for all nicotine metabolites when using 7-point calibration curves over a wide linear dynamic range. Method limits of detection (LOD; *S/N*~3) and quantification (LOQ; *S/N*~10) for all nicotine metabolites ranged between 0.03-0.11 μM and 0.06-0.25 μM respectively, while in sample detection limits were 5-fold higher (**Table S4.2**) due to urine dilution. Additionally, technical precision over six days of data acquisition with daily preventative maintenance and mass tuning demonstrated acceptable reproducibility (CV=12.1%, *n*=152) for 3-hydroxycotinine as shown in the control chart of pooled QC samples randomized in every run (**Figure S4.2C**). Precision for the six additional nicotine metabolites was between 15-34%, with low abundance (*S/N* < 15) metabolites such as nicotine-*N*-oxide and nicotine glucuronide having the highest variance (CV of 34% and 32% respectively; **Table S4.2**). Furthermore, analysis of spiked pooled never smoker urine (in triplicate) demonstrated adequate chemical stability over 24 h at room temperature and following three repeat freeze-thaw cycles

Table 4.1 Summary of characteristics and smoking status of participants from the PURE study.

Variable	Never Smokers (*n*=335)	Light Smokers (<10cigarettes/day) (*n*=324)	Heavy Smokers (>10cigarettes/day) (*n*=341)
HICs (n; %)	11.1% (*n*=111)	12.6% (*n*=126)	13.9% (*n*=139)
MICs (n; %)	11.3% (*n*=113)	14.2% (*n*=142)	12.6% (*n*=126)
LICs (n; %)	11.1% (*n*=111)	5.60% (*n*=56)	7.60% (*n*=76)
Sex (n; %)	-	-	-
F	21.9% (*n*=219)	15.9% (*n*=159)	10.2% (*n*=102)
M	11.6% (*n* =116)	15.5% (*n*=165)	23.9% (*n*=239)
Age (mean)	52 ± 10	53 ± 10	53 ± 9
≥ 50y	59 ± 6 (*n*=191)	59 ± 6 (*n*=200)	59 ± 6 (*n*=204)
< 50y	42 ± 5 (*n*=144)	42 ± 4 (*n*=124)	43 ± 4 (*n*=137)
BMI (mean)	27 ± 6	26 ± 5	26 ± 5
Lean (<25 kg/m^2)	21 ± 2 (*n*=118)	21 ± 2 (*n*=132)	21 ± 2 (*n*=156)
Overweight (25-44 kg/m^2)	30 ± 6 (*n*=208)	29 ± 4 (*n*=185)	30 ± 4 (*n*=183)
# of Cigarettes per day (mean)	0	5±2	19±8
Packs per year	0	9±6	37±21
Tobacco (n; %)	-	-	-
Smokeless Only	-	3.0% (*n*=30)	0.40% (*n*=4)
Cigarettes Only	-	29.0% (*n*=290)	31.7% (*n*=317)
Both Smokeless and Cigarette	-	0.40% (*n*=4)	2.0% (*n*=20)
Location (n; %)	-	-	-
Urban	19.8 % (*n*=198)	18.2% (*n*=182)	20.6% (*n*=206)
Rural	13.7% (*n*=137)	14.2% (*n*=142)	13.5% (*n*=135)
Second-hand Smoke (n; %)	-	-	-
Not exposed	-	-	-
Daily (≥ 1 time/day)	6.50 % (*n*=65)	-	-
Weekly (≥ 1 time/week)	4.60 % (*n*=46)	-	-

[a] *Mean values ± standard deviation are shown unless otherwise indicated*

** Statistical significance at p < 0.0001 after adjusting for age, sex, BMI, and CPD*

(**Figure S4.3**). Overall average change (%) in response between time points was less than 15% for all nicotine metabolites (**Table S4.2**), thus allowing for reliable analysis of urine samples after multiple thawing that can also tolerate delays to freezing.

4.4.3 Total Nicotine Equivalents for Robust Tobacco Smoke Exposure Assessment

Relative distribution of individual urinary nicotine species determined based on mean urinary concentrations measured from current smokers (n = 665) in the PURE study are summarized in **Figure S4.4**. As expected, minor nicotine metabolites including nicotine-*N*-oxide and nicotine glucuronide comprising less than 12% of total nicotine, whereas 3-hydroxycotinine was the most abundant urinary nicotine species overall (~ 40-53% of total nicotine). Interestingly, nicotine, was excreted on average at higher levels (17-26%) than cotinine (11-16%) despite cotinine's lower frequency of non-detects (18.3% *versus* 46.6%; **Figure 4.1A**). As expected, the rate of non-detects among individual nicotine species was dependent on their abundance (~ 15-65%), but they were higher as compared to using a panel of two or more nicotine metabolites (TNE-2-7; ~ 10-12%). As a result, urinary TNE-7, which accounts for over 85% of total excreted nicotine, was used to assess recent tobacco smoke exposure in this study. Overall, TNE-7 demonstrated acceptable long-term precision (CV=14%, n=156; **Figure 4.1B)** with greater sensitivity for biochemical verification of smoking status as it is less susceptible to variations in metabolic rate when comparing multi-ethnic populations in the PURE study. The distribution for urinary TNE-7 for verified current smokers and likely misreported never smokers from the PURE study (n=686) is shown in **Figure 4.1C**. Interestingly, the rate of detectable of urinary TNE-7 among self-identified never smokers was higher than anticipated at 26.5% (n=89).

To determine the potential contribution of second-hand smoke (SHS) exposure on detectable urinary TNE-7 concentrations in never smokers, individuals self-reporting recent SHS exposure were compared against, those who indicated no SHS exposure (**Figure S4.5**). Although an upward trend in TNE-7 was observed with self-reported SHS exposure, no significant difference was observed between sub-groups (p = 0.144, n = 89; **Figure S4.5**). A regional

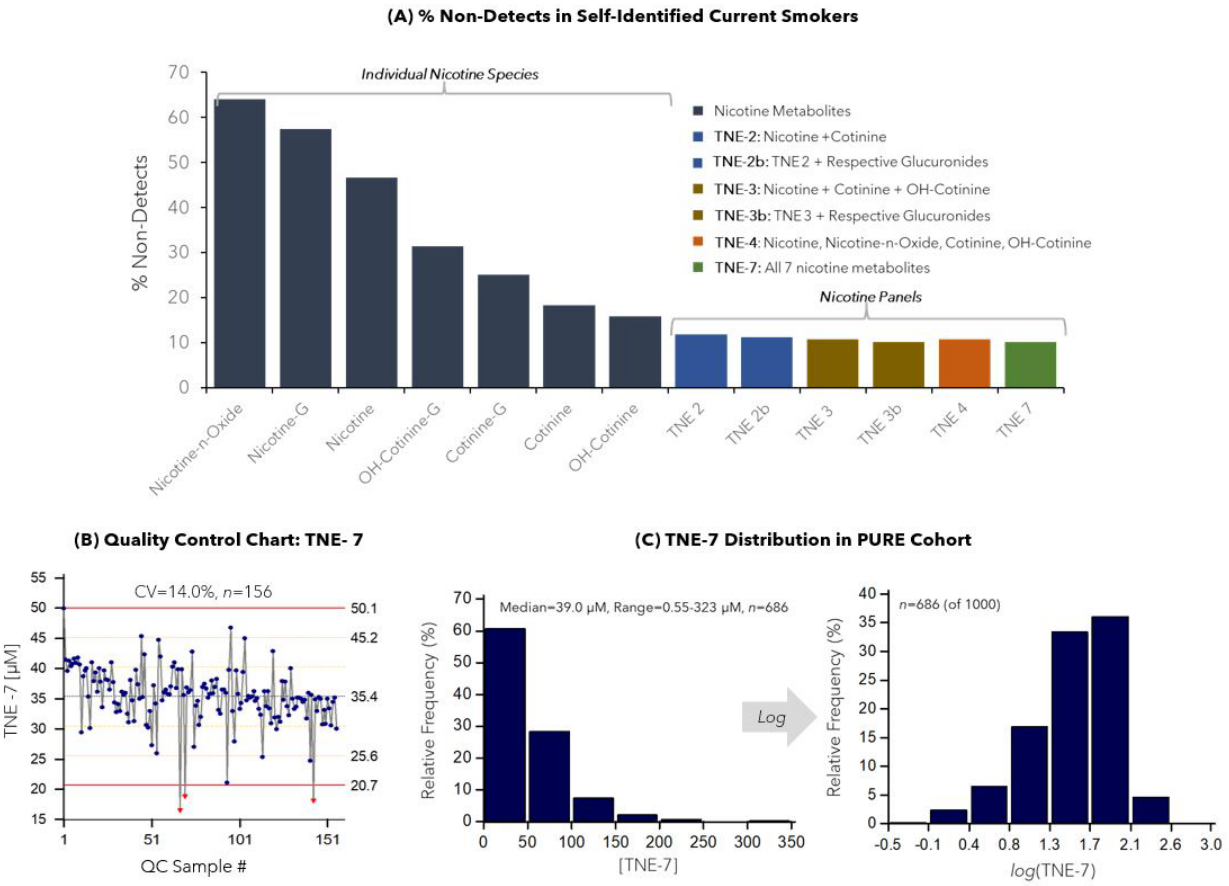


Figure 4.1 (A) Overview of the non-detects in self-identified current smokers (heavy and light smokers) for individual nicotine metabolites compared to the nicotine metabolite panels [TNE (2-7)]. **(B)** A control chart highlighting acceptable long-term (intermediate) precision (CV=14%) based on repeated analysis of TNE-7 (n=156) from a pooled quality control (QC) when using MSI-CE-MS for biochemical verification of smoking status. **(C)** Distribution of TNE-7 concentrations in participants with detectable concentrations (one or more quantifiable nicotine metabolites; n=686) both before and after *log*-transformation where a median 39.0 µM TNE-7, ranging from 0.55-323 µM in the PURE cohort.

comparison of tobacco smoke exposures in the PURE study revealed that heavy smokers exhibited elevated TNE-7 concentrations in HICs, followed by MICs and LICs ($p < 0.001$) after adjusting for age, sex, BMI and CPD (**Figure 4.2A**). In contrast, self-reported light smokers showed no differences in urinary TNE-7 status between country income regions (**Figure 4.2B**). This data also demonstrated greater tobacco smoke exposure among self-identified never smokers in LICs relative to HICs (FC = 6.60, $p = 9.43\times10^{-6}$; **Figure 4.2C**) following adjustments for age, sex, BMI and CPD. In this case, TNE-7 concentrations from these self-reported never

smokers (median of 22.8 μM) were similar in magnitude to light smokers from other regions (median of 24.6 to 30.6 μM, **Figure S4.6**).

A Spearman rank correlation analysis between self-reported smoking intensity (i.e., CPD) and urinary TNE-7 was next applied to evaluate potential misreporting. **Figure S4.7** highlights that there was a moderate correlation between urinary TNE-7 and smoking intensity (CPD) among self-identified current smokers (r =0.391, p = 4.91×10^{-22}, n=564) with the strongest correlation occurring for biochemically verified smokers in HICs (r =0.497, p = $3.72\text{x}10^{-16}$, n=236) followed by MICs (r = 0.425, p = $7.61\text{x}10^{-11}$, n=215). In contrast, there was no correlation observed for urinary TNE-7 with smoking intensity for participant from LICs (r = 0.093, p = 0.328, n=113). An analogous strategy was also used to identify true never smokers from current smokers and misclassified current smokers (**Figure 4.3**). Receiver-operating characteristic (ROC) curves were used to classify regional and sex-dependent cut-off values for urinary TNE-7 in males and females from HICs, MICs and LICs, with missing data replaced by minimum TNE-7 divided by two. Overall, there was good classification of authentic current smokers from never smokers based on the area under the curve (AUC) in all income groups (AUC > 0.77, p < 0.0001). Optimal urinary TNE-7 cut-off values from HICs were 6.58 μM and 4.64 μM for males and females (**Figure 4.3A**), respectively. While in MICs, males had a a significantly lower TNE-7 cut-off (2.09 μM) compared to women smokers (6.99 μM; **Figure 4.3B**). However, lower cut-off values for urinary TNE-7 of 3.65 μM and 2.27 μM were measured for both males and female tobacco smokers from LICs, respectively (**Figure 4.3C**). Importantly, confidence intervals (CI 95%) in LICs, were far more variable compared to MICs and LICs, with females exhibiting the largest variance in recent tobacco smoke exposures. Overall, misreporting was consistently more prevalent for self-reported female never smokers across all regions (**Table**

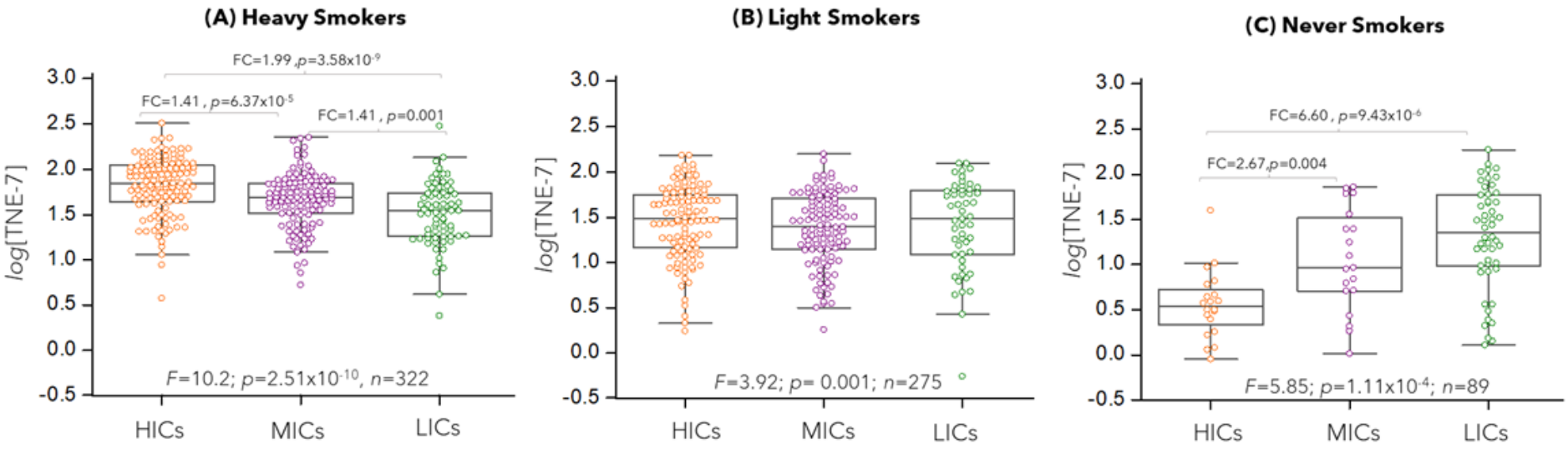


Figure 4.2 Box-whisker plots of TNE-7 comparing HICs, MICs and LICs within self-identified **(A)** heavy smokers, **(B)** light smokers, and **(C)** never smokers, using repeat measures ANCOVA (adjusted for age, sex, BMI, and CPD) to determine regional differences in tobacco exposure. An overall 1.99-fold increase (p=3.58×10^{-9}) in tobacco exposure is observed in HICs relative to LICs. While, among detectable TNE-7 levels in never smokers, LICs demonstrate higher tobacco exposure by 6.6-fold (p=9.43x10^{-6}) compared to HICs.

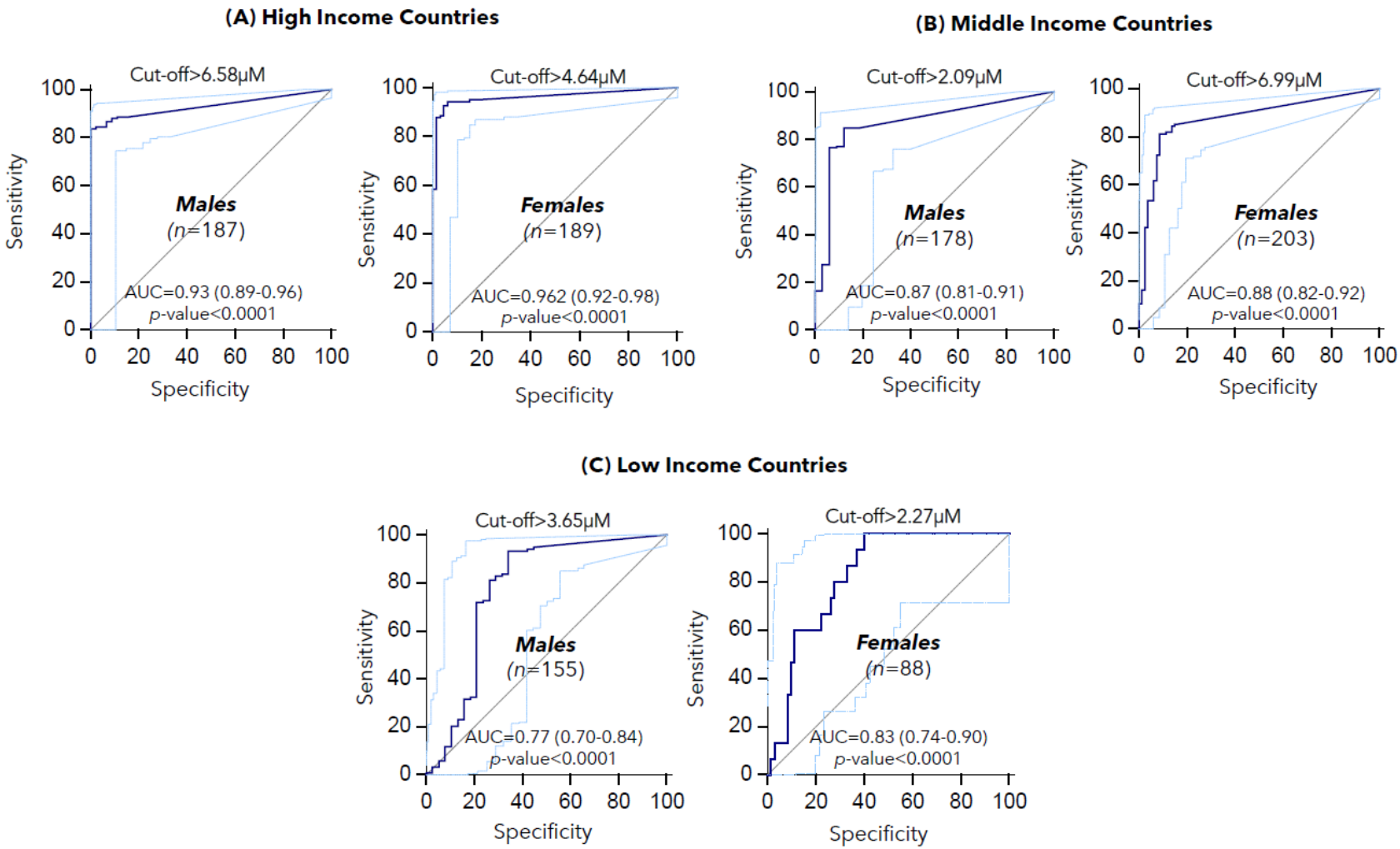


Figure 4.3 Discrimination between current and never smokers was performed to better identify potential misreported smokers from true never smokers using ROC curves for urinary TNE-7 in males and females from **(A)** HICs, **(B)** MICs, and **(C)** LICs. Overall, good discriminatory performance is shown with a 95% confidence interval (AUC > 0.77, $p < 0.0001$) in all regionals for both sexes. Cut-off thresholds for urinary TNE-7 concentrations were dependent on country income status and biological sex, with the highest TNE-7 levels occurring for females from MICs (6.99 μM) followed by males from HICs (6.58 μM). Smokers of both sexes from LICs, such as Bangladesh had consistently lower threshold cut-off values for TNE-7 reflecting reduced tobacco smoke exposures.

S4.3). The highest frequency of misreporting was found in LICs with 45% of never smokers having detectable levels of urinary TNE-7 with a median concentration of 22.8 μM. Conversely, a misreporting rate of 17% and 18% was observed in self-reported never smokers from MICs and HICs respectively, with lower median TNE-7 concentrations ranging from 3.46 to 9.24 μM.

4.4.4 Phenotypic Biomarkers of Nicotine Metabolism

Phenotypic biomarkers of enzyme activity for CYP 2A6 (NMR) and UGT (UGT-cotinine) were determined for verified tobacco smokers in the PURE study using calculated ratios for specific nicotine metabolites as shown in **Figure S4.4**, where a minimum of one metabolite was required in the numerator and denominator. Urinary NMR exhibited the highest frequency of detection (~ 79%) among self-identified current smokers, as well as the widest range of activity (0.23-22; **Table S4.4**). In contrast the less common enzyme ratios UGT nicotine (35,55) was minimally detected (30%, n=201) and demonstrated a smaller range in activity (0.01-2.75; **Table S4.4**). Additionally, technical precision was considered acceptable when measuring NMR, UGT-cotinine and UGT-hydroxycotinine (mean CV=14-19%, n=148) in pooled QC samples. Interestingly, biological variation was highest for UGT 3-hydroxycotinine (173%), as compared to UGT cotinine (87%), and the lowest biological variation occurring for NMR (73%) following partial Pearson correlation analysis between enzyme ratios (after adjustments for age sex, BMI, and CPD) an expected correlation between all glucuronidation related ratios was observed (r = 0.253-0.332, $p < 0.001$; **Table S4.5**). Furthermore, unlike UGT-hydroxycotinine, UGT-cotinine demonstrated a moderate correlation with both urinary TNE-7 (r = 0.397, $p < 0.0001$) and NMR (r = 0.406, $p < 0.001$). Additional regional differences in CYP 2A6 activity were evaluated (ANCOVA adjusted for age, sex, BMI, CPD) based on urinary NMR in HICs, MICs and LICs (**Figure S4.8**). Enzyme activity was significantly elevated for PURE participants from HICs,

followed by MICs, and LICs, respectively (**Figure S4.8A),** with an overall 1.75-fold (p = 9.29×10^{-9}) increase among smokers in HICs relative to LICs. Furthermore, a moderate increase in CYP 2A6 activity was observed for females (FC of 1.10, p = 0.005) compared to males across all fourteen countries after adjustments for age, BMI and CPD (**Figure S4.8B**).

Due to the risk associated with nicotine dependence due to higher CYP 2A6 activity (29), classification of fast *versus* slower nicotine metabolizers was evaluated following stratification of urinary NMR (56) into quartiles (**Figure 4.4A**). Participants in the fourth quartile (NMR > 5.55) were classified as fast metabolizers, while quartiles 1-3 were slower metabolizers, with a progressive stepwise increase observed between all quartiles (**Figure 4.4B**). Overall, urinary NMR was correlated moderately with TNE-7 (r = 0.580, p=2.69×10^{-50}) and fast nicotine metabolizers had elevated tobacco smoke exposures (FC = 1.90, p = 1.10×10^{-33}) following adjustments for age, sex, BMI and CPD (**Figure 4.4D**). Moreover, a TNE-7 threshold of 59.2µM was determined to have excellent performance in discriminating between fast and slower metabolizers of nicotine in the PURE study [AUC = 0.80 (0.76-0.83), p < 0.0001]. Given the previously outlined regional differences in enzyme activity (**Figure S4.8**), urinary NMR was re-stratified within each country income group, with overall fast metabolizers (Q4) having the highest NMR predominately in HICs relative to LICs (**Figure S4.9**). Subsequently, regional urinary TNE-7 cut-offs values were established in this study, with good classification of current smokers from never smokers found for all three country income regions (AUC $\geq$ 0.80, p < 0.0001; Figure 4D). In general, fast metabolizers of nicotine were predominantly from tobacco smokers in HICs (~ 56%), with no significant difference in age or BMI compared to MICs and LICs (~9-33%; **Table 4.2**). Although, no overall difference in self-reported smoking intensity (i.e., CPD) was reported between county income levels, two markers of tobacco smoke exposure

Table 4.2 Stratification of active smokers from PURE (*n*=568) based on their NMR; classified as fast metabolizers (Q4) of nicotine at higher risk for tobacco exposure and nicotine dependence compared to slower metabolizers (Q1-Q3) with lower risk for tobacco derived toxin exposure and biological harm.

Variable[a]	Fast Metabolizers (*n*=142)	Slower Metabolizers (*n*=426)
Sex (n;%)	-	-
F	48% (*n*=69)	39% (*n*=168)
M	51% (*n*=73)	61% (n=258)
Age	55±15	52±15
BMI	32±13	32±12
Income Status (n;%)	-	-
HICs	56% (*n*=80)	31% (n=133)
MICs	35% (*n*=49)	36% (n=155)
LICs	9% (*n*=13)	32% (n=138)
*NMR**	7.86±4.3	2.99±2.1
*UGT Cotinine**	1.71±1.6	0.81±0.8
*TNE-7**	74.1± 55µM	39.2±39 µM
*Thiocyanate**	92.0±104 µM	52.6±87 µM
Smoking Status (n;%)	-	-
Heavy Smokers	61% (*n*=87)	51% (*n*=218)
Light Smokers	36% (*n*=51)	40% (*n*=169)
Misreported Never Smokers	3% (*n*=4)	9% (*n*=39)
# of Cigarettes per day	10±15	10±10
Pack Years	23±32	16±21

[a] *Median values ± interquartile range (IQR) are shown unless otherwise indicated*
** Statistical significance at p <0.0001 after adjusting for age, sex, BMI, and CPD*

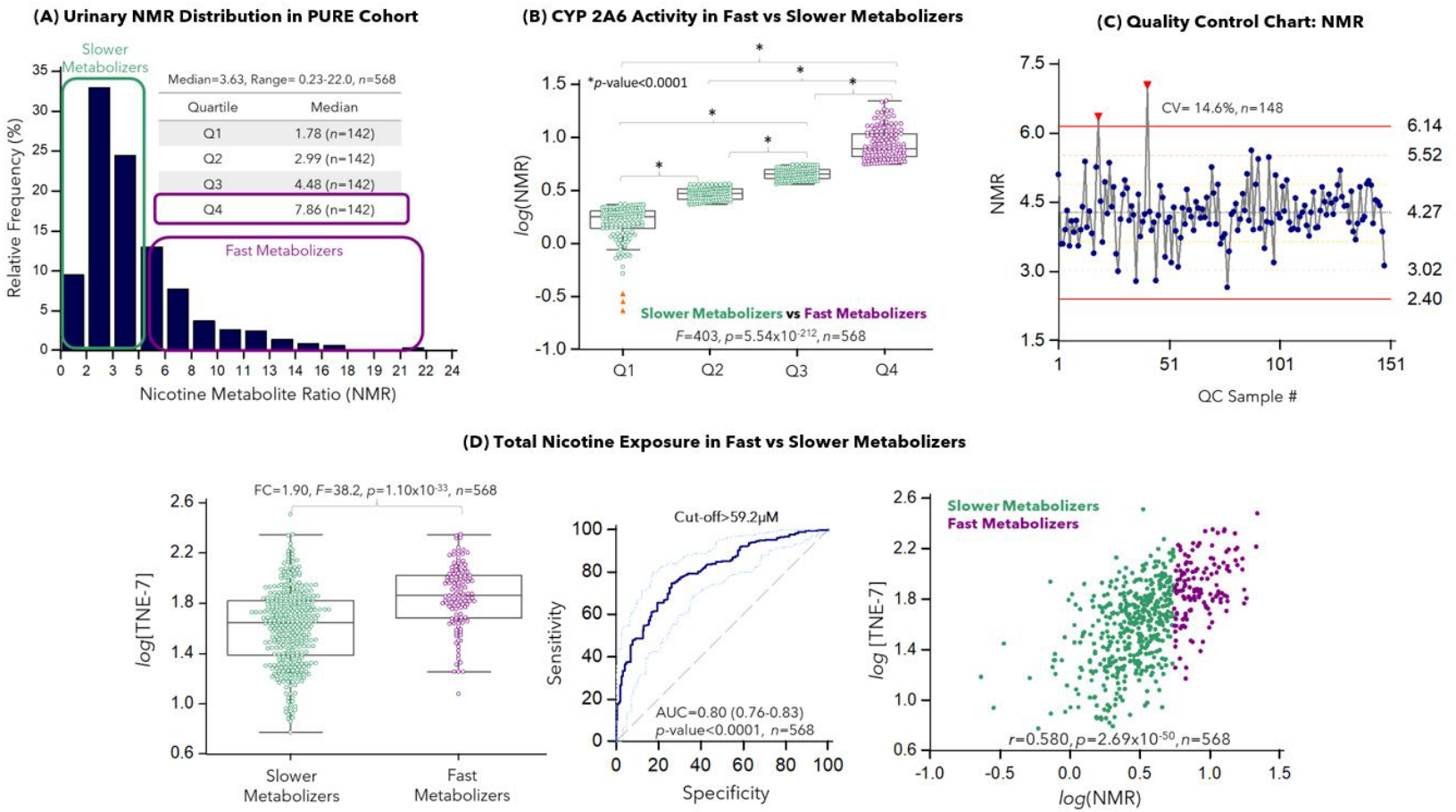


Figure 4.4 (A) Histogram demonstrating the distribution of urinary NMR as a phenotypic biomarker for CYP 2A6 activity in active smokers from PURE (n=568) according to the ratio of total 3-hydroxycotinine to free cotinine. **(B)** Individuals at greater risk for tobacco derived toxicant exposure as a result of increased puff volumes per cigarette are identified as fast metabolizers (Q4) based on stratification of NMR by quartile. **(C)** Quality control (QC) chart indicating excellent technical precision for CYP 2A6 monitoring over the course of a 6-day analysis (n=148). **(D)** Fast metabolizers are most susceptible to biological harm as a result of increased tobacco exposure, determined by TNE-7. Overall, a significant relationship was observed between increasing CYP 2A6 activity and elevated tobacco exposure after correction for age, sex, BMI and CPD as demonstrated by a box-whisker plot with a 1.9-fold (p=1.10×10^{-33}) increase in TNE-7 for fast metabolizers and a strong correlation between NMR and TNE-7 (r=0.580, p=2.69×10^{-50}).

including TNE-7 and thiocyanate, were significantly elevated in fast metabolizers (**Table 4.2**). Following a secondary stratification of fast metabolizers, urinary NMR continued to correlate with TNE-7 (r = 0.356, p=2.38×10^{-5}; **Figure S4.10**), and a sub-set of high-risk tobacco smokers prone to nicotine dependence (NMR > 10.9, n=35) having the highest CYP 2A6 activity within the PURE study were identified. As expected, high-risk fast metabolizers of nicotine consisted of primarily heavy smokers (~ 77%), with significantly elevated smoking intensity and duration (CPD, PY), recent tobacco smoke exposure (TNE-7), cyanide exposure (thiocyanate) and detoxification capacity (UGT-cotinine) (**Table S4.6**). Analogous to NMR, HICs had significantly elevated glucuronidation rate, relative to MICs and LICs (FC > 1.55, p < 0.0001; **Figure S4.11A**). Similarly, female smokers from across all 14 countries in the PURE study also demonstrated elevated UGT-cotinine activity as compared to males (FC = 1.19, p = 8.38×10^{-9}, n =506; **Figure S4.11B**). However, following stratification of UGT-cotinine for verified tobacco smokers in the PURE study, individuals in quartile one were classified as tobacco smokers with slow detoxification capacity at greater risk for potential harm (**Figure S4.11C**) as compared to faster detoxifiers (quartiles 2-4). A lower moderate correlation with urinary TNE-7 was found (r = 0.384, p = 4.73×10^{-19}, n=506; **Figure S4.11D**) relative to urinary NMR following adjustments for age, sex, BMI and CPD.

4.5 Discussion

4.5.1 High-Throughput Analysis of Urinary Nicotine Metabolites

Accurate measurement of nicotine and its complete array of metabolites is critical for accurate biomonitoring of tobacco smoke exposure, as well as nicotine dependence along and smoking topography (23,25,57). However, given the extensive variability in nicotine elimination, most notably for polar glucuronide conjugates (~ 22-40% in urine) (44) the choice of analytical

method is critical. Most commonly, direct, and indirect LC-MS/MS assays (22,35,46) have been reported for the quantification of various nicotine species, with several large-scale epidemiological studies (e.g., PATH and NHANES) currently utilizing both protocols (21,33). Indirect approaches estimate glucuronide concentrations based on a repeat analysis with (total) and without (free) enzyme deconjugation, while direct quantification techniques analyze free nicotine metabolites and their intact glucuronides within a single assay (22,35). For the first time, we demonstrate that MSI-CE-MS offers a simple and cost-effective platform for the direct analysis of urinary nicotine metabolites and their glucuronide conjugates with higher sample throughput than conventional single injection and gradient elution LC-MS/MS methods. Overall, eliminating the need for enzyme hydrolysis and protein denaturation, while avoiding bias associated with inter-batch variations and suboptimal incubation times resulting in incomplete deconjugation (58). In a recent comparison of eleven LC-MS/MS methods (22), direct assays were found to have improved precision and measure consistently higher glucuronide content than indirect assays, with the most notable difference reported for 3-hydroxycotinine glucuronide (CV = 25-52%). However, direct LC-MS/MS methods using reversed-phase chromatographic separations often face limitations caused by poor retention and resolution of polar/ionic glucuronide conjugates requiring longer total analysis times (> 10 min) (22,35). To date, only one prior CE-MS method has been applied in this context (59). Garcia-Perez et al. (60) demonstrate a proof-of principle metabolomics approach towards discriminating smokers from never smokers using cigarettes of varying tar content through urine analysis. However, this was done with modest samples sizes (n = 121), within a single country, and required longer analysis times (~ 15 min).

In this work, method validation outlined by several figures of merit (**Table S4.2**) confirmed the reliable quantification of seven nicotine metabolites with good selectivity, accuracy, technical precision, and long-term chemical stability, consistent with previous studies (22,61). On average CVs range from 9-56%, for LC-MS/MS methods, with minor nicotine metabolites (e.g., nicotine-n-oxide and nicotine glucuronide) and more polar glucuronide conjugates (e.g., 3-hydroxycotinine) often having higher variability (22). Comparatively, MSI-CE-MS demonstrated technical precision within a similar range (12-34%), where low abundance nicotine species had the highest technical precision (31-34%). In contrast, more polar glucuronide conjugates (e.g., cotinine glucuronide, 3-hydroxycotinine glucuronide) had consistently lower CVs (< 21%). As expected, detection and quantification limits in our method are 10 to 100-fold higher than other LC-MS/MS assays (35,46,47,61,62) given the necessary dilution, and lack of sample pre-treatment (e.g., sample preconcentration). Commonly reported method concentration ranges for nicotine, cotinine and 3-hydroxycotinine are between 1-5000 ng/mL (47,61), while we report concentrations above 100 ng/mL in sample. Despite, higher detection limits this method can detect habitual smokers, with measurable levels of cotinine and 3-hydroxycotinine in the range of concentrations found in active smokers (> 200 ng/mL) (22,45,63). Moreover, when comparing mean molar percentages of individual nicotine metabolites (**Figure S4.4**), our analysis demonstrated excretion profiles comparable to other studies (46,61), with 3-hydroxycotinine consistently found to be the most abundance product in urine (35-50%). Comparatively, nicotine was found to be ~10% higher relative to previous reports (29), likely attributed to sampling time for urine collection relative to tobacco exposure.

4.5.2 Robust Biomarkers for Tobacco Smoke Exposure

Given the extensive interindividual variability in nicotine metabolism, the sum of excreted nicotine species (TNE) is considered the gold standard in tobacco exposure assessment (34). Although the definition and nomenclature for TNE varies in the literature, TNE-2/4 (total cotinine, total 3-hydroxycotinine) and TNE-3/6 (total nicotine, total cotinine, total 3-hydroxycotinine) are most frequently used and account for > 75% of excreted metabolites (21,22,33,35). The inclusion of three additional minor nicotine species has been referred to as TNE-9 (64,65), and considered the most complete, however it is minimally utilized given cost and time for analysis. Large-scale studies such as NHANES has reported TNE-2, TNE-3 and TNE-9 in active smokers, with mean (geometric) concentrations of 34 μM, 41 μM, and 47 μM respectively (21). Similarly, self-identified current smokers from HICs in this work had similar mean (geometric) concentrations of 37.8 μM and 42.1 μM for TNE-2 and TNE-3 respectively. Expectedly the geometric mean of TNE-7 reported by MSI-CE-MS was 43.3 μM, just slightly lower than TNE-9 levels described by NHANES, given the exclusion of three additional minor nicotine metabolites (e.g., nornicotine, norcotinine, cotinine-n-oxide) excreted at less than 10% in urine (21). However, it should be noted that accuracy of TNE-2 has been reported to decrease in light smokers and slow metabolizers (26). It is evident that the three additional minor nicotine metabolites have minimal impact on the total sum, while more importantly total nicotine, cotinine and 3-hydroxycotinine make up a substantial portion of the excreted metabolites (> 80%) and are more likely to influence the biochemical verification of smoking status (35). In contrast, for never smokers, NHANES reported TNE-2 levels of 0.013 μM, while our method measured levels at 4.36 μM among never smokers in HICs, indicative that given our detection limits these cases are likely misreported current smokers (21).

Previous studies have evaluated and reported the correlation between self-reported CPD and various urinary biomarkers of tobacco use, with substantial variability depending on the biomarker and ethnicity (66–68). Similarly, amongst all PURE participants a moderate yet significant correlation was determined (ρ = 0.391, p= 4.91×10-22), with the strongest correlations occurring in HICs and MICs ($\rho > 0.425, p < 0.0001$; **Figure S4.7**). Notably, no correlation was observed between CPD and TNE-7 among LICs. Although sampling time may influence correlation, especially among intermittent smokers, socioeconomic status (SES) has been shown to have an impact on misreporting (13,69). As a result, discrimination between current and never smokers was performed to better identify potential misreported smokers from true never smokers, where we observe distinct cut-off thresholds dependent on both country income status and biological sex, with the highest TNE-7 levels occurring for females from MICs and HICs, while smokers of both sexes from LICs, (e.g., Bangladesh) had consistently lower threshold cut-offs (**Figure 4.3**). This highlights potential self-reporting errors as a result regional cultural factors, social desirability, and/or recall biases, more likely to be prevalent in LICs at the time of sampling. This is further supported by a detection rate of ~ 27% among never smokers, where 56% are from LICs (**Table S4.3**), with TNE-7 concentrations in the range of light smokers (22.8 μM; **Figure 4.6**). Similarly, a recent study by Park *et al*. (70), emphasizes the importance of optimal cut-off values, with consideration for age and sex, as they demonstrate a critical role in monitoring the effectiveness tobacco control policies. Moreover, evolving changes in public policies warrant on-going biomonitoring studies using objective biomarkers of tobacco smoke exposure, as relying solely on self-reports may provide inaccurate information especially when comparing tobacco smoking trends between countries globally.

Notable, regional differences were also observed among heavy smokers after adjusting for CPD, age, sex and BMI (**Figure 4.2**). The 2-fold increase in nicotine exposures in smokers from HICs as compared to other regions despite adjustment for smoking intensity may be attributed to differences in metabolism and/or cigarette brands which can influence smoking behavior (17). For example, ventilated and low-yield cigarettes, more common in HICs, (17,71,72) have been associated with increased puff volumes, resulting in elevated nicotine exposure (71,73,74). In addition, increased CYP 2A6 metabolism has been associated with upward of 51% of variance in interindividual puff volume, where fast metabolizers exhibit deeper inhalation per cigarette (6). Discrepancies in biomarker-based tobacco exposure between regions may help to explain an apparent increase in the hazards of tobacco smoking (relative to never smoking), such as greater all-cause mortality and clinical events (e.g., cardiovascular disease, cancer, and respiratory disease), in richer as compared to poorer countries when relying of self-reports alone (17).

