Title

Using Biomarkers to Characterize Exposures to Diesel Exhaust and Smoky Coal

Author

Lu, Sixin Samantha

Publication Date

2024

Peer reviewed|Thesis/dissertation

Using Biomarkers to Characterize Exposures to Diesel Exhaust and Smoky Coal

By

Sixin Samantha Lu

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Molecular Toxicology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Stephen M. Rappaport, Chair
Professor Martyn T. Smith
Professor Evan Williams

Summer 2024

Abstract

Using Biomarkers to Characterize Exposures
to Diesel Exhaust and Smoky Coal

by

Sixin Samantha Lu

Doctor of Philosophy in Molecular Toxicology

University of California, Berkeley

Professor Stephen M. Rappaport, Chair

Ambient air pollution and household air pollution are among the leading causes of mortality, responsible for 6 million premature deaths per year globally. The combustion of fossil fuels, including diesel and coal, constitutes a major source of air pollution. Polycyclic aromatic hydrocarbons (PAHs) are a group of compounds released during incomplete combustion of organic material which has been classified as carcinogenic to humans. Inhalational exposure to combustion emissions can be quantitated either externally such as by determining personal airborne PAH levels, or internally by determining the levels of a biomarker in the exposed population. Given the ease of sample collection, identifying suitable urine and blood biomarkers that accurately reflect exposure to air pollution is of significant interest. This dissertation examines the relations of the levels of biomarkers in urine and plasma with exposure determinants and with personal airborne measurements in humans exposed to emissions from diesel and solid fuel combustion. First, urinary PAH levels were measured using headspace solid-phase microextraction and gas chromatography-mass spectrometry in 28 nonsmokers in two studies before, immediately after and the day after short-term, controlled exposures to diesel exhaust (DE). Likelihood-ratio tests on linear mixed-effects models indicated that, after adjusting for other covariates, DE exposure did not significantly alter urinary PAH levels. Second, urinary PAH levels were reported in 163 Chinese female nonsmokers from Xuanwei and Fuyuan counties, where the highest lung cancer rates in the world are observed due to household air pollution from combustion of smoky coal. Mixed-effects models revealed that personal airborne PAH levels (over a wide range of exposure up to occupational levels experienced by coke oven workers), age and exposure characteristics such as fuel and stove types were significant predictors of urinary PAH levels. Third, using an untargeted nanoflow liquid chromatography and high-resolution mass spectrometry method, 50 adducts were detected at the nucleophilic Cys34 locus of human serum albumin (HSA) in 29 of the aforementioned Xuanwei and Fuyuan women and 10 local controls. Upon adduct annotation and quantitation, Wilcoxon rank sum tests showed that levels of S-glutathione (*S*-GSH) and *S*-γ-glutamylcysteine (*S*-γ-GluCys), adducts related to glutathione, an important antioxidant, were lower in exposed subjects compared to those in controls. After adjustment for age and personal measurements of airborne

benzo(*a*)pyrene (a representative carcinogenic PAH) using multivariate regressions, the largest reductions in levels of *S*-GSH and *S*-γ-GluCys relative to controls were observed for smoky coal users, compared to users of smokeless coal and wood. These studies collectively demonstrate that urinary unmetabolized PAHs may be suitable biomarkers of wide-ranging exposures to fuel combustion emissions and that Cys34 adductomics is a promising tool for discovering biomarkers of longer-term exposures, owing to the one-month residence time of HSA.

This dissertation is dedicated to my parents.

Table of Contents

List of Abbreviations

ACE	acenaphthene
ACY	acenaphthylene
ANT	anthracene
BAA	benzo(*a*)anthracene
BaP	benzo(*a*)pyrene
BBK	benzo(*b*)fluoranthene plus benzo(*k*)fluoranthene
CHR	chrysene
CRE	creatinine
CV	coefficient of variation
Cys	cysteine
Cys34	cysteine residue at position 34 in human serum albumin
CysGly	cysteinylglycine
DE	diesel exhaust
EPA	U.S. Environmental Protection Agency
FDR	false discovery rate
FLE	fluoranthene
FLU	fluorene
FY	Fuyuan County in Yunnan Province of China
GC	gas chromatograph
GSH	glutathione
hCys	homocysteine
HK	"housekeeping peptide"
HRMS	high-resolution mass spectrometry
HSA	human serum albumin
HS-SPME	headspace solid-phase microextraction
IAA-iT3	carbamidomethylated iT3 peptide internal standard
IARC	International Agency for Research on Cancer
ICC	intraclass correlation
iT3	isotopically labeled T3 peptide
LOD	limit of detection
LRT	likelihood-ratio test
MIM	monoisotopic mass
MLE	maximum likelihood estimation
MS	mass spectrometer
NAP	naphthalene
nLC	nanoflow liquid chromatography
OMN	1-methylnaphthalene
PAH	polycyclic aromatic hydrocarbon
PAR	ratio of peak area of the T3 adduct to that of the "housekeeping peptide"
PHE	phenanthrene
PM	particulate matter
$PM_{2.5}$	particulate matter with aerodynamic diameter less than 2.5 μm
PYR	pyrene

REML	restricted maximum likelihood estimation
ROS	reactive oxygen species
r_S	Spearman correlation coefficient
RT	retention time
T3	third largest peptide after tryptic digestion of HSA which includes Cys34
TMN	2-methylnaphthalene
XW	Xuanwei County in Yunnan Province of China
γ-GluCys	γ-glutamylcysteine

Acknowledgments

I extend my profound gratitude to my Ph.D. advisor, Professor Stephen Rappaport, for providing the opportunities to undertake this research and for securing the necessary funding. I am thankful for his insight, guidance, expertise, patience, encouragement, availability, and openness to new ideas. His input in formulating the experimental and statistical methods, interpreting complex results, and effectively communicating findings was vital.

I am grateful to our US and international collaborators for their efforts in planning and conducting the studies, providing specimens and pertinent data, and offering valuable feedback. My sincere appreciation goes to co-supervisors Dr. Gerd Sällsten for the Lund University Study and Dr. Qing Lan for the Xuanwei and Fuyuan Study. Special recognition is due to Dr. Jon Sobus for providing data for the US EPA Study and for teaching me HS-SPME and GC-MS operations, which enabled data collection for Chapters 1 and 2. I also extend my gratitude to others not mentioned here and to the research subjects who participated in the studies.