4.5.3 Tobacco-Related Disease Risk Among Current Smokers

The urinary NMR ratio involving 3-hydroxycotinine and cotinine has been well established as a predictor of CYP 2A6 activity, capturing both genetic and environmental influences on nicotine metabolism (6,16,29,41). In our work, a median urinary NMR of 3.63 was measured with a 96-fold range in CYP 2A6 activity ranging from 0.23-22.0. Given that greater disease risk has been associated with elevated NMR ratios (16,29,75,76), stratification of NMR into quartiles has been recommended to classify fast from slower nicotine metabolizers (56). Although, no standardized cut-offs have been reported, tobacco smokers in the upper quartile have consistently shown greater nicotine/intoxicant exposures, nicotine dependence, and oxidative stress or biological

harm (56). Our urinary NMR data comprised participants from fourteen different countries, where TNE-7 verified current smokers were classified as higher-risk, fast (Q4) as compared to slower (Q1-Q3) nicotine metabolizers (**Figure 4.3**). To evaluate tobacco-related disease risk, urinary TNE-7 was monitored between fast and slower metabolizers, following adjustments for age, sex, BMI and CPD. Although self-reported CPD was not different between NMR sub-groups (**Table 4.2**), tobacco exposure was significantly elevated in fast as compared to slower metabolizers that also demonstrated a dose-response relationship with increasing NMR (**Figure 4.3**). This is indicative of greater inhalation volumes per cigarette and consistent with recent studies, where fast nicotine metabolizers demonstrated elevated puff volumes of about 25% (6). In addition, urinary thiocyanate, a biomarker for cyanide exposure excreted in high levels with active tobacco smoking was also significantly elevated among the fast nicotine metabolizers, indicating an increase in toxicant exposure (**Table 4.2**). Consequently, tobacco smokers classified as fast metabolizers are at greater risk for smoking dependence and increased toxicant exposure as a result of heightened tobacco intake (75). Continued exploration of fast metabolizers (**Figure S4.10**) shows a continued step wise increase in NMR, correlated with TNE-7, emphasizing how tobacco-related disease risk may not plateau, but rather continue to increase. Interestingly, only among these further stratified high-risk individuals is an increase observed for CPD, indicating how self-reports may not capture important smoking behavioral patterns as a result of differences in nicotine metabolism. Overall, high-risk, fast metabolizers of nicotine were primarily from HICs, where elevated median urinary NMR levels were detected in quartile four (**Figure S4.9**, **Figure S4.10**). These results are consistent with Sathish *et al.* (17) who reported that current smokers from HICs had the highest hazard ratios reflected by greater urinary TNE-7 concentrations as compared to smokers from other country income regions. Our

work provides greater insight into these trends based on urinary NMR stratification that likely reflects differences in population genetics and the brands of commercial cigarette products consumed in HICs.

Given the involvement of glucuronidation in secondary nicotine metabolism, and detoxification of toxicants (e.g., nitrosamines), UGT activity may also impact disease risk (31,35) as direct analysis of cotinine and its glucuronide have demonstrate a strong dose relationship with UGT 2B10 (35). As such UGT cotinine activity was evaluated, where quartile one individuals were identified as slow detoxifiers, and are more likely to have a loss of function genetic variant of UGT 2B10 (**Figure S4.11**). These individuals may be at greater risk for retaining bioactivated toxicants from smoke (e.g., tobacco specific nitrosamines) (31,44,77,78). Although a minimal correlation was observed between urinary TNE-7 and UGT cotinine, it likely plays a more important role in detoxification of other toxicants compared to smoking topography. This highlights a secondary mechanism which may put current smokers at greater risk for tobacco-related disease outcomes.

Overall, MSI-CE-MS offers a simple, cost-effective approach with stringent quality control for the rapid analysis of a broad panel of nicotine metabolites that can support large-scale epidemiological and clinical trials for personalized smoking cessation interventions. This method demonstrates rapid analysis times (< 5 min/sample), and acceptable reproducibility compared to direct LC-MS/MS methods, with improved precision for the detection of intact glucuronide conjugates, while mitigating the need for complex sample workup procedures which may introduce bias and lower recovery (e.g., enzyme hydrolysis, solid-phase extraction). Moreover, this approach takes advantage of the multiplexing capabilities as compared to conventional separation techniques, whereby we outline a rigorous validation for high throughput

biomonitoring of nicotine metabolism in an international cohort of current and never smokers. Although limitations in concentration sensitivity did not allow for detection of low levels of SHS exposure in this work, the use of MSI-CE-MS in conjunction with nanospray interfaces and multi-reaction monitoring with tandem mass spectrometry (MS/MS) are recommended. Additionally, future work with prospective clinical data and larger cohort sizes can increase study power and allow for linking elevated toxicant exposures among fast nicotine metabolizers who may be a greater risk for tobacco related deaths and clinical events.

In summary, TNE-7, NMR and UGT provide biochemical verification of regional differences in smoking topography, nicotine metabolism and detoxification, as well as risk for nicotine dependence that are not captured by standardized questionnaires. Quantification of urinary TNE-7 revealed a pattern of a greater misreporting rate among never smokers in LICs, which may be a result of gender and social desirability bias commonly associated with self-reported questionnaires. In such instances, self-identified never-smokers exhibit TNE-7 concentrations similar to light smokers found in other countries, thus demonstrating the limitations of epidemiological survey-based biomonitoring strategies. Furthermore, biochemical verification following adjustments for covariates demonstrated increased tobacco smoke exposure and faster nicotine metabolism among current smokers predominantly from HICs, as compared to MICs and LICs, which may elucidate their apparent elevated tobacco disk risk as compared to other regions that also assumed no disproportionate rate of misreporting in never smokers (17). In summary, this work is the first biomarker-based study to assess tobacco smoke exposure and nicotine clearance from participants across fourteen different countries with diverse multi-ethnic populations for characterization of regional variations in smoking topography, and biochemically verified misreporting. Given the prevalence of tobacco smoking

worldwide, MSI-CE-MS offers a strategy for rapid and reliable biochemical verification of tobacco exposure, ideal for large-scale biomonitoring to better guide evidence based global health policies and improve smoking cessation outcomes in active smokers.

4.6 Acknowledgements

The authors acknowledge funding from the Natural Science and Engineering Research Council of Canada, Genome Canada as well as the Hamilton Health Sciences New Investigator Fund. The authors also acknowledge the contribution of the PURE team and the Population Health Research Institute.

4.7 References

1..World Health Organization. WHO global report on trends in prevalence of tobacco use 2000-2025. World Health Organization, 2019.

2. Doll R, Peto R, Boreham J, Sutherland I. Papers Mortality in relation to smoking: 50 years' observations on male British doctors. British Medical Journal 2004;328:1519.

3. Jung, S.Y., Lee, K. and Kim, S. 2014 surgeon General's report: the health consequences of Smoking-50 years of progress: what we should know and what we can learn. Journal of the Korean Society for Research on Nicotine and Tobacco 2014;5:94-107.

4. Boyd, Andrew T., Susan T. Cookson, Mark Anderson, Oleg O. Bilukha, Muireann Brennan, Thomas Handzel, Colleen Hardy et al. Centers for Disease Control and Prevention public health response to humanitarian emergencies, 2007–2016. Emerging infectious diseases 2017;23:196.

5. Chaiton M, Callard C. Mind the Gap: Disparities in Cigarette Smoking in Canada. Tobacco Use Insights. 2019;12:1179173X1983905.

6. Perez-Paramo YX, Lazarus P. Pharmacogenetics factors influencing smoking cessation success; the importance of nicotine metabolism. Expert Opinion on Drug Metabolism and Toxicology 2021;17:333–49.

7. Department of Health U, Services H, for Disease Control C, Center for Chronic Disease Prevention N, Promotion H, on Smoking O. Smoking Cessation: A Report of the Surgeon General. 2010.

8. Gendall P, Hoek J, Stanley J, Jenkins M, Every-Palmer S. Changes in Tobacco Use during the 2020 COVID-19 Lockdown in New Zealand. Nicotine and Tobacco Research 2021;23:866–71.

9. Reitsma MB, Kendrick PJ, Ababneh E, Abbafati C, Abbasi-Kangevari M, Abdoli A, et al. Spatial, temporal, and demographic patterns in prevalence of smoking tobacco use and attributable disease burden in 204 countries and territories, 1990–2019: a systematic analysis from the Global Burden of Disease Study 2019. The Lancet. 2021;397:2337–60.

10. Casetta B, Videla AJ, Bardach A, Morello P, Soto N, Lee K, et al. Association Between Cigarette Smoking Prevalence and Income Level: A Systematic Review and Meta-Analysis. Nicotine Tob Res. 2017;19:1401–7.

11. Mons U. Strengthening tobacco control must remain a global health priority. The Lancet Global Health. 2020;10:161-162.

12. Palipudi KM, Morton J, Hsia J, Andes L, Asma S, Talley B, et al. Methodology of the global adult tobacco survey—2008–2010. Glob Health Promot. 2016;23:3–23.

13. Jena PK, Kishore J, Jahnavi G. Correlates of digit bias in self-reporting of cigarette per day (CPD) frequency: results from Global Adult Tobacco Survey (GATS), India and its implications. Asian Pacific journal of cancer prevention. 2013;14:3865–9.

14. Soulakova JN, Hartman AM, Liu B, Willis GB, Augustine S. Reliability of adult self-reported smoking history: data from the tobacco use supplement to the current population survey 2002–2003 cohort. Nicotine & Tobacco Research. 2012;14:952–60.

15. Amos A, Greaves L, Nichter M, Bloch M. Women and tobacco: A call for including gender in tobacco control research, policy and practice. Tobacco Control. 2012;21:236–43.

16. Strasser AA, Benowitz NL, Pinto AG, Tang KZ, Hecht SS, Carmella SG, et al. Nicotine metabolite ratio predicts smoking topography and carcinogen biomarker level. Cancer Epidemiology Biomarkers and Prevention. 2011;20:234–8.

17. Sathish T, Teo KK, Britz-McKibbin P, Gill B, Islam S, Paré G, et al. Variations in risks from smoking between high-income, middle-income, and low-income countries: an analysis of data from 179 000 participants from 63 countries. Lancet Global Health 2022;10:216–26.

18. Hyland A, Ambrose BK, Conway KP, Borek N, Lambert E, Carusi C, et al. Design and methods of the Population Assessment of Tobacco and Health (PATH) Study. Tobacco control 2017;26:371-8.

19. Bergen AW, McMahan CS, McGee S, Ervin CM, Tindle HA, le Marchand L, et al. Multiethnic Prediction of Nicotine Biomarkers and Association With Nicotine Dependence. Nicotine Tob Res. 2021;23:2162–9.

20. Murphy SE. Nicotine metabolism and smoking: Ethnic differences in the role of P450 2A6. Chemical Research in Toxicology. 2017;30:410–9.

21. Mazumder S, Shia W, Bendik PB, Achilihu H, Sosnoff CS, Alexander JR, et al. Nicotine Exposure in the U.S. Population: Total Urinary Nicotine Biomarkers in NHANES 2015–2016. International Journal of Environmental Research and Public Health. 2022;19:3660.

22. Wang L, Bernert JT, Benowitz NL, Feng J, Jacob P, McGahee E, et al. Collaborative method performance study of the measurement of nicotine, its metabolites, and total nicotine equivalents in human urine. Cancer Epidemiology Biomarkers and Prevention. 2018;27:1083–90.

23. Habibagahi A, Alderman N, Kubwabo C. A review of the analysis of biomarkers of exposure to tobacco and vaping products. Analytical Methods. 2020;12:4276–302.

24. Feng J, Sosnoff CS, Bernert JT, Blount BC, Li Y, Valle-Pinero D, et al. Urinary Nicotine Metabolites and Self-Reported Tobacco Use Among Adults in the Population Assessment of Tobacco and Health (PATH) Study, 2013–2014. Nicotine & Tobacco Research. 2022;24:768-77

25. Benowitz NL, Hukkanen J, Jacob P. Nicotine chemistry, metabolism, kinetics and biomarkers. Nicotine Psychopharmacology. 2009;29–60.

26. Benowitz NL, St Helen G, Nardone N, Cox LS, Jacob P. Urine Metabolites for Estimating Daily Intake of Nicotine from Cigarette Smoking. Nicotine and Tobacco Research. 2020;22:288–92.

27. Won Hong J, Hyun Noh J, Kim DJ. The prevalence of and factors associated with urinary cotinine-verified smoking in Korean adults: The 2008–2011 Korea national health and nutrition examination survey. PLoS ONE. 2018;13:e0198814.

28. Matsumoto A, Matsumoto A, Ichiba M, Payton NM, Oishi H, Hara M. Simultaneous measurement of urinary total nicotine and cotinine as biomarkers of active and passive smoking among Japanese individuals. Environmental Health and Preventive Medicine. 2013;18:244–50.

29. Murphy SE. Biochemistry of nicotine metabolism and its relevance to lung cancer. Journal of Biological Chemistry; 2021;296.

30.Murphy SE. Nicotine metabolism and smoking: Ethnic differences in the role of P450 2A6. Chemical Research in Toxicology. 2017;30:410–9.

31. Taghavi T, St Helen G, Benowitz NL, Tyndale RF. Effect of UGT2B10, UGT2B17, FMO3, and OCT2 genetic variation on nicotine and cotinine pharmacokinetics and smoking in African Americans. Pharmacogenetics and Genomics. 2017;27:143–54.

32. Borrego-Soto G, Perez-Paramo YX, Chen G, Santuario-Facio SK, Santos-Guzman J, Posadas-Valay R, et al. Genetic variants in CYP2A6 and UGT1A9 genes associated with urinary nicotine metabolites in young Mexican smokers. Pharmacogenomics Journal. 2020;20:586–94.

33. Edwards KC, Naz T, Stanton CA, Goniewicz ML, Hatsukami DK, Smith DM, et al. Urinary cotinine and cotinine + trans-30-hydroxycotinine (TNE-2) cut-points for distinguishing tobacco use from nonuse in the United States: PATH study (2013-2014). Cancer Epidemiology Biomarkers and Prevention. 2021;30:1175–84.

34. Benowitz NL, Bernert JT, Foulds J, Hecht SS, Jacob P, Jarvis MJ, et al. Biochemical Verification of Tobacco Use and Abstinence: 2019 Update. Nicotine and Tobacco Research. 2020;22:1086–97.

35. Taghavi T, Novalen M, Lerman C, George TP, Tyndale RF. A comparison of direct and indirect analytical approaches to measuring total nicotine equivalents in urine. Cancer Epidemiology Biomarkers and Prevention. 2018;27:882–91.

36. Benowitz NL. Clinical pharmacology of nicotine: implications for understanding, preventing, and treating tobacco addiction. Clinical Pharmacology & Therapeutics. 2008;83:531–41.

37. Ray R, Tyndale RF, Lerman C. Nicotine dependence pharmacogenetics: Role of genetic variation in nicotine-metabolizing enzymes. Journal of Neurogenetics. 2009;23:252–61.

38. Fix B v, O RJ, Benowitz N, Heckman BW, Michael Cummings K, Fong GT, et al. Nicotine metabolite ratio (NMR) prospectively predicts smoking relapse: Longitudinal findings from ITC Surveys in five countries. 2017;19:1040-7

39. Giratallah HK, Chenoweth MJ, Addo N, Ahluwalia JS, Cox LS, Lerman C, et al. Nicotine metabolite ratio: Comparison of the three urinary versions to the plasma version and nicotine clearance in three clinical studies. Drug and Alcohol Dependence. 2021;223.

40. Dempsey D, Tutka P, Jacob P, Allen F, Schoedel K, Tyndale RF, et al. Nicotine metabolite ratio as an index of cytochrome P450 2A6 metabolic activity. Clinical Pharmacology and Therapeutics; 2004;76:64–72.

41. Allenby CE, Boylan KA, Lerman C, Falcone M. Precision Medicine for Tobacco Dependence: Development and Validation of the Nicotine Metabolite Ratio. Journal of Neuroimmune Pharmacology. 2016;11:471–83.

42. Taghavi T, Novalen M, Lerman C, George TP, Tyndale RF. A comparison of direct and indirect analytical approaches to measuring total nicotine equivalents in urine. Cancer Epidemiology Biomarkers and Prevention. 2018;27:882–91.

43. Wassenaar CA, Conti D v., Das S, Chen P, Cook EH, Ratain MJ, et al. UGT1A and UGT2B genetic variation alters nicotine and nitrosamine glucuronidation in European and African American smokers. Cancer Epidemiology Biomarkers and Prevention. 2015;24:94–104.

44. Murphy SE, Park SSL, Thompson EF, Wilkens LR, Patel Y, Stram DO, et al. Nicotine N-glucuronidation relative to N-oxidation and C-oxidation and UGT2B10 genotype in five ethnic/racial groups. Carcinogenesis. 2014;35:2526–33.

45. El-Khoury JM, Wang S. Recent advances in MS methods for nicotine and metabolite analysis in human matrices: Clinical perspectives. Bioanalysis. 2014;6:2171–83.

46. Piller M, Gilch G, Scherer G, Scherer M. Simple, fast and sensitive LC-MS/MS analysis for the simultaneous quantification of nicotine and 10 of its major metabolites. Journal of Chromatography B: Analytical Technologies in the Biomedical and Life Sciences. 2014;951–952:7–15.

47. Oh J, Park MS, Chun MR, Hwang JH, Lee JY, Jee JH, et al. A Simple and High-Throughput LC-MS-MS Method for Simultaneous Measurement of Nicotine, Cotinine, 3-OH Cotinine,

Nornicotine and Anabasine in Urine and Its Application in the General Korean Population. Journal of Analytical Toxicology. 2022;46:25–36.

48. Kuehnbaum NL, Kormendi A, Britz-Mckibbin P. Multisegment injection-capillary electrophoresis-mass spectrometry: A high-throughput platform for metabolomics with high data fidelity. Analytical Chemistry. 2013;85:10664–9.

49. Shanmuganathan M, Kroezen Z, Gill B, Azab S, de Souza RJ, Teo KK, et al. The maternal serum metabolome by multisegment injection-capillary electrophoresis-mass spectrometry: a high-throughput platform and standardized data workflow for large-scale epidemiological studies. Nature Protocols. 2021;16:1966–94.

50.Dibattista A, Rampersaud D, Lee H, Kim M, Britz-Mckibbin P. High Throughput Screening Method for Systematic Surveillance of Drugs of Abuse by Multisegment Injection-Capillary Electrophoresis-Mass Spectrometry. Analytical Chemistry. 2017;89:11853–61.

51. Teo KK, Ounpuu S, Hawken S, Pandey MR, Valentin V, Hunt D, et al. Tobacco use and risk of myocardial infarction in 52 countries in the INTERHEART study: a case-control study. The Lancet; 2006;368:647–58.

52.Azab S, Ly R, Britz-Mckibbin P. Robust Method for High-Throughput Screening of Fatty Acids by Multisegment Injection-Nonaqueous Capillary Electrophoresis-Mass Spectrometry with Stringent Quality Control. Analytical Chemistry. 2019;91:2329–36.

53. Saoi M, Li A, McGlory C, Stokes T, von Allmen MT, Phillips SM, et al. Metabolic Perturbations from Step Reduction in Older Persons at Risk for Sarcopenia: Plasma Biomarkers of Abrupt Changes in Physical Activity. Metabolites. 2019;9:134.

54. Wishart DS, Guo AC, Oler E, Wang F, Anjum A, Peters H, et al. HMDB 5.0: The Human Metabolome Database for 2022. Nucleic Acids Research; 2022;50:D622–31.

55. Taghavi T, St Helen G, Benowitz NL, Tyndale RF. Effect of UGT2B10, UGT2B17, FMO3, and OCT2 genetic variation on nicotine and cotinine pharmacokinetics and smoking in African Americans. Pharmacogenetics and Genomics. 2017;27:143–54.

56. Siegel SD, Lerman C, Flitter A, Schnoll RA. The use of the nicotine metabolite ratio as a biomarker to personalize smoking cessation treatment: Current evidence and future directions. Cancer Prevention Research. 2020;13:261–72.

57. Marques H, Cruz-Vicente P, Rosado T, Barroso M, Passarinha LA, Gallardo E. Recent Developments in the Determination of Biomarkers of Tobacco Smoke Exposure in Biological Specimens: A Review. International Journal of Environmental Research and Public Health 2021;18:1768.

58. Gill B, Jobst K, Britz-Mckibbin P. Rapid Screening of Urinary 1-Hydroxypyrene Glucuronide by Multisegment Injection-Capillary Electrophoresis-Tandem Mass Spectrometry: A High-Throughput Method for Biomonitoring of Recent Smoke Exposures. Analytical Chemistry. 2020;92:13558–64.

59. Garcia-Perez I, Lindon JC, Minet E. Application of CE-MS to a metabonomics study of human urine from cigarette smokers and non-smokers. Bioanalysis 2014;6:2733–49.

60. Garcia-Perez I, Lindon JC, Minet E. Application of CE-MS to a metabonomics study of human urine from cigarette smokers and non-smokers. Bioanalysis. Future Science Ltd; 2014;6:2733–49.

61. McGuffey JE, Wei B, Bernert JT, Morrow JC, Xia B, Wang L, et al. Validation of a LC-MS/MS method for quantifying urinary nicotine, six nicotine metabolites and the minor tobacco alkaloids - Anatabine and anabasine - In smokers' urine. PLoS ONE. 2014;9:e101816.

62. Rangiah K, Hwang WT, Mesaros C, Vachani A, Blair IA. Nicotine exposure and metabolizer phenotypes from analysis of urinary nicotine & its 15 metabolites by LC-MS. Bioanalysis. 2011;3:745–61.

63. Fernandes AGO, Santos LN, Pinheiro GP, da Silva Vasconcellos D, de Oliva ST, Fernandes BJD, et al. Urinary Cotinine as a Biomarker of Cigarette Smoke Exposure: A Method to Differentiate Among Active, Second-Hand, and Non-Smoker Circumstances. The Open Biomarkers Journal. 2020;10:60–8.

64. Benowitz NL, St Helen G, Nardone N, Cox LS, Jacob P. Urine Metabolites for Estimating Daily Intake of Nicotine from Cigarette Smoking. Nicotine and Tobacco Research. 2020;22:288–92.

65. Derby KS, Cuthrell K, Caberto C, Carmella SG, Franke AA, Hecht SS, et al. Nicotine metabolism in three ethnic/racial groups with different risks of lung cancer. Cancer Epidemiology and Prevention Biomarkers. 2008;17:3526–35.

66. Bergen AW, McMahan CS, McGee S, Ervin CM, Tindle HA, le Marchand L, et al. Multiethnic Prediction of Nicotine Biomarkers and Association With Nicotine Dependence. Nicotine & Tobacco Research. 2021;23:2162–9.

67. Benowitz NL, St Helen G, Nardone N, Cox LS, Jacob P. Urine Metabolites for Estimating Daily Intake of Nicotine from Cigarette Smoking. Nicotine and Tobacco Research. 2020;22:288–92.

68. Rostron BL, Corey CG, Chang JT, van Bemmel DM, Miller ME, Chang CM. Associations of cigarettes smoked per day with biomarkers of exposure among U.S. Adult cigarette smokers in the population assessment of tobacco and health (PATH) study wave 1 (2013-2014). Cancer Epidemiology Biomarkers and Prevention. 2019;18:1443–53.

69. Bryant J, Bonevski B, Paul C, Lecathelinais C. Assessing smoking status in disadvantaged populations: is computer administered self-report an accurate and acceptable measure? BMC Medical Research Methodology; 2011;11:1–8.

70. Park EY, Lim MK, Park E, Kim Y, Lee D, Oh K. Optimum Urine Cotinine and NNAL Levels to Distinguish Smokers from Non-Smokers by the Changes in Tobacco Control Policy in Korea from 2008 to 2018. Nicotine & Tobacco Research. 2022.

71. US Department of Health and Human Services. The health consequences of smoking—50 years of progress: a report of the Surgeon General. Centers for Disease Control and Prevention; 2014.

72. O'Connor RJ, Wilkins KJ, Caruso R v, Cummings KM, Kozlowski LT. Cigarette characteristic and emission variations across high-, middle-and low-income countries. Public Health. Elsevier; 2010;124:667–74.

73. US Department of Health and Human Services. The health consequences of involuntary exposure to tobacco smoke: a report of the Surgeon General. Centers for Disease Control and Prevention (US); 2006.

74. Burns, David M., and Neal L. Benowitz. Public Health Implications of Changes. Risks associated with smoking cigarettes with low machine-measured yields of tar and nicotine 2001;13:1

75. Carroll DM, Murphy SE, Benowitz NL, Strasser AA, Kotlyar M, Hecht SS, et al. Relationships between the nicotine metabolite ratio and a panel of exposure and effect biomarkers: Findings from two studies of U.S. commercial cigarette smokers. Cancer Epidemiology Biomarkers and Prevention. 2020;29:871–9.

76. Tanner JA, Novalen M, Jatlow P, Huestis MA, Murphy SE, Kaprio J, et al. Nicotine metabolite ratio (3-Hydroxycotinine/Cotinine) in plasma and urine by different analytical methods and laboratories: Implications for clinical implementation. Cancer Epidemiology Biomarkers and Prevention. 2015;24:1239–46.

77. Wassenaar CA, Conti D v., Das S, Chen P, Cook EH, Ratain MJ, et al. UGT1A and UGT2B genetic variation alters nicotine and nitrosamine glucuronidation in European and African American smokers. Cancer Epidemiology Biomarkers and Prevention. 2015;24:94–104.

78. Chen G, Giambrone NE, Dluzen DF, Muscat JE, Berg A, Gallagher CJ, et al. Glucuronidation genotypes and nicotine metabolic phenotypes: Importance of UGT2B10 and UGT2B17 knock-out polymorphisms. Cancer Research 2010;70:7543.

4.8 Supporting Information

Table S4.1 Summary of participants from the PURE study (*n*=1000) based on their self-reported smoking status and country income status.

Country	Never Smokers (*n*=335)	Light Smokers (<10cigarettes/day) (*n*=324)	Heavy Smokers (≥10cigarettes/day) (*n*=341)
High Income Countries (HICs)			
Canada	26.0% (*n*=87)	23.1% (*n*=75)	21.7% (n=74)
Sweden	3.3% (*n*=11)	14.8% (*n*=48)	17.3% (*n*=59)
UAE	4.8% (*n*=13)	0.9% (*n*=3)	1.8% (*n*=6)
Middle Income Countries (MICs)			
South Africa	2.7% (*n*=9)	3.7% (*n*=37)	0.8% (*n*=3)
Brazil	2.4% (*n*=8)	11.4% (*n*=11)	6.5% (*n*=22)
Colombia	9.0% (*n*=30)	7.1% (*n*=23)	2.3% (*n*=8)
Chile	1.2% (*n*=4)	2.2% (*n*=7)	0.6% (*n*=2)
Argentina	10.4% (*n*=35)	16.7% (*n*=54)	19.1% (*n*=65)
Iran	5.4% (*n*=18)	2.5% (*n*=8)	6.5% (*n*=22)
Philippines	2.7% (*n*=9)	0.6% (*n*=2)	1.2% (*n*=4)
Low Income Countries (LICs)			
Bangladesh	29.2% (*n*=98)	13.6% (*n*=44)	21.4% (*n*=73)
Pakistan	0.9% (*n*=3)	2.5% (*n*=8)	0.3% (*n*=1)
Tanzania	3.0% (*n*=10)	0.9% (*n*=3)	0.3% (*n*=1)
Zimbabwe	-	0.3% (*n*=1)	0.3% (*n*=1)

Table S4.2 Key figures of merit for the determination of seven nicotine metabolites in urine by MSI-CE-MS from the PURE study.

Figures of Merit	Nicotine	Cotinine	Nicotine-n-Oxide	OH-Cotinine	Nicotine Glucuronide	Cotinine Glucuronide	OH-Cotinine Glucuronide
Calibration Range [a]	0.25-18.5µM	0.10-17.0µM	0.22-3.37µM	0.21-26.0µM	0.06-1.77µM	0.11-17.0µM	0.11-13.6µM
Linearity [a] (R^2)	0.999	0.988	0.990	0.996	0.985	0.999	0.999
Recovery [b] (Average %)	82.2%	87.6%	80.1%	88.3%	84.3%	92.6%	89.8%
Intra-day Precision [c] (n=21)	4.83%	3.68%	7.94%	4.10%	3.33%	16.7%	11.7%
Inter-day Precision [d] (Average S/N)	18.6% (*S/N*~38)	15.5% (*S/N*~35)	34.2% (*S/N*~11)	12.1% (*S/N*~130)	31.6% (*S/N*~9)	20.5% (*S/N*~60)	15.1% (*S/N*~40)
Freeze Thaw Stability [e] (Average % Change)	7.90%	10.1%	5.34%	2.58%	12.2%	10.9%	2.59%
24 Hour RT Stability [e] (Average % Change)	6.82%	7.68%	5.06%	11.0%	6.92%	12.1%	10.6%
LOD [f] (S/N~3)	0.07 µM	0.10 µM	0.11 µM	0.07 µM	0.03 µM	0.05 µM	0.06 µM
LOQ [f] (S/N~10)	0.25 µM	0.23 µM	0.22 µM	0.21 µM	0.06 µM	0.11 µM	0.11 µM
% Non-Detects in urine [g] (Current Smokers)	46.6%	18.3%	64.1%	15.9%	56.7%	25.1%	31.4%

[a] *External calibration curve concentration range of serially diluted neat standards over 10-170 fold linear dynamic range, used for quantification of all urine samples.*
[b] *Average % recovery is shown following spike studies using authentic nicotine standards spiked into pooled never-smoker urine samples (in triplicate) at three concentration levels (3.9 µM, 13 µM, 130 µM), with an overall CV (%) between replicates of <15%.*
[c] *Intr a-day precision was determined based on CV (%) of repeat calibrants runs over the course of a day in triplicate, with 7 injections per run (n=21).*
[d] *CV(%) of pooled quality control (QC) samples run during the 6-day study (n=156), with the average S/N of each nicotine metabolite.*
[e] *Acceptable stability (<15% change) was determined based on both repeat freeze-thaws (n=3) and a 24-hour room temperature study (n=6) conducted on spiked pooled never-smoker urine in triplicate, with average % change determined according to the difference between the initial time point and all follow up time points.*
[f] *Instrumental LOD and LOQ were based on serial dilution of neat standards.*
[g] *Analysis of 1000 single-spot urine samples with % non-detects calculated for individual nicotine metabolites according to the number of self-identified current smokers (n=665).*

Table S4.3 Characteristics of never smokers with detectable levels of urinary TNE-7 (*n*=89) indicative of misreporting within each income group from the PURE study.

Variable[a]	HICs (*n*=20/111)	MICs (*n*=19/112)	LICs (*n*=50/111)
% Misreported Never Smokers	18%	17%	45%
TNE-7*	3.46±3.7 µM	9.24±33 µM	22.8±49 µM
Sex (n;%)	-	-	-
F	65% (*n*=13)	68% (*n*=12)	66% (*n*=33)
M	35% (*n* =7)	32% (*n*=6)	34% (*n* =17)
Age*	55±17	58±9	50±16
BMI*	29±6	26±8	22±6
PM 2.5*	9±2	27±23	78±14
Location (n,%)	-	-	-
Urban	75% (*n*=15)	79% (*n*=15)	50% (*n*=25)
Rural	25% (*n*=5)	21% (*n*=4)	50% (*n*=25)
Second-hand Smoke (n;%)	-	-	-
Exposed (>1 time/week)	20% (*n*=4)	32% (*n*=6)	48% (*n*=24)

[a] Median values ± interquartile range (IQR) are shown unless otherwise indicated
**Significant at p<0.05*

Table S4.4. Summary of key figures of merit for ratiometric biomarkers of enzyme activity (CYP 2A6, UGT) measured in verified current smokers from the PURE study.

Figures of Merit	NMR	UGT Cotinine	UGT OH-Cotinine	UGT Nicotine
Range (mM) [a]	0.23-22.0	0.01-7.85	0.04-8.67	0.01-2.75
Technical Precision[b] (CV)	14.6%	19.4%	16.7%	41.1%
Biological Variation (CV)	73%	87%	173%	147%
ICC	0.95	0.94	0.99	0.92
Frequency of Detection [c]	79% (*n*=525)	71% (*n*=470)	67% (*n*=443)	30% (*n*=201)

[a] Enzyme activity range in the PURE cohort for participants with quantifiable levels of individual nicotine metabolites necessary for calculating the ratio.
[b] Precision was monitored through pooled quality control samples over the course of the 6-day study (n=156) based on mean CV
[c] Frequency of detection (%) among self-identified current smokers for those with a ratio that could be mathematically calculated.

Table S4.5 Partial Pearson correlation matrix between ratiometric biomarkers of enzyme activity (CYP 2A6, UGT) and TNE-7 for biochemically current smokers from the PURE study.

	NMR	**UGT Cotinine**	**UGT OH-Cotinine**	**UGT Nicotine**	**TNE-7**
NMR	1	**0.406***	0.095	0.002	**0.580***
	UGT Cotinine	1	**0.332***	**0.253***	**0.396***
		UGT OH-Cotinine	1	0.095	-0.029
			UGT Nicotine	1	-0.024
				TNE-7	1

**A signification correlation based on partial Pearson correlation analysis with adjustments for age, sex, BMI, CPD, when using pairwise deletion (*p*-value<0.001)*

Table S4.6 Stratification of urinary NMR in fast nicotine metabolizers from PURE (*n*=142) classified as high-risk tobacco smokers (Q4) due to their greater toxicant exposure compared to slower nicotine metabolizers (Q1-Q3) (*n*=426) at lowest risk for biological harm.

Variable[a]	High Risk Metabolizers (*n*=35)	Fast Metabolizers (*n*=107)	Slower Metabolizers (*n*=426)
Sex (n;%)	-	-	-
F	51% (*n*=17)	48% (*n*=51)	39% (*n*=168)
M	49% (*n*=18)	52% (*n*=56)	61% (n=258)
Age	56±10	53±12	52±15
BMI	32±20	32±15	32±12
Income Status (n;%)	-	-	-
HICs	54% (*n*=19)	57% (*n*=61)	31% (n=133)
MICs	40% (*n*=14)	33% (*n*=35)	36% (n=155)
LICs	6% (*n*=2)	10% (*n*=11)	32% (n=138)
NMR*	13.3 ±3.7	7.15 ±2.1	2.99±2.1
UGT Cotinine*	2.39 ± 3.2	1.61 ± 1.4	0.81±0.8
TNE-7*	108 ± 88 μM	72.5 ± 56 μM	39.2±39 μM
Thiocyanate*	107 ± 149 μM	84.9 ±98 μM	52.6±87 μM
Smoking Status (n;%)	-	-	-
Heavy Smokers	77% (*n*=27)	56% (*n*=60)	51% (*n*=218)
Light Smokers	23% (*n*=8)	40% (*n*=43)	40% (*n*=169)
Misreported Never Smokers	0%	4% (*n*=4)	9% (*n*=39)
# of Cigarettes per day	18±10	10±11	10±10
Pack Years	31±31	20±24	16±21

[a] *Median values ± interquartile range (IQR) are show unless otherwise indicated.*
**Significant at p<0.0001, after adjusting for age, sex, BMI, and CPD.*

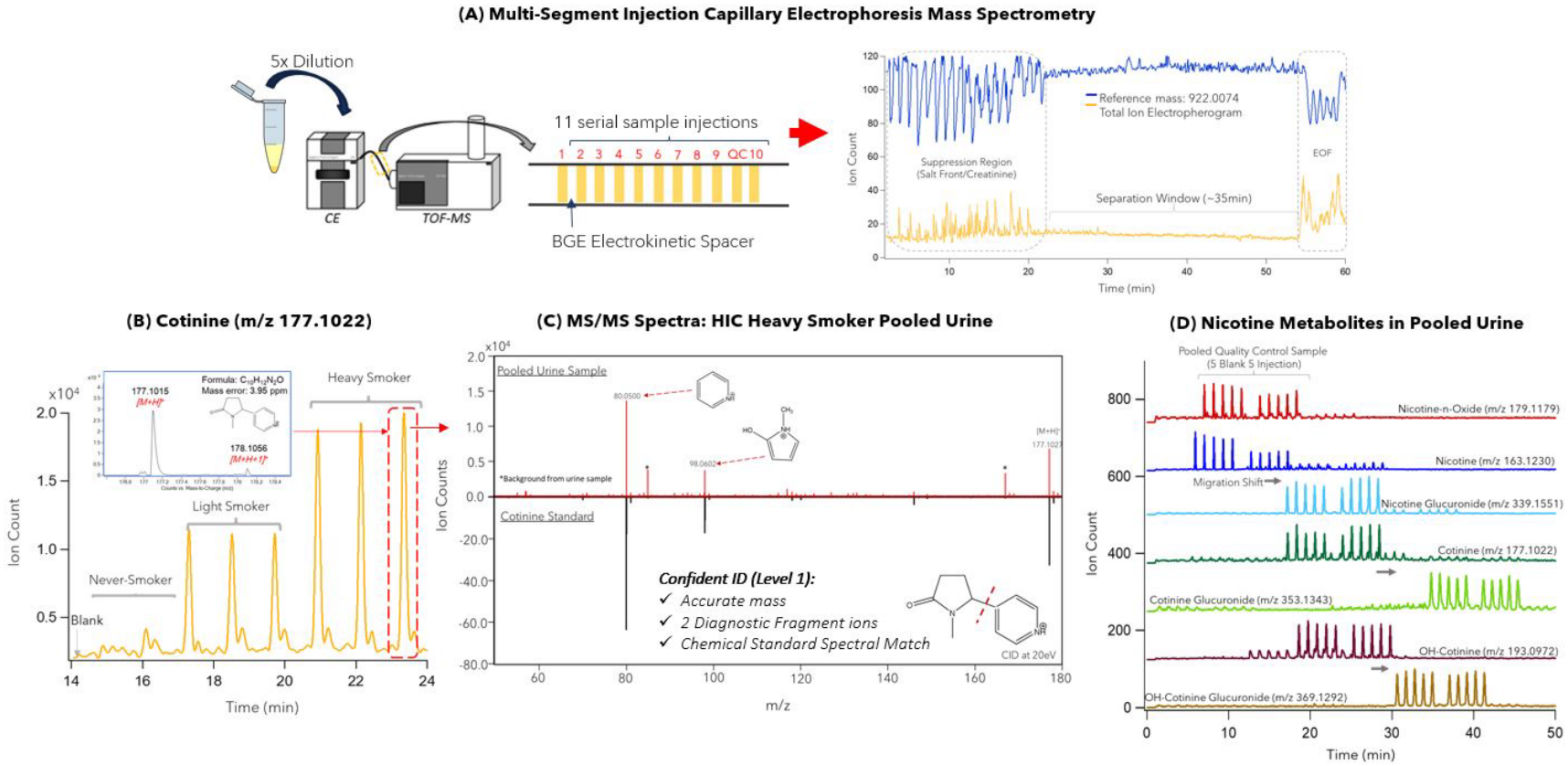


Figure S4.1 (A) Schematic of the high throughput workflow used for comprehensive urine metabolic phenotyping by MSI-CE-MS, demonstrating the multiplexing capabilities with 11 sample plugs and the inclusion of a pooled quality control sample (QC) within each run (< 5 min/sample). Subsequent data output indicates regions of suppression due to involatile salts and abundant creatinine in urine monitored with a reference mass, which can be corrected for when using a matched deuterated internal standard. **(B)** Extracted ion electropherogram of cotinine displays the unique injection pattern sequencing possible with MSI-CE-MS, where pooled subgroup QCs are measured in triplicate with a blank. As expected, cotinine shows no signal in the blank and never smokers, but is measured at progressively higher concentrations (e.g., larger peak area) from light to heavy current smokers. **(C)** Unambiguous (*Level 1*) confirmation of cotinine is confirmed by MS/MS in a separate analysis with a matched authentic standard, run at the same collision energy (CID at 20 V). **(D)** Separation based on unique electrophoretic mobilities demonstrates distinct migration time shifts for individual nicotine metabolites, with secondary identification based on their accurate mass (*m/z*).

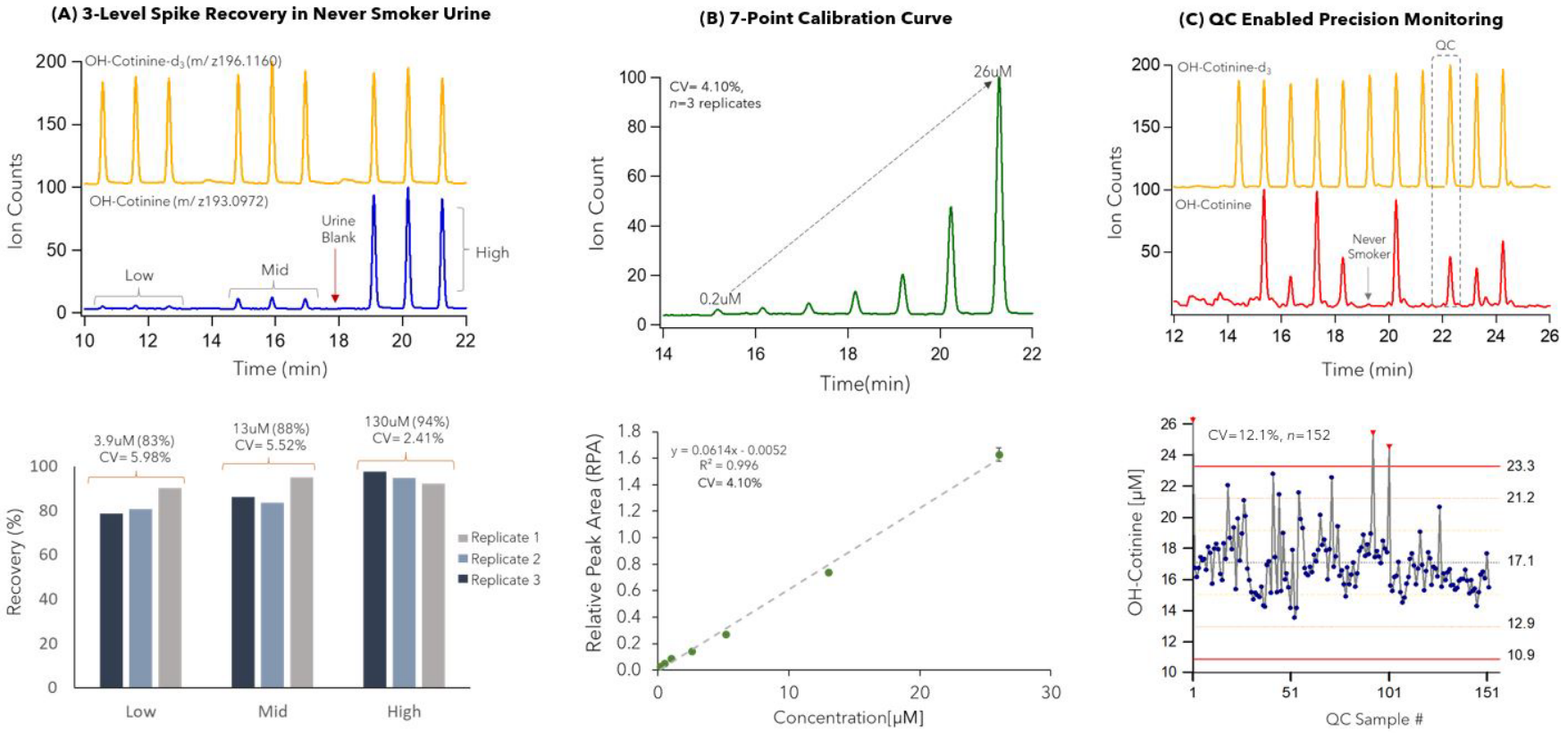


Figure S4.2 An overview of the method validation with **(A)** an extracted ion electropherogram (EIE) depicting a spike-recovery study performed in never pooled smoker urine for 3-hydroxycotinine and a matching comigrating stable-isotope internal standard (3-hydroxycotinine-d_3) for normalization. Acceptable recoveries (>80%) at three concentration levels was observed. **(B)** Additionally good linearity (R^2=0.996) with acceptable technical precision (CV=4.1%) is demonstrated by a seven-point calibration curve over a 125-fold concentration range, analyzing within a single run. **(C)** Subsequent precision monitoring using randomized pooled quality control samples within each run over the course of the 6-day study showed acceptable technical precision (CV=12.1%, n=152) following normalization to 3-hydroxycotinine-d_3

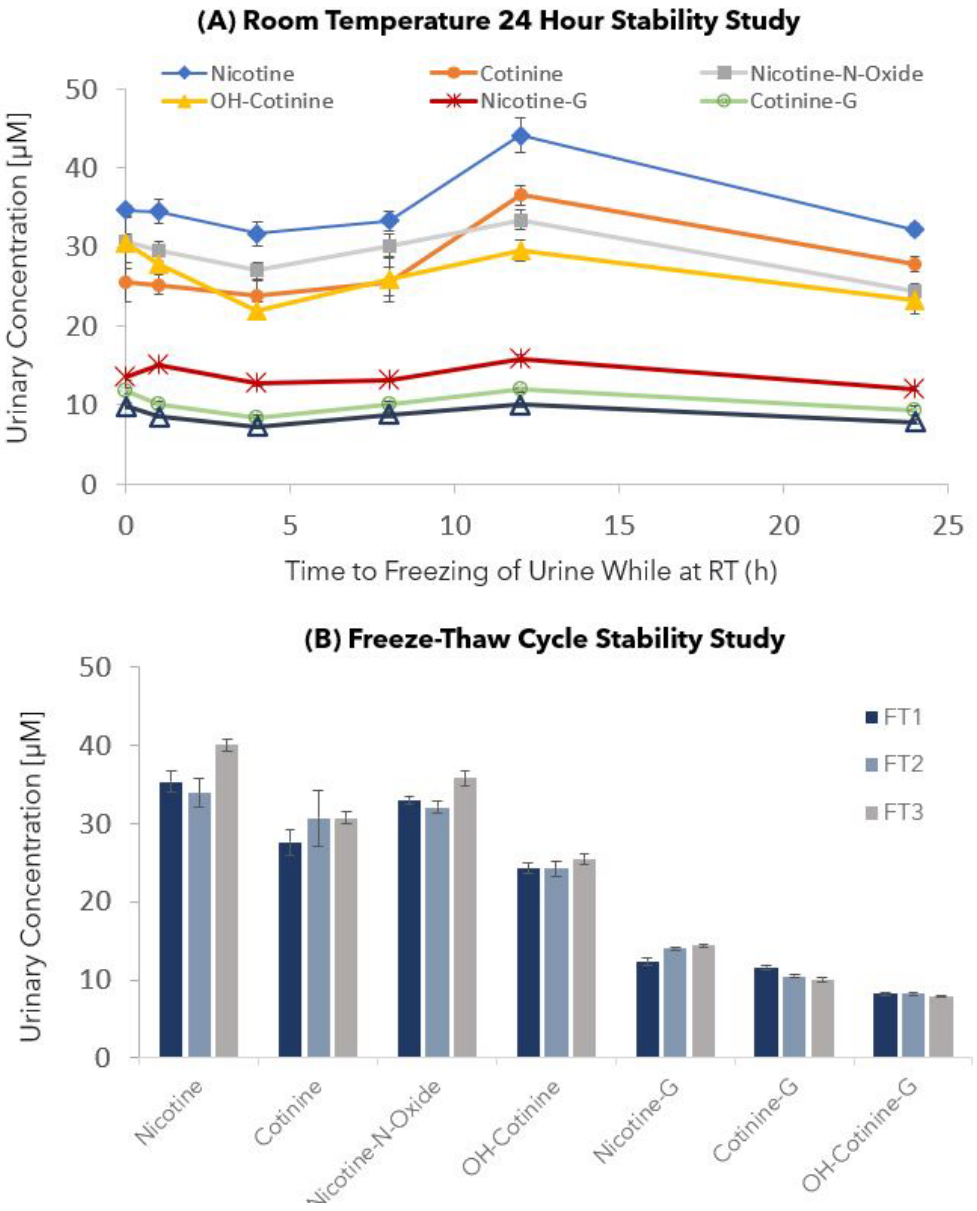


Figure S4.3 Stability studies were performed by spiking all seven nicotine metabolites in pooled never smoker urine, including **(A)** delays to freezing at room temperature and **(B)** repeat freeze-thaw cycles. All nicotine metabolites were measured in triplicate (CV <15%) by MSI-CE-MS, which confirmed no significant change (%change <15%) in concentration between all-time points.

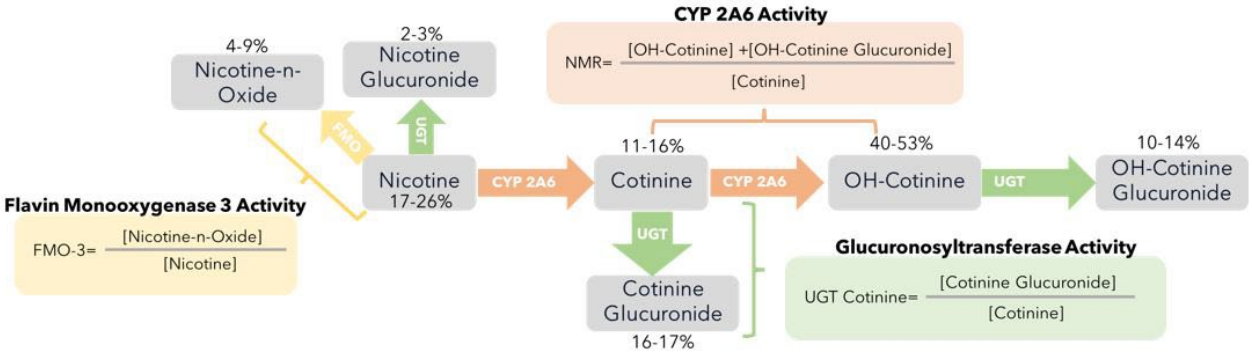


Figure S4.4 Schematic of nicotine metabolism as mediated by CYP 2A6, UGT, and FMO and their phenotypic biomarkers of smoking behavior and nicotine dependence (e.g., NMR). The sum of all seven nicotine metabolites in urine is referred to the total nicotine equivalents (TNE-7), which is a robust biomarker of recent tobacco smoke exposure independent of metabolic rate.

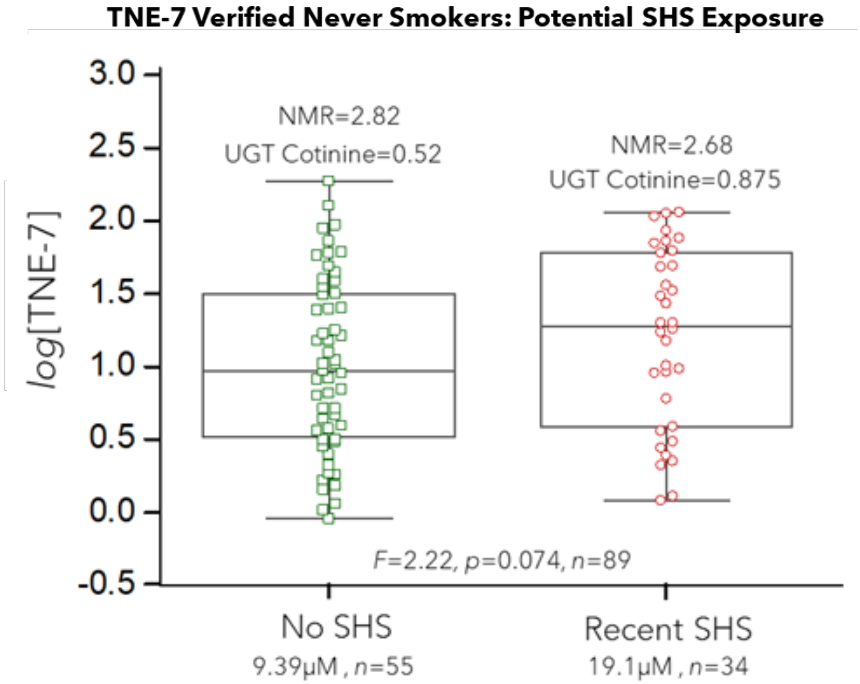


Figure S4.5 TNE-7 detected among never smokers (*n*=89) in the PURE cohort was compared between individuals self-reporting recent SHS exposure, as compared to individuals indicating no exposure. No significant difference in TNE-7, NMR, and UGT demonstrates the current instrumental detection limits likely cannot distinguish between current smokers and those exposed to second-hand smoke, but rather a more likely contribution of misreports among never smokers.

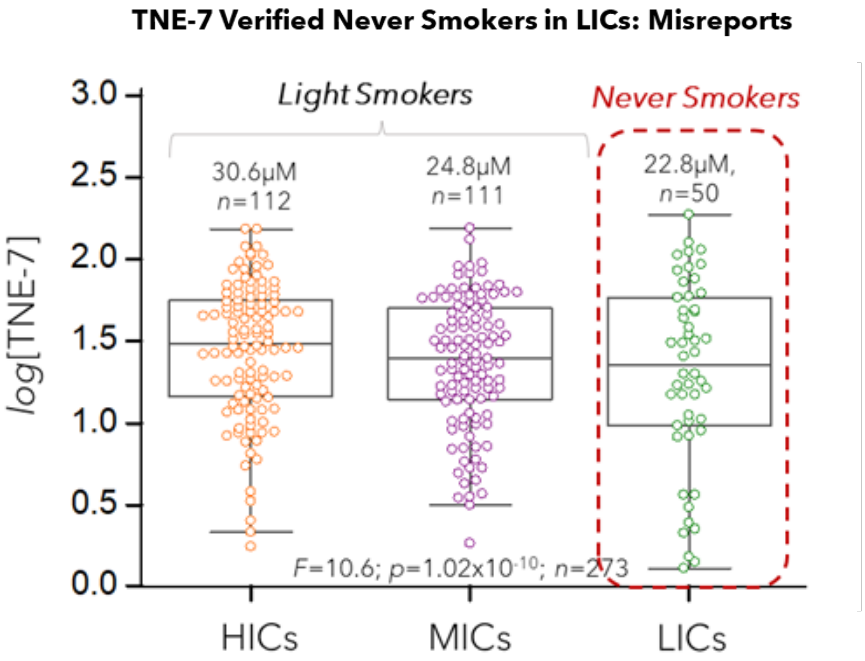


Figure S4.6. TNE-7 detected among never smokers (*n*=50) in LICs, demonstrates median levels (22.8 µM) within the range of biochemically verified light smokers found in HICs and MICs (24.8-30.6 µM), likely indicated misreporting in LICs.

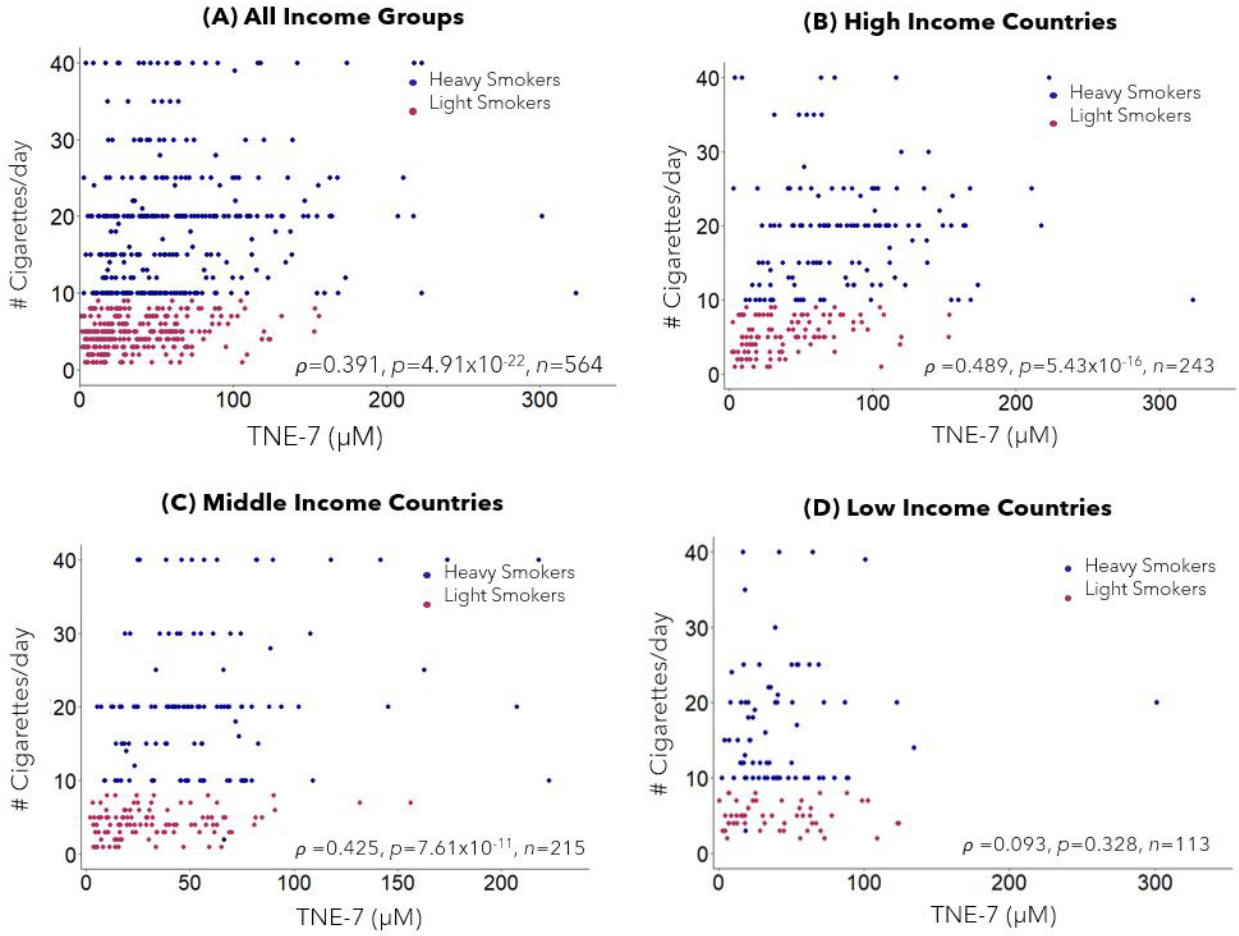


Figure S4.7 (A) A moderate association was found between self-reported smoking intensity (CPD) to recent tobacco smoke exposure (TNE-7) among all self-identified current smokers from PURE when using a Spearman rank correlation analysis. Overall, the strongest correlation was demonstrated in **(B)** HICs, followed by **(C)** MICs, with no significant correlation to self-reports of smoking intensity in **(D)** LICs indicative of misreporting discrepancies between regions.

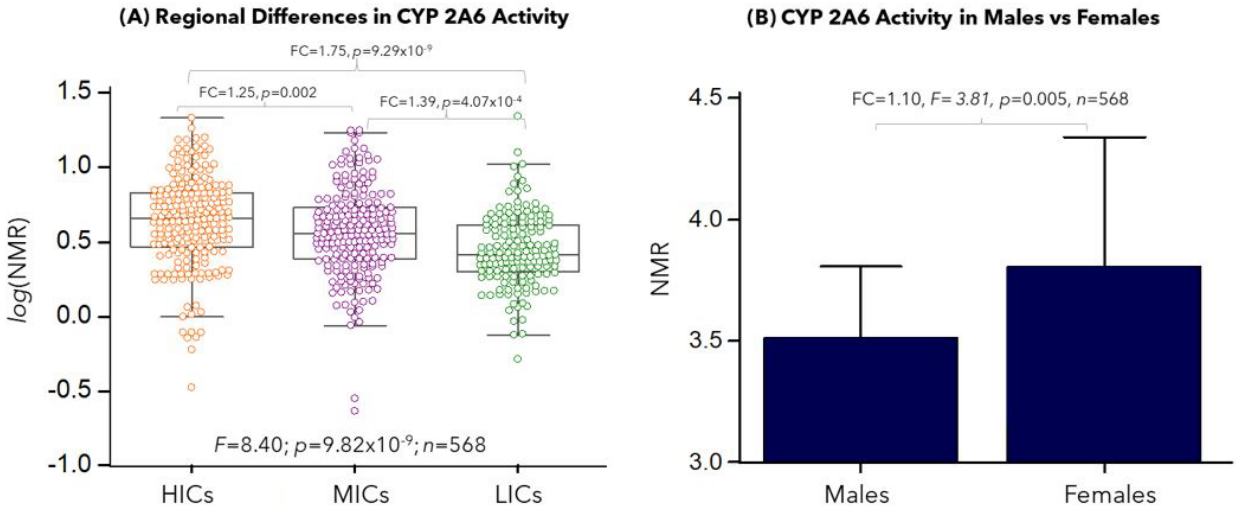


Figure S4.8 Regional and sex differences were identified for CYP 2A6 activity amongst active smokers in PURE (*n*=568) based on urinary NMR. **(A)** Following ANCOVA with adjustments for age, sex, BMI and CPD enzyme activity is highest in HICs relative to MICs and LICs. This is indicative of regional metabolism differences that be a result of genetics or environmental factors (e.g., diet, medication) **(B)** While overall females tend to have slightly elevated CYP 2A6 activity compared to males after correction for age, BMI and CPD due to the role of elevated estrogen in the induction of CYP 2A6 metabolism.

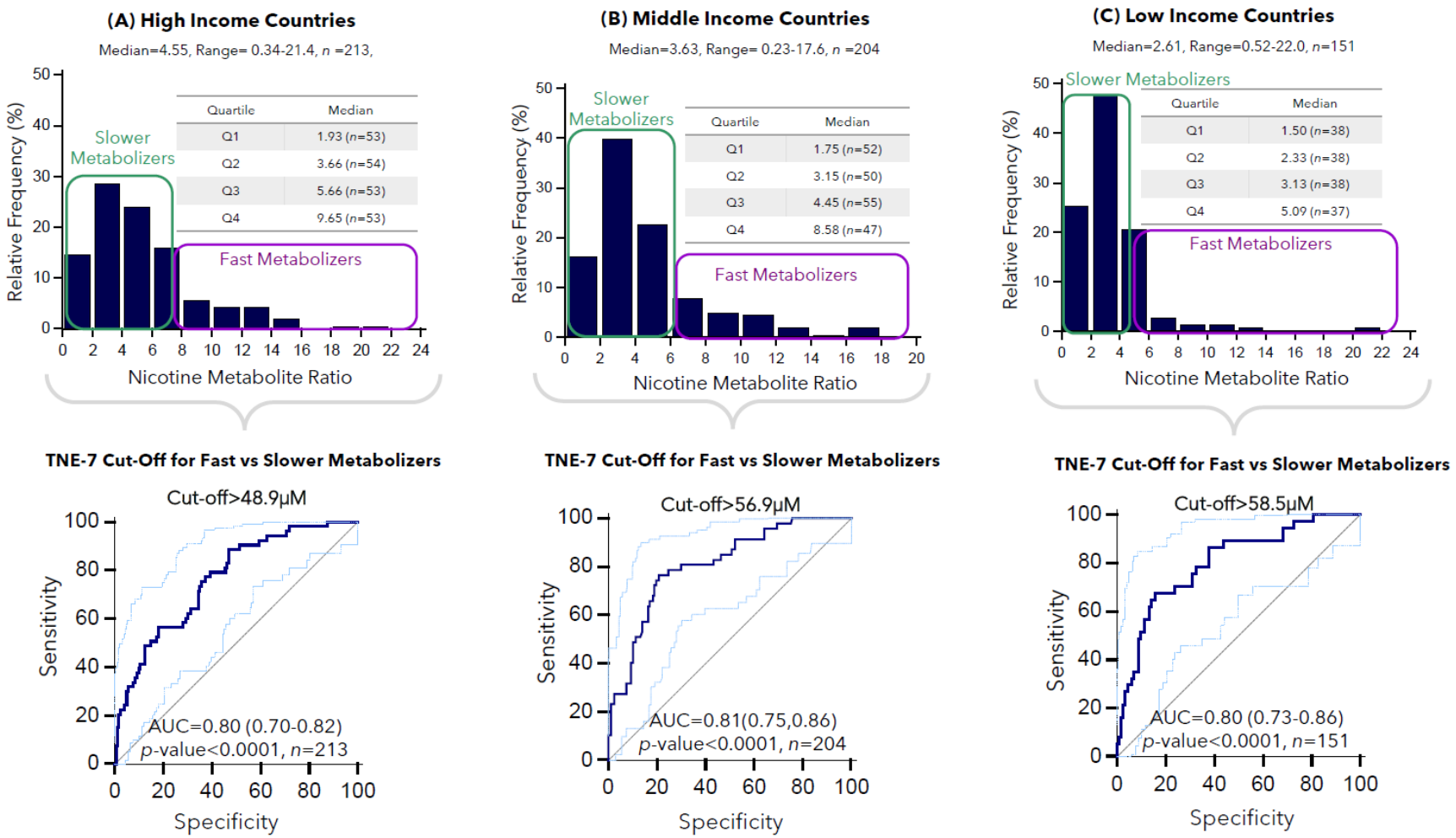


Figure S4.9 Stratification of urinary NMR by quartiles within different country income groups, **(A)** HICs **(B)** MICs and (**C**) LICs. This was performed to more accurately identify smokers at greater risk for toxicant exposure from tobacco smoke and potential biological harm. Fast nicotine metabolizers (Q4) were determined for each income group, following slower metabolizers (Q1-Q3), and cut-off concentrations were identified within each income group using ROC curves for urinary TNE-7. Overall consistent and strong discrimination was achieved between fast and slower metabolizers (AUC > 0.801, $p < 0.0001$) using urinary TNE-7 with significantly higher cut-off levels noted for current smokers from LICs (58.5 µM) as compared to HICs (48.9 µM).

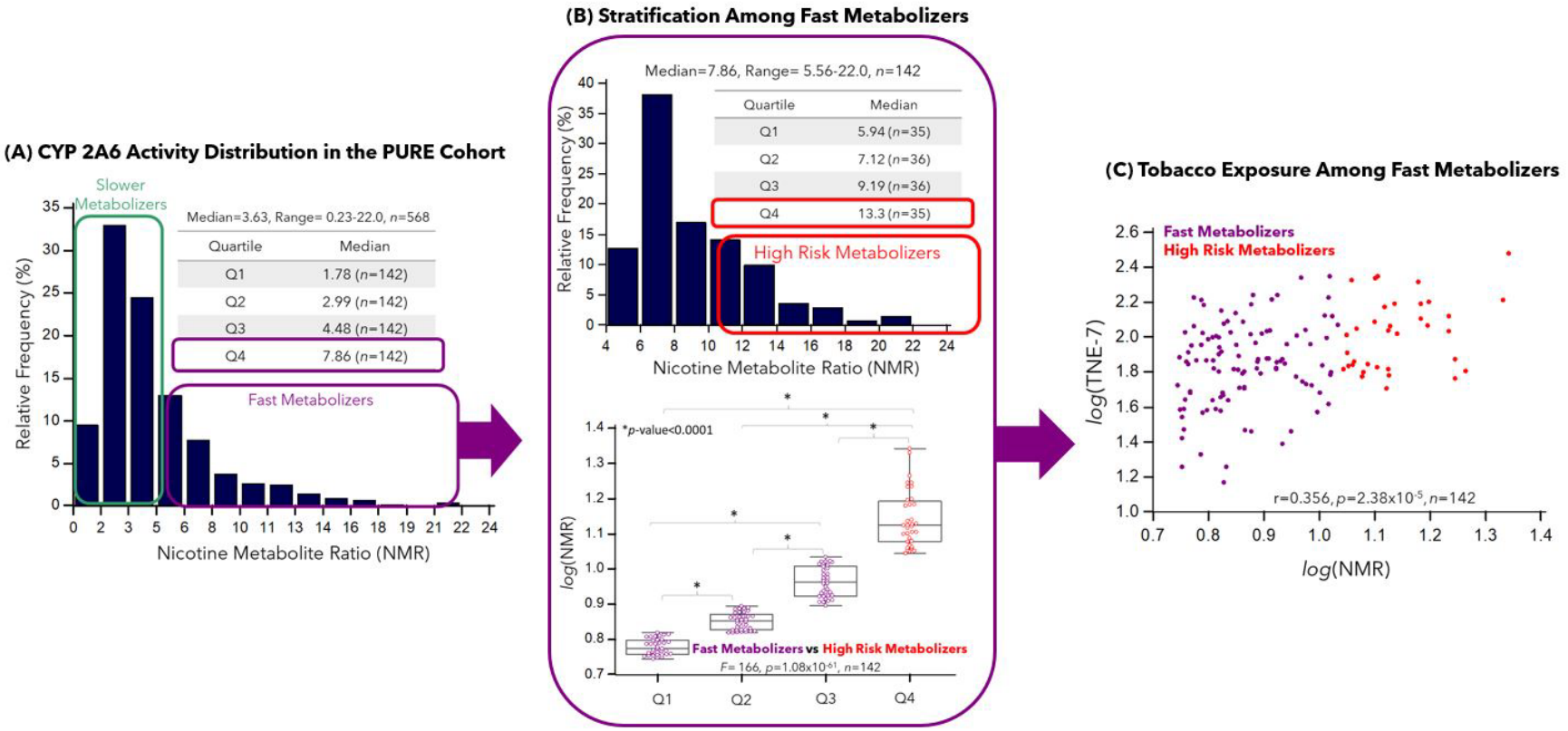


Figure S4.10 (A) Stratification of urinary NMR among current smokers (*n*=568) was preformed to identify fast nicotine metabolizers classified within quartile four (Q4). **(B)** Fast metabolizers were further stratified by quartiles to identify the highest risk group (*n*=35) with the fastest CYP 2A6 metabolism. These tobacco smokers were likely at greatest risk for nicotine dependence with greater toxicant exposure and biological harm. **(C)** Increased risk was inferred by a linear correlation of urinary NMR with TNE-7 demonstrating a positive association for smokers with faster nicotine metabolism also suffer from greater tobacco smoke exposure.

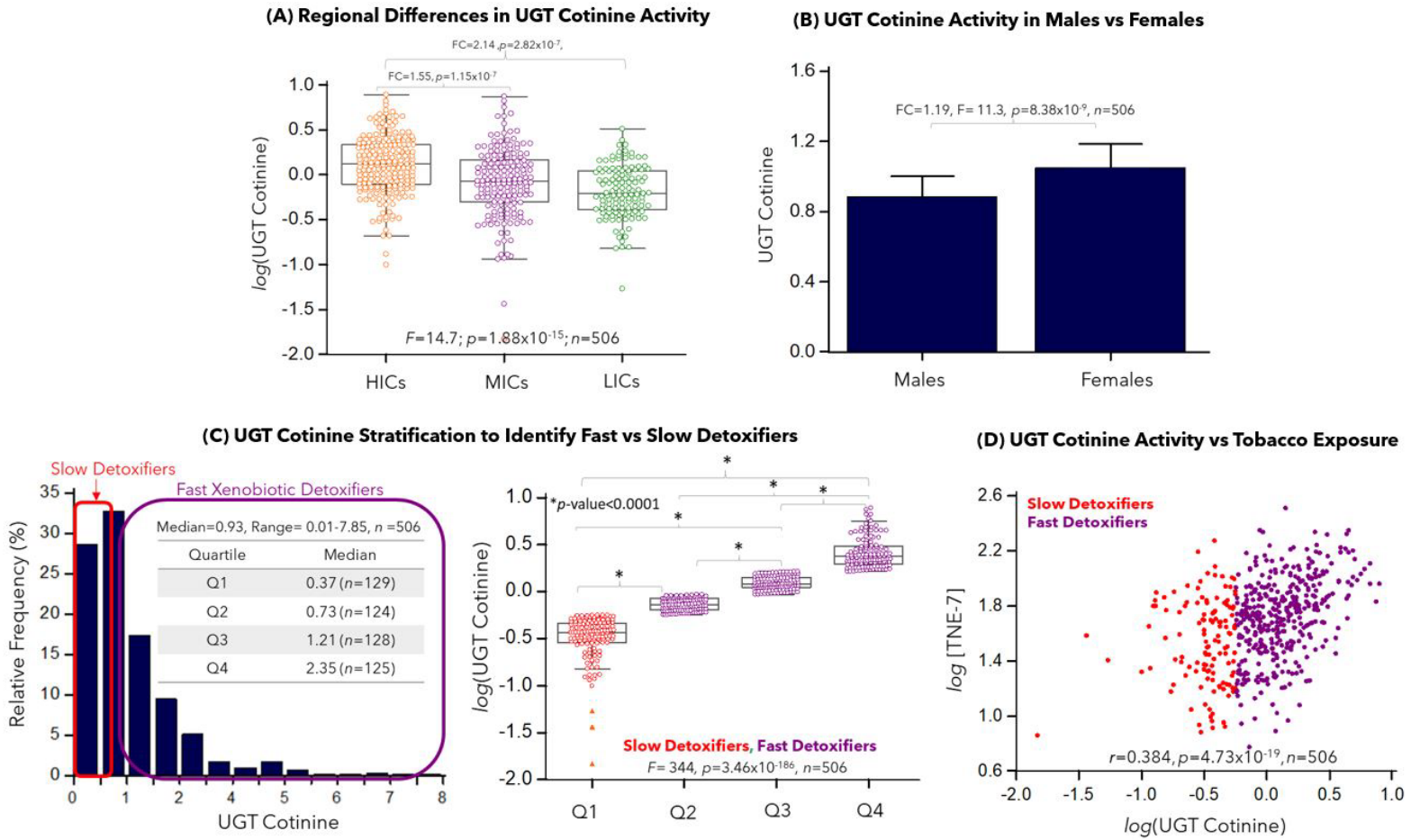


Figure S4.11 UGT cotinine is indicative of phase II metabolism and has been linked with detoxification capacity for tobacco-specific nitrosamines. **(A)** A regional comparison in enzyme activity based on the ratiometric biomarker (cotinine glucuronide: cotinine), regional differences in glucuronidation were identified using an ANCOVA analysis following correction for age, sex, BMI and CPD. Overall, HICs had significantly elevated glucuronidation rates, relative to MICs and LICs (FC > 1.55, $p < 0.0001$). **(B)** Similarly, females also demonstrated elevated UGT cotinine activity after correction for age, BMI and CPD. **(C)** Stratification of UGT among current smokers in PURE (n=506) allowed for the identification of slow detoxifiers (Q1) who may retain biologically activated toxicants for longer periods of time resulting in increased biological harm compared to faster detoxifiers (Q2-Q4). **(D)** Although glucuronidation does not directly relate to cigarette consumption, an overall moderate correlation with urinary TNE-7 was observed following adjustments for age, sex, BMI and CPD.

Chapter V:

High-Throughput Metabolomics for the Identification of Robust Dietary Biomarkers of Food Intake from the Prospective Urban and Rural Epidemiological (PURE) Study

Biban Gill,[1] Vanessa Martinez, Erick Helmeczi,[1] Stellena Mathiaparanam,[1] Sathish Thirunavukkarasu,[2] Guillaume Paré,[2,3] Koon Teo,[2] Salim Yusuf,[2] Philip Britz-McKibbin*[1]

[1]Department of Chemistry & Chemical Biology, McMaster University, Hamilton, ON, Canada
[2]Population Health Research Institute, McMaster University, Hamilton, ON, Canada
[3] Department of Pathology and Molecular Medicine, McMaster University, Canada

* *Manuscript in preparation for submission*

B.G. performed all experiments, including sample preparation and data acquisition, statistical analysis, and interpretation. B.G. wrote the initial manuscript draft for publication. V.M. assisted with data processing and statistical analysis. E.H. assisted with data processing and the R code used for statistical analysis. S.M. provided data for select urinary metabolites included in the statistical analysis. S.Y. is the principal investigator for the PURE study, K.T. was involved in the study design for PURE, and G.P., aided in the design and supervision of the PURE urine sub study. S.T. was involved in the selection of samples for the PURE urine sub study and provided participant information including diet records used for the statistical analysis. B.G. wrote the initial draft for publication. P.B.M. was involved in the design of the PURE urine sub study and provided revisions for the final manuscript draft.

Chapter V: High-Throughput Metabolomics for the Identification of Robust Dietary Biomarkers of Food Intake from the Prospective Urban and Rural Epidemiological (PURE) Study

5.1 Abstract

A suboptimal diet is characterized by the habitual consumption of high caloric and low nutrient processed foods, which is an important modifiable risk factor for chronic diseases worldwide. Current methods for elucidating the health effects from food exposures largely rely on standardized food frequency questionnaires (FFQs) that may be prone to bias and misreporting. Although there is growing interest in using dietary biomarkers as more objective indicators of food intake in nutritional epidemiology, sparse studies have validated their generalizability in diverse populations from developing and developed countries. Herein, we applied a nontargeted metabolomics workflow by multisegment injection-capillary electrophoresis-mass spectrometry (MSI-CE-MS) in conjunction with multivariate and univariate statistical analysis methods for identifying robust dietary biomarkers in urine samples collected from participants in the Prospective Urban and Rural Epidemiology (PURE) study. A total of 116 urinary metabolites were reliably quantified by MSI-CE-MS (CV < 30%) from participants (n=1000) from fourteen countries with varying income status and compared to semi-quantitative food intake measures from FFQs (n=60). A total of 215 significant correlations were initially classified ($r > 0.20$, $p < 0.05$) with subsequent filtering used to identify 19 dietary biomarker candidates in urine associated with five distinct food classes and overall diet quality based on the Alternative Healthy Eating Index score. Overall, eight top-ranked candidates were considered as robust dietary biomarkers based on their dose response, plausibility and generalizability, such as urinary trigonelline which had the strongest correlation ($r = 0.662$, $p = 1.20\times10^{-31}$) to self-reported coffee intake while also reflecting regional trends in coffee consumption and smoking habits in the

PURE study. Similarly, urinary proline betaine, carnitine, 3-methylhistidine, saccharin, acesulfame K, uracil, and the sum of quinic acid and trigonelline (r = 0.272-0.308, $p < 6.42\times 10^{-6}$) were replicated across a large multiethnic cohort reflecting variable and regional dietary patterns associated with the consumption of fruits and vegetables, red meat, total sugar, processed foods, nuts and soy, and raw vegetables. Our study highlights that a panel of urinary dietary biomarkers of food intake may provide more reliable assessment of dietary patterns globally while providing new insights into chronic disease risk across diverse populations.

5.2 Introduction

A global epidemic of chronic non-communicable diseases (NCDs) is having a disproportionate burden on morbidity and mortality notably in developing countries due to increases in population growth and urbanization [1]. NCDs are now responsible for over 70% of deaths globally [2], including cardiovascular diseases (CVD), obesity, type 2 diabetes, and certain cancers that share several common modifiable risk factors associated with lifestyle, such as tobacco smoking, alcohol consumption, physical inactivity and unhealthy eating patterns [3]. Suboptimal diet linked to excessive intake of sodium, carbohydrates and red meat, as well as low consumption of fruits, vegetables, legumes, whole grain and fiber is a leading preventable risk factor to chronic disease burden and disability-adjusted life-years worldwide [4–8]. Contemporary unhealthy diets that contribute to cardiometabolic diseases are also associated with increased consumption of nutrient-poor, ultra-processed foods containing artificial sweeteners [9], whereas adherence to an Alternative Healthy Eating Index (AHEI) dietary pattern can reduce the risk of all-cause mortality [10]. Indeed, dietary modifications are often utilized for both the prevention and management of NCDs, particularly in instances of multiple comorbidities [4]. Nevertheless, there

remains on-going controversy on the what exactly constitutes a healthy diet to better guide evidence-based public health policies and guidelines for chronic disease prevention [11].

Evaluating the synergistic effects of food exposures is important when assessing diet quality [12,13] since nutrients are not consumed in isolation and many non-nutrients (e.g., phytochemicals, fiber) play a myriad of roles in human health. For instance, individuals scoring in the top quintile of the AHEI exhibit a lower risk for multiple NCDs, including coronary heart disease, stroke, and type 2 diabetes [14–16]. While scoring methods such as AHEI have proved effective and low cost to implement in large-scale nutritional epidemiological studies, they are limited by their reliance on self-reported dietary intake measures using 24 h diet records or food frequency questionnaires (FFQs) [17]. These surveys are prone to random and systematic errors related to underestimated energy intake, inaccurate portion sizes, as well as gender and social desirability bias when comparing dietary patterns in certain demographic groups and multi-ethnic populations [18–21]. For instance, a positive social approval bias may overestimate the self-reported intake of fruits and vegetables [22], which have long troubled the validity of FFQs in nutritional epidemiology. Additionally, such reporting methods do not fully capture the complex chemical composition of foods, variations in preparation and cooking methods, as well as habitual consumption of stimulants with meals, such as coffee [23]. Moreover, survey questions are unable to directly assess the interactions of food exposures with genetics and the gut microbiome that impact nutrient bioavailability and disease susceptibility [17,24].

Comprehensive metabolic profiling of non-invasive biological fluids, such as urine may offer more objective dietary biomarkers of food intake that are complementary to self-reported diet recall methods [21]. Modern nontargeted metabolomic techniques enable the concurrent and unbiased determination of hundreds of small molecules to provide unexpected insights into

complex food exposures and better decipher underlying diet-disease relations in nutritional epidemiology [25,26]. Overall, an ideal dietary biomarker of food intake is not extensively metabolized, exhibits a distinct temporal and dose dependence, demonstrates specificity to a distinct food item or food category, while also being applicable to diverse populations globally [17,27]. To date, recent studies have focused on validating the utility of dietary biomarkers in well-controlled clinical trials and observational studies, while assessing the bias associated with standardized FFQs [28,29]. However, most epidemiological studies have focused on targeted metabolite-FFQ associations defined *a priori* in high-income countries (HICs) with few multi-ethnic and multi-center cohorts involving participants from middle- (MICs) or low-income countries (LICs) [17,30].

Herein, we outline a high throughput and nontargeted metabolomics data workflow using multisegment injection-capillary electrophoresis-mass spectrometry (MSI-CE-MS) to characterize food exposures in urine samples. A cohort of 1000 participants recruited from the Prospective Urban and Rural Epidemiology (PURE) study [31] were examined to classify and validate dietary biomarkers of food intake in urine that were correlated to semi-quantitative food intake and diet quality measures. Importantly, this work was able to validate the utility of urinary dietary biomarkers generalizable to participants from 14 countries at different socioeconomic developments, reflecting large regional variations in dietary patterns, including tobacco smoking.