Acknowledgment with gratitude is due to the members of the Rappaport lab and the UC Berkeley research community: Dr. He “Harry” Li for introducing me to HPLC, LC-MS, and early adductomics experiments; Dr. Hasmik Grigoryan for her essential contributions to Chapter 3, including teaching me the innovative nLC-Orbitrap-HRMS adductomics method developed by her and others, and for instructing me in the thorough inspection and annotation of data; the late Dr. William Edmands for developing in-house software to accurately identify adducts, which greatly streamlined what would have otherwise been a painstaking manual process; Dr. Anthony Iavarone of QB3-Berkeley for sharing his expertise in mass spectrometry; and many others for engaging in fruitful discussions and collaborations that enriched my research approach.

I extend my appreciation to my qualifying exam committee for their guidance in refining my research aims and my dissertation committee for reviewing this dissertation. Special thanks to Marcela Cardona and Professor Daniel Nomura of the NST Department for facilitating my reenrollment and the filing of this dissertation.

I owe a deep debt of gratitude to my parents for their unwavering love and continuous support throughout my studies and prolonged hiatus. Despite being deprived of opportunities for higher education during an extenuating historical context, my parents are my role models for creative problem-solving and grit, always encouraging me to explore and pursue my interests. I thank my friends and roommates, especially those I met at IGSM Berkeley, for their encouragement, help, companionship, and visits during my time away. I am blessed to have mentors who have tirelessly provided wise and timely counsel. I am grateful to God for granting me the strength, peace, and hope to persevere and overcome setbacks.

This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE-0946797.

Chapter 1. Evaluating the Effect of Controlled Exposure to Diesel Exhaust on Levels of PAHs in Human Urine[a]

1.1 Introduction

Diesel exhaust (DE) was classified in 2012 as a group 1 carcinogen by the International Agency for Research on Cancer (IARC)[1] based on long-term occupational exposures. DE contributes to ambient particulate matter that has been linked to chronic cardiopulmonary and vascular effects.[2, 3] Yet, it has been difficult to quantify exposure–response relationships for DE due to the lack of quantitative data regarding exposures, which have primarily been classified by observational descriptors such as job title.[4-6] This reflects the complex nature of DE, which is comprised of both gaseous and fine particulate constituents.

Although gaseous DE contains predominantly small molecules such as nitrogen oxides and aldehydes, the particulate phase consists of elemental carbon coated with organic compounds (organic carbon), including particle-bound polycyclic aromatic hydrocarbons (PAHs).[2, 3, 7] The class of PAHs contains hundreds of chemicals with two or more fused-aromatic rings that are formed from incomplete combustion of hydrocarbons. As a class, PAHs have been associated with human lung and bladder cancers, and several of the five-ring PAHs, such as benzo(*a*)pyrene (BAP), are potent animal carcinogens.[8] Volatile or semi-volatile PAH molecules (two or three rings) are found primarily in the gas phase, while the larger compounds (four to six rings) reside primarily in the particulate phase.

Concentrations of gas-phase PAHs are much greater in DE than those of particulate-phase PAHs. For example, Sobus *et al.*[7] reported air concentrations of naphthalene (NAP, two rings) and phenanthrene (PHE, three rings) during controlled human exposure to DE that were about three orders of magnitude greater than the BAP concentration. Because air levels of NAP and/or PHE were highly correlated with those of organic carbon as well as semivolatile and particulate PAHs, the authors speculated that NAP and PHE could be suitable surrogates for exposures to all DE-derived PAHs –and to DE more generally – in studies of health effects.[7] In a separate analysis, Sobus *et al.*[9] showed that the levels of unmetabolized NAP and PHE in urine were also positively correlated in workers exposed to coke-oven emissions, asphalt fumes and DE.

Urinary PAHs are biomarkers of exposure, which offer attractive alternatives to air monitoring for determining exposure–response relationships.[10] Many PAHs have been detected in urine from urban populations exposed to air pollutants[9, 11-15] as well as from workers exposed to emissions containing PAHs.[9, 11-18] Given the abundance of volatile PAHs in DE, we hypothesized that urinary levels of gas-phase PAHs would elevate upon short-term exposure to air concentrations of DE that were well above ambient values. Since urinary PAHs are remarkably simple to measure by headspace gas chromatography-mass spectrometry (GC-MS), they would be particularly useful biomarkers if they were found at higher urine concentrations following DE exposure.

---

[a]This chapter is based on a publication. Reprinted with permission from Lu SS, Sobus JR, Sallsten G, Albin M, Pleil JD, Gudmundsson A, Madden MC, Strandberg B, Wierzbicka A, Rappaport SM. Are urinary PAHs biomarkers of controlled exposure to diesel exhaust? *Biomarkers* **2014,** *19* (4), 332-9. DOI: 10.3109/1354750X.2014.910553.

In this study, we analyzed urinary levels of 14 two-ring to five-ring PAHs in volunteer subjects from two chamber studies of DE exposure, namely, an investigation conducted at the US Environmental Protection Agency in Chapel Hill, NC (hereafter "EPA study", a 2-h exposure to 106 μg/m^3 DE measured as $PM_{2.5}$)[7] and the second conducted at Lund University, Sweden (hereafter "Lund study", a 3-h exposure to 276 μg/m^3 DE measured as PM_1).[19] Our purpose was to determine whether levels of urinary PAH significantly increased after controlled exposure to DE and thus whether urinary PAHs can be used as biomarkers of DE exposure.

1.2 Materials and Methods

Chamber Studies

All subjects were exposed with informed consent under protocols approved by institutional committees for the protection of human subjects at the relevant institutions (EPA study: US EPA and University of North Carolina, Chapel Hill, study duration 2006–2007 and Lund study: Lund University, study duration 2009–2010). In both studies, urine specimens were collected immediately before exposure (time 1), immediately after exposure (time 2) and the morning following exposure (time 3). Because the estimated half-lives of urinary NAP and PHE are about 8–10 h,[18] these collection times should allow effects of DE on urinary concentrations of volatile PAHs to be estimated at baseline (time 1), following chamber exposure (time 2) and from residual chamber exposure (time 3). Different combinations of PAHs were measured in urine from the two studies as shown in Table 1.1¸which also lists the numbers of measurements that were above the limits of detection (LOD). In the EPA study, each LOD was three times the standard deviation of replicate controls. In the Lund study, each LOD was the upper 95% confidence limit of the linear least-squares intercept of the respective calibration curve.

Details of the EPA study[7, 20, 21] and Lund study[19] have been reported. Briefly, all subjects were nonsmokers, and gas-phase PAHs were measured by adsorption on Tenax followed by thermal desorption/GC-MS). The EPA study exposed four males and six females for 2-h during two sessions (diesel or no-diesel) at least three weeks apart with subjects exposed during moderate exercise. During diesel sessions, DE was from an idling commercial vehicle with a six-cylinder, 5.9-L diesel engine, which generated $PM_{2.5}$ concentrations of 106 ± 9 (SD) μg/m^3. Urine voids were combined for each subject over 2-h periods for time 2.