5.3 Materials and Methods

5.3.1 PURE Study Cohort & Dietary Self Reporting

This PURE study comprises an international cohort of current and never smokers that is on a nested-case control study design as described elsewhere [31]. In a sub-study involving 1000

participants from 14 countries, morning void urine samples were collected and stored frozen (-20 to -70 °C), and shipped at -160 °C using liquid nitrogen to Hamilton, Ontario. Urine samples were selected at random in approximately equal proportions of never smokers (*n*=335), as well as light (< 10 cigarettes per day; *n*=324), and heavy (≥10 cigarettes per day; *n*=341) current smokers, with a maximum of one freeze-thaw cycle and 7 h thaw time. Detailed questionnaires were provided, with information regarding age, sex, education, location, smoking status, disease history, alcohol consumption, and current medication use as previously described [32]. Standardized FFQs were also used to collect detailed information on dietary patterns and nutritional status, with subsequent calculations to determine AHEI score provided based on self-reports [32]. To account for variations in locally eaten foods, a master international nutrient database was used and modified to local food composition tables and supplemented with recipes of locally eaten dishes [32].

5.3.2 Chemicals and Reagents

Ultra HPLC grade LC-MS solvents (water, methanol, acetonitrile) were obtained from Caledon laboratories (Georgetown, ON, Canada) for preparation of buffers, sheath liquid and stock solutions. Additional chemicals were all obtained from Sigma-Aldrich Inc. (St Louis, MO, USA).

5.3.3 Nontargeted Metabolite Profiling of Urine by MSI-CE-MS

Prior to analysis, urine aliquots were thawed on ice, vortexed for 30 seconds, and centrifuged (14,000 *g* for 5 min) to remove particulates. Samples were then prepared for nontargeted metabolite profiling with a 5-fold dilution using HPLC-grade water and an internal standard

mixture containing 3-chloro-l-tyrosine (Cl-Tyr), D-glucose-^{13}C, choline-d_9, gaba-d_6, and 2-napthalenesulfonic acid (NMS) at a final concentration of 20 μM in sample. All diluted urine aliquots were then analyzed by MSI-CE-MS [33], on an Agilent 7100 capillary electrophoresis (CE) instrument (Agilent Technologies Inc. Mississauga, ON, Canada) coupled with an Agilent 6230 Time of Flight Mass Spectrometer (TOF-MS), equipped with a coaxial sheath liquid electrospray ionization (ESI) source with heated nitrogen gas. An Agilent 1260 infinity isocratic pump and a 1260 infinity degasser were used to deliver 60 % volume MeOH with 0.1 % volume formic acid sheath liquid for positive ion mode, while 70% volume MeOH was used for negative ion mode at a flow rate of 10 μL/min. For real time mass correction, reference ions including purine and hexakis (2,2,3,3-tetrafluoropropoxy)phosphazine (HP-921) were spiked into the sheath liquid at 0.02 % volume. The nebulizer spray was turned off during serial sample injections to avoid suctioning effects and current instabilities [34]. It was then subsequently turned on at 4 psi following voltage application with the source temperature set to 300 °C, and the drying gas delivered at 8 L/min. The Vcap was set to 2000 V, and the nebulizer gas was delivered at 10 psi, while the sheath gas was set to 195 °C and delivered at 3.5 L/min. Additionally, MS voltage settings were set to a fragmentor of 120 V, the skimmer at 65 V, and Oct1 RF at 750 V. The TOF-MS system was operated with full scan data acquisition over a *m/z* range of 50-1700 and an acquisition rate of 500 ms/spectrum.

Separations were performed on an uncoated fused silica capillary (Polymicro Technologies, Phoenix Az, USA) with an inner diameter of 50 μm, outer diameter of 360 μm, and total length of 135 cm with a voltage of 30 kV at 25 °C. The polyimide coating was removed from both ends (7 mm) to avoid sample carry-over and swelling effects [35]. All diluted urine samples were analyzed under both positive ion mode (cationic/basic or zwitterionic metabolites)

and negative ion mode (anionic/acidic metabolites) conditions with a background electrolyte (BGE) of 1M formic acid with 15% volume acetonitrile (pH 1.8) and 50 mM ammonium bicarbonate (pH 8.5), respectively. Separations were performed under normal polarity, with a pressure gradient of 2 mbar/min from 0 to 40 min for negative ion mode. Serial sample injections consisted of 11 discrete hydrodynamically injected sample plugs, (5 s at 100 mbar) followed by background electrolyte BGE spacers injected electrokinetically at 30 kV for 75 s for cationic metabolites and 45 s for anionic metabolites [36]. A five min flush time with BGE at 950 mbar was used at the start and end of each run. Preventative maintenance included daily cleaning of the ion source with 50% volume isopropanol, CE electrode cleaning, and capillary flushing for 15 min at the start of the day. A daily quality control (QC) run was also performed with a blank to assess stability prior to running samples, with the additional inclusion of a pooled QC within each individual run. High resolution tandem mass spectrometry (MS/MS) spectra were acquired for structural elucidation of unknown metabolites of significance from a pooled QC urine sample on an Agilent G7100A CE system with a coaxial sheath liquid Jetstream electrospray ion source connected to an Agilent 6550 iFunnel QTOF instrument. A single urine sample was injected hydrodynamically at 100 mbar for 90 s, followed by a BGE spacer at 100 mbar for 5 s. Precursor ions were then selected for collision-induced dissociation experiments at varying collision energies (10-40 V). The ESI conditions were set to a Vcap of 3500 V, with a nozzle voltage at 2000 V, nebulizer gas set to 8 psi, drying gas delivered at 14 L/min at 225 °C, while the MS voltage settings included a fragmentor of 380 V and Oct1 RF set to 750 V. Subsequently, acquired MS/MS spectra were annotated based on characteristic product ions, neutral mass losses, and compared to open-access public databases, such as the Human Metabolome Database (HMDB) [37].

5.3.4 Metabolomic Data Processing and Statistical Analysis

Raw data was converted to a mzXML format using MSconvert (Proteowizard) [38] prior to importation in MZmine 2.0 for nontargeted metabolite profiling [39]. Subsequently, settings were set to a minimum signal intensity of 200 counts for spectra generation to extract molecular features by exact mass. The generation of extracted ion electropherograms (EIEs) was preformed based on a minimum signal intensity of 200 counts, a minimum time span of 0.01 min, and a *m*/*z* tolerance of 5 ppm. Adduct filtering was then applied within MZmine2 based on a migration time tolerance of 1.0%, *m*/*z* tolerance of 10 ppm, and maximum relative adduct peak height of 50%. Molecular features were processed using Agilent Mass Hunter Workstation Software (Qualitative Analysis, version B.07.00) after extracting in profile mode using a 10-ppm mass window. EIE's were integrated after smoothing using a quadratic/cubic Savitzky Golay function (15 points) with the peak areas and migration times being transferred to Excel (Microsoft Office, Redmond, WA, USA) for the calculation of relative peak areas (RPAs) and relative migration times (RMTs). In all cases, the integrated ion response (i.e., peak area) for each metabolite was normalized to Cl-Tyr in positive ion mode, and NMS in negative ion mode. Next, the determination of a coefficient of variation (CV) from QC urine samples was calculated, where urinary metabolites with high technical variance (CV > 30%), and low detection frequency (< 60%) were removed from the data matrix. Remaining urinary metabolites were annotated based on their characteristic accurate mass (*m/z*) for their protonated [$M+H^+$], or deprotonated [$M-H^-$] molecular ion, relative migration time (RMT) and detection mode (positive:p or negative:n). Overall, 116 urinary metabolites satisfied selection criteria in our nontargeted metabolomics workflow, which were classified based on their technical precision (CV%) in QCs, biological variance (CV%) in samples, and interclass coefficients (ICC) as

summarized in **Table S5.2**. Given the positive skew among dietary variables due to a substantial number of zero values reported to indicate a lack of dietary intake, a cubic root transformation was performed on all metabolomic and FFC data prior to further analysis in R version 4.0.3 (R Foundation for Statistical Computing, Vienna, Austria). [40–42]. Food records were then correlated against metabolites using a partial Pearson correlation in R with the "ppcor" package [43] following listwise adjustments for age, sex, body mass index (BMI), alcohol intake (current vs former/never), education (secondary and above vs primary education), total energy (kcal), smoking status (current vs never smoker), and history of comorbidities (hypertension and diabetes). Metabolites were filtered based on a significance threshold ($p < 0.05$) and a correlation cut-off ($r > 0.2$). A hierarchical correlation analysis (HCA) plot for FFQ data was subsequently performed in R with the "corrplot" package [44]. ANCOVA for determining dose response and regional comparisons, as well as univariate statistical analyses (t-test, Mann Whitney U test) and normality testing (Shapiro-Wilk test) was conducted using the Statistical Package for Social Sciences (SPSS, version 18.0), with adjustments for covariates (age, sex, BMI, alcohol intake, education, total energy (kcal), smoking status, comorbidities), including specific food categories (e.g., soft drinks, tea, raw vegetables, fish, fruit, coffee). Receiver-operating characteristic curves (ROC), control charts, histograms, and box plots were created using Medcalc version 12.5 (MedCalc Software). In addition, autoscaling and missing value imputation with minimum ion response divided by two was performed prior to multivariate statistical analysis in MetaboAnalyst 5.0 [45], including a principal component analysis (PCA) and HCA/2D heat map for data visualization.

5.4 Results

5.4.1 Cohort Characteristics and Metabolomics Workflow

The participants selected for the PURE pilot study (*n*=1000) were comprised of current (65%) and never smokers (35%) recruited from HICs (Canada, Sweden, United Arab Emirates), MICs (Argentina, Columbia, Brazil, Chile, South Africa, Iran, Philippines) and LICs (Bangladesh, Pakistan Zimbabwe, Tanzania) as described elsewhere [31]. The mean (± SD) age of the study population was 52 (± 9.6) years, with a mean BMI of 26 (± 5.8) kg/m^2, and a mean AHEI score of 32.3 (± 9.1). A nearly equal number of males (52%) and females (48%) were included, with ~ 38% (*n*=376) of participants coming from HICs, ~ 38% (*n*=381) from MICs, and ~24% (*n*=243) from LICs. The cohort was subsequently stratified according to AHEI score (**Figure 5.1A**), with participants in the top quintile (Q5) classified as eating a "healthy" diet (*n*=196), compared to individuals in the bottom four quintiles (Q1-Q4) categorized as an "unhealthy" diet (*n*=782) [46]. The mean diet score (± SD) amongst individuals consuming a healthy diet was found to be (44.7 ± 4.3), while participants eating a poorer diet had a lower score of (29.2 ± 7.1). No significant difference ($p > 0.05$) in age, sex or BMI was measured between these diet quality sub-groups (**Table S5.1**). However, a greater frequency of participants in the unhealthy eating group were from HICs (35%) and MICs (41%) relative to LICs (23%). Importantly, 71% of participants consuming a suboptimal diet self-identified as current tobacco smokers (**Table S5.1**). Although there were no differences ($p > 0.05$) in self-reported dairy, coffee, or total lipids between the two diet quality sub-groups from PURE, daily average consumption of red meat, processed meat, and processed foods, were all elevated ($p < 0.05$) in participants consuming an unhealthy diet (**Table S5.1**). In contrast, participants with a high diet quality based on their AHEI classification reported a greater consumption of fiber, nuts, fruits, vegetables, carbohydrates, and total sugar,

with an increased overall total caloric intake ($p < 0.001$; **Table S5.1**). In addition, diet quality also demonstrated regional variations in the PURE study, with significantly higher scores found in LICs (40.4 ± 4) as compared to similar dietary quality patterns in MICs (30.5 ± 5), and HICs (31.6 ± 6) (**Figure 5.1B**).

Comprehensive metabolite profiling of urine samples (*n*=1000) by MSI-CE-MS (**Figure 5.1C**) characterized a total of 116 urinary metabolites (71 cations, 45 anions) with adequate technical precision (CV < 30%) and frequency of detection (> 60%) with the exception of acesulfame K (ASK) and saccharin that were detected in only a subset of participants from HICs and MICs. A summary of all authenticated urinary metabolites is listed in **Table S5.2**, where each metabolite was annotated based on their accurate mass and relative migration time under positive (*m/z*: RMT:p) or negative (*m/z*:RMT:n) ion mode along with their most likely molecular formula, compound name, mass error, technical precision (CV%), biological variation (CV%), frequency of detection (% FOD) and intraclass coefficient (ICC). Overall, acceptable technical precision (mean CV=15.6%, *n*=160) was achieved based on the repeated analysis of pooled QC urine samples, whereas the mean biological variance in all urine samples analyzed was 94.7% (*n*=1000) as shown in a 2D scores plot when using PCA (**Figure 5.1D**).

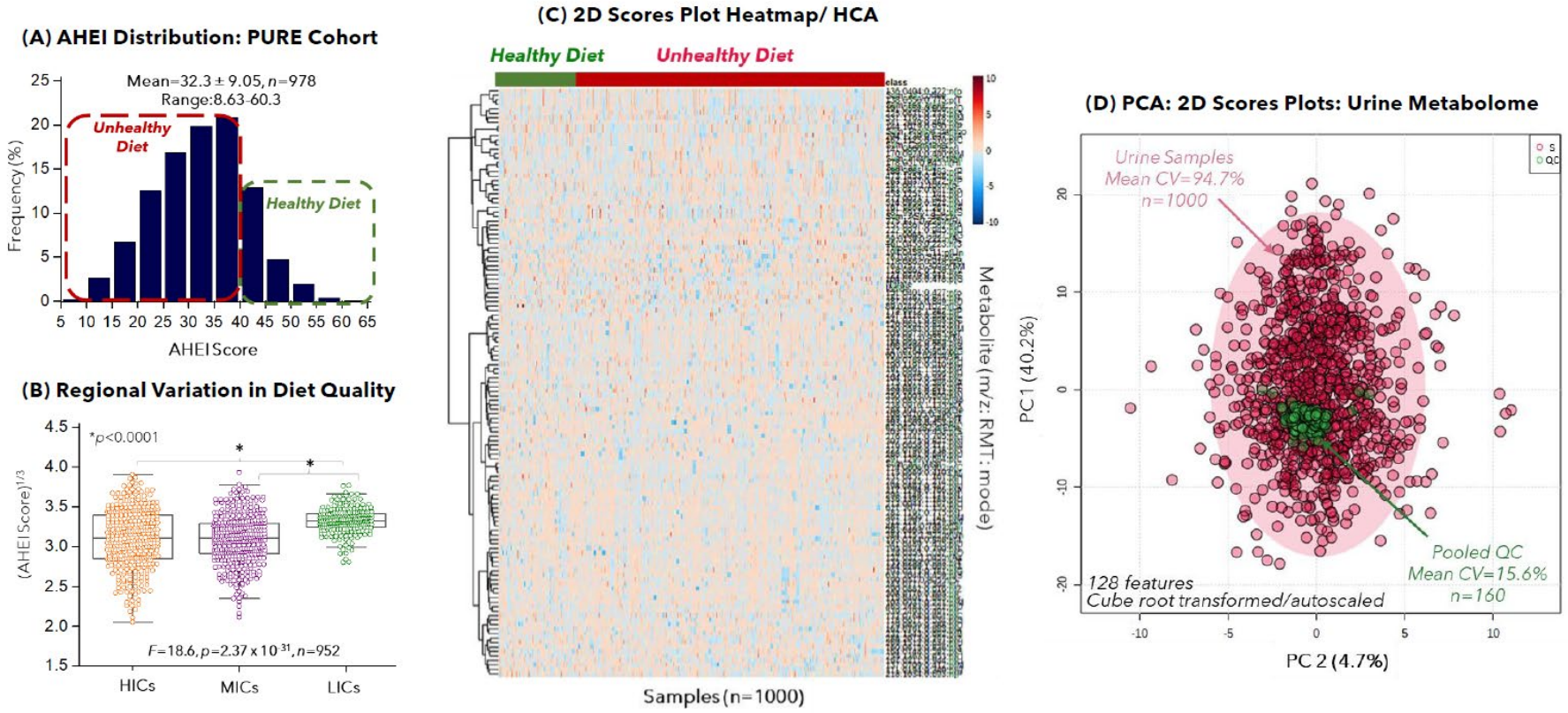


Figure 5.1 (A) Distribution of overall diet quality determined based on the Alternative Healthy Eating Index (AHEI) in the Prospective Urban and Rural Epidemiology (PURE) cohort (*n*=978), with healthy participants (*n*=196) indicated by the top quintile as compared to individuals with an unhealthy diet (*n*=782; quintiles 1-4). **(B)** Variation in diet quality between regions based on self-reported diet records used to determine AHEI scores, with a significant increase found in LICs, relative to HICs and MICs (p=2.37 x 10^{-31}). **(C)** A 2D heatmap with hierarchical cluster analysis (HCA) of the urine metabolome based on nontargeted analysis of polar/ionic metabolites by MSI-CE-MS, following a cubic root transformation and autoscaling. (D) 2D scores plots from PCA highlighting acceptable technical precision (CV =15.6%) as compared to biological variation (CV=94.7%) for 129 features including urine metabolites and their sums.

5.4.2 Correlation between Urine Metabolites and Food Records

Self-reported dietary intake for 60 food groups or specific food items was obtained from FFQs collected from PURE participants, with individual descriptions summarized in **Table S5.3** [32]. Following a partial Pearson correlation analysis between food records (n=60) and urinary metabolites (n=116), metabolite sums (n=9) and previously reported iodide, thiocyanate, and nitrate concentrations with adjustments for age, sex, BMI, smoking status, total energy (kcal), alcohol usage, education, and comorbidities (hypertension and diabetes), a resultant 7998 correlations were observed (**Figure 5.2**). With the application of a significance threshold ($p<$ 0.05), 47.6% of correlations were removed, resulting in 4195 significant associations. Subsequently, a more stringent cut-off of $r > 0.20$ [47] was applied, resulting in a total of 215 meaningful pairings between urinary metabolites and self-reported food intake measures (**Figure 5.2**). Among those remaining, 29 food records were correlated with ≥ 1 metabolite (n=48 metabolites). An HCA of the 29 food records indicated five distinct food classes, including coffee, processed foods and sugars, meats and fat, nuts and soy, along with fruits and vegetables (**Figure 5.2**). In instances with more than one urinary metabolite per food record, as many as five metabolites were selected based on the rank order of correlation coefficients ($r \geq 0.25$), leaving a 19 candidate dietary biomarkers of food intake excreted in urine. **Table S5.4** summarizes these 19 candidate urinary biomarkers, including predetermined sums (e.g., trigonelline and quinic acid) and their three strongest FFQ associations. Overall, meat and fat intake generated the most correlations, with a total of seven urinary metabolites, and two sums satisfying statistical thresholds ($p < 0.05$, $r > 0.20$; **Table S5.4**). Next, coffee consumption was significantly associated with two distinct metabolites (trigonelline and quinic acid), in addition their sum (r =0.32-0.41; **Table S5.4**). Notably, all three urinary biomarkers associated with coffee were

also correlated with self-reported raw vegetable intake (r = 0.26-0.31; **Table S5.4**). Comparatively, only urinary proline betaine and the sum of proline betaine with pantothenic acid were associated with fruits, and the sum of fruits and vegetables (r = 0.21-0.24). In addition, food records related to nuts and soy (n = 3) all showed a significant relationship to uracil (r >0.20, p < 0.0001; **Table S5.4**). Lastly, ASK, saccharin and their tabulated sum displayed a significant correlation and high specificity for processed foods (r = 0.24), soft drinks and total sugar (r = 0.23-0.29), and both metabolites did not demonstrate a relationship with any additional food categories (**Table S5.4**).

Selection of the strongest associations within each food class resulted in a total of eight distinct urinary biomarker-diet pairings (**Table 5.1**). Subsequent validation according to dose response and discriminatory performance were performed as outlined in the data workflow in **Figure 5.3**. After checking for adequate reproducibility in QC samples (median CV < 15%), data from semi-quantitative self-reported food records were correlated with metabolomic data following further corrections for specific dietary covariates. For instance, the correlation between coffee intake and urinary trigonelline underwent an additional adjustment for tea, soft drinks, and raw vegetables, while carnitine and red meat were adjusted for fish intake. Similarly, the correlation of raw vegetables to the sum of urinary quinic acid and trigonelline was adjusted for coffee and fruit consumption. Next, self-reports were stratified and categorized by consumption patterns in a binary manner (i.e., high *versus* low intake) corresponding to either quartiles or literature thresholds indicative of health benefits (e.g., < 1cup/day vs ≥ 1cup/day for coffee) [48,49]. Accordingly, dose response and discriminatory performance were evaluated based on an ANCOVA with covariate adjustments and ROC analysis, respectively. A significance threshold (p < 0.05) and median fold-change cut-off (FC > 1.5) were applied while classification

performance was determined according to the area under the curve (AUC > 0.65). Amongst the eight selected dietary biomarkers of food intake from the PURE study, seven apart from uracil met these criteria. Overall, adjustments for covariates did not result in the loss of significant correlations, or fold changes, however a moderate drop in correlation strength was observed for six out eight top candidates, with the exception of saccharin and the sum of quinic acid and trigonelline, where the unadjusted values are outlined in (**Table S5.5**).

5.4.3 Top Candidate Biomarkers for Habitual Food Intake

Urinary metabolites surpassing the established significance threshold ($p < 0.05$), correlation coefficient cut-off ($r > 0.20$), and minimum discriminatory performance (AUC > 0.65, $p < 0.05$) included candidate biomarkers for coffee ($n = 1$), fruits and vegetables ($n = 2$), meats and fats ($n = 2$), as well as processed foods and total sugar ($n = 2$) as summarized in (**Table 5.1)**. Next, these top-ranked dietary biomarkers in urine were compared with diet quality (AHEI score), where an ANCOVA and partial Pearson correlation were employed with adjustments for covariates (**Table S5.6**). Overall, the eight initially selected candidates demonstrated significant differences between self-reported healthy diet and unhealthy diet quality with the exception of ASK. However, urinary ASK, carnitine, trigonelline, saccharin, and the sum of quinic acid and trigonelline, showed a significant negative correlation with AHEI (r = -0.07 to -0.16, $p < 0.002$), whereas urinary proline betaine exhibited a modest positive correlation ($r = 0.07$) for participants in the PURE study. Comparatively, urinary 3-methylhistidine and uracil revealed no signification association (**Table S5.6**). Of the seven top-ranked correlations found between self-reported food records and urinary metabolites, trigonelline established the most robust biomarker-food record pairing as illustrated in **Figure 5.3**.

Table 5.1. Top-ranked candidate biomarkers associated for specific food records when using partial Pearson correlation analysis followed by ANCOVA for dose response between high and low self-reported intake.

Biomarker		**Correlation to Food Record** [a]			**Dose Response** [b]			
Tentative ID	**Identifier (*m/z*:RMT:mode)**	**Food Record**	***r***	***p*-value (*n*)**	***F*-value**	***p*-value (*n*)**	**High:Low**	**AUC ($p<0.0001$)**
Trigonelline	138.0550:0.775:p $C_7H_7NO_2$	Coffee (mL/day)	0.374	4.47×10^{-16} (n=430)	7.08	1.55×10^{-12} (n=450)	3.4	0.80
Saccharin	181.9917:1.154:n $C_7H_5NO_3S$	Total Sugar (g/day)	0.308	2.80×10^{-13} (n=548)	6.61	2.08×10^{-10} (n=555)	2.8	0.68
Sum (Quinic acid & Trigonelline)	-	Raw Vegetables (g/day)	0.296	3.27×10^{-16} (n=737)	34.0	1.05×10^{-69} (n=945)	2.6	0.70
Acesulfame K	161.9869:1.267:n $C_4H_5NO_4S$	Processed Foods (g/day)	0.285	6.42×10^{-6} (n=251)	2.87	0.001 (n=252)	4.1	0.70
3-Methylhistidine	170.0924:0.440:p $C_7H_{11}N_3O_2$	MUFA (g/day)	0.284	2.74×10^{-18} (n=914)	27.8	1.53×10^{-46} (n=914)	3.3	0.83
Carnitine	162.1125:0.570:p $C_7H_{15}NO_3$	Red Meat (g/day)	0.272	3.05×10^{-16} (n=903)	16.8	2.33×10^{-30} (n=903)	2.4	0.68
Proline Betaine	144.1019:0.930:p $C_7H_{13}NO_2$	Fruits & Vegetables (g/day)	0.243	1.81×10^{-13} (n=903)	12.9	3.51×10^{-21} (n=903)	2.7	0.68
Uracil	111.0194:0.340:n $C_4H_4N_2O_2$	Nuts & Soy (g/day)	0.223	1.12×10^{-10} (n=827)	12.1	1.01×10^{-19} (n=827)	1.4	0.66

[a] Food records from self-reported food frequency questionnaires with the strongest association to biomarker response ($r>0.2$, $p<0.05$) were selected. A partial Pearson correlation analysis of urinary metabolites to specific food records following a cubic root transformation with listwise deletion when adjusted for age, sex, BMI, smoking status (current vs never smoker), alcohol intake (current user vs former/never user), total energy (kcal), education (secondary education and above vs primary), and history of comorbidities (hypertension and diabetes) was performed. Additional adjustments for soft drinks, tea and raw vegetables were applied to coffee intake, while the raw vegetable correlation underwent additional corrections for fruit and coffee consumption, and the red meat correlation to metabolite response was further corrected for fish intake. [b] ANCOVA analysis between high and low intake was determined with the previously described corrections, where fold changes are reported based on median response of the respective biomarker in each group (high vs low), followed by a receiver-operating characteristic (ROC) curves to obtain an area under the curve (AUC) at a significance threshold of $p<0.0001$ to examine discrimination performance between high and low intake.

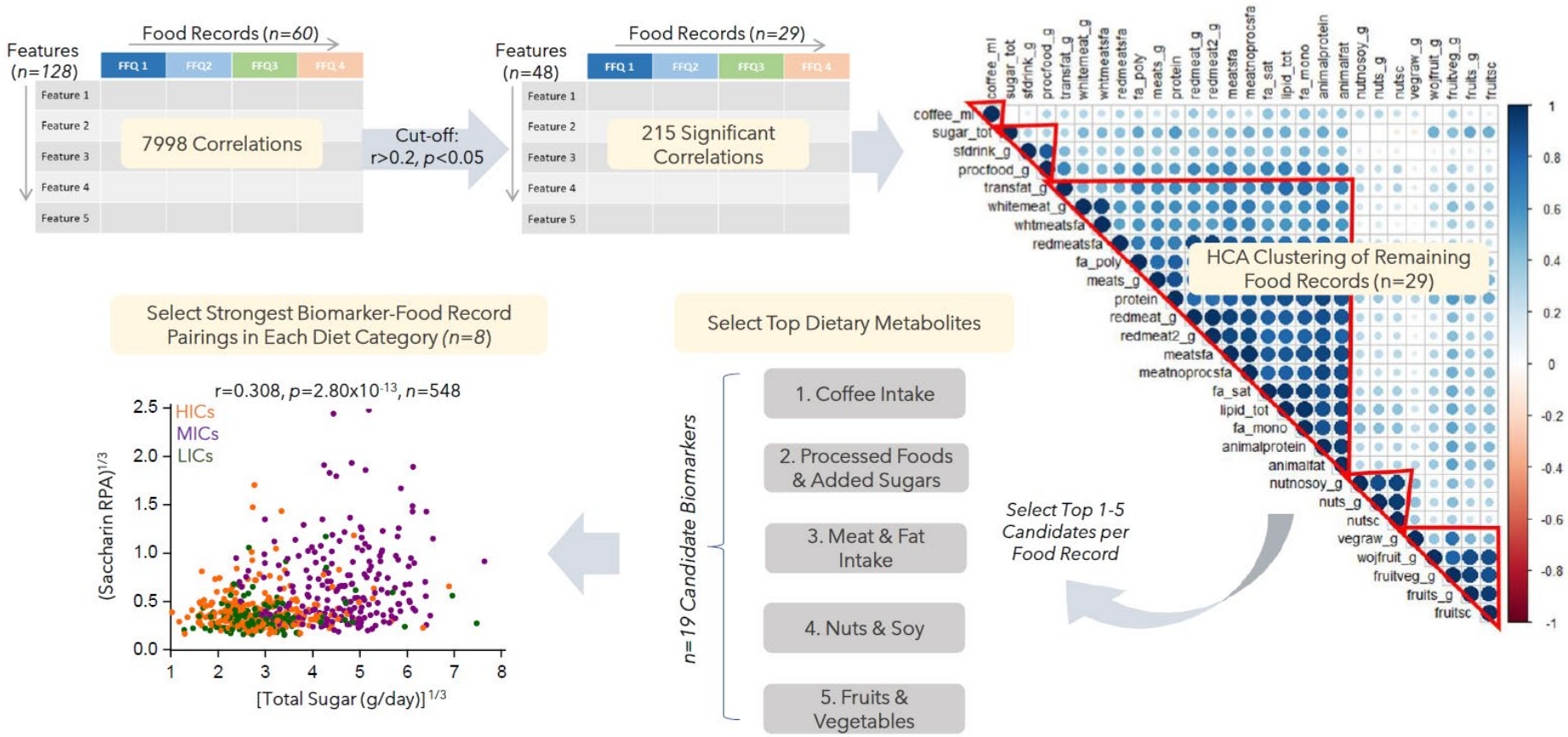


Figure 5.2 Overview of statistical analysis workflow for the identification of robust dietary biomarker-food record associations. Steps include a partial Pearson correlation analysis with covariate adjustment between all food records (n=60) and metabolite features including calculated sums (n=128), results in 1998 correlations. Following a cut-off threshold of r>0.2, and p<0.05, significant features were reduced to 215 significant correlations including 29 food records and 48 metabolites. Hierarchical cluster analysis (HCA) of self-reported food records reveals 5 distinct food categories for which the top 1-5 candidate biomarkers were selected based on ranked correlation coefficients revealing 19 potential biomarkers. The strongest biomarker-food record pairings (n=8) were chosen within each food category for further validation (e.g. dose response, predictive capacity, generalizability).

An overall 3.4-fold increase in urinary trigonelline excretion, was measured between high (≥ 1cup/day) versus low (< 1cup/day) consumers of coffee, with a strong classification capacity (AUC=0.80, $p < 0.0001$). In addition, regional trends in coffee consumption were suggested by a 3.4-fold increase in trigonelline for participants from HICs, compared to LICs, and a 2.7-fold increase as compared to MICs (**Table S5.6**). In contrast, only a 1.2-fold increase was seen between MICs and LICs, with 86% (*n*=210) of participants from LICs (most from Bangladesh) reporting no coffee intake.

Urinary trigonelline, along with quinic acid and their sum also displayed a significant association with raw vegetable intake specifically (r = 0.26-0.31, $p < 0.0001$). Following a covariant adjustment for self-reported fruit and coffee consumption, the significant associations remained, the strongest being between raw vegetables and the sum of urinary quinic acid and trigonelline in the adjusted model (r = 0.296). Accordingly, a significant dose response (*FC*=2.6, $p < 0.0001$), and moderate discrimination performance between high and low consumers of vegetables (AUC=0.70, $p < 0.0001$) was achieved using a cut-off value of 75 g/day of raw vegetable intake, equivalent to that of one serving per day (**Figure S5.1A**) [6]. As expected, a regional variation was also noted, where HICs exhibited the highest levels of raw vegetable intake as compared to MICs and LICs (**Figure S5.1B**). In comparison, the sum of fruits of vegetables displayed the strongest association to urinary proline betaine, while carnitine was selected as the top-ranked dietary biomarker for red meat intake. Established thresholds of 375 g/day for fruits and vegetables, as well as 70 g/day for red meat (**Figure 5.4**) were selected in accordance with previously reported health benefits associated with reduced CVD risk [6,50,51] Consequently, both dietary biomarkers in urine demonstrated a considerable 2.4- and 2.7-fold increase in biomarker response for red meat as well as fruits and vegetables respectively (**Figure**

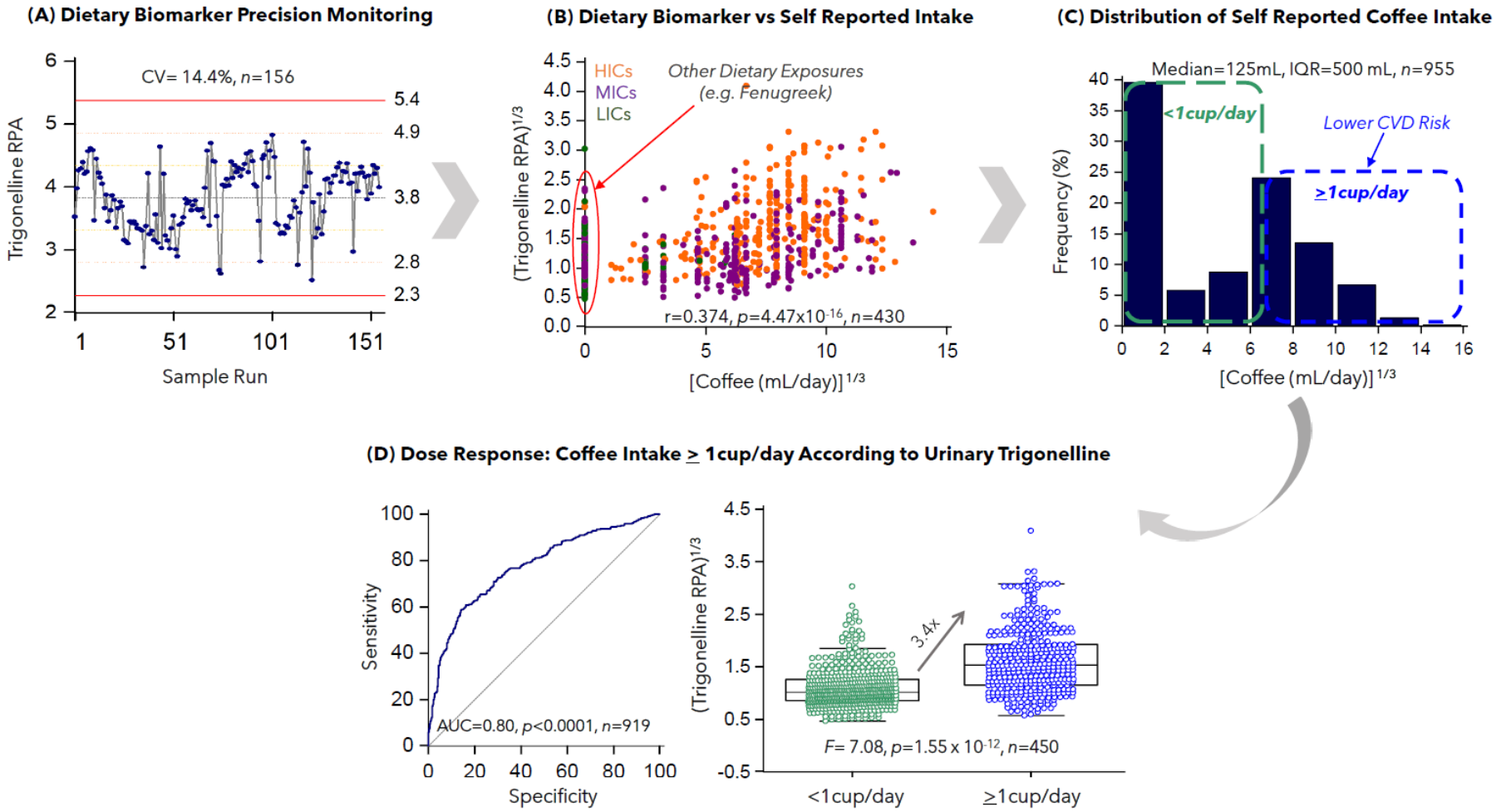


Figure 5.3 Overview of the validation workflow for individual candidate biomarkers (e.g. Trigonelline), including **(A)** acceptable technical precision based on repeat analysis of pooled QCs (CV<15%). **(B)** With subsequent correlation of metabolite signal (RPA) relative to self-reported food records using partial Pearson correlation analysis of cubic root transformed data with listwise adjustment for covariates. **(C-D)** Distribution of self-reports demonstrates the stratification of food records and their subsequent categorization based on previously reported health benefits (e.g., ≥1cup/day coffee intake) with an ANCOVA and receiver operating cure (ROC) performed to monitor dose response and discriminatory capacity of select candidate biomarkers.

5.4), both having a good classification performance (AUC = 0.68; **Table 5.1**). In addition, regional variations in urinary proline betaine and carnitine between country income status was evident, with more than a 5-fold increase in abundance was measured between HICs and LICs. Comparatively, urinary carnitine only displayed a 1.1-fold elevation when comparing HICs to MICs, while proline betaine was found to be on average 1.9 times greater (**Table S5.6**). A significant correlation was also observed between self-reported intake of monounsaturated fats (MUFA) and 3-methylhistidine (**Figure S5.2**). In accordance with literature, formal intake recommendations fall within a narrow range of 16-22 g/day or 15-20% of total daily caloric intake [52,53]. As such, self-reports were stratified and categorized by low intake (Q1; < 10 g/day) compared to moderate/high intake (Q2-4; ≥ 10 g/day). A notable 3.3- fold increase along with a strong discrimination performance (AUC=0.83, $p < 0.0001$) between predefined exposure levels was achieved, indicating a significant relationship between urinary metabolite excretion and dietary MUFA intake. Moreover, regional trends in urinary 3-methylhistidine were similar to that of carnitine, with a 1.2- and 3.6-fold increase in urinary metabolite levels in HICs relative to MICs and LICs, respectively (**Table S5.6**). In contrast, nut and soy intake according to the AHEI nut score (0-10) was shown to be associated with urinary uracil, whose chemical structure was putative identified based on an annotated MS/MS spectrum (CID 10-40 V) (**Figure S5.3**) [54]. In this case, there was a modest correlation of urinary uracil with self-reported nut and soy intake ($r = 0.223$) having a dose response of 1.4-fold between high and low intake scores corresponding to AHEI guidelines [55]. However, despite regional variations in urinary uracil excretion between HICs, MICs and LICs (**Figure S5.3**), a poor discriminatory performance was found as reflected by an AUC of 0.662 (**Table 5.1**).

Dietary biomarkers in urine for self-reported processed food and total sugar were found to be ASK and saccharin, respectively. While a low FOD (%) was observed for both exogenous/artificial sweeteners, (< 60%) significant correlations with their respective food consumption patterns (processed food and total sugar) were evident ($r > 0.25$; **Table 5.1**). While ASK exhibited the fewest detects (22%), MS/MS confirmed its putative identification (Level 2; **Figure 5.5**), based on a comparison to previously published MS/MS spectra at an optimal CID of 10 V [56]. As expected, urinary ASK was predominantly measured in participants from HICs and MICs with few detects (n=16, ~7%) in LICs (**Figure 5.5**). While specific dietary thresholds are not currently recognized for these artificial sweeteners, self-reports were stratified and categorized based on low (Q1-2) and high (Q3-4) intake coinciding with a cut-off of 75 g/day. In this case, a notable 4.1-fold dose response was determined with good discriminatory performance based on ROC curves (AUC = 0.70; **Table 5.1**). Similarly, urinary saccharin was associated with total sugar intake (**Table 5.1**). With recommended guidelines of 50 g/day [57], a median fold change of 2.8 was measured, whereby a moderate discrimination was achieved based on an AUC of 0.68. Urinary saccharin was also correlated with ASK ($r = 0.193$, $p < 0.05$) and follows a similar regional distribution, where HICs and MICs are elevated relative to LICs, with MICs displaying the greatest overall saccharin exposures (**Figure S5.4**).

Given the strong association found for urinary tetrahydroaldosterone glucuronide and dihydrotestosterone glucuronide with raw vegetable intake, fruits and vegetables, as well as red and processed meats, these metabolites were explored as biomarkers of physiological effect. According to previous reports of the relationships between tetrahydroaldosterone glucuronide and potassium [58], along with their role in cardiovascular health, the correlation between potassium, extrapolated from food frequency questionnaire data [32], was correlated with both

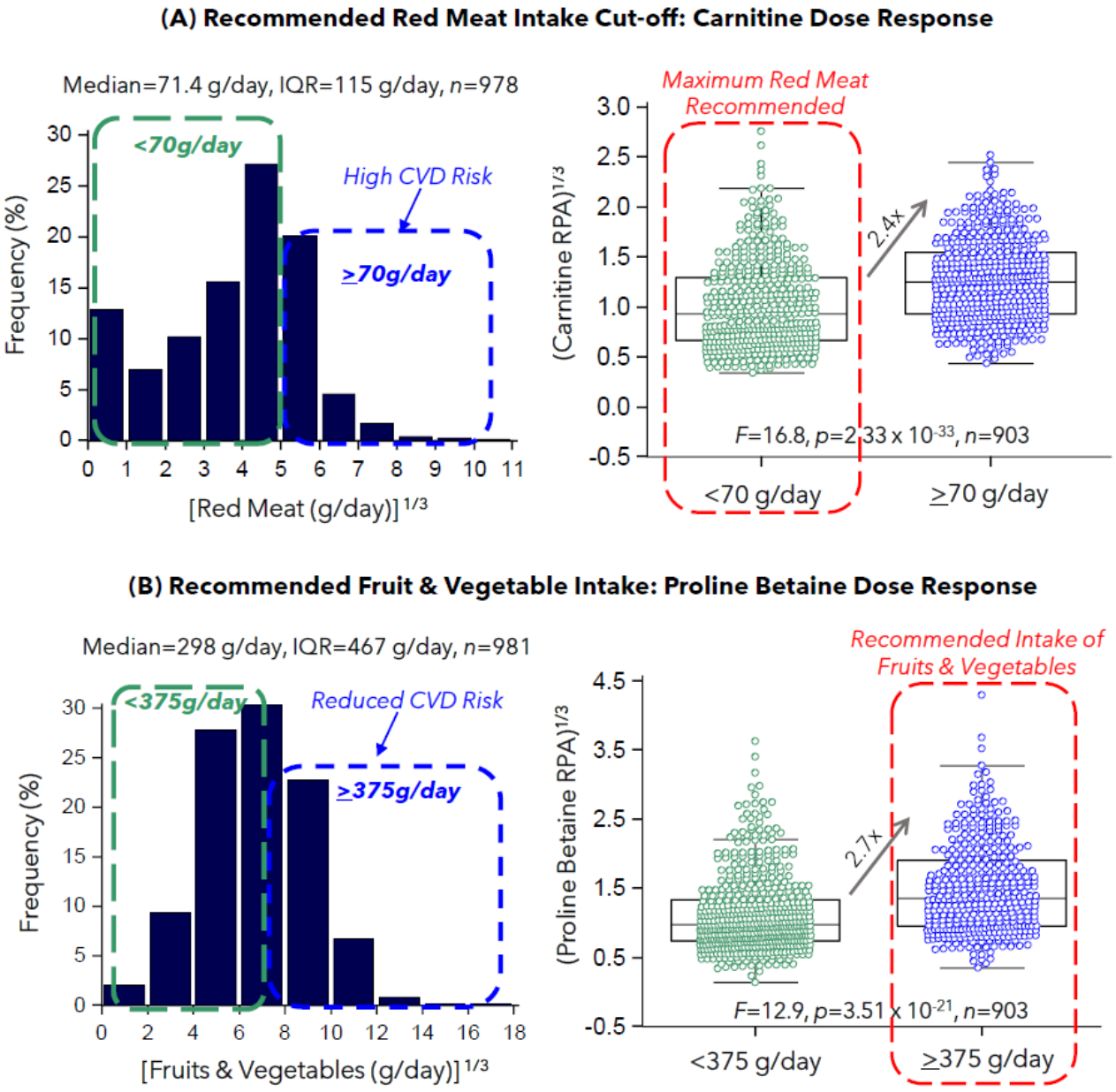


Figure 5.4 Comparison of dose response based on recommended health guidelines for **(A)** red meat along with **(B)** total fruit and vegetable consumption following stratification of self-reports. An overall fold change of 2.4- and 2.7- fold biomarker response (RPA) from low to high self-reported intake is observed.

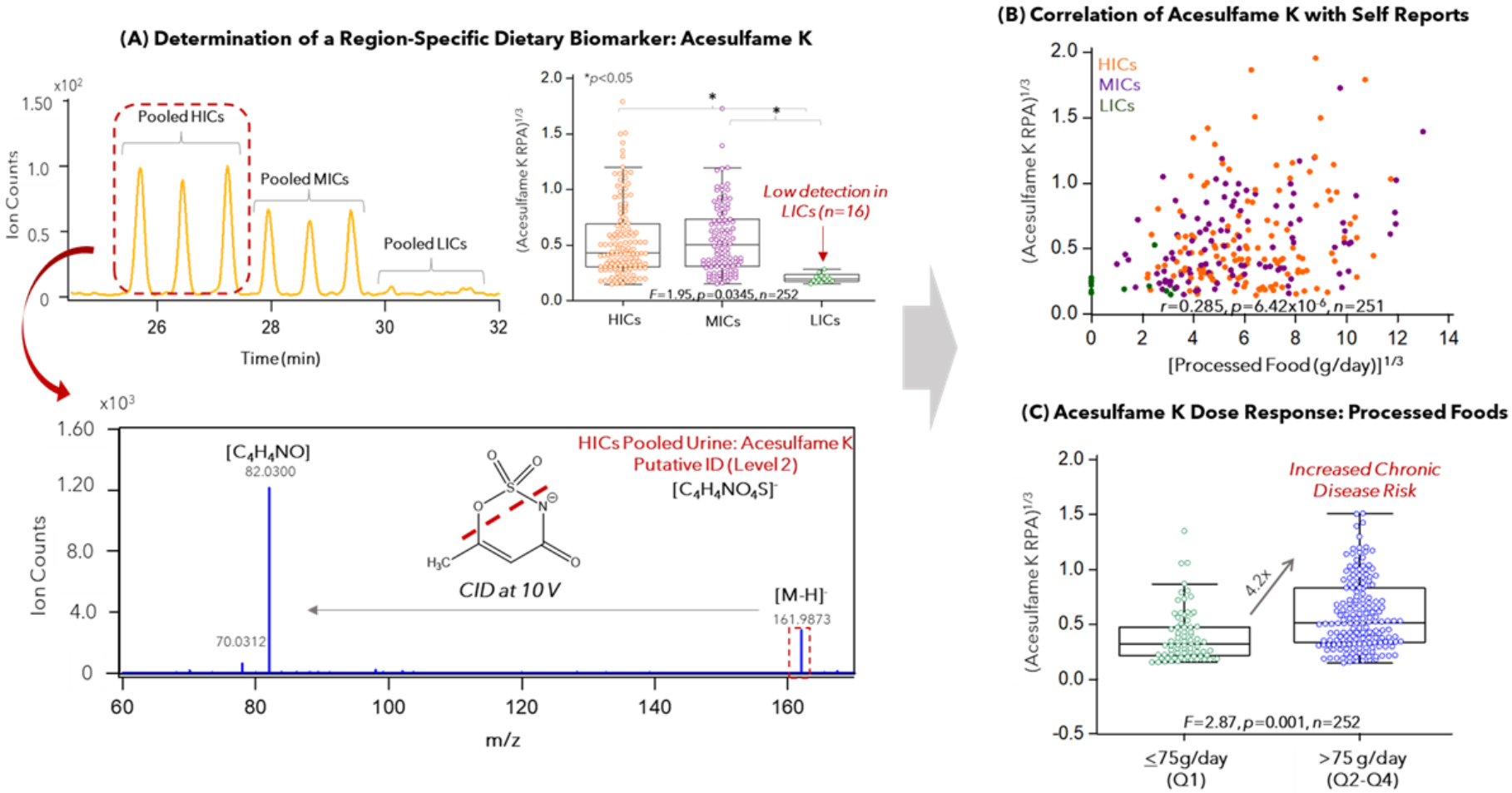


Figure 5.5 (A) Selection of a biomarker found in primarily HICs and MICs with few (n=16) detects amongst LICs, indicative of regional specificity and limited intake in certain countries. Putative identification based on a characteristic MS/MS spectrum (CID at 10 V) in a pooled HIC sample provides confidence (Level 2) that this is not a spurious signal and is likely the artificial sweetener Acesulfame K (ASK). **(B)** Correlation to self-reports (r= 0.285, p < 0.05) and **(C)** dose response between low and high intake (FC=4.2,p<0.05) further demonstrates a significant relationship to self-reported processed foods.

steroid hormones, along with diastolic and systolic blood pressure (**Table S5.7**). A significant correlation was observed between both steroid conjugates and potassium excretion following adjustments for age, sex, BMI, smoking, education, alcohol, and comorbidities, (r = 0.232-0.224, p<0.05; **Figure S5.5).** As expected, both urinary metabolites were also highly co-linear with each other (r = 0.764, **Figure S5.5**). However, only dihydrotestosterone glucuronide showed a weak negative correlation with diastolic blood pressure (r = -0.112, **Figure S5.5**). In addition, both metabolites demonstrate notable relationships to hypertension progression, defined as normal blood rpessure (< 120 mmHg systolic, < 80 mmHg diastolic), elevated/stage 1 hypertension (systolic: 120-139 mmHg, diastolic: 80-89 mmHg), and stage 2 hypertension (systolic > 140 mmHg, diastolic> 90 mmHg). A significant fold change was observed between normal blood pressure levels and stage 2 for both tetrahydroaldosterone glucuronide (FC = 1.3, p < 0.05), and dihydrotestosterone glucuronide (FC = 1.2, p < 0.05).

Lastly, HCA clustering of the eight top-ranked dietary biomarkers, along with their associated food records and two additional potential biomarkers of biological effect are depicted in **Figure S5.6**. Overall, three distinct clusters are outlined pertaining to features associated with healthy dietary patterns, smoking habits, and unhealthy diet choices. As expected, metabolites including proline betaine and uracil were found amongst the healthy diet cluster along with AHEI, potassium, fruits and vegetables, as well as nuts and soy. In contrast, biomarkers representing meat and fat intake, including urinary carnitine and 3-methylhistidine were clustered with ASK and saccharin. These biomarkers were associated with an unhealthy diet and have a distinct negative association with AHEI diet scores as shown in **Figure S5.6**. Moreover, smoking related features including self-reported cigarettes per day (CPD), total nicotine equivalents (TNE-7) were clustered with features associated with coffee consumption. Upon

further investigation, seven out of eight candidate biomarkers in urine demonstrated a significant association with smoking status following adjustments for age, sex, BMI, education, alcohol use, total energy, co-morbidities, and specific food covariates (**Table S5.8**). However, urinary ASK was the exception, likely a result of its low detection rate in participants from the PURE study (n=252). Urinary proline betaine that was associated with fruits and vegetable consumption, demonstrated a lower intake amongst current smokers, whereas urinary biomarkers for nuts and soy, raw vegetables, coffee, red meat, MUFA, and total sugar were consumed to a greater extent relative to never smokers (**Table S5.8**). However, habitual coffee intake as reflected by urinary trigonelline excretion had the strongest association to current smoking (r = 0.364, p < 0.0001). Comparatively, raw vegetables, red meat, MUFA, along with nuts and soy also showed modest positive correlations (r > 0.10, p < 0.014) with urinary (TNE-7), which is a robust biomarker of recent tobacco smoke exposure (**Table S5.8**).

5.4 Discussion

5.4.1 Identification of dietary biomarkers of food intake

This works presents a high-throughput and nontargeted data workflow for the identification of robust dietary biomarkers of food intake in urine that are applicable to diverse populations with variable dietary patterns and lifestyle. Top-ranked urinary metabolites correlated with specific food categories were subsequently validated based on dose-response, discriminatory performance, and generalizability in participants from 14 different countries in the PURE study. In addition, we revealed regional trends for five important food classes relevant to diet quality scores based on AHEI, that likely impact chronic disease risk and potentially exacerbate the unadjusted hazards of tobacco smoking. Overall, 19 urinary metabolites exhibited meaningful

associations with 29 food intake records (**Table S5.4**), irrespective of age, sex, BMI, smoking, education, alcohol intake, total energy, and comorbidities. Overall, eight top-ranked dietary biomarkers of food intake were selected in accordance with the strength of their correlation to self-reports of daily food consumption (**Table 5.1**), which were further evaluated for dose-response and classification performance. While we have independently replicated previously reported food intake urinary biomarkers using an untargeted metabolomics data workflow, including trigonelline, proline betaine, carnitine, 3-methylhistidine, ASK, and saccharin, this work demonstrates for the first time their generalized utility in an international cohort of participants from 14 different countries for routine biomonitoring of habitual diet [23,56,59].

Diet quality according to AHEI was shown to differ across regions globally, with lower median scores found in HICs and MICs as compared to LICs (**Figure 5.1**), likely a consequence of the predominantly Western diet in these countries (e.g., Canada, Argentina, Sweden) reflected by greater average consumption of red meat, processed foods, added sugars, and unhealthy fats [60]. Of the top eight candidates (**Table 5.1**) all selected urinary biomarkers did exhibit moderate fold-changes (*FC* = 0.71-1.36) between stratified healthy and unhealthy diet quality, in addition to nominal correlations with diet quality scores (**Table S5.6**). Urinary metabolites associated with fruit and vegetable intake, nuts and soy, coffee consumption and MUFA were elevated in a healthy diet, while those associated with red meat, processed foods and total sugar were consumed to a greater extent for participants consuming a suboptimal diet. Given that AHEI is determined from cumulative ranked scores involving multiple food categories, these results reveal how a single biomarker may not reflect overall diet quality [14]. However, despite AHEI having been previously used to assess chronic disease risk [14,55], it does not capture nutrient bioavailability and metabolism, cooking methods or certain food sources used in distinct

regions or cultures [23]. For instance, coffee intake which has shown an inverse association with both total and cause-specific mortality is not included in the AHEI assessment score [61].

5.4.2 Urinary Trigonelline Reflects Coffee Consumption Patterns

Coffee is the most widely consumed beverage worldwide [61], with a higher prevalence among HICs and MICs, with tea being more common in LICs [62]. Often coffee consumption is mistakenly considered an unhealthy habit due to the presence of caffeine, a popular stimulant [61,63]. However, it has been shown to contribute a large proportion of the recommended daily intake of dietary antioxidants, even greater than tea, fruits and vegetables [61,64,65]. Likewise, decaffeinated coffee is compositionally similar to regular coffee and has also shown beneficial health effects [61,66]. Comparable to previous studies, trigonelline demonstrated the strongest correlation ($r = 0.662$) with self-reports, in addition to quinic acid ($r = 0.323$) and their respective sum ($r = 0.401$) [67]. Aside from caffeine, trigonelline is a bitter alkaloid naturally present in high abundance in coffee beans, and has been reported as a candidate biomarker of exposure up to three days following ingestion [68]. Trigonelline is also able to better capture intake of both decaffeinated and caffeinated coffee unlike other excreted metabolites (e.g., 1-methyluric acid), with little to no contribution from other caffeine sources (e.g., energy drinks, soft drinks), making it a more specific biomarker in both acute and habitual dietary exposure. Moreover, its direct excretion into urine following coffee consumption without extensive biotransformation, is an ideal trait for a dietary biomarker of food intake. Conversely, chlorogenic acids (e.g., quinic acid), a class of polyphenols, are the most abundant antioxidants found in coffee beans [61,69]. However, they are found in other plant sources consumed and are strongly influenced by host metabolism and gut microbiome activity, resulting in greater inter-individual variations [68].

Amongst PURE participants (*n*=1000), those found in HICs had the highest intake of coffee according to urinary trigonelline levels, as compared to MICs and LICs (**Tables S5.6**). However, a large variation in signal produced by trigonelline was observed from those reporting no coffee intake (**Figure 5.3B**). This may be a result of survey misreporting, as well as the contribution of trigonelline from other dietary sources such as fenugreek, a commonly used spice in South Asian cuisine, likely contributing to background levels found in LICs who reported no coffee consumption (e.g., Bangladesh) [70–72]. In addition, regional disparities may be a consequence of coffee preparation and roasting processes, where increased roasting resulting in lower trigonelline and chlorogenic acids [73,74]. Overall, coffee consumption ≥ 1 cup/day has been linked to reduced CVD risk, with the most salutary dose reported between 3-4 cups/day [61,63]. Specifically, trigonelline and quinic acid have known antioxidant and anti-inflammatory properties contributing to the reduced risk of cardiovascular disease and type 2 diabetes [61]. Overall, urinary trigonelline represents a robust urinary biomarker of coffee intake in the PURE study with excellent discriminatory performance (AUC = 0.80) between < 1 cup/day and ≥ 1 cup/day that may provide new insights into the potential health benefits of coffee consumption not considered in the AHEI diet quality score.

5.4.3 Urinary Biomarkers for Fruit and Vegetable Intake

The adequate intake of fruits and vegetables has been reported to have protective effects against NCDs due to the high concentrations of dietary fiber, vitamins, minerals, and antioxidants [75–77]. Although the World Health Organization (WHO) recommends a daily intake of 400 g/day [6,78], existing guidelines are largely based on European and US data [6]. Currently, most countries report consumption below the WHO recommendations, partly due to the availability

and cost of fresh fruits and vegetables in many LICs [79]. Based on recent work in the PURE cohort by Yusuf *et al.* [6] intake as low as 375 g/day has demonstrated similar benefits against risk for CVD risk and all-cause mortality. In addition, while fruits are commonly eaten fresh, vegetables are often consumed both raw and cooked in HICs, while they are primarily ingested after cooking in LICs (e.g., Bangladesh) [6]. Surprisingly, few studies have investigated the intake of raw vegetables alone, despite potential differences in nutrient absorption and bioavailability, resulting in a lack of established dietary guidelines [80]. In this work, we identified the sum of urinary trigonelline and quinic acid as a biomarker for raw vegetables, following adjustment for coffee and fruit consumption (**Figure S5.1**). Sources of quinic acid include various plant-based foods containing chlorogenic acids (e.g., carrots, artichoke) [81]. Similarly, trigonelline has been found in many plant seeds, lettuce as well as peas [82]. Although these metabolites are less specific, they do demonstrate a moderate correlation with self-reported raw vegetable intake (r = 0.296), as well as a significant fold-change between low and high vegetable intake (FC=2.6-fold). This combination of urinary metabolites has not been previously reported as a dietary biomarker for raw vegetables likely due to the strong influence of coffee consumption. However, when FFQ data is available, these measurements may be used synergistically to monitor raw vegetable consumption with adjustments for confounding food exposures.

In contrast, urinary proline betaine was shown to be a more specific biomarker for intake of fruits and vegetables (r = 0.243). Although previous reports have implicated proline betaine as a robust measure for citrus fruit intake, it has also been reported in Chinese artichoke, as well as to a small degree in fresh vegetables [83]. Consequently, we reveal a correlation to self-reported fruit and vegetable intake, with a significant dose response (median FC = 2.7) based on recent

dietary recommendations in the PURE study by Yusuf *et al.* [6] (> 375 g/day). Notably, an increase in fruit and vegetable intake was prevalent among participants from HICs and MICs according to biomarker response (**Table S5.6**). These observed regional patterns may be associated with the diversity of this international cohort of participants, differences in food preparation (e.g., cooked vs raw) and the types of foods comprising the various diet categories. Moreover, South Asian diets (e.g., Bangladesh) consist of more whole grains, fewer raw fruits and vegetables along with less alcohol and red meat consumption compared to Western diets [24]. Importantly, although individual food categories often considered to be a staple of healthy diet (e.g., fruits and vegetables) may be consumed in higher amounts in HICs, Western diet quality is considered low overall when unhealthy food categories (e.g., processed foods, red meat, added sugar) are also taken into account for AHEI scores (**Figure 5.1B**).

5.4.4 Urinary Biomarkers for Meat and Fat Consumption

The consumption of certain meats and dietary fats (e.g., MUFA, PUFA) can be a valuable source of macro- and micronutrients in human health [84], if consumed in the correct form and quantity with minimal processing, chemical preservation or charring when cooking. For instance, increased amounts of red and processed meats have been associated with certain cancers (e.g., colorectal cancer), CVD, and type 2 diabetes risk [84,85], as compared to the consumption of white meat (e.g., chicken). Similarly, dietary sources of MUFA (e.g., fish, olive oil) have exhibited health benefits if consumed within recommended ranges of 15-20% of total energy intake (16-22 g/day) [46]. In contrast, red meat provides an excellent source of iron, vitamin B12, and zinc, however consumption below 70 g/day is recommended [51], as elevated intake may contribute to atherogenic effects [84]. Several dietary biomarkers previously reported for

meat intake have been replicated in the PURE study, including urinary carnitine, acetylcarnitine, imidazole propionic acid, and 3-methylhistidine (**Table S5.4**) [23,86,87]. Overall, urinary carnitine demonstrated the strongest association with red meat (r = 0.272), followed by red meat without the inclusion of processed meat (r = 0.247). While carnitine and acetylcarnitine are naturally more abundant in red meat as compared to white meat, previous studies have shown several other sources of dietary carnitine, including fish and dairy products [87–89]. Although a significant increase was observed between low (< 70 g/day) and high (> 70 g/day) intake of red meat, a moderate classification performance in ROC curves may be explained by the endogenous production of carnitine and other dietary sources [23,86]. Comparatively, urinary 3-methylhistidine was found to correlate with MUFA, PUFA and meat related fatty acids (**Table S5.4**), with the strongest relationship observed with total MUFA. A notable dose response (median *FC*=3.3) and discriminatory performance (AUC = 0.83) between low (Q1) and high (Q2-4) MUFA consumption was also observed. Since 3-methylhistidine is reported as a biomarker of muscle protein turnover, this association with self-reported fats is likely a consequence of meat consumption [86]. Given, MUFA and PUFA are both classified as healthy fats, this indicates that the association of urinary 3-methylhistidine to a higher diet quality is likely associated with a health promoting Prudent diet as compared to a Western diet pattern [56]. Based on carnitine and 3-methylhistidine responses, a significant increase in meat and fat intake was observed for participants from HICs (median *FC* = 3.6-5.1), as compared to MICs and LICs. For instance, Ranbhate *et al.* [90] have reported a negative correlation of red meat with life expectancy, specifically in HICs and MICs indicating potential for increased NCD risk for these demographics. Our work also revealed regional variation in eating habits, where despite substantial increases in global meat consumption in the last 30 years (500%) [91], individuals in

LICs have a reduced intake of specific meat products (e.g., red meat, processed meat) [91]. However, both carnitine and 3-methylhistidine are normal constituents in urine as result of muscle metabolism, and thus may lack specificity as they are also influenced by host physiology. [23,56] This demonstrates a major challenge in nutritional epidemiology when validating robust and food-specific dietary biomarkers. As a result, biomonitoring of compounds that are not produced endogenously are preferred as they can indicate food specific dietary exposures with greater certainty with less potential for confounding.

5.4.5 Biomonitoring Processed Food and Sugar Consumption in PURE

Ultra-processed foods and added sugars are considered to be another major risk factor for chronic diseases and increased gut inflammation, resulting in its incorporation to the AHEI diet score [92,93]. Several previously reported biomarkers including sucralose, cyclamates, steviol glycosides, saccharin, and ASK [92] have been identified in various processed foods. We reported the detection of urinary saccharin and ASK in the PURE study, whereby we found they were significantly associated with total sugar and processed food respectively (**Table 5.1**). Given the widespread use of artificial sweeteners found in not only soft drinks, but also several processed foods including, low-sugar snacks and ready to eat meals [92,94,95], an expected, highly specific correlation between ASK and both processed foods (r = 0.285) as well as soft drinks (r = 0.233) was observed in our work. While specific recommendations for processed foods are not available, the consensus of restricted intake is emphasized. As such a cut-off between low Q1-2 (< 75 g/day) and high Q3-4 (> 75 g/day) consumers was used to determine a dose response since increased risk of inflammation was previously reported at high intake levels [96,97]. Overall, a significant 4-fold dose response between low (< 70 g/day) and high (> 75

g/day) intake was identified in the PURE study. Comparatively, recommendations of < 50g/day have been made for total sugar, with more specific guidelines available for added sugar (<10%) by the WHO [98]. Since saccharin has reported a greater widespread use internationally as a result of its low cost, along with its prevalence in various low-calorie sweeteners, artificially sweetened beverages and desserts [92], a more robust correlation was shown for urinary saccharin with total sugar (r = 0.308). A significant dose response between < 50 g/day and > 50 g/day (FC=2.8) was also demonstrated in the PURE study. However, moderate classification performance was noted for both candidate biomarkers (AUC = 0.68-0.70), likely attributed to the lower frequency of detection (< 60%), especially among LICs (**Figure S5.4**). Currently, South American countries which make up most of the MICs in this study have reported to have the highest consumption of artificial sweeteners, followed by HICs [95]. This is also suggested by the regional trends observed in PURE for ASK and saccharin, where median signals were highest among MICs as compared to LICs which were either relatively low or undetectable (i.e., below method detection limit). Despite an infrequent detection rate in certain regions, urinary ASK and saccharin are ideal dietary biomarkers of food intake given they are produced exogenously, are not metabolized extensively and their presence is a clear indicator of dietary exposure. While artificial sweeteners have been advertised as low calorie, healthy alternatives to sugar, their presence in processed foods, soft drinks and refined sweetened foods make them a likely contributor to the increased inflammatory bowel disease risk associated with processed food consumption [99]. In addition, the consumption of these foods has been recently linked to increased urinary concentrations of several endocrine disrupting plasticizers (e.g., phthalates), which may further exacerbate chronic disease risk [100]. Additional evidence of their potential deleterious effects on health outcomes was reported in a recent epidemiological study

highlighting the connection between artificial sweeteners, such as aspartame, and ASK with elevated cancer risk [101]. Thus, the consequences of unhealthy eating patterns may be compounded by the inadvertent consumption of environmental contaminants, emphasizing the need for reliable dietary biomarkers of exposure to predict disease risk.

5.4.6 Candidate Biomarkers of Biological Effect

A biomarker related to diet can either provide an objective measure of specific foods or conversely, they can reveal the impact of dietary exposures on host physiology and potential implications toward disease risk [17]. For instance, tetrahydroaldosterone-3-glucuronide, a major aldosterone metabolite excreted in urine, was found to have a considerable correlation with raw vegetable intake (r = 0.262; **Table S5.4**) as well as the sum of fruits and vegetables (r = 0.203). Tetrahydroaldosterone-3-glucuronide is a primary mineralocorticoid, responsible for promoting sodium and potassium transport, with an important role in blood pressure regulation [102]. Specifically, elevated potassium levels, have been reported with a reduction in blood pressure and cardiovascular disease risk [103]. A previous study by Mente *et al.* [58] implicating potassium as a clinically useful predicter of diet quality found a positive relationship to vegetables, fruits, whole grains, low-fat diary, and fish, with a significant inverse relationship to blood pressure [58]. Similarly, this work demonstrates a moderate positive correlation between tetrahydroaldosterone glucuronide and estimated potassium intake based on self-reports (r = 0.232, **Figure S5.5**), with a notable decrease in metabolite levels observed for participants with stage 2 hypertension (**Table S5.7**). Given that increased potassium levels are typically found in fruits and vegetables, this is likely suggestive of a surrogate physiological measure of potassium intake [58,102,104].

Interestingly, dihydrotestosterone glucuronide, a major testosterone conjugate excreted in urine, also revealed moderate correlations with the dietary intake of raw vegetables in addition to MUFA, as well as red and processed meat (r = 0.212-0.213; **Table S5.4**) suggestive of a biomarker of biological effect or physiological response Previous studies pertaining to high consumption of animal fat and red meat have shown an upregulation in testosterone levels, indicative of potential health risks [105]. However, the role of testosterone on cardiovascular disease risk is largely inconclusive, as the effects of exogenous and endogenous testosterone have produced contradicting results [106]. In this work, the excretion of dihydrotestosterone glucuronide was found to be significantly correlated to both potassium, as well as tetrahydroaldosterone-3-glucuronide (r =0.224-0.764; **Figure S5.5**), with a notable decrease in excretion among individuals at stage 2 hypertension (median FC = 1.2**; Table S5.7**). In addition, a modest negative association was also observed with diastolic blood pressure in the PURE study (r = -0.112; **Table S5.7**). Consequently, this indicates a possible protective effect against hypertension, which may be a result of either increased cholesterol metabolism or an indirect association with an overall healthy diet given the significant relationship observed with raw vegetable intake (r = 0.212, **Table S5.4**).

Similarly, urinary uracil which was found to be correlated consistently with food records indicative of nut and soy intake, demonstrating a moderate dose-response of 1.4-fold (**Table 5.1**). However, this was likely a result of increased purine/pyrimidine metabolism commonly observed after ingestion of purine-rich foods (e.g., nuts, seafood, soy protein), which is associated with a healthier diet for those consuming more nuts and soy proteins [107,108]. As observed in **Figure S5.6**, where the clustering of uracil, along with the excreted conjugates of testosterone and aldosterone with both self-reported food records and metabolite biomarkers

relating to healthy diet demonstrates the complex associations between dietary intake and host physiology. Thus, the relationships shown by select metabolites and dietary exposures are likely biomarkers of effect owing to an overall higher diet quality, as compared to a specific markers of food intake, with potential implications towards cardiovascular disease risk.

5.4.7 Identifying Relationships Between Smoking and Diet

Among the top risk factors for NCDs, tobacco smoking is considered the largest contributor to preventable death worldwide [3]. Current smokers often also consume a poor diet quality as compared to never smokers, which can further exacerbate chronic disease risk synergistically [109,110]. In addition, tobacco smoking is commonly associated with elevated coffee intake with current smokers having a greater propensity for caffeine consumption [111]. As outlined in **Figure S5.6**, three distinct clusters demonstrate the patterns of dietary metabolites, smoking and self-reported food records. Overall, a clear relationship with coffee intake is observed based on HCA assigned clusters, however a negative association is showed with AHEI, self-reported fruits and vegetables, raw vegetables, and potassium (**Figure S5.6**). More specifically **Table S5.8** illustrates the significant relationship of all dietary biomarkers with smoking status, where fruits and vegetables intake were lower in current smokers (median FC = 0.9) while meats, fats, nuts, processed foods and coffee consumption were consistently higher compared to never smokers (median FC=1.1-1.4). Moreover, urinary biomarkers for coffee, red meat, MUFA, raw vegetables and nut intake also had a significant correlation with urinary TNE-7 (**Table S5.8**), a robust biomarker for recent tobacco smoke exposure that is independent on metabolic rate [112,113]. Given that urinary uracil was used to monitor nut and soy consumption, this correlation is more likely a consequence of increased purine/pyrimidine metabolism common

among smokers, rather than a diet specific relationship [114]. Similarly, raw vegetable intake was associated to the sum of urinary quinic acid and trigonelline, following adjustment for coffee and fruit intake, where quinic acid is also found in tobacco leaves and may be a confounder in this association [115]. In contrast, coffee shows the strongest correlation based on urinary trigonelline (r = 0.380). Although coffee has putative health benefits, the presence of caffeine, a known stimulant has been known to increase metabolic activity amongst smokers and increase restlessness which can impact smoking patterns [111]. Consequently, the complex relationship between diet and smoking requires further investigation with controlled feeding studies and prospective clinical data to better elucidate the impact of specific food items on smoking habits, and overall tobacco related disease risk.

5.4.8 Limitations and Future Work

Overall, this work demonstrates a cost-effective, high-throughput approach towards the comprehensive assessment of the urine metabolome to determine the potential utility of several dietary biomarkers in a diverse international population. However, limitations include the use of FFQs for the identification and validation of candidate biomarkers. Self-reported survey data is prone to misreporting, as well as gender and social desirability bias, often exacerbated in lower socioeconomic regions. This, compounded with the delay between sample collection and FFQ implementation, may also explain the weaker correlations observed by urinary metabolites detected in this study. Moreover, analysis of urinary electrolytes can provide a more accurate depiction of essential minerals and their role in cardiac health (e.g., sodium, potassium), as compared to estimated values determined based on self-reported data. In addition, nontargeted urine profiling by MSI-CE-MS was limited to polar ionic metabolites and was unable to detect

lower abundance food exposures. Thus, future investigations with expanded metabolome coverage on both urine and blood samples, will utilize nonaqueous multi-segment injection capillary electrophoresis (MSI-NACE-MS) [34], to provide a larger panel of candidate biomarkers better suited for certain food groups not ascertained in this work (e.g., fish intake, dairy). Lastly, future studies with expanded cohort sizes and prospective clinical data will allow for increased study power, while also enabling a more accurate assessment of hazard ratios due to compounded modifiable risk factors (e.g., suboptimal diet, heavy smoking, poor air quality), on all-cause mortality and clinical events over an average 8-year time frame, such as incidence of cardiovascular disease, IBS, diabetes, and specific cancers.

5.5 Conclusions

Our work provides compelling evidence for the utility of a panel of urinary biomarkers of food intake in a diverse free-living population, that also reflects large regional variations in dietary patterns in the PURE study. Overall, while the AHEI diet score can depict general dietary habits and overall differences in diet quality between regions, it is prone to self-reporting bias, exacerbated in LICs, and is not able to capture food preparation methods, nutrient absorption, dose response or coffee consumption. In this context, we demonstrate the applicability of more specific dietary biomarkers for select food items, while also addressing potential confounding sources of exposure which may be unique to certain regions (e.g., fenugreek vs coffee in LICs). Comparatively, dietary biomarkers indicating broader food categories (e.g., proline betaine, carnitine, quinic acid, trigonelline), can provide insight on habitual eating patterns, more closely reflecting overall diet quality (e.g., meats and fats). The findings in this work have critical implications towards the management and prevention of NCDs. Through the elucidation of

robust and generalizable dietary biomarkers we can provide more accurate assessments of food intake and nutritional status across diverse populations worldwide. With majority of current interventions and recommendations for nutrition relying on data from HICs, our work addresses understudied populations in LICs at greater risk for misreporting and chronic disease burden. Thus, objective dietary biomarkers that have been replicated in multi-ethnic populations are critical in guiding future evidence based public health policies, which are also applicable across regions of varying income status.

5.6 References:

1. Hajat, C.; Stein, E. The Global Burden of Multiple Chronic Conditions: A Narrative Review. *Prev. Med. Reports* **2018**, *12*, 284-293.
2. Malta, D. C., Bernal, R. T. I., Gomes, C. S., Cardoso, L. S. D. M., Lima, M. G., & Barros, M. B. D. A. Inequalities in the use of health services by adults and elderly people with and without noncommunicable diseases in Brazil, 2019 National Health Survey. *Revista Brasileira de Epidemiologia,* **2021**, 24, e210003.
3. Ng, R.; Sutradhar, R.; Yao, Z.; Wodchis, W.P.; Rosella, L.C. Smoking, Drinking, Diet and Physical Activity—Modifiable Lifestyle Risk Factors and Their Associations with Age to First Chronic Disease. *Int. J. Epidemiol.* **2020**, *49*, 113–130.
4. Ojo, O. Nutrition and Chronic Conditions. *Nutrients* **2019**, *11*, 459.
5. Tucker, K.L. The Role of Diet in Chronic Disease. *Present Knowl. Nutr.* **2020**, 329–345.
6. Miller, V.; Mente, A.; Dehghan, M.; Rangarajan, S.; Zhang, X.; Swaminathan, S.; Dagenais, G.; Gupta, R.; Mohan, V.; Lear, S.; et al. Fruit, Vegetable, and Legume Intake, and Cardiovascular Disease and Deaths in 18 Countries (PURE): A Prospective Cohort Study. *Lancet* **2017**, *390*, 2037–2049.
7. Liu, W.; Hu, B.; Dehghan, M.; Mente, A.; Wang, C.; Yan, R.; Rangarajan, S.; Tse, L.A.; Yusuf, S.; Liu, X.; et al. Fruit, Vegetable, and Legume Intake and the Risk of All-Cause, Cardiovascular, and Cancer Mortality: A Prospective Study. *Clin. Nutr.* **2021**, *40*, 4316–4323.
8. Afshin, A.; Sur, P.J.; Fay, K.A.; Cornaby, L.; Ferrara, G.; Salama, J.S.; Mullany, E.C.; Abate, K.H.; Abbafati, C.; Abebe, Z.; et al. Health Effects of Dietary Risks in 195 Countries, 1990–2017: A Systematic Analysis for the Global Burden of Disease Study 2017. *Lancet* **2019**, *393*, 1958–1972.
9. Elizabeth, L.; Machado, P.; Zinöcker, M.; Baker, P.; Lawrence, M. Ultra-Processed Foods

and Health Outcomes: A Narrative Review. *Nutrients* **2020**, *12*, 1–36.

10. Onvani, S.; Haghighatdoost, F.; Surkan, P.J.; Larijani, B.; Azadbakht, L. Adherence to the Healthy Eating Index and Alternative Healthy Eating Index Dietary Patterns and Mortality from All Causes, Cardiovascular Disease and Cancer: A Meta-Analysis of Observational Studies. *J. Hum. Nutr. Diet.* **2017**, 30, 216-226.

11. Seidelmann, S.B.; Claggett, B.; Cheng, S.; Henglin, M.; Shah, A.; Steffen, L.M.; Folsom, A.R.; Rimm, E.B.; Willett, W.C.; Solomon, S.D. Dietary Carbohydrate Intake and Mortality: A Prospective Cohort Study and Meta-Analysis. *Lancet Public Heal.* **2018**, *3*, 419–428.

12. Schwingshackl, L.; Hoffmann, G. Diet Quality as Assessed by the Healthy Eating Index, the Alternate Healthy Eating Index, the Dietary Approaches to Stop Hypertension Score, and Health Outcomes: A Systematic Review and Meta-Analysis of Cohort Studies. *J. Acad. Nutr. Diet.* **2015**, *115*, 780-800.

13. Guillermo, C.; Boushey, C.J.; Franke, A.A.; Monroe, K.R.; Lim, U.; Wilkens, L.R.; Le Marchand, L.; Maskarinec, G. Diet Quality and Biomarker Profiles Related to Chronic Disease Prevention: The Multiethnic Cohort Study. *J. Am. Coll. Nutr.* **2020**, *39*, 216–223.

14. Chiuve, S.E.; Fung, T.T.; Rimm, E.B.; Hu, F.B.; McCullough, M.L.; Wang, M.; Stampfer, M.J.; Willett, W.C. Alternative Dietary Indices Both Strongly Predict Risk of Chronic Disease. *J. Nutr.* **2012**, *142*, 1009–1018.

15. McCullough, M.L.; Willett, W.C. Evaluating Adherence to Recommended Diets in Adults: The Alternate Healthy Eating Index. *Public Health Nutr.* **2006**, *9*, 152–157.

16. Al-Ibrahim, A.A.; Jackson, R.T. Healthy Eating Index versus Alternate Healthy Index in Relation to Diabetes Status and Health Markers in U.S. Adults: NHANES 2007-2010. *Nutr. J.* **2019**, *18*, 1–18.

17. Maruvada, P.; Lampe, J.W.; Wishart, D.S.; Barupal, D.; Chester, D.N.; Dodd, D.; Djoumbou-Feunang, Y.; Dorrestein, P.C.; Dragsted, L.O.; Draper, J.; et al. Perspective: Dietary Biomarkers of Intake and Exposure—Exploration with Omics Approaches. *Adv. Nutr.* **2020**, *11*, 200–215.

18. Bell, W.; Colaiezzi, B.A.; Prata, C.S.; Coates, J.C. Scaling up Dietary Data for Decision-Making in Low-Income Countries: New Technological Frontiers. *Adv. Nutr.* **2017**, *8*, 916–932.

19. Archer, E.; Marlow, M.L.; Lavie, C.J. Controversy and Debate: Memory-Based Methods Paper 1: The Fatal Flaws of Food Frequency Questionnaires and Other Memory-Based Dietary Assessment Methods. *J. Clin. Epidemiol.* **2018**, *104*, 113–124.

20. Amoutzopoulos, B.; Steer, T.; Roberts, C.; Cade, J.E.; Boushey, C.J.; Collins, C.E.; Trolle, E.; Boer, E.J.D.; Ziauddeen, N.; van Rossum, C.; et al. Traditional Methods v. New Technologies – Dilemmas for Dietary Assessment in Large-Scale Nutrition Surveys and Studies: A Report Following an International Panel Discussion at the 9th International Conference on Diet and Activity Methods (ICDAM9), Brisban. *J. Nutr. Sci.* **2018**, 1–10.

21. Brennan, L.; Hu, F.B.; Sun, Q. Metabolomics Meets Nutritional Epidemiology:

Harnessing the Potential in Metabolomics Data. *Metabolites* **2021**, *11, 709.*

22. Miller, T.M.; Abdel-Maksoud, M.F.; Crane, L.A.; Marcus, A.C.; Byers, T.E. Effects of Social Approval Bias on Self-Reported Fruit and Vegetable Consumption: A Randomized Controlled Trial. *Nutr. J.* **2008**, *7*, 1–7.

23. Rafiq, T.; Azab, S.M.; Teo, K.K.; Thabane, L.; Anand, S.S.; Morrison, K.M.; De Souza, R.J.; Britz-Mckibbin, P. Nutritional Metabolomics and the Classification of Dietary Biomarker Candidates: A Critical Review. *Adv. Nutr.* **2021**, *12*, 2333–2357.

24. de Souza, R.J.; Shanmuganathan, M.; Lamri, A.; Atkinson, S.A.; Becker, A.; Desai, D.; Gupta, M.; Mandhane, P.J.; Moraes, T.J.; Morrison, K.M.; et al. Maternal Diet and the Serum Metabolome in Pregnancy: Robust Dietary Biomarkers Generalizable to a Multiethnic Birth Cohort. *Curr. Dev. Nutr.* **2020**, *4*, 1–12.