The Lund study exposed nine males and nine females at rest during four sessions at least five days apart. Exposures included diesel and no-diesel sessions, with low or high recorded traffic noise at 46 and 75 dBA, respectively. During diesel sessions, the chamber was supplied with DE from an idling passenger vehicle at PM_1 concentrations of 276 ± 27 (SD) μg/m^3, and during no-diesel sessions, filtered air was supplied (PM_1 ~ 2 μg/m^3). Particulate PAHs were collected on Teflon filters using $PM_{2.5}$ cyclones and were analyzed by GC-MS.[22] Time 1 and time 3 urine samples were collected at home (before 7:00 in the morning) and time 2 urine samples were collected within 20 min of the end of chamber exposure. Two urine samples of 216 were not available for analysis.

Urinary Creatinine Measurements

Urinary creatinine (CRE) concentrations were measured with a colorimetric method employing alkaline picrate.[23] Briefly, the three following solutions were prepared: A – 0.05 M sodium borate, 0.05 M sodium phosphate, adjusted to pH 12.7 with sodium hydroxide, B – 4% SDS and C – 1% picric acid. Fifty microliters of CRE standard or 20 × diluted sample were loaded into a 96-well plate (Corning Inc., Corning, NY). Then 100 μL of a mixture of solutions

A:B:C at ratio 2:2:1 v/v were added and mixed. After 30 min, the absorbance at 490 nm was read against reference wavelength 580 nm using a PowerWave XS Microplate Reader (BioTek Instruments, Inc., Winooski, VT). Sample CRE concentrations were calculated based on a calibration curve of CRE standards (3.1–100 mg/L).

Urinary PAH Measurements

For the EPA study, urinary NAP, 2-methylnaphthalene (TMN), 1-methylnaphthalene (OMN), fluorene (FLU), PHE and anthracene (ANT) were measured as described by Sobus *et al.*[9] Briefly, urine samples were stored on ice for up to 4 h following collection and were then transferred to vials and stored at −80 °C for up to one year prior to analysis. For each sample, a 0.7-mL aliquot of urine was spiked with isotopic PAH internal standards [($^2H_8$)NAP, ($^2H_{10}$)TMN, ($^2H_{10}$)FLU, ($^2H_{10}$)PHE and ($^2H_{10}$)ANT from Sigma-Aldrich Co. LLC., St. Louis, MO] and analyzed by headspace solid-phase microextraction (HS-SPME) using a CombiPal auto-sampler (CTC Analytics, Zwingen, Switzerland) and a 10 mm, 100-μm film thickness polydimethylsiloxane fiber with adsorption at 55 °C for 30 min and desorption in the GC inlet at 250 °C for 20 min. The PAHs were measured with a Model 6890N GC coupled to a Model 5973N mass spectrometer (Agilent Technologies, Inc., Santa Clara, CA). The following ions were monitored in selected-ion monitoring mode: *m/z* 128 (NAP), *m/z* 136 [($^2H_8$)NAP], *m/z* 142 (TMN and OMN), *m/z* 152 [($^2H_{10}$)TMN], *m/z* 166 (FLU), *m/z* 178 (PHE and ANT) and *m/z* 188 [($^2H_{10}$)PHE and ($^2H_{10}$)ANT]. Calibration curves were prepared using pooled urine from human volunteers spiked with PAHs (giving concentrations of 4, 20, 100, 250, 500, 750 and 1000 ng/L) and the internal standards (giving concentrations of 500 ng/L).

For the Lund study, urine was stored at 4 °C for up to 25 h and then at −20 °C for 3–8 months prior to shipment to the University of California, Berkeley, where it was maintained at −80 °C for up to four months prior to analysis. Urine samples were analyzed for NAP, TMN, OMN, acenaphthylene (ACY), PHE, fluoranthene (FLE), pyrene (PYR), benzo(*a*)anthracene (BAA), chrysene (CHR), benzo(*b*)fluoranthene plus benzo(*k*)-fluoranthene (BBK) and BAP. Two milliliter amber-glass autosampler vials and NaCl were baked in a vacuum oven at 160 °C overnight to remove any residual target compounds. Upon thawing at room temperature, the urine was mixed by inversion and 0.8 mL was centrifuged for 5 min at 10 000× g to remove sediment. Then 0.7 mL of the supernatant was transferred to an autosampler vial containing 0.3 g NaCl. Two microliters of a mixture containing deuterium-labeled PAH internal standards (Sigma-Aldrich Co. LLC.) were injected into the urine to give concentrations of 1 μg/L of ($^2H_8$)NAP, 200 ng/L of ($^2H_{10}$)PHE, 200 ng/L of ($^2H_{10}$)PYR and 200 ng/L of ($^2H_{12}$)BAP. HS-SPME and GC-MS employed the same instruments as the EPA study. Samples were adsorbed at 85 °C for one hour and desorption was carried out at 275 °C for 20 min. A DB-5 (Agilent Technologies, Inc.) fused silica column (60-m length, 250-μm diameter and 0.25-μm film thickness) was used with He carrier gas at a flow rate of 1.1 mL/min. The GC was programmed from 45 °C to 320 °C, and the following molecular ions were monitored in the selected-ion monitoring mode: *m/z* 128 (NAP), *m/z* 136 [($^2H_8$)NAP], *m/z* 142 (TMN and OMN), *m/z* 152 (ACY), *m/z* 178 (PHE), *m/z* 188 [($^2H_{10}$)PHE], *m/z* 202 (FLE and PYR), *m/z* 212 [($^2H_{10}$)PYR], *m/z* 228 (BAA and CHR), *m/z* 252 (BBK and BAP) and *m/z* 264 [($^2H_{12}$)BAP (Benzo(*b*)fluoranthene and benzo(*k*)fluoranthene could not be chromatographically separated and were thus reported as BBK). The urine for calibration curves was created by pooling urine from 10 healthy non-smoking volunteers and then filtering through EnvirElut PAH cartridges (Agilent Technologies Inc.) to reduce background levels of PAHs. The calibration samples were generated by spiking urine with 2 μL of acetonitrile-diluted PAH calibration mix 47 940-U

(Sigma-Aldrich Co. LLC.) to give final concentrations of 0, 0.2, 2, 20, 200, 2000 or 20 000 ng/L of each PAH, with triplicates at each concentration. The calibration curves (R^2>0.96) were generated by ordinary least squares linear regression of calibration data (averages of triplicates) from a range of concentrations that was close to that of sample concentrations. The urine PAH levels from the study were estimated by fitting the response ratios of PAHs to their corresponding (same ring number) internal standards on the calibration curves.

Statistical Methods

Descriptive statistics and Student's *t*-tests were generated using Microsoft Excel 2010 (Microsoft Corp., Redmond, WA) and Stata Statistical Software release 10 (StataCorp LP, College Station, TX). The level of statistical significance was $\alpha = 0.05$ unless otherwise stated. After adjusting urinary PAH levels for CRE, median values were estimated using all observations (including imputed values) within each study. Trends and correlation coefficients were generated using ordinary least squares linear regression of logged (base 10) levels. If a PAH was detected in a sample, the observed concentration was used even if it was below the LOD. When a PAH was not detected, its urine concentration was imputed using the operative LOD/ $\sqrt{2}$ (Table 1.1). The following non-detected observations were imputed: EPA study – FLU (n = 16); Lund study – NAP (n = 10), TMN (n = 2), ACY (n = 23), BAA (n = 4), CHR (n = 2), BBK (n = 31) and BAP (n = 105). Non-detected observations for ANT (n = 29) were not imputed due to the unavailability of a LOD.

Post-exposure background ratios were calculated from CRE-adjusted PAH levels by dividing the urinary PAH levels measured immediately after exposure (time 2) by those collected prior to exposure (time 1). Since subjects from the Lund study had repeated diesel and no-diesel measurements, subject-specific means of ratios were used. Likewise, morning-after-exposure background ratios were calculated by dividing CRE-adjusted PAH levels measured the following morning (time 3) by those collected at time 1. Wilcoxon matched-pairs signed-ranks tests were performed to compare the background ratios between diesel and no-diesel sessions after Bonferroni correction of significance levels (EPA study: $\alpha = 0.05/5 = 0.01$; Lund study: $\alpha = 0.05/11 = 0.0045$). In the EPA study, ANT was excluded since many of its background ratios were unavailable.

Linear mixed-effects models for each urinary PAH (log-transformed, ng/L) were generated using the *xtmixed* command of Stata, using exposure status (diesel or no-diesel), time of urine collection (time 1, time 2 or time 3), exposure status × time interactions, gender and the urinary CRE concentration (g/L, log-transformed) as fixed effects and subject as a random effect. Only detected observations were used. Models that included imputed values produced similar results; models that excluded values below LODs had lower rates of convergence, but those that converged produced essentially the same results. Dummy variables were assigned using no-diesel, time 1 and female as reference groups. For the Lund study, we found no effect of added traffic noise on PAH levels (data not shown) and, therefore, excluded noise as a categorical variable, while retaining the diesel and no-diesel data. The following four models were fit using maximum likelihood estimation for each PAH:

$$\ln(PAH_{ijk}) = \beta_0 + (\beta_1 diesel_j) + (\beta_2 time2_k) + [\beta_3(diesel_j \times time2_k)] + (\beta_4 gender_i) + [\beta_5 \ln(creatinine_{ijk})] + \mu_{0_i} + \varepsilon_{ijk} \quad (1)$$

$$\ln(PAH_{ijk}) = \beta_0 + (\beta_1 diesel_j) + (\beta_2 time2_k) + (\beta_3 gender_i) + [\beta_4 \ln(creatinine_{ijk})] + \mu_{0_i} + \varepsilon_{ijk} \quad (2)$$

$$\ln(PAH_{ijk}) = \beta_0 + (\beta_1 diesel_j) + (\beta_2 time3_k) + [\beta_3(diesel_j \times time3_k)] + (\beta_4 gender_i) + [\beta_5 \ln(creatinine_{ijk})] + \mu_{0_i} + \varepsilon_{ijk} \quad (3)$$

$$\ln(PAH_{ijk}) = \beta_0 + (\beta_1 diesel_j) + (\beta_2 time3_k) + (\beta_3 gender_i) + [\beta_4 \ln(creatinine_{ijk})] + \mu_{0_i} + \varepsilon_{ijk} \quad (4)$$

where PAH_{ijk} is the concentration and ε_{ijk} is the residual error of a given urinary PAH for the i^{th} subject, the j^{th} exposure and the k^{th} urine sample; β_0 (intercept) through β_5 are the fixed effects and μ_{0i} is the subject-specific random effect for the i^{th} subject. For the EPA study, $i = 1,\ldots, 10$ and for the Lund study, $i = 1,\ldots, 18$. Note that models (1) and (2) were fit using only time 1 (reference) and time 2 observations and models (3) and (4) were fit using only time 1 and time 3 observations. Because models (1) and (3) included diesel × time interaction terms, statistical evidence for better fits of model (1) relative to model (2) or of model (3) relative to model (4) would indicate that DE significantly affected urinary levels of a given PAH either immediately following exposure or on the morning after exposure, respectively. Goodness-of-fit between models (1) and (2) and between models (3) and (4) were compared with likelihood-ratio tests (LRTs) using the *lrtest* command of Stata with Bonferroni-adjusted significance (EPA study: $\alpha = 0.05/6 = 0.0083$; Lund study post-exposure: $\alpha = 0.05/9 = 0.0056$ and morning-after-exposure $\alpha = 0.05/8 = 0.0063$).

1.3 Results

Urinary PAH Levels

In the Lund study, a higher HS-SPME extraction temperature was used than in the earlier EPA study to permit detection of several four-ring and five-ring PAHs. However, the higher temperature used in the Lund study also raised the LOD and decreased sensitivity for the two-ring and three-ring PAHs and, in the case of the more volatile PAHs (especially NAP), may have contributed to imprecise measurements. Table 1.1 lists the numbers of observations exceeding the respective LODs in both studies. Detection rates decreased from nearly 100% for some two-ring and three-ring compounds to about 10% for five-ring PAHs. As shown by summary statistics in Table 1.2, the median levels of two-ring PAHs (NAP, TMN or OMN) were greater than those of the three-ring PAHs (ACY, FLU, PHE or ANT) in both studies, while median concentrations of three-ring compounds exceeded those of four-ring and five-ring PAHs (FLE, PYR, BAA, CHR, BBK or BAP) in the Lund study.