25. Guertin, K.A.; Moore, S.C.; Sampson, J.N.; Huang, W.Y.; Xiao, Q.; Stolzenberg-Solomon, R.Z.; Sinha, R.; Cross, A.J. Metabolomics in Nutritional Epidemiology: Identifying Metabolites Associated with Diet and Quantifying Their Potential to Uncover Diet-Disease Relations in Populations. *Am. J. Clin. Nutr.* **2014**, *100*, 208–217.

26. Guasch-Ferre, M.; Bhupathiraju, S.N.; Hu, F.B. Use of Metabolomics in Improving Assessment of Dietary Intake. *Clin. Chem.* **2018**, *64*, 82–98.

27. Picó, C.; Serra, F.; Rodríguez, A.M.; Keijer, J.; Palou, A. Biomarkers of Nutrition and Health: New Tools for New Approaches. *Nutrients* **2019**, *11*, 1–30.

28. Brennan, L.; Hu, F.B. Metabolomics-Based Dietary Biomarkers in Nutritional Epidemiology-Current Status and Future Opportunities. *Mol. Nutr. Food Res.* **2019**, *63,* 1701064.

29. Brennan, L. Moving toward Objective Biomarkers of Dietary Intake. *J. Nutr.* **2018**, *148*, 821–822.

30. Turner, C.; Kalamatianou, S.; Drewnowski, A.; Kulkarni, B.; Kinra, S.; Kadiyala, S. Food Environment Research in Low- and Middle-Income Countries: A Systematic Scoping Review. *Adv. Nutr.* **2020**, *11*, 387–397.

31. Sathish, T.; Teo, K.K.; Britz-McKibbin, P.; Gill, B.; Islam, S.; Paré, G.; Rangarajan, S.; Duong, M.L.; Lanas, F.; Lopez-Jaramillo, P.; et al. Variations in Risks from Smoking between High-Income, Middle-Income, and Low-Income Countries: An Analysis of Data from 179 000 Participants from 63 Countries. *Lancet Glob. Heal.* **2022**, *10*, 216–226.

32. Teo, K.; Chow, C.K.; Vaz, M.; Rangarajan, S.; Yusuf, S.; Islam, S.; Zhang, M.; Dehghan, M.; DeJesus, J.; Mackie, P.; et al. The Prospective Urban Rural Epidemiology (PURE) Study: Examining the Impact of Societal Influences on Chronic Noncommunicable Diseases in Low-, Middle-, and High-Income Countries. *Am. Heart J.* **2009**, *158*, 1-7.

33. Kuehnbaum, N.L.; Kormendi, A.; Britz-Mckibbin, P. Multisegment Injection-Capillary Electrophoresis-Mass Spectrometry: A High-Throughput Platform for Metabolomics with High Data Fidelity. *Anal. Chem.* **2013**, *85*, 10664–10669.

34. Azab, S.; Ly, R.; Britz-Mckibbin, P. Robust Method for High-Throughput Screening of

Fatty Acids by Multisegment Injection-Nonaqueous Capillary Electrophoresis-Mass Spectrometry with Stringent Quality Control. *Anal. Chem.* **2019**, *91*, 2329–2336.

35. Yamamoto, M.; Ly, R.; Gill, B.; Zhu, Y.; Moran-Mirabal, J.; Britz-McKibbin, P. Robust and High-Throughput Method for Anionic Metabolite Profiling: Preventing Polyimide Aminolysis and Capillary Breakages under Alkaline Conditions in Capillary Electrophoresis-Mass Spectrometry. *Anal. Chem.* **2016**, 88, 10710-10719.

36. Saoi, M.; Li, A.; McGlory, C.; Stokes, T.; von Allmen, M.T.; Phillips, S.M.; Britz-Mckibbin, P. Metabolic Perturbations from Step Reduction in Older Persons at Risk for Sarcopenia: Plasma Biomarkers of Abrupt Changes in Physical Activity. *Metabolites* **2019**, *9*, 134.

37. Wishart, D.S.; Guo, A.C.; Oler, E.; Wang, F.; Anjum, A.; Peters, H.; Dizon, R.; Sayeeda, Z.; Tian, S.; Lee, B.L.; et al. HMDB 5.0: The Human Metabolome Database for 2022. *Nucleic Acids Res.* **2022**, *50, D622-D631*.

38. Adusumilli, R.; Mallick, P. Data Conversion with ProteoWizard MsConvert. In *Proteomics* **2017**, 339–368.

39. Pluskal, T.; Castillo, S.; Villar-Briones, A.; Orešič, M. MZmine 2: Modular Framework for Processing, Visualizing, and Analyzing Mass Spectrometry-Based Molecular Profile Data. *BMC Bioinformatics* **2010**, *11*, 1–11.

40. R Core Team R: A Language and Environment for Statistical Computing, 2020.

41. van den Berg, R.A.; Hoefsloot, H.C.J.; Westerhuis, J.A.; Smilde, A.K.; van der Werf, M.J. Centering, Scaling, and Transformations: Improving the Biological Information Content of Metabolomics Data. *BMC Genomics* **2006**, *7*, 1–15.

42. Thompson, F.E.; Midthune, D.; Subar, A.F.; McNeel, T.; Berrigan, D.; Kipnis, V. Dietary Intake Estimates in the National Health Interview Survey, 2000: Methodology, Results, and Interpretation. *J. Am. Diet. Assoc.* **2005**, *105*, 352–363.

43. Kim, S. Ppcor: Partial and Semi-Partial (Part) Correlation. **2015**.

44. Taiyun Wei and Viliam Simko R Package "Corrplot": Visualization of a Correlation Matrix. **2021**.

45. Pang, Z.; Chong, J.; Zhou, G.; De Lima Morais, D.A.; Chang, L.; Barrette, M.; Gauthier, C.; Jacques, P.É.; Li, S.; Xia, J. MetaboAnalyst 5.0: Narrowing the Gap between Raw Spectra and Functional Insights. *Nucleic Acids Res.* **2021**, *49,* 388-396.

46. Schwingshackl, L.; Hoffmann, G. Monounsaturated Fatty Acids, Olive Oil and Health Status: A Systematic Review and Meta-Analysis of Cohort Studies. *Lipids Health Dis.* **2014**, *13*, 1–15.

47. Wang, Y.; Gapstur, S.M.; Carter, B.D.; Hartman, T.J.; Stevens, V.L.; Gaudet, M.M.; McCullough, M.L. Untargeted Metabolomics Identifies Novel Potential Biomarkers of Habitual Food Intake in a Cross-Sectional Study of Postmenopausal Women. *J. Nutr.* **2018**, *148*, 932–943.

48. Miranda, A.M.; Goulart, A.C.; Benseñor, I.M.; Lotufo, P.A.; Marchioni, D.M. Coffee Consumption and Risk of Hypertension: A Prospective Analysis in the Cohort Study. *Clin. Nutr.* **2021**, *40*, 542–549.

49. Kouli, G.M.; Panagiotakos, D.B.; Georgousopoulou, E.N.; Mellor, D.D.; Chrysohoou, C.; Zana, A.; Tsigos, C.; Tousoulis, D.; Stefanadis, C.; Pitsavos, C. J-Shaped Relationship between Habitual Coffee Consumption and 10-Year (2002–2012) Cardiovascular Disease Incidence: The ATTICA Study. *Eur. J. Nutr.* **2018**, *57*, 1677–1685.

50. Diet, Activity and Cancer - WCRF International, Available online: https://www.wcrf.org/diet-activity-and-cancer/ (accessed on 25 May 2022).

51. Clinton, S.K.; Giovannucci, E.L.; Hursting, S.D. The World Cancer Research Fund/American Institute for Cancer Research Third Expert Report on Diet, Nutrition, Physical Activity, and Cancer: Impact and Future Directions. *J. Nutr.* **2020**, *150*, 663.

52. Baxter, C.S.; Hoffman, J.D.; Knipp, M.J.; Reponen, T.; Haynes, E.N. Exposure of Firefighters to Particulates and Polycyclic Aromatic Hydrocarbons. *J. Occup. Environ. Hyg.* **2014**, *11,* 85-91.

53. FAO/WHO. Expert Consultation on Fats and Fatty Acids in Human Nutrition Interim Summary of Conclusions and Dietary Recommendations on Total Fat & Fatty Acids Summary of Total Fat and Fatty Acid Requirements for Adults , Infants (0-24 Months) and Children (2-18 Years). **2008**.

54. MassBank of North America. Available online: https://mona.fiehnlab.ucdavis.edu/spectra/display/MoNA035782 (accessed on 26 May 2022).

55. Van Duong, T.; Tseng, I.H.; Wong, T.C.; Chen, H.H.; Chen, T.H.; Hsu, Y.H.; Peng, S.J.; Kuo, K.L.; Liu, H.C.; Lin, E.T.; et al. Adaptation and Validation of Alternative Healthy Eating Index in Hemodialysis Patients (AHEI-HD) and Its Association with All-Cause Mortality: A Multi-Center Follow-Up Study. *Nutrients* **2019**, *11*, 1407.

56. Wellington, N.; Shanmuganathan, M.; De Souza, R.J.; Zulyniak, M.A.; Azab, S.; Bloomfield, J.; Mell, A.; Ly, R.; Desai, D.; Anand, S.S.; et al. Metabolic Trajectories Following Contrasting Prudent and Western Diets from Food Provisions: Identifying Robust Biomarkers of Short-Term Changes in Habitual Diet. *Nutrients* **2019**, *11*, 2407.

57. Erickson, J.; Slavin, J. Total, Added, and Free Sugars: Are Restrictive Guidelines Science-Based or Achievable? *Nutrients* **2015**, *7*, 2866.

58. Mente, A.; Irvine, E.J.; Honey, R.J.D.A.; Logan, A.G. Urinary Potassium Is a Clinically Useful Test to Detect a Poor Quality Diet. *J. Nutr.* **2009**, *139*, 743–749.

59. McNamara, A.E.; Brennan, L. Potential of Food Intake Biomarkers in Nutrition Research. In Proceedings of the Proceedings of the Nutrition Society; Cambridge University Press, November 1 2020; Vol. 79, pp. 487–497.

60. Peeters, A.; Blake, M.R.C. Socioeconomic Inequalities in Diet Quality: From Identifying the Problem to Implementing Solutions. *Curr. Nutr. Rep.* **2016**, *5*, 150–159.

61. Freedman, N.D.; Park, Y.; Abnet, C.C.; Hollenbeck, A.R.; Sinha, R. Association of Coffee Drinking with Total and Cause-Specific Mortality. *N. Engl. J. Med.* **2012**, *366*, 1891.

62. DeSilver, Drew. Chart of the Week: Coffee and tea around the world. **2013**.

63. Poole, R.; Kennedy, O.J.; Roderick, P.; Fallowfield, J.A.; Hayes, P.C.; Parkes, J. Coffee Consumption and Health: Umbrella Review of Meta-Analyses of Multiple Health Outcomes. *BMJ* **2017**, *359*, 5024.

64. Komes, D.; Bušić, A. Antioxidants in Coffee. *Process. Impact Antioxidants Beverages* **2014**, 25–32.

65. Yashin, A.; Yashin, Y.; Wang, J.Y.; Nemzer, B. Antioxidant and Antiradical Activity of Coffee. *Antioxidants* **2013**, *2*, 230.

66. Jeszka-Skowron, M.; Sentkowska, A.; Pyrzyńska, K.; De Peña, M.P. Chlorogenic Acids, Caffeine Content and Antioxidant Properties of Green Coffee Extracts: Influence of Green Coffee Bean Preparation. *Eur. Food Res. Technol.* **2016**, *242*, 1403–1409.

67. Rothwell, J.A.; Keski-Rahkonen, P.; Robinot, N.; Assi, N.; Casagrande, C.; Jenab, M.; Ferrari, P.; Boutron-Ruault, M.C.; Mahamat-Saleh, Y.; Mancini, F.R.; et al. A Metabolomic Study of Biomarkers of Habitual Coffee Intake in Four European Countries. *Mol. Nutr. Food Res.* **2019**, *63*, 1900659.

68. Madrid-Gambin, F.; Garcia-Aloy, M.; Vázquez-Fresno, R.; Vegas-Lozano, E.; de Villa Jubany, M.C.R.; Misawa, K.; Hase, T.; Shimotoyodome, A.; Andres-Lacueva, C. Impact of Chlorogenic Acids from Coffee on Urine Metabolome in Healthy Human Subjects. *Food Res. Int.* **2016**, *89*, 1064–1070.

69. Liang, N.; Kitts, D.D. Role of Chlorogenic Acids in Controlling Oxidative and Inflammatory Stress Conditions. *Nutr. 2016, Vol. 8, Page 16* **2015**, *8*, 16.

70. Nair, K.P. Fenugreek. *Minor Spices Condiments* **2021**, 79–87.

71. Wani, S.A.; Kumar, P. Fenugreek: A Review on Its Nutraceutical Properties and Utilization in Various Food Products. *J. Saudi Soc. Agric. Sci.* **2018**, *17*, 97–106.

72. Das, S.; Anjeza, C.; Mandal, S. *Synergistic or Additive Antimicrobial Activities of Indian Spice and Herbal Extracts against Pathogenic, Probiotic and Food-Spoiler Micro-Organisms*; 2012; Vol. 19.

73. Yildirim, S.; Demir, E.; Gok, I.; Hassan, •; Aboul-Enein, Y. Use of Electrochemical Methods to Determine the Effect of Brewing Techniques (Espresso, Turkish and Filter Coffee) and Roasting Levels on the Antioxidant Capacity of Coffee Beverage. *J. Food Sci. Technol.* **2022**, 1–11.

74. Górecki, M.; Hallmann, E. The Antioxidant Content of Coffee and Its In Vitro Activity as an Effect of Its Production Method and Roasting and Brewing Time. *Antioxidants* **2020**, *9*, 308.

75. Samtiya, M.; Aluko, R.E.; Dhewa, T.; Moreno-Rojas, J.M. Potential Health Benefits of

Plant Food-Derived Bioactive Components: An Overview. *Foods* 2021, *10*.

76. Liu, R.H. Health Benefits of Fruit and Vegetables Are from Additive and Synergistic Combinations of Phytochemicals. *Am. J. Clin. Nutr.* **2003**, *78*, 517S-520S.

77. Eichholzer, M.; Lüthy, J.; Gutzwiller, F.; Stähelin, H.B. The Role of Folate, Antioxidant Vitamins and Other Constituents in Fruit and Vegetables in the Prevention of Cardiovascular Disease: The Epidemiological Evidence. *Int. J. Vitam. Nutr. Res.* **2001**, *71*, 5–17.

78. Healthy Diet. Available online: https://www.who.int/news-room/fact-sheets/detail/healthy-diet (accessed on 25 May 2022).

79. Miller, V.; Yusuf, S.; Chow, C.K.; Dehghan, M.; Corsi, D.J.; Lock, K.; Popkin, B.; Rangarajan, S.; Khatib, R.; Lear, S.A.; et al. Availability, Affordability, and Consumption of Fruits and Vegetables in 18 Countries across Income Levels: Findings from the Prospective Urban Rural Epidemiology (PURE) Study. *Lancet. Glob. Heal.* **2016**, *4*, e695–e703.

80. Oude Griep, L.M.; Verschuren, W.M.M.; Kromhout, D.; Ocké, M.C.; Geleijnse, J.M. Raw and Processed Fruit and Vegetable Consumption and 10-Year Stroke Incidence in a Population-Based Cohort Study in the Netherlands. *Eur. J. Clin. Nutr.* **2011**, *65*, 791–799.

81. Meinhart, A.D.; Damin, F.M.; Caldeirão, L.; de Jesus Filho, M.; da Silva, L.C.; da Silva Constant, L.; Teixeira Filho, J.; Wagner, R.; Teixeira Godoy, H. Study of New Sources of Six Chlorogenic Acids and Caffeic Acid. *J. Food Compos. Anal.* **2019**, *82*, 103244.

82. Tice, R. Trigonelline [535-83-1] Review of Toxicological Literature, Ph.D.,Research Triangle Park, North Carolina, 1997.

83. Lang, R.; Lang, T.; Bader, M.; Beusch, A.; Schlagbauer, V.; Hofmann, T. High-Throughput Quantitation of Proline Betaine in Foods and Suitability as a Valid Biomarker for Citrus Consumption. *J. Agric. Food Chem.* **2017**, *65*, 1613–1619.

84. Richi, E.B.; Baumer, B.; Conrad, B.; Darioli, R.; Schmid, A.; Keller, U. Health Risks Associated with Meat Consumption: A Review of Epidemiological Studies. *Int. J. Vitam. Nutr. Res* **2015**, *85,* 70-78.

85. Chai, W.; Morimoto, Y.; Cooney, R. V.; Franke, A.A.; Shvetsov, Y.B.; Le Marchand, L.; Haiman, C.A.; Kolonel, L.N.; Goodman, M.T.; Maskarinec, G. Dietary Red and Processed Meat Intake and Markers of Adiposity and Inflammation: The Multiethnic Cohort Study. *J. Am. Coll. Nutr.* **2017**, *36*, 378–385.

86. Dragsted, L.O. Biomarkers of Meat Intake and the Application of Nutrigenomics. *Meat Sci.* **2010**, *84*, 301–307.

87. Cuparencu, C.; Praticó, G.; Hemeryck, L.Y.; Sri Harsha, P.S.C.; Noerman, S.; Rombouts, C.; Xi, M.; Vanhaecke, L.; Hanhineva, K.; Brennan, L.; et al. Biomarkers of Meat and Seafood Intake: An Extensive Literature Review. *Genes Nutr. 2019 141* **2019**, *14*, 1–30.

88. Koeth, R.A.; Wang, Z.; Levison, B.S.; Buffa, J.A.; Org, E.; Sheehy, B.T.; Britt, E.B.; Fu, X.; Wu, Y.; Li, L.; et al. Intestinal Microbiota Metabolism of L-Carnitine, a Nutrient in

Red Meat, Promotes Atherosclerosis. *Nat. Med. 2013 195* **2013**, *19*, 576–585.

89. Wang, Z.; Bergeron, N.; Levison, B.S.; Li, X.S.; Chiu, S.; Xun, J.; Koeth, R.A.; Lin, L.; Wu, Y.; Tang, W.H.W.; et al. Impact of Chronic Dietary Red Meat, White Meat, or Non-Meat Protein on Trimethylamine N-Oxide Metabolism and Renal Excretion in Healthy Men and Women. *Eur. Heart J.* **2019**, *40*, 583–594.

90. Ranabhat, C.L.; Park, M.B.; Kim, C.B. Influence of Alcohol and Red Meat Consumption on Life Expectancy: Results of 164 Countries from 1992 to 2013. *Nutrients* **2020**, *12*, 459.

91. González, N.; Marquès, M.; Nadal, M.; Domingo, J.L. Meat Consumption: Which Are the Current Global Risks? A Review of Recent (2010–2020) Evidences. *Food Res. Int.* **2020**, *137*, 109341.

92. Muli, S.; Goerdten, J.; Oluwagbemigun, K.; Floegel, A.; Schmid, M.; Nöthlings, U. *A* Systematic Review of Metabolomic Biomarkers for the Intake of Sugar-Sweetened and Low-Calorie Sweetened Beverages; *Metabolites*, **2021**, 11, 546.

93. Marques, H.; Cruz-Vicente, P.; Rosado, T.; Barroso, M.; Passarinha, L.A.; Gallardo, E. Recent Developments in the Determination of Biomarkers of Tobacco Smoke Exposure in Biological Specimens: A Review. *Int. J. Environ. Res. Public Health* **2021**, *18*, 1–23.

94. Logue, C.; Roy, L.; Dowey, C.; Verhagen, H.; Strain, J.J.; O'mahony, M.; Kapsokefalou, M.; Athanasatou, A.; Gallagher, A.M. A Novel Urinary Biomarker Approach Reveals Widespread Exposure to Multiple Low-Calorie Sweeteners in Adults. *The Journal of Nutrition* 2020, 150, 2435-2441,

95. Sylvetsky, A.C.; Rother, K.I. Trends in the Consumption of Low-Calorie Sweeteners. *Physiol. Behav.* **2016**, *164*, 446–450.

96. Rauber, F.; Louzada, M.L.D.C.; Martinez Steele, E.; De Rezende, L.F.M.; Millett, C.; Monteiro, C.A.; Levy, R.B. Ultra-Processed Foods and Excessive Free Sugar Intake in the UK: A Nationally Representative Cross-Sectional Study. *BMJ Open* **2019**, *9, e027546*.

97. Yousuf, H.; Hofstra, M.; Tijssen, J.; Leenen, B.; Lindemans, J.W.; van Rossum, A.; Narula, J.; Hofstra, L. Estimated Worldwide Mortality Attributed to Secondhand Tobacco Smoke Exposure, 1990-2016. *JAMA Netw. open* **2020**, *3*, e201177–e201177.

98. Tasevska, N. Urinary Sugars—a Biomarker of Total Sugars Intake. *Nutrients* 2015, *7*, 5816–5833.

99. Narula, N.; Wong, E.C.L.; Dehghan, M.; Mente, A.; Rangarajan, S.; Lanas, F.; Lopez-Jaramillo, P.; Rohatgi, P.; Lakshmi, M.; Varma, R.P.; et al. Association of Ultra-Processed Food Intake with Risk of Inflammatory Bowel Disease: Prospective Cohort Study. *BMJ* **2021**, *374*, 1554.

100. Buckley, J.P.; Kim, H.; Wong, E.; Rebholz, C.M. Ultra-Processed Food Consumption and Exposure to Phthalates and Bisphenols in the US National Health and Nutrition Examination Survey, 2013–2014. *Environ. Int.* **2019**, *131*, 105057.

101. Debras, C.I.; Chazelas, E.I.; Srour, B.I.; Druesne-Pecollo, N.; EsseddikID, Y.; Szabo de EdelenyiID, F.; dric Agaë sse, C.; De Sa, A.; Lutchia, R.; phane GigandetID, S.; et al.

Artificial Sweeteners and Cancer Risk: Results from the NutriNet-Santé Population-Based Cohort Study. *PLOS Med.* **2022**, *19*, e1003950.

102. O'Donnell, M.; Mente, A.; Rangarajan, S.; McQueen, M.J.; O'Leary, N.; Yin, L.; Liu, X.; Swaminathan, S.; Khatib, R.; Rosengren, A.; et al. Joint Association of Urinary Sodium and Potassium Excretion with Cardiovascular Events and Mortality: Prospective Cohort Study. *BMJ* **2019**, *364*, 772.

103. Staruschenko, A. Beneficial Effects of High Potassium: Contribution of Renal Basolateral K+ Channels. *Hypertension* **2018**, *71*, 1015–1022.

104. Rosário, R.; Santos, R.; Lopes, L.; Agostinis-Sobrinho, C.; Moreira, C.; Mota, J.; Póvoas, S.; Oliveira, A.; Padrão, P.; Moreira, P.; et al. Fruit, Vegetable Consumption and Blood Pressure in Healthy Adolescents: A Longitudinal Analysis from the LabMed Study. *Nutr. Metab. Cardiovasc. Dis.* **2018**, *28*, 1075–1080.

105. Dorgan, J.F.; Judd, J.T.; Longcope, C.; Brown, C.; Schatzkin, A.; Clevidence, B.A.; Campbell, W.S.; Nair, P.P.; Franz, C.; Kahle, L.; et al. Effects of Dietary Fat and Fiber on Plasma and Urine Androgens and Estrogens in Men: A Controlled Feeding Study. *Am. J. Clin. Nutr.* **1996**, *64*, 850–855.

106. Kaur, H.; Werstuck, G.H. The Effect of Testosterone on Cardiovascular Disease and Cardiovascular Risk Factors in Men: A Review of Clinical and Preclinical Data. *CJC Open* **2021**, *3*, 1238.

107. Jakše, B.; Jakše, B.; Pajek, M.; Pajek, J. Uric Acid and Plant-Based Nutrition. *Nutr.* **2019**, *11*, 1736.

108. Torralba, K.D.; De Jesus, E.; Rachabattula, S. The Interplay between Diet, Urate Transporters and the Risk for Gout and Hyperuricemia: Current and Future Directions. *Int. J. Rheum. Dis.* **2012**, *15*, 499–506.

109. Dallongeville, J.; Maré, N.; Fruchart, J.-C.; Amouyel, P. Community and International Nutrition Cigarette Smoking Is Associated with Unhealthy Patterns of Nutrient Intake: A Meta-Analysis. *The Journal of Nutrition* **1998**, *128*, 1450-1457.

110. Heydari, G.; Heidari, F.; Yousefifard, M.; Hosseini, M. Smoking and Diet in Healthy Adults: A Cross-Sectional Study in Tehran*, Iranian Journal of Public Health, **2010***; 2014, 43, 485.

111. Treur, J.L.; Taylor, A.E.; Ware, J.J.; Mcmahon, G.; Hottenga, J.J.; Baselmans, B.M.L.; Willemsen, G.; Boomsma, D.I.; Munafò, M.R.; Vink, J.M. Associations between Smoking and Caffeine Consumption in Two European Cohorts. *Addiction* **2016**, *111*, 1059–1068.

112. Benowitz, N.L.; Bernert, J.T.; Foulds, J.; Hecht, S.S.; Jacob, P.; Jarvis, M.J.; Joseph, A.; Oncken, C.; Piper, M.E. Biochemical Verification of Tobacco Use and Abstinence: 2019 Update. *Nicotine Tob. Res.* **2020**, *22*, 1086–1097.

113. Benowitz, N.L.; St Helen, G.; Nardone, N.; Cox, L.S.; Jacob, P. Urine Metabolites for Estimating Daily Intake of Nicotine from Cigarette Smoking. *Nicotine Tob. Res.* **2020**, *22*, 288–292.

114. Cruickshank-Quinn, C.I.; Mahaffey, S.; Justice, M.J.; Hughes, G.; Armstrong, M. Transient and Persistent Metabolomic Changes in Plasma Following Chronic Cigarette Smoke Exposure in a Mouse Model. *PLoS One* **2014**, *9*, 101855.

115. Zhang, L.; Zhang, X.; Ji, H.; Wang, W.; Liu, J.; Wang, F.; Xie, F.; Yu, Y.; Qin, Y.; Wang, X. Metabolic Profiling of Tobacco Leaves at Different Growth Stages or Different Stalk Positions by Gas Chromatography–Mass Spectrometry. *Ind. Crops Prod.* **2018**, *116*, 46–55.

5.7 Supporting Information

Table S5.1 Participant summary characteristics for the PURE pilot study (*n*=1000) reflecting self-reported food records and clinical meta data between individuals classified as consuming a healthy diet (Q5) based on their AHEI score as compared to unhealthy dietary patterns (Q1-Q4).

Variable	Unhealthy Diet (*n*=782)	Healthy Diet (*n*=196)	*p*-value
AHEI (mean)	29.2 ± 7.1	44.7 ± 4.3	**6.43×10^{-136}**
Age (mean)	52.2 ± 9.6	52.9 ± 9.8	0.311
≥ 50y	42.5 ± 4.2	41.8 ± 4.2	0.211
< 50y	58.9 ± 5.7	59.2 ± 5.7	0.588
Sex (%; n)	-	-	0.847
Female	48% (*n*=377)	49% (*n*=96)	-
Male	53 % (*n*=405)	51% (*n*=100)	-
BMI (mean)	26.3 ± 5.8	25.5 ± 5.8	0.106
Lean (<25 kg/m^2)	21.1 ± 2.4	21.0 ± 9.7	0.641
Overweight (25-44 kg/m^2)	29.8 ± 4.6	29.1 ± 5.1	0.128
Current Smokers (%; n)	71% (*n*=552)	52% (*n*=101)	-
Average TNE-7 (μM)	47.9 ± 40	49.1 ± 51	0.797
HICs (%; n)	35% (*n*=279)	43% (*n*=84)	-
MICs (%; n)	41% (*n*=323)	29% (*n*=56)	-
LICs (%; n)	23% (*n*=180)	29% (*n*=56)	-
Current Alcohol Users (%; n)	71% (*n*=387)	65% (*n*=80)	-
Average Fiber (total/day)	20.6 ± 12	35.5 ± 19	**2.67×10^{-40}**
Average Nut Intake (g/day)	5.03 ± 9.8	28.0 ± 37	**5.99×10^{-39}**
Average Raw Vegetable Intake (g/day)	259 ± 193	509 ± 315	**2.04×10^{-38}**
Average Fruit Intake (g/day)	239 ± 254	462 ± 421	**4.05×10^{-22}**

Average Carbohydrates (g/day)	296 ± 121	374 ± 144	**1.14×10^{-14}**
Average Processed Meat intake (g/day)	19.9 ± 24	8.70 ± 10	**1.62×10^{-7}**
Total Sugar (g/day)	57.5 ± 62	73.6 ± 72	**5.20×10^{-5}**
Average Caloric Intake (kcal)	2148 ± 965	2460 ± 970	**5.60×10^{-5}**
Average Red Meat Intake (g/day)	81.6 ± 101	56.8 ± 63	**2.12×10^{-4}**
Average Processed Food (g/day)	223 ± 355	131 ± 191	**1.69×10^{-3}**
Total Lipids (g/day)	64.7 ± 48	65.9 ± 44	0.493
Average Coffee Intake (mL/day)	303 ± 423	318 ± 410	0.620
Dairy (g/day)	268 ± 313	272 ± 296	0.621

Self-reported diet quality according to the alternative healthy eating index (AHEI) was used to categorize individuals as healthy (Quintile 5) vs unhealthy (Quintiles 1-4), where participant characteristics were found within each group. Means are all reported with standard deviation while statistical comparisons assuming equal (t-test) or unequal variance (Welch's t-test) were performed on cubic root transformed data.

Table S5.2 Summary of reliably measured urinary metabolites (*n*=116) detected in PURE participants following nontargeted analysis by MSI-CE-MS, annotated based on accurate mass (*m/z*), relative migration time (RMT), ionization mode (*p* = ESI+, *n* = ESI-), tentative ID, most likely molecular formula, mass error (ppm), technical precision (%CV) from repeat QC measurements (*n*=160), biological variation (%CV) from all urine samples, frequency of detection (FOD), and interclass coefficient (ICC).

***m/z*: RMT: mode**	**Tentative ID**	**Molecular Formula**	**Monoisotopic Mass**	**Mass Error (ppm)**	**Technical Precision (%CV)**	**Biological Variation (%CV)**	**FOD (%)**	**ICC**
62.0601:0.282: p	Ethanolamine	C_2H_7NO	62.0601	9.67	16.0	65.8	95.0	0.941
76.0393:0.543: p	Glycine	$C_2H_5NO_2$	76.0393	6.58	14.7	85.5	89.0	0.971
76.0757:0.309: p	Trimethylamine N-oxide (TMAO)	C_3H_9NO	76.0757	10.1	15.6	80.6	85.2	0.963
90.0550:0.624: p	Alanine	$C_3H_7NO_2$	90.0550	3.33	11.8	73.9	95.8	0.974
104.0706:0.478: p	γ-Aminobutyric acid (GABA)	$C_4H_9NO_2$	104.0706	7.69	22.3	191.4	96.0	0.986
104.1075:0.344: p	Choline	$C_5H_{14}NO$	104.1075	2.88	17.0	76.6	97.0	0.951
106.0499:0.759: p	Serine	$C_3H_7NO_3$	106.0499	1.89	9.90	71.9	98.7	0.981
108.0411:0.396: p	*Unknown*		108.0411	8.81	17.7	70.7	88.5	0.937
114.0606:0.402: p	Carnosine	$C_9H_{14}N_4O_3$	114.0606	9.53	15.0	58.6	99.0	0.935
115.0669:0.4103: p	Creatinine	$C_4H_7N_3O$	115.0669	8.69	15.4	57.9	99.0	0.929
118.0617:0.534: p	Guanidinoacetic acid	$C_3H_7N_3O_2$	118.0617	3.39	15.5	85.0	97.0	0.967
118.0868:0.738: p	Valine	$C_5H_{11}NO_2$	118.0868	1.69	20.3	84.9	58.4	0.943
118.0868:0.918: p	Betaine	$C_5H_{11}NO_2$	118.0868	1.69	12.9	145	69.4	0.992
120.0652:0.823: p	Threonine	$C_4H_9NO_3$	120.0652	0.83	13.5	77.4	94.5	0.970
122.0270:0.911: p	Cysteine	$C_3H_7NO_2S$	122.0270	10.7	46.9	86.4	84.1	0.706
129.0659:0.588: p	Dihydrothymine	$C_5H_8N_2O_2$	129.0659	1.55	14.5	66.6	96.9	0.952
132.0768:0.604: p	Creatine	$C_4H_9N_3O_2$	132.0768	4.54	13.6	190	95.1	0.995
133.0597:0.200: p	Thiamine	$C_{12}H_{17}N_4OS$	133.0597	8.96	18.2	71.8	95.8	0.936
137.0457:1.167: p	Hypoxanthine	$C_5H_4N_4O$	137.0457	5.11	10.3	114	99.7	0.992
137.0709:0.412: p	Methyl Nicotinamide	$C_7H_9N_2O$	137.0709	5.84	33.0	78.1	95.9	0.821

***m/z*: RMT: mode**	**Tentative ID**	**Molecular Formula**	**Monoisotopic Mass**	**Mass Error (ppm)**	**Technical Precision (%CV)**	**Biological Variation (%CV)**	**FOD (%)**	**ICC**
141.0660:0.528: p	Imidazole propionic acid	$C_6H_8N_2O_2$	141.0660	0.71	19.6	90.8	95.9	0.953
144.1019:0.926: p	Proline betaine	$C_7H_{13}NO_2$	144.1019	2.78	14.8	167	94.8	0.992
147.0764:0.862: p	Glutamine	$C_5H_{10}N_2O_3$	147.0764	2.72	8.40	63.9	98.6	0.983
147.1128:0.353: p	Lysine	$C_6H_{14}N_2O_2$	147.1128	0.68	21.2	152	98.3	0.980
150.0583:0.835: p	Methionine	$C_5H_{11}NO_2S$	150.0583	0.67	16.5	80.7	93.6	0.958
156.0768:0.417: p	Histidine	$C_6H_9N_3O_2$	156.0768	7.05	16.7	82.9	97.0	0.959
162.0761:0.883: p	α-Aminoadipate	$C_6H_{11}NO_4$	162.0761	1.23	14.0	66.6	85.6	0.956
162.1125:0.570: p	Carnitine	$C_7H_{15}NO_3$	162.1125	3.70	21.0	114	95.0	0.966
164.0748:0.592: p	S-Propyl-L-Cysteine	$C_6H_{13}NO_2S$	164.0748	1.83	16.7	92.2	84.8	0.967
170.0924:0.439: p	Methyl-histidine	$C_7H_{11}N_3O_2$	170.0924	1.76	19.0	113	96.0	0.972
175.1077:0.386: p	Arginine	$C_6H_{14}N_4O_2$	175.1077	12.6	25.4	72.9	97.2	0.879
176.0658:0.777: p	N-Acetyl Aspartic Acid	$C_6H_9NO_5$	176.0658	8.52	10.1	79.0	95.5	0.984
182.0810:0.933: p	Tyrosine	$C_9H_{11}NO_3$	182.0810	3.30	8.60	72.7	100	0.986
188.0692:0.885: p	*Unknown*	$C_9H_7N_4O$	188.0692	3.72	11.6	72.3	99.7	0.974
189.1598:0.386: p	Trimethyl Lysine	$C_9H_{20}N_2O_2$	189.1598	1.06	14.8	76.8	95.8	0.963
190.1191:0.910: p	Homocitrulline	$C_7H_{15}N_3O_3$	190.1191	1.05	15.7	80.5	86	0.962
191.0661:1.023: p	L-β-Aspartyl-L-Glycine	$C_6H_{10}N_2O_5$	191.0661	6.80	11.1	65.1	97.9	0.971
194.0458:1.11: p	*Unknown*	$C_7H_5N_4O_3$	194.0458	1.03	15.0	77.7	80.3	0.963
195.0764:0.846: p	4-Aminohippuric acid	$C_9H_{10}N_2O_3$	195.0764	7.82	13.7	95.4	94.0	0.979
198.0862:0.742: p	*Unknown*	$C_6H_9N_6O_2$	198.0862	2.02	13.2	72.6	99.6	0.967
203.1503:0.470: p	Symmetric Dimethyl Arginine	$C_8H_{18}N_4O_2$	203.1503	3.94	20.6	58.6	95.5	0.877

***m/z*: RMT: mode**	**Tentative ID**	**Molecular Formula**	**Monoisotopic Mass**	**Mass Error (ppm)**	**Technical Precision (%CV)**	**Biological Variation (%CV)**	**FOD (%)**	**ICC**
203.1503:0.453: p	Asymmetric Dimethyl Arginine	$C_8H_{18}N_4O_2$	203.1503	3.94	22.6	59.2	95.5	0.855
204.1230:0.637: p	Acetyl Carnitine	$C_9H_{17}NO_4$	204.123	8.33	18.3	172	90.2	0.989
205.0972:0.885: p	Tryptophan	$C_{11}H_{12}N_2O_2$	205.0972	2.93	12.9	68.5	96.6	0.964
212.0044:1.197: p	*Unknown*	$C_4H_5NO_9$	212.0040	6.60	12.5	64.9	97.8	0.963
217.1294:0.766: p	Acetyl-Arginine	$C_8H_{16}N_4O_3$	217.1294	7.83	13.5	68.1	97.1	0.961
222.0796:0.738: p	5-δ-carboxyl-butyl-homocysteine	$C_8H_{15}NO_4S$	222.0796	4.05	14.1	74.7	85.6	0.964
226.083:0.715: p	Deoxycytidine	$C_9H_{13}N_3O_4$	226.0830	6.63	12.8	149	99.5	0.993
232.1543:0.692: p	Butyryl-L-Carnitine	$C_{11}H_{19}NO_4$	232.1543	6.89	14.8	75.3	90.6	0.961
235.0928:1.096: p	*Unknown*	$C_{11}H_{19}ClOS$	235.0928	5.53	12.1	66.8	89.6	0.967
238.0916:1.135: p	Xylosyl-Serine	$C_8H_{15}NO_7$	238.0916	4.20	14.2	66.4	95.3	0.954
241.0311:0.895: p	Cystine	$C_6H_{12}N_2O_4S_2$	241.0311	0.41	11.1	84.9	100	0.983
242.0145:1.274: p	*Unknown*	$C_{12}H_4N_2O_2S$	242.0145	2.07	11.9	83.1	95.8	0.980
259.0918:0.812: p	Ribothymidine	$C_{10}H_{14}N_2O_6$	259.0918	3.09	12.8	87.3	100	0.979
268.1040:1.239: p	Deoxyguanosine	$C_{10}H_{13}N_5O_4$	268.1040	3.73	22.3	65.3	92.8	0.884
276.1441:0.795: p	Glutaryl carnitine	$C_{12}H_{21}NO_6$	276.1441	9.51	21.8	75.5	92.8	0.916
282.1197:0.775: p	3-Methyladenosine	$C_{11}H_{15}N_5O_4$	282.1197	9.22	12.8	59.9	98.4	0.954
283.0403:1.360: p	*Unknown*	$C_{15}H_5N_2O_2S$	283.0403	4.59	14.9	71.8	95.5	0.957
288.1181:0.646: p	*Unknown*		288.1181	3.47	14.3	68.3	93.6	0.956
291.1299:0.684: p	L-Argininosuccinate	$C_{10}H_{18}N_4O_6$	291.1299	5.50	15.0	61.2	95.1	0.940
298.1146:1.171: p	1-Methylguanosine	$C_{11}H_{15}N_5O_5$	298.1146	3.35	12.9	62.9	96.8	0.958
302.2325:0.827: p	Nonyl-L-Carnitine	$C_{16}H_{31}NO_4$	302.2325	0.99	14.7	107	92.6	0.981