Background Ratios

Univariate statistics employed background ratios to determine whether DE exposure affected PAH levels either immediately following (time 2) or on the morning after exposure (time 3). As shown in Figures 1.1 and 1.2, ranges of background ratios for a given PAH spanned about 100-fold (from 0.1 to 10), indicating that urinary PAH levels varied greatly across individual subjects in each study. In both studies, background ratios were marginally greater immediately after exposure, but this was the case for both diesel and no-diesel sessions. In fact, no difference in background ratios was observed between diesel and no-diesel sessions for any PAH in either study, using Bonferroni-adjusted *p*-values. When we repeated the analyses using unadjusted PAH levels to calculate background ratios, the results and plots were essentially the same (not shown).

Mixed Models

Results from the mixed-effects models are summarized in Table 1.3 along with *p*-values from the LRTs (comparisons were not available for six pairs of models because of non-convergence). By comparing the fits between models (1) and (2) and models (3) and (4), we tested whether inclusion of the diesel × time interaction term was justified, and by extension, whether exposure to DE affected urinary PAH levels. Of the 28 LRTs, only OMN in the post-exposure samples from the EPA study showed a significant interaction term (*p*-value = 0.008). (Results for OMN from the Lund study could not be tested because the models did not converge.) This indicates that, with the possible exception of OMN, DE was not significantly associated with increased levels of any measured PAH in subjects' urine. Of the 29 models without interaction terms (models 2 and 4), significant effects of *diesel, time 2, time 3* and *gender* were only sporadically observed and the signs of *diesel* and *time 2* effects were inconsistent across PAHs. In fact, CRE concentrations and intercepts were the only fixed effects consistently associated with urinary PAH levels. Thus, we conclude that DE sessions did not alter urinary PAH levels after adjustment for other covariates, consistent with the univariate results summarized in Figures 1.1 and 1.2. This probably reflects masking of DE exposure by other exposure sources, including the diet and the background environment. Since diet is the main source of PAH exposure for nonsmokers,[8] dietary sources could have overwhelmed the contribution of DE during two or three hours of experimental exposure.

Linear mixed-effects models have been used previously to investigate urinary PAHs as potential biomarkers in a study of workers exposed to asphalt.[24] In that study, workers were exposed to PAHs both by inhalation and dermal absorption and both routes of exposure produced significant effects on urinary levels of NAP and PHE (the only PAHs reported). As in this investigation, Sobus *et al.*[24] observed significant effects of urinary CRE on urinary NAP and PHE levels but also detected significant effects of work activities and the timing of urine collection.

Urinary versus Airborne PAHs

An unexpected result from our studies was the strong positive correlations between median levels of CRE-adjusted urinary PAHs and mean levels of airborne PAHs to which subjects were exposed in the chambers. As shown by the log-scale plots in Figure 1.3 between 90% and 98% of the variation in urinary PAH levels was explained by the corresponding PAH air concentrations. Since the mixed-model results (Table 1.3) showed that DE sessions did not increase urinary PAH levels, the relationships shown in Figure 1.3 suggest that profiles of urinary PAHs also reflect background PAH exposures received by the subjects in the respective EPA and Lund studies. In fact, each set of the CRE-adjusted urinary PAHs, stratified by exposure status and time, had correlation coefficients with airborne PAHs between 86% and 99% (not shown).

1.4 Discussion

Table 1.4 lists baseline urinary concentrations of NAP and PHE from several urban populations and occupationally exposed populations at the beginning of a work week. The fold ranges of urinary NAP and PHE concentrations across studies were 72 and 44, respectively. This highlights the importance of having geographically matched controls in biomarker studies for ubiquitous contaminants such as PAHs. The EPA study had the lowest values of urinary NAP and PHE shown in Table 1.4 while the Lund study had values at the higher end. (Note that the

Lund study NAP result may have been anomalously high due to reduced sensitivity of the analytical method.)

Although urinary levels of unmetabolized PAHs were not significantly affected by DE sessions, PAH metabolites that are present at much greater abundance than their parent compounds in urine may be more robust biomarkers of DE. For example, 1-hydroxypyrene (a metabolite of PYR), has been a significant indicator of occupational DE exposure studies, many of which also included smoking status and/or genotypic expression of metabolizing enzymes.[25, 26] Although our controlled experiments with non-smokers avoided the confounding effect of smoking, we found no association between DE exposure and levels of urinary PYR, the precursor for 1-hydroxypyrene. Another PAH biomarker is 1-aminopyrene,[27, 28] a metabolite of 1-nitropyrene that is at least 10-fold more abundant in DE than in ambient air.[29] Following controlled exposure of 55 subjects for 1 h to DE containing 1-nitropyrene at 2.68 ng/m^3, the median 24-h time-weighted concentration of urinary 1-aminopyrene was 139 ng/g CRE compared with 21.7 ng/g CRE following exposure to clean air.[27]

We measured several unmetabolized two-ring to five-ring PAHs in 274 urine samples from 28 subjects after controlled exposures to DE in two chamber studies. There was no increase in urinary PAH levels associated with DE after adjusting for other covariates. This unexpected result probably reflects masking of PAH from DE by contributions from the diet and background environment. Since the diet is the main source of PAH exposure for most nonsmokers,[8] dietary sources could have overwhelmed the contribution of DE during two or three hours of experimental exposure even though the DE levels (106–276 μg/m^3) were high compared to ambient sources of DE, and the EPA study increased the alveolar ventilation rate – and DE absorption – by exposing subjects under moderate exercise. In any case, our results from two controlled chamber studies indicate that urinary PAHs are not promising biomarkers of short-term DE exposure.

Chapter 1 Figures

Figure 1.1. Background ratios of urinary PAH levels for the EPA study, using creatinine-adjusted data (imputed observations are included; squares and triangles represent medians of background ratios from no-diesel and diesel sessions, respectively; solid-line and dashed bars represent the ranges of background ratios from no-diesel and diesel sessions, respectively). *P*-values refer to significance of comparisons of background ratios between diesel and no-diesel sessions. FLU, fluorene; NAP, naphthalene; OMN, 1-methylnaphthalene; PHE, phenanthrene; TMN, 2-methylnaphthalene.

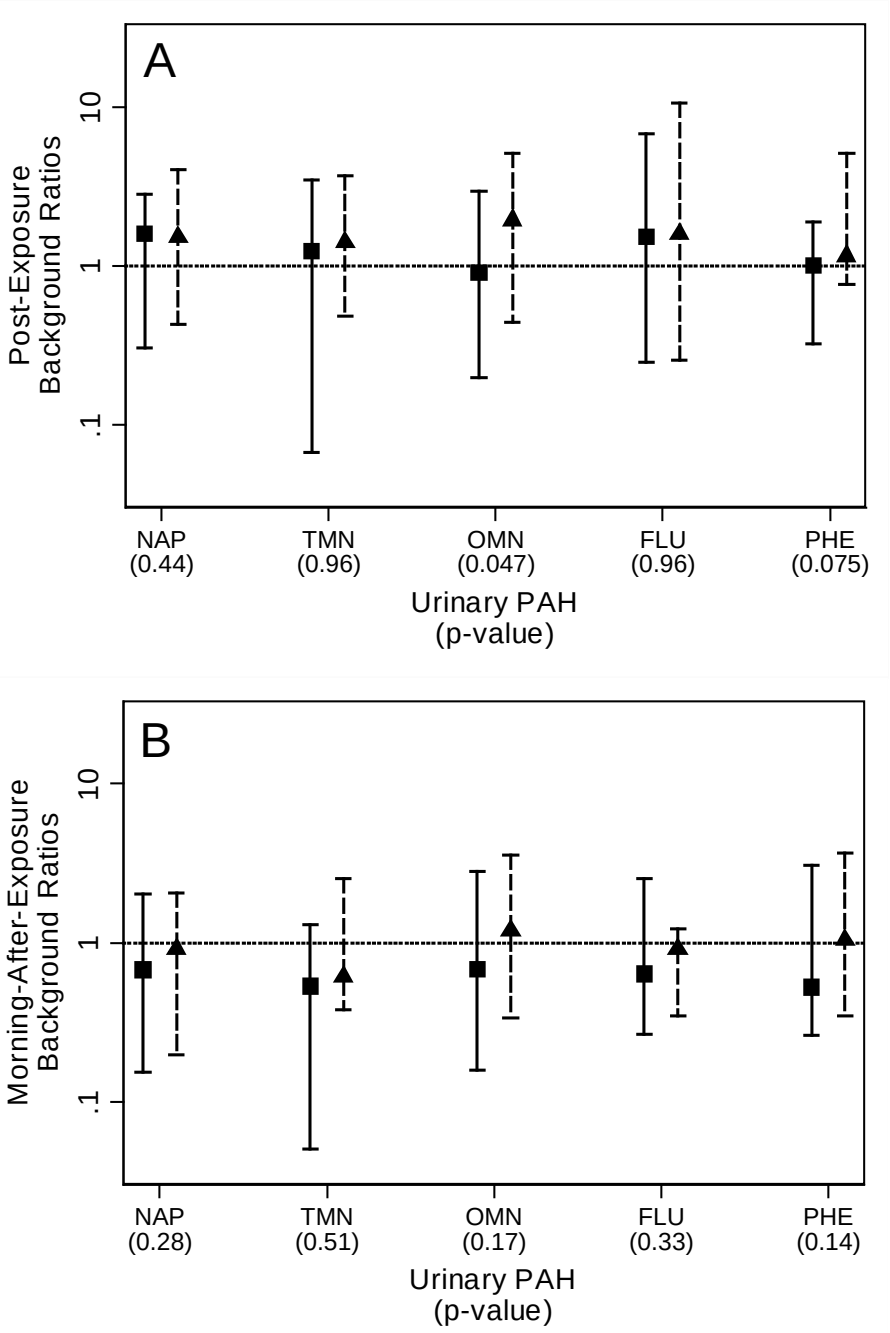


Figure 1.2. Background ratios of urinary PAH levels for the Lund study, using creatinine-adjusted data (imputed observations are included; squares and triangles represent medians of background ratios from no-diesel and diesel sessions, respectively; solid-line and dashed bars represent the ranges of background ratios from no-diesel and diesel sessions, respectively). *P*-values refer to significance of comparisons of background ratios between no-diesel and diesel sessions. ACY, acenaphthylene; BAA, benzo(*a*)anthracene; BAP, benzo(*a*)pyrene; BBK, benzo(*b*)fluoranthene plus benzo(*k*)fluoranthene; CHR, chrysene; FLE, fluoranthene; NAP, naphthalene; OMN, 1-methylnaphthalene; PHE, phenanthrene; PYR, pyrene; TMN, 2-methylnaphthalene.