***m/z*: RMT: mode**	**Tentative ID**	**Molecular Formula**	**Monoisotopic Mass**	**Mass Error (ppm)**	**Technical Precision (%CV)**	**Biological Variation (%CV)**	**FOD (%)**	**ICC**
312.1297:1.129: p	N2,N2-Dimethylguanosine	$C_{12}H_{17}N_5O_5$	312.1297	8.01	9.01	59.8	99.4	0.977
336.1386:1.183: p	*Unknown*	$C_{19}H_{19}N_4S$	336.1386	10.7	13.3	64.8	97.9	0.958
341.0489:1.435: p	*Unknown*		341.0489	7.53	18.8	63.8	82.4	0.913
367.1509:1.140: p	Mannosyl Tryptophan	$C_{17}H_{22}N_2O_7$	367.1509	6.81	11.7	61.3	99.5	0.964
384.1150:1.522: p	*Unknown*	$C_{13}H_{11}N_{12}O_3$	384.1150	7.03	16.3	79.7	92.0	0.958
453.1257:1.140: p	*Unknown*	$C_{28}H_{14}N_5O_2$	453.1257	1.77	13.5	105	84.5	0.983
487.2120:0.735: p	Glucosyl galactosyl hydroxylysine	$C_{18}H_{34}N_2O_{13}$	487.2120	1.85	11.9	63.1	99.5	0.965
627.2071:1.155: p	*Unknown*	$C_{12}H_{40}NO_{18}S$	627.2071	9.89	17.6	62.2	97.7	0.919
89.0244:1.241: n	Lactic acid	$C_3H_6O_3$	89.0244	10.3	20.0	267	25.8	0.996
111.0194:0.340: n	Uracil	$C_4H_4N_2O_2$	111.0194	9.01	19.5	74.7	58.1	0.999
115.0401:1.063: n	α-Ketoisovaleric acid	$C_5H_8O_3$	115.0401	6.94	16.7	58.8	72.3	0.986
128.0353:1.032: n	Oxo-proline	$C_5H_7NO_3$	128.0353	9.37	8.60	63.2	74.6	0.991
132.0302:1.045: n	Aspartic acid	$C_4H_7NO_4$	132.0302	8.33	14.6	65.6	74.8	0.983
135.0299:0.993: n	Threonic acid	$C_4H_8O_5$	135.0299	8.89	8.80	55.8	75.0	0.992
136.0404:0.996: n	p-Aminobenzoic acid	$C_7H_7NO_2$	136.0411	5.15	12.6	56.9	78.2	0.997
144.0299:1.003: n	2-Hydroxyoxoproline	$C_5H_7NO_4$	144.0299	3.47	7.70	57.6	82.8	0.981
151.0401:0.906: n	Hydroxyphenylacetic acid	$C_8H_8O_3$	151.0415	9.27	18.9	94.7	84.2	0.953
160.0615:0.836: n	α-Aminoadipic acid	$C_6H_{11}NO_4$	160.0615	1.25	15.4	73.3	84.4	0.960
161.9869:1.267: n	Acesulfame-K	$C_4H_5NO_4S$	161.9869	1.85	13.5	226	85.0	0.791
167.0211:0.944: n	Uric acid	$C_5H_4N_4O_3$	167.0211	8.98	12.7	47.1	86.3	0.932

m/z: RMT: mode	Tentative ID	Molecular Formula	Monoisotopic Mass	Mass Error (ppm)	Technical Precision (%CV)	Biological Variation (%CV)	FOD (%)	ICC
171.0068:1.527: n	Glycerol phosphate	$C_3H_9O_6P$	171.0068	1.17	17.5	80.7	87.8	0.994
172.9912:1.242: n	Phenyl sulfate	$C_6H_6O_4S$	172.9912	9.25	12.2	133	87.9	0.987
178.0510:0.845: n	Hippuric acid	$C_9H_9NO_3$	178.0510	1.68	9.10	115	88.9	0.929
181.0367:0.841: n	1-Methyluric acid	$C_9H_8O_4$	181.0355	9.55	35.9	118	89.3	0.978
181.9917:1.154: n	Saccharin	$C_7H_5NO_3S$	181.9917	5.49	8.80	331	90.4	0.956
187.0070:1.106: n	p-Cresol sulfate	$C_6H_6N_4O_3$	187.0070	9.09	9.10	104	90.9	0.919
188.9865:1.183: n	Pyrocatechol sulfate	$C_7H_8O_4S$	188.9865	8.89	21.0	89.1	92.7	0.995
190.0510:0.805: n	5-Hydroxy indole acetic acid	$C_{10}H_9NO_3$	190.0510	1.05	19.4	60.4	92.7	0.944
191.0552:0.806: n	Quinic acid	$C_7H_{12}O_6$	191.0552	9.94	13.4	145	92.8	0.967
193.0373:0.781: n	Glucuronic acid	$C_6H_{10}O_7$	193.0373	1.55	9.70	79.7	92.9	0.956
195.0524:0.790: n	1,7-dimethyluric acid	$C_7H_8N_4O_3$	195.0524	3.59	20.3	69.7	93.0	0.907
212.0023:1.043: n	Indoxyl sulfate	$C_8H_7NO_4S$	212.0023	4.72	7.90	83.0	93.7	0.985
218.1034:0.693: n	Pantothenic acid	$C_9H_{17}NO_5$	218.1034	5.04	14.4	105	94.5	0.915
222.9921:0.945: n	*Unknown*	$C_5H_{10}N_3OS_3$	222.9921	2.69	13.0	108	94.6	0.985
225.0631:0.733: n	AAMU	$C_8H_{10}N_4O_4$	225.0631	6.66	18.2	141	95.0	0.535
243.0771:0.344: n	Indolyl acryloyl glycine	$C_{13}H_{12}N_2O_3$	243.0771	7.41	22.5	70.0	95.0	0.939
255.1363:0.653: n	*Unknown*		255.1363	7.73	16.0	109	95.5	0.897
263.1037:0.672: n	Phenylacetylglutamine	$C_{13}H_{16}N_2O_4$	263.1037	7.98	14.9	90.1	95.7	0.983
269.1512:0.634: n	*Unknown*	$C_{14}H_{18}N_6$	269.1512	3.34	13.0	101	95.8	0.897
283.0827:0.641: n	p-Cresol glucuronide	$C_{13}H_{16}O_7$	283.0827	4.95	15.0	130	95.9	0.931
287.0236:0.849: n	Dihydroxyphenyl-g sulfate	$C_{11}H_{12}O_7S$	287.0236	3.48	14.4	150	96.0	0.888

m/z: RMT: mode	Tentative ID	Molecular Formula	Monoisotopic Mass	Mass Error (ppm)	Technical Precision (%CV)	Biological Variation (%CV)	FOD (%)	ICC
290.0882:0.616: n	2,3-Dehydro-2-deoxy-N-acetylneuraminic acid	$C_{11}H_{17}NO_8$	290.0882	1.03	14.9	56.7	96.0	0.973
308.0776:0.605: n	Indoxyl glucuronide	$C_{14}H_{15}NO_7$	308.0776	8.81	14.7	63.9	96.2	0.951
319.0823:0.622: n	Hydroxynaphthalene glucuronide	$C_{16}H_{16}O_7$	319.0823	10.2	16.1	116	96.5	0.947
324.0936:0.621: n	N-glycolylneuraminic acid	$C_{11}H_{19}NO_{10}$	324.0936	7.79	14.8	81.9	96.7	0.987
331.1762:0.578: n	*Unknown*	$C_{16}H_{28}O_7$	331.1759	3.32	19.7	164	96.7	0.994
350.1093:0.599: n	N-Acetyl-acetylneuraminic acid	$C_{13}H_{21}NO_{10}$	350.1093	8.91	13.6	121	96.9	0.981
375.1310:0.225: n	Riboflavin	$C_{17}H_{20}N_4O_6$	375.1310	10.1	23.6	88.1	97.0	0.971
465.2483:0.522: n	Dihydrotestosterone glucuronide	$C_{25}H_{38}O_8$	465.2483	1.93	12.0	93.2	97.4	0.983
473.1453:0.548: n	Enterolactone-3' glucuronide	$C_{24}H_{26}O_{10}$	473.1453	4.65	14.3	260	98.0	0.982
539.2493:0.491: n	Terthydroaldosterone-3-glucuronide	$C_{27}H_{40}O_{11}$	539.2493	6.49	13.9	80.8	98.7	0.982
541.2649:0.484: n	Cortolone glucuronide	$C_{27}H_{42}O_{11}$	541.2649	4.99	19.9	80.4	98.9	0.975
632.2044:0.459: n	Sialyllactose	$C_{23}H_{39}NO_{19}$	632.2044	2.53	14.2	67.4	99.4	0.991

Table S5.3 Food records collected from Food Frequency Questionnaires (FFQ) used for correlation analysis against urine metabolites (*n*=60).

FFQ Food Record	Description
tea_ml	Tea intake in mL/day
coffee_ml	Coffee in mL/day
sfdrink_g	Soft drinks in grams/day
fruitjuice_g	Fruit juice in grams/day
wojfruit_g	Without fruit juice in grams/day
fruits_g	Fruit in grams/day
vegets_g	Vegetables in grams/day
vegraw_g	Raw vegetables in grams/day
otherveg_g	Other vegetables in grams/day
glveg_g	Green leafy vegetables in grams/day
dyveg_g	Dark yellow vegetables in grams/day
crucveg_g	Cruciferous vegetables in grams/day
vegcooked_g	Cooked vegetables in grams/day
Sum_Fruit_Veg	Sum of fruit in and vegetables in grams/day
nutnosoy_g	Nuts without soybean in grams/day
nuts_g	Nuts in grams/day
legumes_g	Legumes in grams/day
fiber_td	Total fibre in grams
fibcereal	Cereal fibre in grams/day
pakbred_g	Packaged bread in grams/day
frshbred_g	Fresh bread in grams/day
carbohydrt	Carbohydrates in grams
starch_g	Breads and cereals in grams/day
whitemeat_g	White meat in grams/day
redmeat2_g	Red meat without processed meat in grams/day
redmeat_g	Red meat in grams/day
wtredr	Ratio of white to red meat
protein	Protein in grams
meats_g	Meats in grams/day
fishonly_g	Fish only in grams/day
fish_g	Fish in grams/day
animalprotein	Animal protein
procmeat_g	Processed meat in grams/day
whtmeatsfa	White meat saturated fatty acids in grams/day
transfat_g	Trans fats in grams/day

FFQ Food Record	Description
fa_sat	Total saturated fatty acids
fa_poly	Total polyunsaturated fatty acids (PUFA)
fa_mono	Total monounsaturated fatty acids (MUFA)
lipid_tot	Total fat in grams
redprocsfa	Red processed meat saturated fatty acids in grams/day
redmeatsfa	Red meat saturated fatty acids in grams/day
psfat	Ratio of polyunsaturated fat to saturated fat
meatnoprocsfa	Not processed meat saturated fatty acids in grams/day
meatsfa	Meat saturated fatty acids in grams/day
fishsfa	Fish saturated fatty acids in grams/day
dairysfa	Dairy saturated fatty acids in grams/day
cholestrl	Cholesterol
animalfat	Animal fat
procfood_g	Processed food in grams/day
dairy_g	Dairy in grams/day
protdary	Protein from dairy in grams/day
alcohol_g	Alcoholic beverages in grams/day
aheiscore	Alternative healthy eating index (AHEI) diet score
vegs	Score for vegetables in AHEI Diet Score (/10)
fruitsc	Score for fruits in AHEI Diet Score (/10)
frys	Score for fried foods in AHEI Diet Score (/10)
nutsc	Score for nuts in AHEI Diet Score (/10)
fibs	Score for cereal fiber in AHEI Diet Score (/10)
wtredrs	Score for ratio of white to red meat in AHEI Diet Score (/10)
psfats	Score for ratio of P:S in AHEI Diet Score (/10)

Table S5.4 Selected candidate biomarkers meeting significance criteria ($p < 0.05$) and correlation thresholds ($r > 0.20$) following partial Pearson correlation analysis for 5 diet categories (Coffee, Fruits and Vegetables, Nuts and Soy, Meats and Fat, Processed Food and Total Sugar).

Candidate Biomarker [a]	Identifier (*m/z*:RMT:mode)	*r* [b] (*n*=)	*p*-value	Food Records [c]
Trigonelline	138.0550:0.775: p $C_7H_7NO_2$	0.230 (*n*=917) 0.258 (*n*=713) **0.408 (*n*=896)**	2.17×10^{-12} 3.34×10^{-12} **8.02×10^{-37}**	Total Saturated FA (g/day) Raw Vegetables (g/day) **Coffee (mL/day)**
Sum (Trigonelline and Quinic acid)	-	0.232 (*n*=862) 0.310 (*n*=737) **0.401 (*n*=924)**	6.97×10^{-12} 1.07×10^{-17} **1.23×10^{-36}**	Fish Saturated FA (g/day) Raw Vegetables (g/day) **Coffee (mL/day)**
Quinic acid	191.0552:0.806: n $C_7H_{12}O_6$	0.286 (*n*=621) **0.323 (*n*=702)**	5.83×10^{-13} **2.78×10^{-18}**	Raw Vegetables (g/day) **Coffee (mL/day)**
Saccharin	181.9917:1.154: n $C_7H_5NO_3S$	**0.308 (*n*=548)**	**2.80×10^{-13}**	**Total sugar (g/day)**
Sum (Trigonelline, Quinic acid, 1-methyluric acid, and 1,7-dimethyluric acid)	-	0.272 (*n*=952) 0.273 (*n*=938) **0.294 (*n*=931)**	2.00×10^{-17} 2.15×10^{-17} **6.84×10^{-20}**	Total Saturated FA (g/day) Animal Fat (g/day) **Coffee (mL/day)**
Acesulfame-K	161.9869:1.267: n $C_4H_5NO_4S$	0.233 (*n*=252) **0.285 (*n*=251)**	2.55×10^{-4} **6.42×10^{-6}**	Soft Drinks (g/day) **Processed Food (g/day)**
3-Methylhistidine	170.0924:0.440: p $C_7H_{11}N_3O_2$	0.261 (*n*=899) 0.274 (*n*=914) **0.284 (*n*=914)**	2.48×10^{-15} 4.25×10^{-17} **2.74×10^{-18}**	Meat Saturated FA (g/day) Total PUFA (g/day) **Total MUFA (g/day)**
Carnitine	162.1125:0.570: p $C_7H_{15}NO_3$	0.227 (*n*=878) 0.247 (*n*=878) **0.272 (*n*=903)**	8.94×10^{-12} 1.57×10^{-13} **1.37×10^{-16}**	Meat saturated FA (g/day) Red meat w/o processed meat (g/day) **Red meat (g/day)**
Sum (Methyl histidine, Acetyl carnitine, and Carnitine)	-	0.252 (*n*=952) 0.263 (*n*=952) **0.265 (*n*=952)**	3.82×10^{-15} 1.98×10^{-16} **1.26×10^{-16}**	Total PUFA (g/day) Red meat (g/day) **Total MUFA (g/day)**
Tetrahydroaldosterone-3-Glucuronide	539.2493:0.491: n $C_{27}H_{40}O_{11}$	**0.203 (*n*=923)** **0.262 (*n*=725)**	**1.03×10^{-9}** **1.17×10^{-12}**	**Sum (Fruits & Veg.) (g/day)** **Raw Vegetables (g/day)**

Candidate Biomarker [a]	Identifier (*m/z*:RMT:mode)	*r* [b] (*n*=)	*p*-value	Food Records [c]
Sum (Carnitine and Acetyl carnitine)	-	0.215 (*n*=925) 0.238 (*n*=925) **0.261 (*n*=950)**	4.86×10^{-11} 2.96×10^{-13} **3.60×10^{-16}**	Red Meat Saturated FA (g/day) Red Meat w/o Processed Meat (g/day) **Red Meat (g/day)**
Proline Betaine	144.1019:0.930: p $C_7H_{13}NO_2$	0.219 (*n*=903) 0.224 (*n*=903) **0.243 (*n*=903)**	3.30×10^{-11} 1.23×10^{-11} **1.81×10^{-13}**	Score for Fruits (AHEI) Fruits (g/day) **Sum (Fruits & Veg.) (g/day)**
Imidazole Propionic acid	141.0660:0.528: p $C_6H_8N_2O_2$	0.221 (*n*=898) 0.225 (*n*=913) **0.242 (*n*=913)**	2.78×10^{-11} 7.52×10^{-12} **1.81×10^{-13}**	Meat Saturated FA (g/day) Total PUFA (g/day) **White Meat (g/day)**
Sum (Acesulfame-k and Saccharin)	-	0.235 (*n*=614) **0.238 (*n*=624)**	3.53×10^{-9} **3.21×10^{-9}**	Processed Food (g/day) **Total Sugar (g/day)**
Hippuric Acid	178.0510:0.845: n $C_9H_9NO_3$	0.225 (*n*=920) 0.233 (*n*=905) **0.236 (*n*=906)**	6.21×10^{-12} 1.57×10^{-12} **8.82×10^{-13}**	Red Meat (g/day) Total Sugar (g/day) **Animal Fat** (g/day)
Sum (Proline betaine and Pantothenic acid)	-	0.205 (*n*=948) 0.212 (*n*=948) **0.234 (*n*=948)**	2.38×10^{-10} 5.08×10^{-11} **3.82×10^{-13}**	Score for Fruits (AHEI) Fruits (g/day) **Sum (Fruits & Vegetables) (g/day)**
Uracil	111.0194:0.340: n $C_4H_4N_2O_2$	0.213 (*n*=732) 0.222 (*n*=756) **0.223 (*n*=827)**	7.71×10^{-9} 9.01×10^{-10} **1.12×10^{-10}**	Nuts w/o Soybean (g/day) Total Nuts (g/day) **Score for Nuts (AHEI)**
Dihydrotestosterone Glucuronide	465.2483:0.522: n $C_{25}H_{38}O_8$	0.212 (*n*=734) 0.212 (*n*=928) **0.213 (*n*=928)**	7.87×10^{-9} 8.19×10^{-11} **7.47×10^{-11}**	Raw Vegetables (g/day) Red and Processed Meat (g/day) **Total MUFA (g/day)**
Betaine	118.0862:0.918: p $C_5H_{11}NO_2$	0.203 (*n*=664) **0.207 (*n*=654)**	1.62×10^{-7} **1.18×10^{-7}**	Total MUFA (g/day) **Processed Food (g/day)**

[a] Top ranked food records from food frequency questionnaires surpassing the initial thresholds ($r>0.2$, $p<0.05$), with subsequent selection according to ranked correlation coefficients, where 1-5 biomarkers per diet category were chosen resulting in 19 total candidate biomarkers. [b] A partial Pearson correlation analysis of urinary metabolites to specific food records following a cubic root transformation with listwise deletion when adjusted for age, sex, BMI, smoking status (current vs never smoker), alcohol intake (current user vs former/never user), total energy (kcal), education (secondary education and above vs primary), and history of comorbidities (hypertension and diabetes) was performed. The strongest correlation coefficient for each metabolite is displayed in bold, with all metabolites ordered from strongest to weakest association to their respective food record. [c] The top three self-reported food records with a significant correlation are reported.

Table S5.5 Top-ranked candidate biomarkers associated with specific food records when using unadjusted Pearson correlation analysis and ANOVA for dose response between high and low self-reported intake.

Biomarker		**Correlation to Food Record** [a]			**Dose Response** [b]			
Tentative ID	**Identifier (*m/z*:RMT:mode)**	**Food Record**	***r***	***p*-value (*n*=)**	***F*-value**	***p*-value (*n*=)**	**High : Low**	**AUC (*p*<0.0001)**
Trigonelline	138.0550:0.775:p $C_7H_7NO_2$	Coffee (mL/day)	0.498	4.22×10^{-59} (*n*=919)	259	1.69×10^{-51} (*n*=919)	3.4	0.80
Saccharin	181.9917:1.154:n $C_7H_5NO_3S$	Total Sugar (g/day)	0.352	3.58×10^{-18} (*n*=563)	25.2	3.36×10^{-11} (*n*=581)	2.8	0.68
Sum (Quinic acid & Trigonelline)	-	Raw Vegetables (g/day)	0.352	1.38×10^{-23} (*n*=748)	103	1.64×10^{-41} (*n*=992)	2.6	0.70
Acesulfame K	161.9869:1.267:n $C_4H_5NO_4S$	Processed Foods (g/day)	0.348	6.52×10^{-9} (*n*=253)	10.9	2.90×10^{-5} (*n*=258)	4.1	0.70
3-Methylhistidine	170.0924:0.440:p $C_7H_{11}N_3O_2$	MUFA (g/day)	0.419	1.51×10^{-41} (*n*=939)	114	5.64×10^{-45} (*n*=960)	3.3	0.83
Carnitine	162.1125:0.570:p $C_7H_{15}NO_3$	Red Meat (g/day)	0.421	1.73×10^{-41} (*n*=928)	39.8	2.51×10^{-17} (*n*=950)	2.4	0.68
Proline Betaine	144.1019:0.930:p $C_7H_{13}NO_2$	Fruits & Vegetables (g/day)	0.342	3.19×10^{-27} (*n*=928)	87.9	5.18×10^{-20} (*n*=929)	2.7	0.68
Uracil	111.0194:0.340:n $C_4H_4N_2O_2$	Nuts & Soy (g/day)	0.292	7.46×10^{-17} (*n*=768)	33.0	1.52×10^{-14} (*n*=863)	1.4	0.66

[a] Food records from self-reported food frequency questionnaires with the strongest association to biomarker response (r>0.2, p<0.05) were selected. A Pearson correlation analysis of urinary metabolites to specific food records following a cubic root transformation was performed with no adjustment for covariates. [b] ANOVA analysis between high and low intake was determined with fold changes reported based on median response (RPA) of the respective biomarker in each group (high vs low), while receiver operating curves (ROC) were used to obtain an area under the curve (AUC) at a significance threshold of p<0.0001 to examine discrimination performance between high and low intake.

Table S5.6 Top-ranked candidate biomarkers and their relationship to overall diet quality based on the Alternative Healthy Eating Index (AHEI), and regional trends between HICs, MICs and LICs when using ANCOVA and partial Pearson correlation analysis.

			Diet Quality [a]					Regional Trends [b]			
Food Record	**Biomarker**	***r***	***p*-value (*n*=)**	***F*-value**	***p*-value (*n*=)**	**Healthy: Unhealthy**	***F*-value**	***p*-value (*n*=)**	**HIC:MIC**	**MIC:LIC**	**HIC:LIC**
Red Meat (g/day)	Carnitine	**-0.161**	1.34×10^{-6} (*n*=903)	16.5	6.54×10^{-30} (*n*=903)	0.71	22.0	7.47×10^{-43} (*n*=903)	1.1	4.6	5.1
Processed Foods (g/day)	Acesulfame K	**-0.161**	1.19×10^{-2} (*n*=252)	1.85	0.0536 (*n*=252)	0.8	1.95	0.0345 (*n*=252)	0.6	17	10
Coffee (mL/day)	Trigonelline	**-0.138**	4.81×10^{-4} (*n*=450)	7.18	9.78×10^{-13} (*n*=450)	1.2	6.88	8.46×10^{-13} (*n*=450)	2.7	1.2	3.4
Total Sugar (g/day)	Saccharin	**-0.095**	2.57×10^{-2} (*n*=555)	4.86	9.71×10^{-7} (*n*=555)	0.9	12.8	1.22×10^{-21} (*n*=555)	0.3	4.6	1.4
Raw Vegetables (g/day)	Sum (Quinic acid & Trigonelline)	**-0.078**	1.91×10^{-2} (*n*=924)	32.6	6.25×10^{-63} (*n*=945)	1.1	36.6	1.78×10^{-74} (*n*=945)	2.7	1.5	4.1
Fruits & Vegetables (g/day)	Proline Betaine	**0.067**	4.50×10^{-2} (*n*=903)	9.48	4.22×10^{-15} (*n*=903)	1.4	14.6	2.94×10^{-26} (*n*=903)	1.9	2.7	5.1
MUFA (g/day)	3-Methylhistidine	-0.052	0.115 (*n*=914)	17.7	1.28×10^{-29} (*n*=914)	1.1	29.2	8.40×10^{-53} (*n*=914)	1.2	3.0	3.6
Nuts & Soy (g/day)	Uracil	0.052	0.137 (*n*=827)	9.34	9.07×10^{-15} (*n*=827)	1.2	19.1	2.11×10^{-34} (*n*=827)	1.5	1.3	1.9

[a] Food records from self-reported food frequency questionnaires with the strongest association to biomarker response ($r>0.2$, $p<0.05$) were selected and compared to diet quality based on the alternative healthy eating index (AHEI). A partial Pearson correlation analysis of urinary metabolites to the alternative healthy eating index score (AHEI) following a cubic root transformation with listwise deletion when adjusted for age, sex, BMI, smoking status(current vs never smoker), alcohol intake (current user vs former/never user), total energy (kcal), education (secondary education and above vs primary), and history of comorbidities (hypertension and diabetes) was preformed. Additional adjustments for raw vegetables, soft drinks and tea were applied to coffee intake, while the raw vegetable correlation underwent additional corrections for fruit and coffee consumption, and the red meat correlation to metabolite response was further corrected for fish intake. All significant correlations are indicated in bold based on a significance threshold of $p<0.05$. A subsequent ANCOVA with the previously described adjustments was preformed between healthy participants determined based on stratified AHEI where healthy (Quintile 5) and unhealthy (Quintile1-4) groups were compared, with fold changes determined based on median biomarker responses (RPA) within each group. [b] ANCOVA analysis between regions was investigated with the above-mentioned adjustments, where fold changes were reported according to median biomarker response (RPA) within each subgroup (HICs, MICs, LICs).

Table S5.7 Candidate biomarkers of biological effect and their and their association with self-reported potassium, and blood pressure to determined physiological response to dietary exposures.

Biomarker		**Potassium** [a]		**Systolic (mm Hg)**		**Diastolic (mm Hg)**		**Blood Pressure Stage** [b]				
Tentative ID	**Identifier (*m/z*:RMT:mode)**	***r***	***p*-value (*n*=)**	***r***	***p*-value (*n*=)**	***r***	***p*-value (*n*=)**	***F*-value**	***p*-value (*n*=)**	**N: 1**	**1: 2**	**N:2**
Tetrahydroaldosterone-3-Glucuronide	539.2493:0.491: n $C_{27}H_{40}O_{11}$	**0.232**	$1.01x10^{-12}$ (*n*=895)	-0.0568	0.051 (n=911)	-0.0714	0.156 (*n*=911)	4.55	$9.80x10^{-7}$ (*n*=923)	1.1	1.1	1.3
Dihydrotestosterone Glucuronide	465.2483:0.522: n $C_{25}H_{38}O_8$	**0.224**	$5.32x10^{-12}$ (*n*=900)	-0.0550	0.050 (n=914)	**-0.112**	$3.26 x10^{-4}$ (*n*=914)	3.85	$1.90x10^{-4}$ (*n*=926)	1.2	1.0	1.2

[a]Partial Pearson correlation of potassium, estimated based on self-reported food records, and biomarker response in urine (RPA), following cubic root transformation with listwise deletion when adjusted for age, sex, BMI, alcohol intake (current user vs former/never user), smoking status (current vs never), total energy (kcal), education (secondary education and above vs primary), and history of comorbidities (hypertension and diabetes) was preformed and repeated for correlation of biomarker response (RPA) with systolic and diastolic blood pressure All significant correlations are displayed in bold (p<0.05). [b] An ANCOVA analysis between blood pressure stages was determined with the previously described corrections, where fold changes are reported based on median response (RPA) of the respective dietary biomarker in each group. N, indicates normal blood pressure (systolic<120 mm Hg, diastolic<80 mm Hg), while 1 represents elevated and stage 1 of hypertension (systolic120-129 mm Hg, diastolic80-89 mm Hg), and 2 indicating stage 2 of hypertension (systolic>140 mm Hg, diastolic>90 mm Hg).

Table S5.8 Top-ranked candidate biomarkers and their association with tobacco exposure, determined by TNE-7, and self-reported smoking status using an ANCOVA and partial Pearson correlation analysis.

Food Record	Biomarker	*r* [a]	*p*-value [a] (*n*=)	*F*-value [b]	*p*-value [b] (*n*=)	Current:Never [b]
Coffee (mL/day)	Trigonelline	**0.321**	3.88×10^{-9} (*n*=307)	7.68	5.43×10^{-13} (*n*=450)	1.2
Raw Vegetables (g/day)	Sum (Quinic acid & Trigonelline)	**0.279**	1.02×10^{-12} (*n*=640)	35.3	2.37×10^{-63} (*n*=945)	1.2
Nuts & Soy (g/day)	Uracil	**0.192**	5.03×10^{-6} (*n*=563)	9.26	1.85×10^{-13} (*n*=827)	1.1
MUFA (g/day)	3-Methylhistidine	**0.176**	1.07×10^{-5} (*n*=627)	19.6	2.82×10^{-30} (*n*=914)	1.2
Processed Foods (g/day)	Acesulfame K	0.117	0.140 (*n*=168)	1.88	0.056 (*n*=252)	1.2
Red Meat (g/day)	Carnitine	**0.099**	0.014 (*n*=619)	17.2	7.44×10^{-29} (*n*=903)	1.4
Total Sugar (g/day)	Saccharin	0.054	0.297 (*n*=382)	5.31	5.78×10^{-7} (*n*=555)	1.3
Fruits & Vegetables (g/day)	Proline Betaine	0.010	0.812 (*n*=624)	10.1	6.26×10^{-15} (*n*=903)	0.9

[a]Top candidate biomarkers demonstrated the strongest correlation to indicated food records and were selected for subsequent analysis against smoking status. A partial Pearson correlation analysis of candidate biomarkers to TNE-7 following a cubic root transformation with listwise deletion when adjusted for age, sex, BMI, alcohol intake (current user vs former/never user), total energy (kcal), education (secondary education and above vs primary), and history of comorbidities (hypertension and diabetes) was preformed. Additional adjustments for raw vegetables, soft drinks and tea were applied to coffee intake, while the raw vegetable correlation underwent additional corrections for fruit and coffee consumption, and the red meat correlation to metabolite response was further corrected for fish intake. All significant correlations are displayed in bold ($p<0.05$). [b] An ANCOVA analysis between current and never smokers was determined with the previously described corrections, where fold changes are reported based on median response (RPA) of the respective dietary biomarker in each group (current vs never).

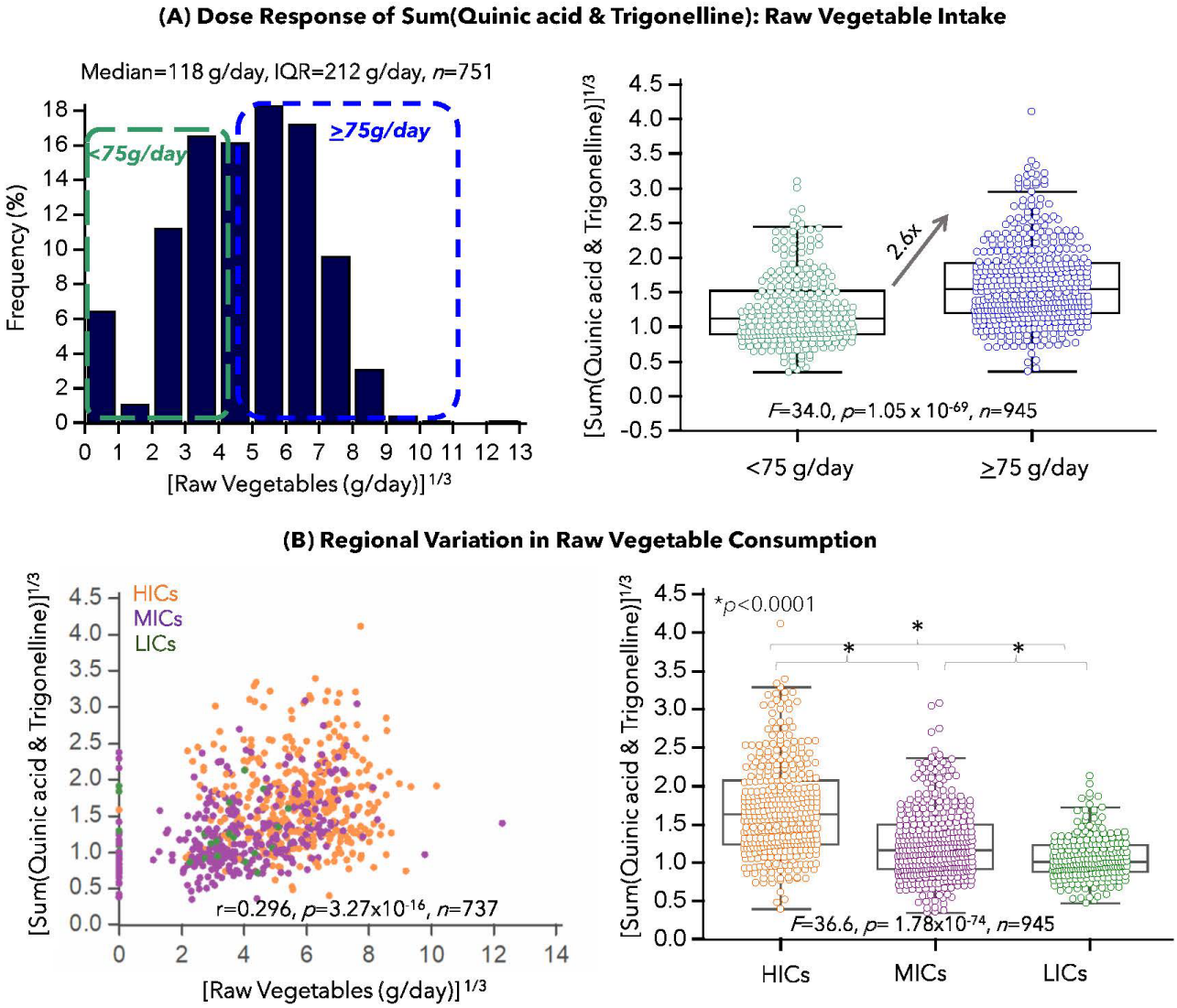


Figure S5.1 Validation of a dietary biomarker for Raw vegetable intake based on the sum of responses (RPA) for quinic acid and trigonelline. (A) Distribution of self-reports and subsequent categorization of low and high intake according to a single serving of raw vegetable intake demonstrates a 2.6-fold increase in biomarker response. (B) Significant regional trends were observed, with correlation of 0.296 amongst all PURE participants, where HICs were found to have elevated raw vegetable consumption relative to MICs and LICs according to detected biomarker levels in single spot urine samples.

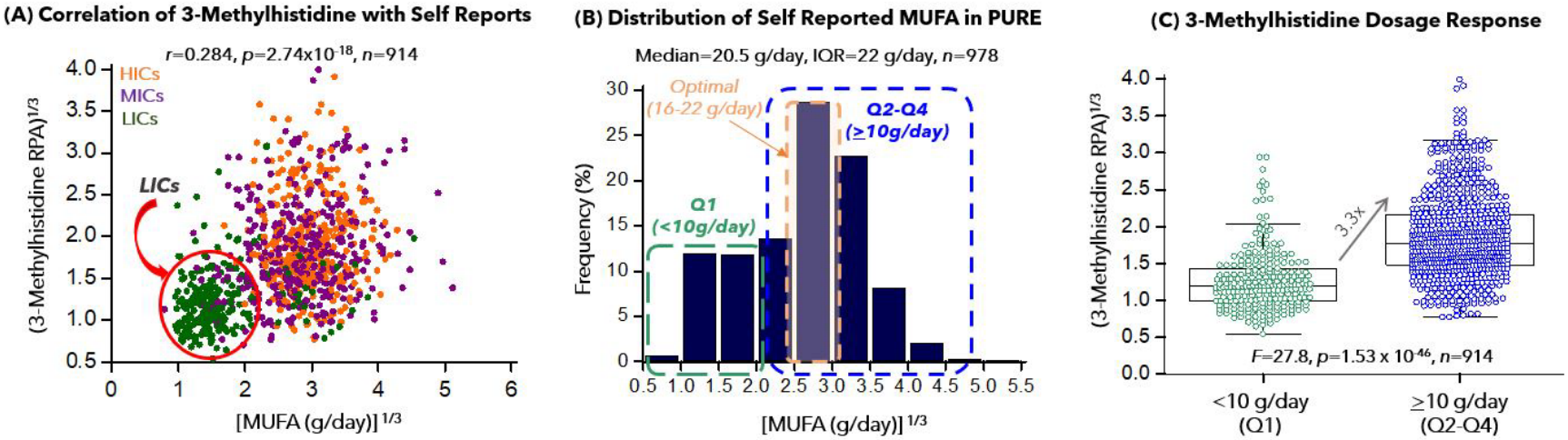


Figure S5.2 (A) Partial Pearson correlation of 3-methylhistidine with self-reported intake of monounsaturated fatty acids (MUFA) was determined, with LICs, showing significantly, lower levels of both biomarker response and self-reports. **(B)** Subsequent distribution of MUFA intake displays an optimal range of consumption recommended for health. (**C)** A comparison of low intake (quartile 1) relative to moderate and high intake (quartiles 2-4) indicates a strong dose response (3.3-fold).

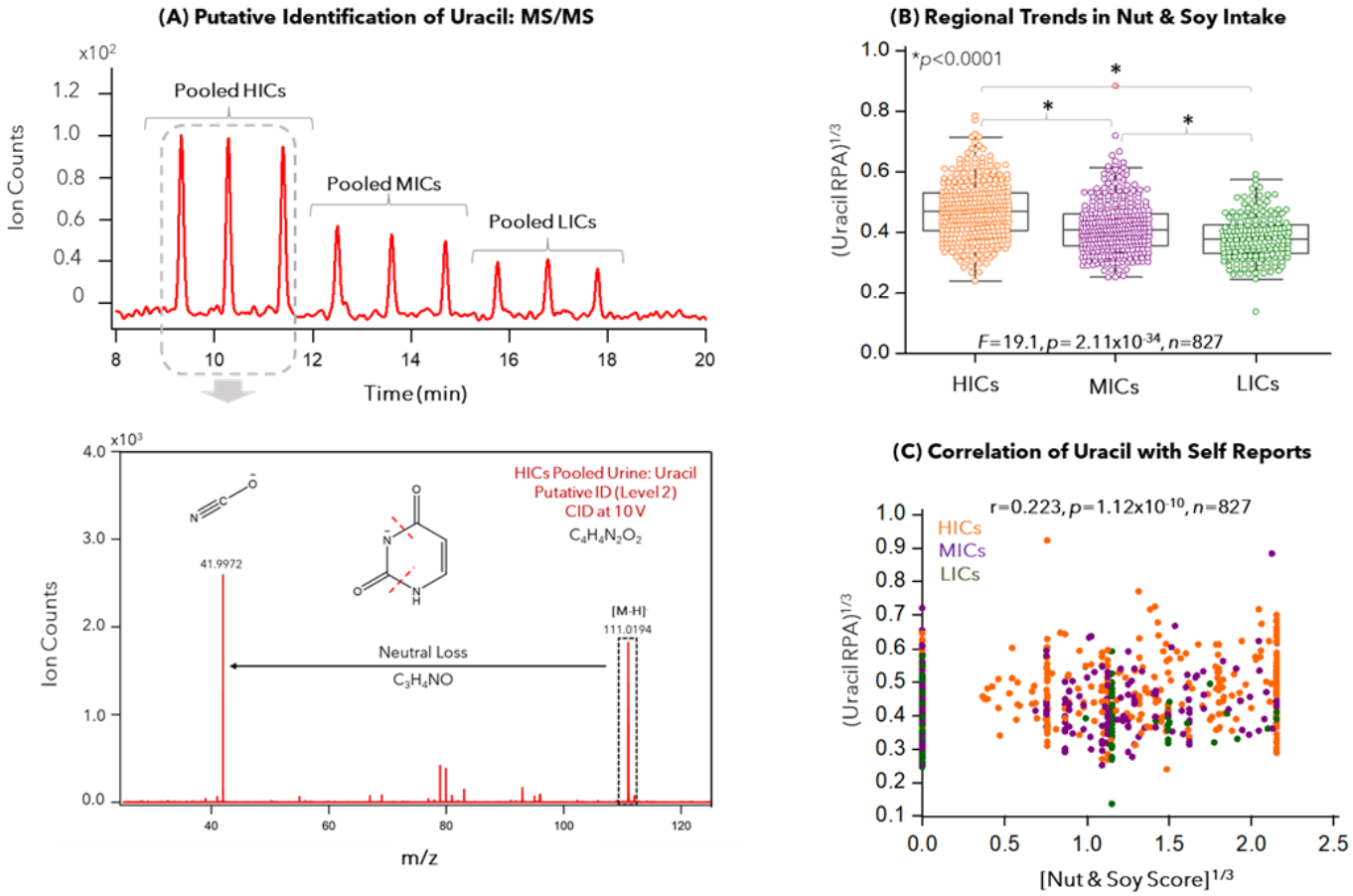


Figure S5.3 Putative identification (Level 2) of an unknown anion detected in urine samples from PURE participants, with a significant increase found in HICs compared to MICs and LICs demonstrates a moderate association with self-reported nut and soy intake based on a partial Pearson correlation analysis (r=0.223) with a characteristic MS/MS spectrum revealing a likely ID of Uracil at a CID of 10 V under negative ion mode conditions.