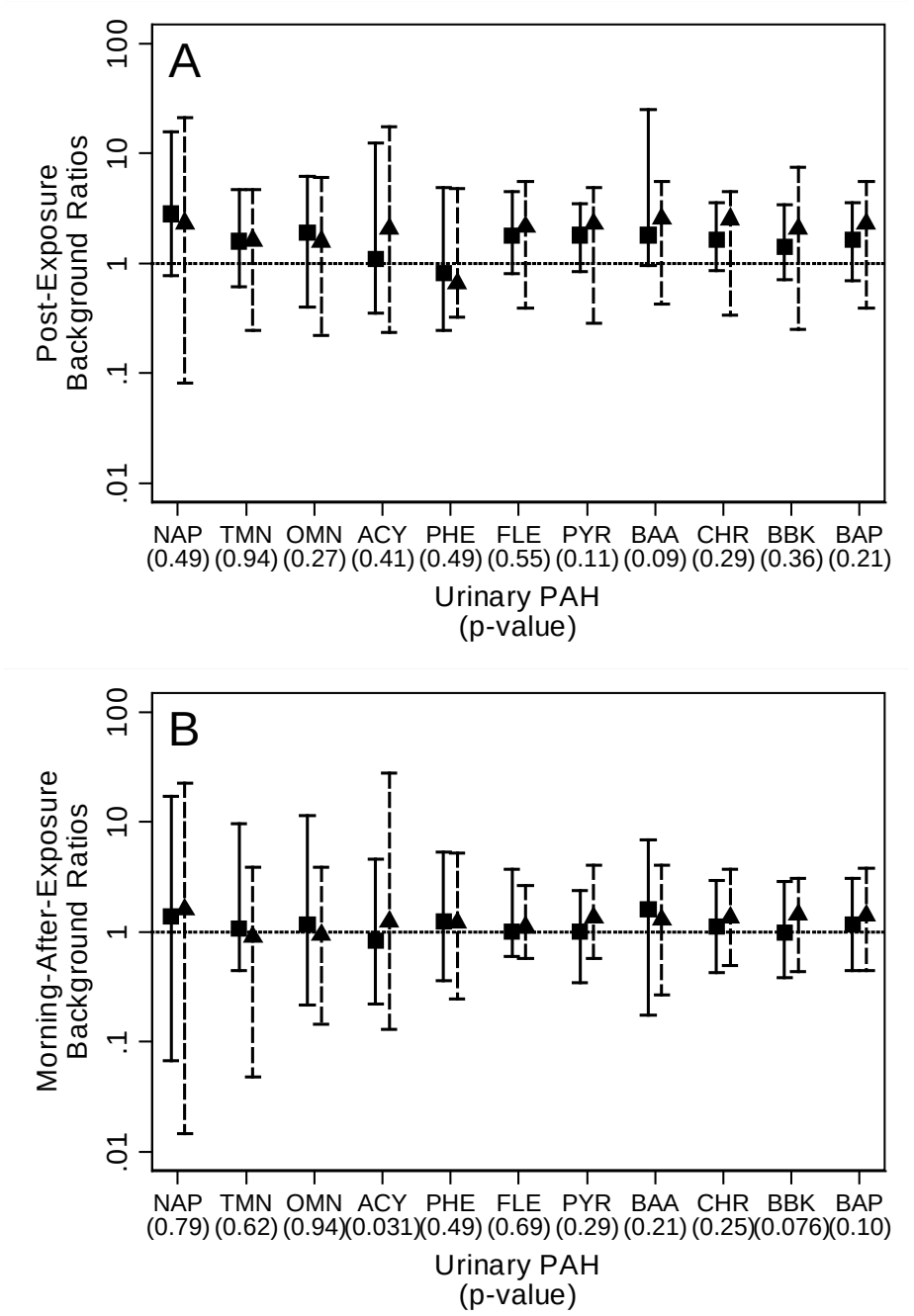
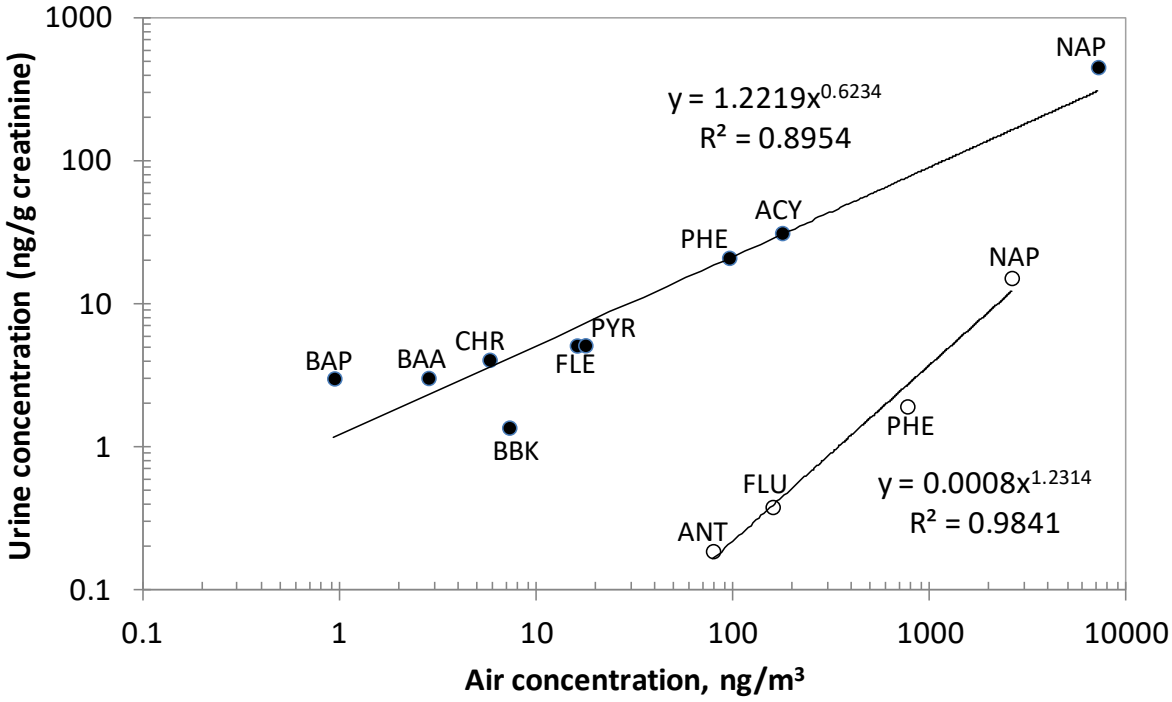


Figure 1.3. Log-scale plots of median urinary-PAH levels versus mean airborne-PAH levels during diesel-exhaust exposure (closed symbols represent the Lund study and open symbols represent the EPA study). (Air levels of TMN and OMN were not measured). ACY, acenaphthylene; ANT, anthracene; BAA, benzo(*a*)anthracene; BAP, benzo(*a*)pyrene; BBK, benzo(*b*)fluoranthene plus benzo(*k*)fluoranthene; CHR, chrysene; FLE, fluoranthene; FLU, fluorine; NAP, naphthalene; PHE, phenanthrene; PYR, pyrene.



Chapter 1 Tables

Table 1.1. Summary of PAHs detected in urine and limits of detection (LOD) and numbers of measurements above LOD (*n* = 60 for EPA study and *n* = 214 for Lund study).

Study	Polycyclic aromatic hydrocarbon (PAH)	LOD (ng/L)[a]	Number above LOD (%)
EPA	Naphthalene (NAP)	3.7	60 (100)
	2-Methylnaphthalene (TMN)	0.71	60 (100)
	1-Methylnaphthalene (OMN)	0.49	60 (100)
	Fluorene (FLU)	0.27	35 (58)
	Phenanthrene (PHE)	1.3	60 (100)
	Anthracene (ANT)	nd	nd[b]
Lund	Naphthalene (NAP)	54	190 (89)
	2-Methylnaphthalene (TMN)	12	212 (99)
	1-Methylnaphthalene (OMN)	8.7	194 (91)
	Acenaphthylene (ACY)	6.1	189 (88)
	Phenanthrene (PHE)	2.6	213 (100)
	Fluoranthene (FLE)	1.6	213 (100)
	Pyrene (PYR)	1.6	213 (100)
	Benzo(*a*)anthracene (BAA)	5.5	42 (20)
	Chrysene (CHR)	2.7	212 (99)
	Benzo(*b*)fluoranthene plus	1.5	27 (13)
	Benzo(*k*)fluoranthene (BBK)		
	Benzo(*a*)pyrene (BAP)	3.3	21 (10)

[a]Limit of detection.
[b]Not determined.

Table 1.2. Summary statistics for urinary PAHs, ng/L (n = 60 for EPA study and n = 214 for Lund study, including imputed values for non-detected observations).