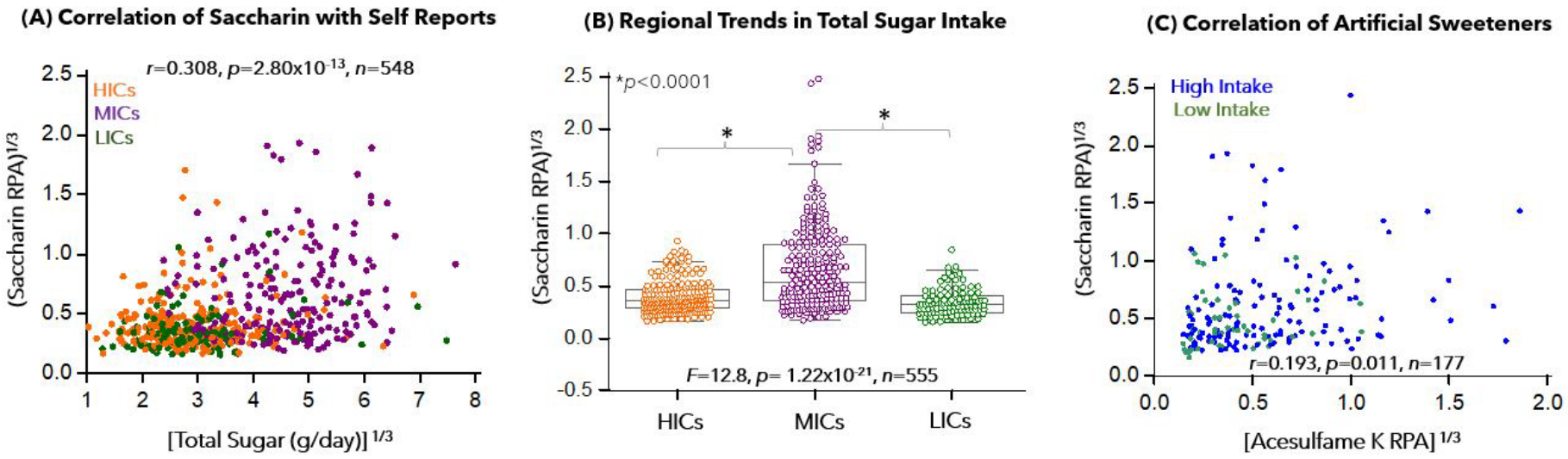


Figure S5.4 (A) Partial Pearson correlation of saccharin response (RPA) from the urine of PURE participants with self-reported intake of total sugar, demonstrates an overall regional pattern of increased total sugar consumption amongst **(B)** MICS as compared to HIC s and LICs. **(C)** A partial Pearson correlation with Acesulfame K (ASK), indicates the likely relationship between artificial sweeteners amongst participants consuming sugar and processed foods.

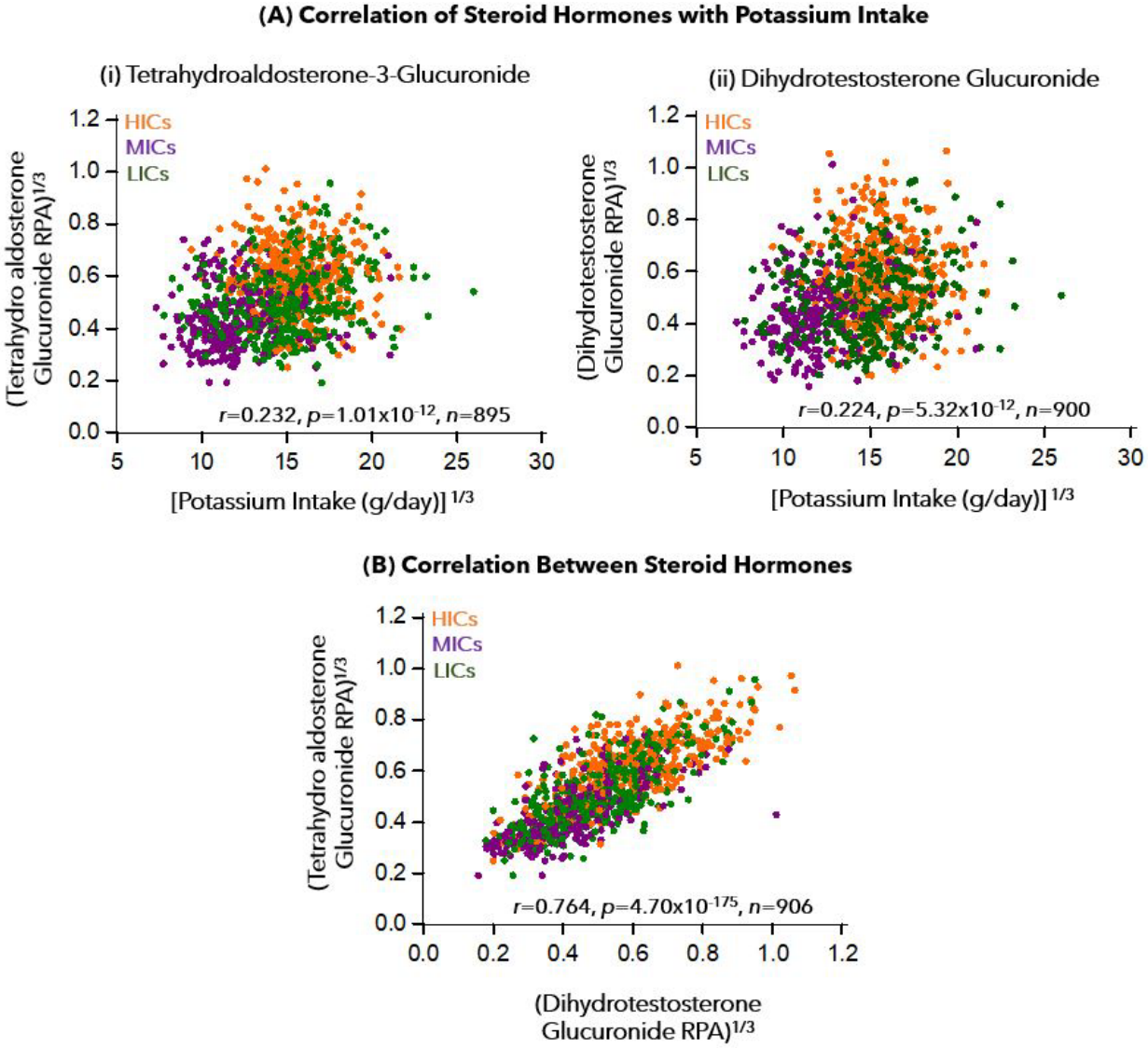


Figure S5.5 (A) Partial Pearson correlation of steroid hormones detected in urine with potassium determined from self-reported food records, demonstrates a significant (p<0.05) association of potassium intake with both (i) tetrahydroaldosterone-3-glucuronide and (ii) dihydrotestosterone glucuronide, where **(B)** both steroid hormones also show a strong correlation with another (r=0.764, p<0.05).

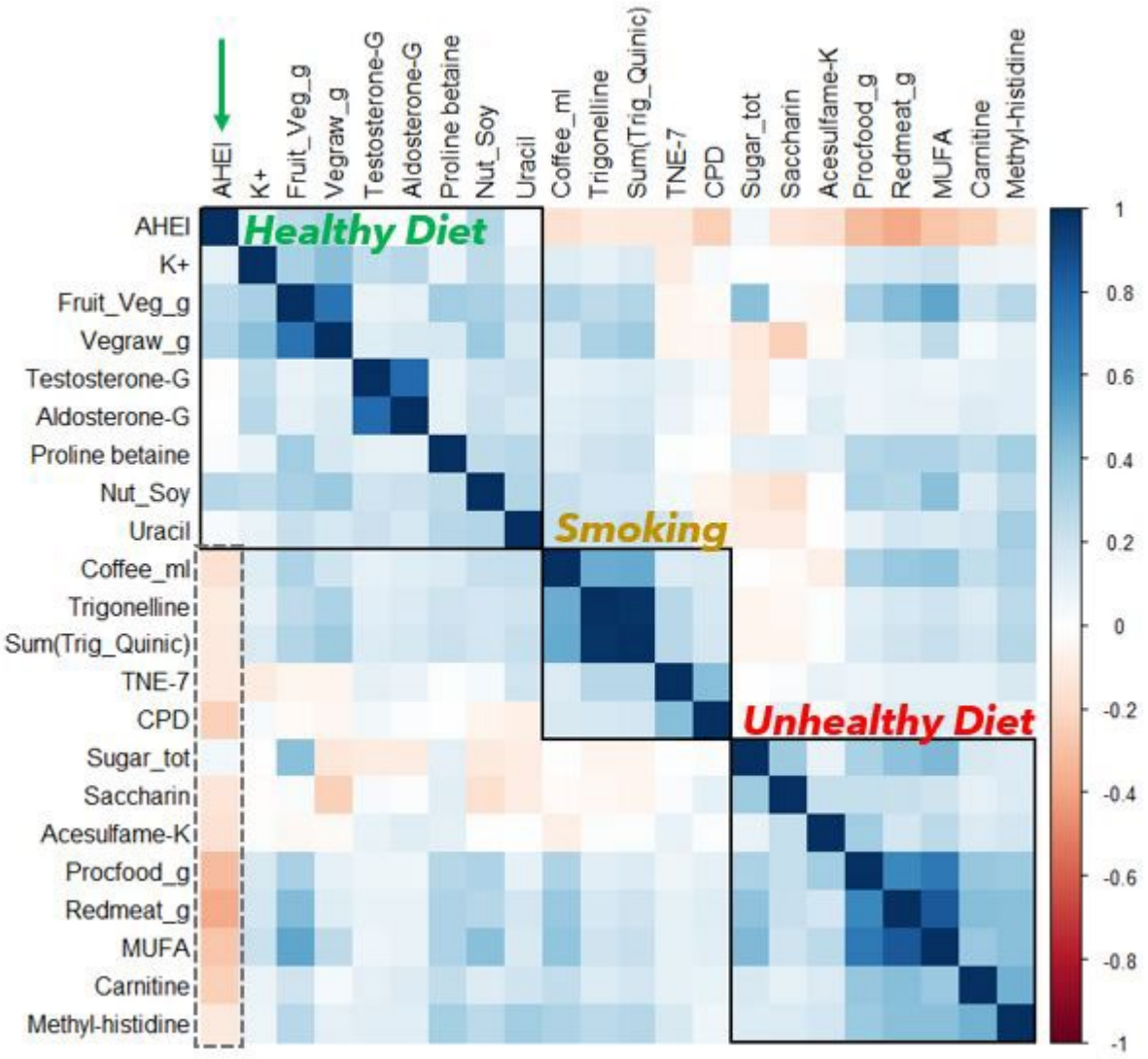


Figure S5.6 A 2D heatmap based on partial Pearson correlation coefficients and Hierarchical Cluster Analysis (HCA), where clusters are outlined in black indicating groupings associated with heathy diet, smoking and unhealthy diet. Features include diet score (AHEI), top ranked self-reported food records, metabolite responses for top eight candidate biomarkers, in addition to the two candidate biomarkers of biological effect, self-reported cigarettes per day and biochemically verified smoking status TNE-7.

Chapter VI:

Exposomics for Population Health and Chronic Disease Risk Assessment

Chapter VI: Exposomics for Population Health and Chronic Disease Risk Assessment

6.1 Overview of Major Thesis Contributions

The work presented in this thesis has contributed to the development of high throughput analytical strategies for characterization of the human exposome in urine samples, including: (1) targeted biomonitoring of environmental smoke exposures in firefighters, (2) comprehensive nicotine metabolite profiling for improved assessment of tobacco smoke exposure and nicotine dependence, and (3) nontargeted workflows for the discovery and validation of robust dietary biomarkers of food intake that may contribute to chronic disease burden. Thus, this thesis lays the foundation for more accurate and reliable risk assessment in occupational and population health, while minimizing reliance on self-reports to estimate food and/or smoke exposures that are prone to bias. Urine metabolic phenotyping offers an important tool in exposomic studies to discover novel biomarkers that can reflect exposure, biological effect, and susceptibility to modifiable risk factors that contribute to mortality and morbidity worldwide.[1] Given the well-recognized impact of environmental smoke and poor diet quality on chronic disease risk,[2,3] this thesis has examined different cohorts for the targeted and nontargeted exploration of the human urine exposome in high-risk professions and diverse global populations.

Chapter I of this thesis describes a comprehensive overview of recent progress in exposure analysis, from traditional human biomonitoring of toxicants to a more recent paradigm shift outlining more comprehensive exposome analyses for more reliable risk assessment. Current omics-based technologies, followed by the major steps and challenges of an untargeted metabolomics data workflow are outlined to demonstrate the importance of rigorous study design coupled with pre-analytical, analytical, and post-analytical considerations in exposure science.

Chapter II demonstrates a rigorous interlaboratory method comparison for the validation of urinary 1-hydroxypyrene (HP), a widely used biomarker for evaluating exposure to polycyclic aromatic hydrocarbons (PAHs) from smoke exposure and other sources (e.g., diet).[4] PAHs are formed from the incomplete combustion of organic materials, and comprise the largest class of cancer-causing chemicals, with several natural (e.g. wood smoke) and anthropogenic (e.g. tobacco smoke) sources of pollution.[5] Occupational exposure to PAHs, some which are carcinogenic upon bioactivation, increases risk for chronic disease events, including several cancers and cardiovascular disease.[6] Specifically, firefighters deployed under emergency conditions often experience heightened exposure with minimal access to protective equipment (e.g. self-contained breathing apparatus) and proper hygiene practices relevant to mitigating dermal uptake of toxicants from environmental smoke.[7] *Chapter II* compared two complementary instrumental protocols for the analysis of urinary HP among a cohort of firefighters deployed at the Fort McMurray wildfire. In this case, analysis of urinary HP was performed following enzyme deconjugation, which allowed for rapid screening without complex sample pretreatment by liquid chromatography-tandem mass spectrometry (LC-MS/MS). Alternatively, a solid-phase extraction (SPE) and chemical derivatization protocol, with subsequent analysis by gas chromatography with high resolution mass spectrometry (GC-HRMS) allowed for a 15-fold lower detection limit. The GC-HRMS procedure enabled simultaneous analysis of several OH-PAHs and their isomers revealing distinct urinary excretion profiles in firefighters following their deployment, highlighting differences in the expression, induction, and inhibition of cytochrome P450 (CYP P450) enzymes responsible for their metabolism (e.g., CYP 1A1, 1A2, 1B1, 3A4). Although both methods exhibited adequate technical and acceptable accuracy, poor mutual agreement was found as reflected by a bias of

39%. Further studies confirmed that this bias was largely attributed to incomplete enzymatic hydrolysis (i.e., deglucuronidation) prior to LC-MS/MS due to suboptimal incubation protocols recommended by manufacturer (2 h at 60°C versus 18 h at 37°C as used in GC-HRMS). This work highlights the importance of inter-laboratory standardization especially for pre-analytical sample workup procedures to minimize bias and false discoveries following rigorous method validation.

Chapter III of this thesis introduces a simple yet high throughput protocol based on multisegment injection-capillary electrophoresis-tandem mass spectrometry (MSI-CE-MS/MS) following a simple acidified ether extraction, to directly analyze the glucuronide conjugation of HP (HP-G) in urine samples collected from firefighters prone to smoke exposure.[8] As discussed in *Chapter II,* current methods often rely on complex, time-consuming sample preparation, enzyme deconjugation and chemical derivatization steps which may contribute to greater variability and bias resulting in potential misreporting of "true exposures". For the first time, we demonstrated a multiplexed separation method for the quantification of urinary HP-G as a convenient surrogate for total HP in occupationally exposed individuals. Analysis of up to 13 urine extracts within a single run (<3 min/sample) was achieved with stringent quality control (QC), which offers a cost-effective and rapid screening method for risk assessment of PAH exposure. Rigorous validation demonstrated excellent reproducibility and recovery, with good mutual agreement when compared to the previously reported GC-HRMS protocol for total urinary HP determination following quantitative enzyme deconjugation. Overall, firefighters were determined to be below the recommended biological exposure index (BEI) for health hazards according to creatinine normalized HP-G concentrations in urine. This unexpected result was mainly attributed to the challenges associated with delayed and variable urine collection at

early stages of firefighting under emergency conditions. However, this work offers a rapid and reliable measure of recent environmental smoke exposure with potential applications for large-scale population-based studies.

Chapter IV transitions to verifying tobacco smoke exposure in the Prospective Urban and Rural Epidemiology (PURE) study to provide insight on the hazards associated with smoking relative to never smoking.[9] Unlike less specific biomarkers of smoke exposure (e.g., PAHs), this work focused on comprehensive analysis of tobacco-specific nicotine metabolites excreted in urine as a robust measure of recent tobacco smoke exposure. Importantly, we directly analyzed the total nicotine equivalent (TNE-7) in urine samples after a simple dilution step by MSI-CE-MS, which is defined as the sum of up to seven major nicotine metabolites.[10] In addition, the nicotine metabolite ratio (NMR) offers a complementary phenotypic biomarker for enzyme activity to better ascertain smoking behaviors and nicotine dependence.[11] Current methods for the analysis of nicotine and its metabolites in urine typically rely on two-step enzyme deconjugation protocols, which are prone to bias, have limited sample throughput and suffer from high operational costs, and thus are not amenable to large-scale epidemiological studies.[12] MSI-CE-MS was validated for the rapid analysis of a comprehensive panel of urinary nicotine metabolites and their intact conjugates, which demonstrated excellent stability along with acceptable technical precision and accuracy. Importantly, this approach was applied for robust biochemical verification of smoking status and smoking behaviors in an international cohort of participants from 14 different countries (n=1000). Moreover, urinary TNE-7 exhibited a moderate correlation with self-reported smoking intensity (cigarettes per day, CPD), with the strongest association occurring in high-income countries (HICs), followed by middle-income countries (MICs), while no relationship to self-reports was found in low-income countries

(LICs), likely indicative of increased misreporting. Importantly, this work also discovered elevated tobacco smoke exposure among heavy smokers from HICs relative to MICs and LICs after covariate adjustments, including CPD.[9] Additionally, with the subsequent evaluation of urinary NMR, 'fast' metabolizers of nicotine were identified in the PURE study and compared against 'slow' metabolizers. In this case, fast metabolizers with high NMR scores were found to have greater tobacco exposure based on TNE-7 following adjustments for CPD and other covariates. This likely indicates that 'fast' metabolizers are prone to increased puff volumes and puff frequency associated with compensatory smoking behavior when using ventilated and low yield cigarettes prevalent in HICs. As a result, urinary TNE-7 and NMR provide new insights when assessing the exposure to and the potential harm from tobacco smoking as compared to self-reports that are prone to gender and social desirability bias. Furthermore, this work is urgently required for improved risk assessment of tobacco smokers more susceptible to nicotine dependence, greater toxicant exposure, and more likely to suffer from tobacco related chronic disease burden and clinical events.

Lastly, *Chapter V* describes a nontargeted data workflow for characterization of the urine metabolome by MSI-CE-MS, for the identification of objective dietary biomarkers in the PURE cohort. Given the important roles that diet plays as a modifiable risk factor of chronic disease, robust urinary biomarkers indicative of recent food intake, which are also generalizable across diverse populations are critical.[13] As such, a dietary biomarker should ideally be specific, not extensively metabolized, and demonstrate dose dependency, while also remaining applicable to multiethnic cohorts with variable cooking methods and food exposures.[14] In this work, semi-quantitative estimates of food intake using a standardized food frequency questionnaire in the PURE study were compared against 130 authenticated urinary metabolites to identify candidate

biomarkers of habitual diet, which were generalizable across countries of varying income status. Overall, eight top-ranked candidates were identified and underwent further validation for their dose-response. Urinary trigonelline revealed the strongest association with self-reported coffee intake, which is not currently incorporated in diet quality scoring systems, such as alternative healthy eating index (AHEI), yet has been reported to provide protective health benefits.[15,16] Additional urinary metabolites including carnitine, 3-methylhistidine, saccharin, acesulfame K and proline betaine were also determined to be moderately correlated with self-reports of intake for meats and fat, total sugar and processed food, as well as fruits and vegetables respectively. This panel of urinary metabolites also exhibited significant fold-changes between high versus low consumption with moderate classification performance based on recommended healthy diet guideline thresholds for specific foods (when available). Moreover, exogenous metabolites measured in urine, such as saccharin and acesulfame K, revealed the strongest specificity for total sugar, soft drinks, and processed foods, with distinct regional trends highlighting their disproportionate greater consumption in MICs, such as Argentina. Moreover, a moderate relationship of individual dietary biomarkers with overall diet quality was noted, where biomarkers of healthy foods, such as fruits and vegetables (e.g., proline betaine) were elevated amongst participants with higher diet quality scores, while indicators of unhealthy foods, such as red meat (e.g., carnitine) were reduced. Importantly, urinary metabolites associated with suboptimal diet patterns revealed a significant relationship to smoking, where TNE-7 confirmed current smokers relative to never smokers had increased levels of metabolites related to red meat, processed foods, and total sugar. Notably, coffee intake, which may have an impact on nicotine metabolism and smoking behavior exhibited a considerable positive correlation with recent tobacco smoke exposure (i.e., TNE-7). Urinary dietary biomarkers of food intake identified in

this work can complement self-reports to better capture lifestyle and drinking habits (e.g., coffee consumption) not currently monitored by diet quality scores, for a more reliable assessment of food exposures, and to better elucidate complex relationships between dietary habits and chronic disease risk across different countries.

6.2 Future Advancements for Biomonitoring Occupational Smoke Exposure in Firefighters

While both *Chapter II* and *III* present a rigorous interlaboratory method validation for urinary HP and HP-G determination in a cohort of firefighters respectively, a key study limitation was the delay in urine collection under emergency conditions, which resulted in an underreporting of smoke exposures. Consequently, future studies that enable simplistic sampling/storage methods such as spot urine specimens (e.g., filter paper cards), or wearable devices including silicone wristbands and non-stimulated sweat collection systems (e.g., Macroduct), can provide alternative approaches towards more accurate and real-time biomonitoring for occupational exposure (e.g., firefighters).[17,18] For instance, a recent study utilized silicone wristbands to monitor on and off-duty firefighters for exposure to 134 chemicals including PAHs, flame retardants and several plasticizers (e.g., phthalates).[19] This also highlights the constraints of personalized protective equipment, where due to a lack of standardized hygiene practices, coupled with extreme heat stress when using bunker gear, dermal uptake of numerous harmful contaminants is often exacerbated.[7,17] Thus, standardized hygiene protocols and proper cleaning techniques are urgently needed to mitigate the harmful impact of smoke exposure from dermal uptake. In addition, *Chapter II* demonstrates limitations in current analytical protocols, where a lack of consistency in enzyme hydrolysis procedures can introduce bias and misreporting of "true exposures". Accordingly, future studies using simple, cost-effective technologies, such as those developed in *Chapter III* are recommended to eliminate the need for enzyme deconjugation

altogether. However, simple strategies including further optimization of multiple reaction monitoring (MRM) transitions, and the utilization of nanospray designs to improve ionization efficiency and lower detection limits may allow for more comprehensive analysis of OH-PAH glucuronides when using MSI-CE-MS/MS as introduced in *Chapter III*.

Overall, current biomonitoring studies in occupational health have largely focused on small cohorts of male participants, where variables such as tobacco smoke, alcohol consumption, medication use and diet are either not routinely monitored, or these important modifiable risk factors for chronic disease are estimated via self-reported questionnaires.[17] As shown in *Chapter II,* different OH-PAH isomers and species may be not colinear as expected, likely due to differences in CYP P450 isoform specificity, often influenced by diet, lifestyle and genetic factors. Consequently, future work using nontargeted data workflows with stringent quality control, similar to those applied in *Chapter V,* may offer a more comprehensive understanding of the relationship between smoke exposures and chronic disease risk. Expanded metabolome coverage will also allow for the quantification of biomarkers of smoke exposure and biological effect (e.g., inflammation), while also aiding in the discovery of unknown metabolites of clinical significance. If further combined with prospective studies, future studies can better evaluate and identify modifiable risk factors that reduce clinical events among high-risk firefighters.

6.3 Future Directions in Determining Chronic Disease Risk among Current Smokers

Chapter IV of this thesis outlined a robust analysis of a comprehensive panel of nicotine metabolites in urine (i.e., TNE-7), where their total sum is less impacted by interindividual differences in metabolism and therefore provides a more reliable measure of recent tobacco smoke exposure than single nicotine species alone (e.g., cotinine).[20] Accordingly, this work

enabled the reliable determination of smoking status for participants from HICs, MICs and LICs. However, given the higher detection limits, this study was unable to characterize individuals subject to self-reported passive/second-hand smoke exposure (SHS). Importantly, majority of smoke-free polices to mitigate SHS have been passed in HICs, while 89% of deaths due to SHS exposure occur in LICs and MICs.[21] With these current disparities in public health policy worldwide, further improvements in sensitivity, either through the use multiple reaction monitoring by MS/MS, nanospray interfaces for CE-MS to enhance ionization efficiency, or on-line sample preconcentration protocols in MSI-CE-MS, would provide deeper insights into the extent of low background levels of SHS exposure among never smokers in HICs, MICs and LICs. Improved sensitivity could also expand the analysis of lower abundance nicotine alkaloids (e.g., anatabine) which can help distinguish between types of tobacco products (e.g. smokeless tobacco, cigarettes) and nicotine replacement therapies.[22] In a recent study by Lorkiewicz *et al*.[22] anatabine was not detected in e-cigarettes and may be a useful biomarker for identifying combustible cigarettes from e-cigarettes. Also, the advent of 'heat-not-burn' devices promoted as a safer non-combustible option for tobacco use than cigarettes or e-cigarettes may represent a gateway to increase cigarette smoking in the population rather than a replacement.[23] In addition to public health policies related to reducing active smoking and SHS exposures, air quality also varies drastically across countries especially in urban environments.[24] As described in *Chapters I-IV*, environmental smoke exposure from air pollution is another major source of harmful chemicals (e.g. PAHs), and may have additive effects when compounded with smoking, resulting in greater disease burden among current smokers in areas with poor air quality. Consequently, future studies should aim to determine the contribution of environmental smoke exposure to better understand the lifestyle and local environmental factors contributing to disease risk. PAHs

present in both environmental and tobacco smoke can increase CYP 2A6 activity and thus impact smoking habits, such as increasing nicotine dependence. Therefore, applications of methods such as MSI-CE-MS/MS from *Chapter III*, with advancements for expanded OH-PAH coverage can be applied to monitor PAH metabolism and its impact on smoking habits (e.g., NMR). Moreover, the increased presence of PAHs can also help decipher between combustible and smokeless cigarettes in conjunction with other toxic classes of chemicals in smoke vapor, such a 2-ethylpyridine and pyrazine.[25]

Chapter IV monitored nicotine metabolism based on phenotypic biomarkers determined from the ratios of specific nicotine metabolites as substrates and products of enzyme activity (e.g., NMR). While ratios such as NMR have proven to be useful tools in clinical trials for guiding personalized smoking cessation interventions,[26] they are unable to differentiate between genetic and environmental (e.g., diet, alcohol, medication use) factors which may contribute to differences in metabolism and future clinical events.[11] Future studies which incorporate genetic data for the calculation of a polygenic risk score based on proteins involved in nicotine metabolism and nicotinic receptors can help decipher genetic risk factors from modifiable lifestyle factors which may be useful in mitigating tobacco related deaths.[27] Analysis of an additional 2000 urine samples is currently underway, with genetic and clinical follow-up data available over an eight year period from the PURE study. The incorporation of prospective data and additional participants will help to improve study power and validate urinary biomarkers of exposure and/or harm as prognostic indicators of tobacco-related disease risk that are associated with specific incident clinical events (e.g., lung cancer, stroke, myocardial infarction, COPD).

As described in *Chapter V*, a panel of urinary dietary biomarkers were found to be associated with self-reported intake of specific food categories as well as smoking status, while

also exhibiting extensive regional variation between HICs, MICs, and LICs. Follow-up analysis for the quantification of validated dietary biomarkers of food intake will establish regional reference ranges for urinary metabolites of significance which can better reflect dietary habits, nutritional uptake, and cooking practices, specific to country income levels. However, to properly address the contribution of diet on smoking, future controlled feeding studies are required to elucidate the underlying mechanisms contributing to these associations. While unhealthy fat intake is one of the largest contributors to poor diet quality (e.g., omega-6 fatty acids from vegetable seed oils and *trans*-fats from hydrogenated fats), consumption of healthy fats in moderation (e.g., omega-3 fatty acids from fish oil) can provide health benefits. For instance, omega-3 fatty acid supplementation from fish oil and pharmaceutical grade synthetic analogs of eicosapentaenoic acid (EPA) have been explored for prevention and treatment of CVD.[28] In *Chapter V*, self-reported monounsaturated fatty acids (MUFA) were associated with urinary 3-methylhistidine excretion, however this is likely a consequence of meat intake and thus lacks specificity. Consequently, the incorporation of alternative dietary biomarkers specific to fat intake can complement those found in *Chapter V* of this thesis. By incorporating this study with paired serum samples, which are currently available for future analysis, determination of non-esterified fatty acids (NEFAs) by multisegment injection-nonaqueous capillary electrophoresis-mass spectrometry (MSI-NACE-MS) will allow for expanded lipid analysis, such as dietary NEFA biomarkers for dairy fat (e.g., C15:0) and fish oil (e.g., EPA).[29,30] Thus, a more diverse range of food specific dietary biomarkers from both urine and serum samples can be determined to comprehensively evaluate diet quality and potential chronic disease risk in diverse populations.

6.4 General Conclusions

In summary, this thesis outlines the development and validation of novel multiplexed MSI-CE-MS techniques for targeted and nontargeted biomonitoring of smoke exposure in occupational and population health. Rigorous interlaboratory method validation of previously reported methods revealed limitations in current analytical protocols, while novel, direct analysis technologies mitigated the need for sample workup procedures as required for large-scale epidemiological studies. Additionally, this thesis also contributed to innovative approaches for reliable assessment of tobacco smoke exposure, smoking behavior, and food exposures as compared to self-reports. Metabolite coverage was expanded through nontargeted analysis to identify robust and generalizable dietary biomarkers, which may influence smoking habits and overall chronic disease risk. Overall, this thesis contributed targeted and nontargeted strategies for characterization of the urinary exposome, which sets the stage for more comprehensive and reliable risk assessment to guide personalized interventions for chronic disease prevention, including effective smoking cessation and optimal diets that are health promoting.

6.5 References

(1) Maria Vitale, C.; Price, E. J.; Miller, G. W.; David, A.; Antignac, J.-P.; Barouki, R.; Klá nová, J. Analytical Strategies for Chemical Exposomics: Exploring Limits and Feasibility. *Exposome* **2021**, *1* (1).

(2) Tucker, K. L. The Role of Diet in Chronic Disease. *Present Knowl. Nutr.* **2020**, 2, 329–345.

(3) Ng, R.; Sutradhar, R.; Yao, Z.; Wodchis, W. P.; Rosella, L. C. Smoking, Drinking, Diet and Physical Activity—Modifiable Lifestyle Risk Factors and Their Associations with Age to First Chronic Disease. *Int. J. Epidemiol.* **2020**, *49* (1), 113–130.

(4) Gill, B.; Mell, A.; Shanmuganathan, M.; Jobst, K.; Zhang, X.; Kinniburgh, D.; Cherry, N.; Britz-McKibbin, P. Urinary Hydroxypyrene Determination for Biomonitoring of Firefighters Deployed at the Fort McMurray Wildfire: An Inter-Laboratory Method Comparison. *Anal. Bioanal. Chem.* **2019**, *411* (7), 1397–1407.

(5) Stec, A. A.; Dickens, K. E.; Salden, M.; Hewitt, F. E.; Watts, D. P.; Houldsworth, P. E.; Martin, F. L. Occupational Exposure to Polycyclic Aromatic Hydrocarbons and Elevated Cancer Incidence in Firefighters. *Sci. Rep.* **2018**, 8 (1), 1-8.

(6) Manisalidis, I.; Stavropoulou, E.; Stavropoulos, A.; Bezirtzoglou, E. Environmental and Health Impacts of Air Pollution: A Review. *Front. Public Heal.* **2020**, *8*, 14.

(7) Cherry, N.; Aklilu, Y. A.; Beach, J.; Britz-Mckibbin, P.; Elbourne, R.; Galarneau, J. M.; Gill, B.; Kinniburgh, D.; Zhang, X. Urinary 1-Hydroxypyrene and Skin Contamination in Firefighters Deployed to the Fort McMurray Fire. *Ann. Work Expo. Heal.* **2019**, 63 (4), 448-458.

(8) Gill, B.; Jobst, K.; Britz-Mckibbin, P. Rapid Screening of Urinary 1-Hydroxypyrene Glucuronide by Multisegment Injection-Capillary Electrophoresis-Tandem Mass Spectrometry: A High-Throughput Method for Biomonitoring of Recent Smoke Exposures. *Anal. Chem.* **2020**, *92* (19), 13558–13564.

(9) Sathish, T.; Teo, K. K.; Britz-McKibbin, P.; Gill, B.; Islam, S.; Paré, G.; Rangarajan, S.; Duong, M. L.; Lanas, F.; Lopez-Jaramillo, P.; Mony, P. K.; Pinnaka, L.; Kutty, V. R.; Orlandini, A.; Avezum, A.; Wielgosz, A.; Poirier, P.; Alhabib, K. F.; Temizhan, A.; Chifamba, J.; Yeates, K.; Kruger, I. M.; Khatib, R.; Yusuf, R.; Rosengren, A.; Zatonska, K.; Iqbal, R.; Lui, W.; Lang, X.; Li, S.; Hu, B.; Dans, A. L.; Yusufali, A. H.; Bahonar, A.; O'Donnell, M. J.; McKee, M.; Yusuf, S. Variations in Risks from Smoking between High-Income, Middle-Income, and Low-Income Countries: An Analysis of Data from 179 000 Participants from 63 Countries. *Lancet Glob. Heal.* **2022**, *10* (2), e216–e226.

(10) Benowitz, N. L.; St Helen, G.; Nardone, N.; Cox, L. S.; Jacob, P. Urine Metabolites for Estimating Daily Intake of Nicotine from Cigarette Smoking. *Nicotine Tob. Res.* **2020**, *22* (2), 288–292.

(11) Allenby, C. E.; Boylan, K. A.; Lerman, C.; Falcone, M. Precision Medicine for Tobacco Dependence: Development and Validation of the Nicotine Metabolite Ratio. *Journal of Neuroimmune Pharmacology*. **2016**, 11 (3), 471–483.

(12) Taghavi, T.; Novalen, M.; Lerman, C.; George, T. P.; Tyndale, R. F. A Comparison of Direct and Indirect Analytical Approaches to Measuring Total Nicotine Equivalents in Urine. *Cancer Epidemiol. Biomarkers Prev.* **2018**, *27* (8), 882–891.

(13) Rafiq, T.; Azab, S. M.; Teo, K. K.; Thabane, L.; Anand, S. S.; Morrison, K. M.; De Souza, R. J.; Britz-Mckibbin, P. Nutritional Metabolomics and the Classification of Dietary Biomarker Candidates: A Critical Review. *Adv. Nutr.* **2021**, *12* (6), 2333–2357.

(14) de Souza, R. J.; Shanmuganathan, M.; Lamri, A.; Atkinson, S. A.; Becker, A.; Desai, D.; Gupta, M.; Mandhane, P. J.; Moraes, T. J.; Morrison, K. M.; Subbarao, P.; Teo, K. K.; Turvey, S. E.; Williams, N. C.; Britz-McKibbin, P.; Anand, S. S. Maternal Diet and the Serum Metabolome in Pregnancy: Robust Dietary Biomarkers Generalizable to a Multiethnic Birth Cohort. *Curr. Dev. Nutr.* **2020**, *4* (10), 1–12.

(15) Miranda, A. M.; Goulart, A. C.; Benseñor, I. M.; Lotufo, P. A.; Marchioni, D. M. Coffee Consumption and Risk of Hypertension: A Prospective Analysis in the Cohort Study. *Clin. Nutr.* **2021**, *40* (2), 542–549.

(16) Poole, R.; Kennedy, O. J.; Roderick, P.; Fallowfield, J. A.; Hayes, P. C.; Parkes, J. Coffee Consumption and Health: Umbrella Review of Meta-Analyses of Multiple Health Outcomes. *BMJ.* **2017**, *359*, 5024.

(17) Gill, B.; Britz-McKibbin, P. Biomonitoring of Smoke Exposure in Firefighters: A Review. *Curr. Opin. Environ. Sci. Heal.* **2020**, 15, 57-65.

(18) Nalbant, A. A.; Boyacı, E. Advancements in Non-Invasive Biological Surface Sampling and Emerging Applications. *Separations* **2019**, *6* (4), 52.

(19) Levasseur, J. L.; Hoffman, K.; Herkert, N. J.; Cooper, E.; Hay, D.; Stapleton, H. M. Characterizing Firefighter's Exposure to over 130 SVOCs Using Silicone Wristbands: A Pilot Study Comparing on-Duty and off-Duty Exposures. *Sci. Total Environ.* **2022**, *834*, 155237.

(20) Mazumder, S.; Shia, W.; Bendik, P. B.; Achilihu, H.; Sosnoff, C. S.; Alexander, J. R.; Luo, Z.; Zhu, W.; Pine, B. N.; Feng, J.; Blount, B. C.; Wang, L. Nicotine Exposure in the U.S. Population: Total Urinary Nicotine Biomarkers in NHANES 2015–2016. *Int. J. Environ. Res. Public Health* **2022**, *19* (6), 3660.

(21) Byron, M. J.; Cohen, J. E.; Frattaroli, S.; Gittelsohn, J.; Drope, J. M.; Jernigan, D. H. Implementing Smoke-Free Policies in Low- and Middle-Income Countries: A Brief Review and Research Agenda. *Tob. Induc. Dis.* **2019**, *17 (60)*.

(22) Lorkiewicz, P.; Riggs, D. W.; Keith, R. J.; Conklin, D. J.; Xie, Z.; Sutaria, S.; Lynch, B.; Srivastava, S.; Bhatnagar, A. Comparison of Urinary Biomarkers of Exposure in Humans Using Electronic Cigarettes, Combustible Cigarettes, and Smokeless Tobacco. *Nicotine Tob. Res.* **2019**, *21* (9), 1228–1238.

(23) Fried, N. D.; Gardner, J. D. Heat-Not-Burn Tobacco Products: An Emerging Threat to Cardiovascular Health. *Am. J. Physiol. - Hear. Circ. Physiol.* **2020**, *319* (6), H1234–H1239.

(24) Gulia, S.; Khanna, I.; Shukla, K.; Khare, M. Ambient Air Pollutant Monitoring and Analysis Protocol for Low and Middle Income Countries: An Element of Comprehensive Urban Air Quality Management Framework. *Atmos. Environ.* **2020**, *222*, 117120.

(25) Liu, M.; Poo, W. K.; Lin, Y. L. Pyrazine, 2-Ethylpyridine, and 3-Ethylpyridine Are Cigarette Smoke Components That Alter the Growth of Normal and Malignant Human Lung Cells, and Play a Role in Multidrug Resistance Development. *Exp. Mol. Pathol.* **2015**, *98* (1), 18–26.

(26) Siegel, S. D.; Lerman, C.; Flitter, A.; Schnoll, R. A. The Use of the Nicotine Metabolite Ratio as a Biomarker to Personalize Smoking Cessation Treatment: Current Evidence and Future Directions. *Cancer Prev. Res.* **2020**; 13 (3), 261–272.

(27) Quaak, M.; van Schayck, C. P.; Knaapen, A. M.; van Schooten, F. J. Implications of Gene-Drug Interactions in Smoking Cessation for Improving the Prevention of Chronic Degenerative Diseases. *Mutat. Res. - Fundam. Mol. Mech. Mutagen.* **2009**, *667* (1–2), 44–57.

(28) Hu, Y.; Hu, F. B.; Manson, J. A. E. Marine Omega-3 Supplementation and Cardiovascular

Disease: An Updated Meta-Analysis of 13 Randomized Controlled Trials Involving 127 477 Participants. *J. Am. Heart Assoc.* **2019**, *8* (19), e013543.

(29) Azab, S. M.; De Souza, R. J.; Teo, K. K.; Anand, S. S.; Williams, N. C.; Holzschuher, J.; McGlory, C.; Philips, S. M.; Britz-McKibbin, P. Serum Non-Esterified Fatty Acids Have Utility as Dietary Biomarkers of Fat Intake from Fish, Fish Oil, and Dairy in Women. *J. Lipid Res.* **2020**, *61* (6), 933–944.

(30) Azab, S. M.; de Souza, R. J.; Ly, R.; Teo, K. K.; Atkinson, S. A.; Morrison, K. M.; Anand, S. S.; Britz-McKibbin, P. Non-Esterified Fatty Acids as Biomarkers of Diet and Glucose Homeostasis in Pregnancy: The Impact of Fatty Acid Reporting Methods: NEFA Reporting Methods Affect Dietary and Cardiometabolic Endpoints. *Prostaglandins Leukot. Essent. Fat. Acids* **2022**, *176*, 102378.