Study	PAH	No. of rings	Minimum	Median	Interquartile range	Maximum
EPA	NAP	2	4.9	12	9.6–14	50
	TMN	2	1.5	3.8	2.7–5.8	16
	OMN	2	0.52	2.7	1.4–5.1	22
	FLU	3	0.17	0.30	0.19–0.39	1.7
	PHE	3	0.55	1.4	0.91–2.5	13
	ANT	3	n/a	0.24	n/a–0.62	2.1
Lund	NAP	2	0.89	310	170–580	3200
	TMN	2	8.2	99	57–180	750
	OMN	2	1.4	24	14–37	110
	ACY	3	4.3	20	12–40	84
	PHE	3	1.8	18	7.9–33	400
	FLE	4	1.1	3.6	3.0–4.7	17
	PYR	4	1.2	4.3	3.9–4.9	7.4
	BAA	4	0.016	2.6	1.2–4.9	30
	CHR	4	1.9	3.1	3.0–3.3	5.4
	BBK	5	0.41	1.0	0.82–1.2	9.5
	BAP	5	1.3	2.3	2.3–2.3	21

ACY, acenaphthylene; ANT, anthracene; BAA, benzo(*a*)anthracene; BAP, benzo(*a*)pyrene; BBK, benzo(*b*)fluoranthene plus benzo(*k*)fluoranthene; CHR, chrysene; FLE, fluoranthene; FLU, fluorene; NAP, naphthalene; OMN, 1-methylnaphthalene; PHE, phenanthrene; PYR, pyrene; TMN, 2-methylnaphthalene.

Table 1.3. Results from linear mixed-effects models for urinary PAHs.

		Post-exposure (model 2) parameters						Morning-after-exposure (model 4) parameters					
Study	PAH	Diesel	Time 2	Gender	ln(CRE)	Int.	LRT	Diesel	Time 3	Gender	ln(CRE)	Int.	LRT
EPA	NAP				+[b]	+[b]	0.206				+[b]	+[b]	0.118
	TMN				+[a]	+[b]	0.101				+[a]	+[b]	0.021
	OMN				+[b]	+[b]	0.008[c]	–			+[b]	+[b]	0.056
	FLU					–[b]	0.599				+[b]	–[b]	0.149
	PHE			–[a]	+[a]	+[b]	0.049			–	+[b]	+[b]	0.055
	ANT				+		n/a[d]				+	–[a]	0.628
Lund	NAP					+[b]	0.036	n/a	n/a	n/a	n/a	n/a	n/a[d]
	TMN				+[b]	+[b]	0.345					+[b]	0.032
	OMN	n/a	n/a	n/a	n/a	n/a	n/a[d]		–			+[b]	0.392
	ACY	n/a	n/a	n/a	n/a	n/a	n/a[d]	n/a	n/a	n/a	n/a	n/a	n/a[d]
	PHE		–[b]		+[b]	+[b]	0.432			–[a]	+[b]	+[b]	0.686
	FLE		+[b]		+[b]	+[b]	0.412				+[b]	+[b]	0.117
	PYR	–			+	+[b]	0.198	n/a	n/a	n/a	n/a	n/a	n/a[d]
	BAA		+[a]		+[b]		0.690				+[b]		0.144
	CHR	+	+		+[a]	+[b]	0.986				+[a]	+[b]	0.312
	BBK	–			+[a]	+	0.831					+	0.775
	BAP			–	+[a]	+[b]	0.152	–[a]		–[a]	+[b]	+[b]	0.475

[a] $0.0005 \leq p < 0.01$

[b] $p < 0.0005$

[c] Statistically significant LRT using the Bonferroni-corrected α.

[d] The corresponding model including diesel × time interaction did not converge.

Structure of tests: "diesel" - diesel exposure compared to no-diesel exposure (reference); "time 2"- post-exposure compared to pre-exposure (reference); "time 3" - morning after exposure compared to pre-exposure (reference); "gender" - male compared to female (reference); "CRE", urinary creatinine (continuous variable); "Int.", intercept; LRT, likelihood-ratio test comparing models (1) to (3) or models (2) to (4).

ACY, acenaphthylene; ANT, anthracene; BAA, benzo(*a*)anthracene; BAP, benzo(*a*)pyrene; BBK, benzo(*b*)fluoranthene plus benzo(*k*)fluoranthene; CHR, chrysene; FLE, fluoranthene; FLU, fluorene; NAP, naphthalene; OMN, 1-methylnaphthalene; PHE, phenanthrene; PYR, pyrene; TMN, 2-methylnaphthalene.

Plus indicates a significant positive association and minus indicates a significant negative association. Models that did not converge are indicated by n/a. The *p*-values are given for likelihood-ratio tests (LRT) comparing models (1) and (2) or models (3) and (4). A significant LRT would indicate a significant diesel × time interaction term.

Table 1.4. Comparison of urinary naphthalene (NAP) and phenanthrene (PHE) concentrations (ng/L) from the current studies with published data from urban populations or pre-shift urine from workers.

NAP median (minimum–maximum)	PHE median (minimum–maximum)	No. of subjects	No. of smokers	Type of subjects	Country	References
859 (107–3170)	62 (8–690)	22	3	Controls	China	Waidyanatha *et al.*[14]
310 (0.89–3200)	18 (1.8–400)	18	0	DE exposed and controls	Sweden	Current Lund study
155 (99.4–191)	25.7 (20.7–58.6)	5	5	Controls	Italy (assumed)	Campo *et al.*[12]
69 (32–225)	22 (10–109)	100	44	Baseline, asphalt workers	Italy	Campo *et al.*[11]
61 (25–175)	21 (9–97)	47	30	Baseline, construction workers	Italy	Campo *et al.*[11]
36.2 (21.0–63.1)	15.5 (3.89–62.0)	6	1	Baseline, millers	US	Sobus *et al.*[18]
24.5 (7.46–190)	7.64 (0.45–44.5)	20	5	Baseline, pavers	US	Sobus *et al.*[18]
27 (GM) 2.46 (GSD)	n/a	26	26	Control military personnel, pre-shift	US	Serdar *et al.*[13]
14 (GM) 4.79 (GSD)	n/a	48	0	Control military personnel, pre-shift	US	Serdar *et al.*[13]
22.0 (4.71–46.1)	3.53 (0.82–11.2)	27	11	Pre-shift, dock workers	US	Sobus *et al.*[9]
16.3 (8.50–22.0)	3.61 (1.92–11.6)	4	1	Pre-shift, office workers	US	Sobus *et al.*[9]
14.0 (6.89–22.2)	4.54 (1.54–11.8)	8	1	Pre-shift, shop workers	US	Sobus *et al.*[9]
12 (4.9–50)	1.4 (0.55–13)	10	0	DE exposed and controls	US	Current EPA study

GM, geometric mean; GSD, geometric standard deviation.

Chapter 2. Levels of Urinary Polycyclic Aromatic Hydrocarbons in Female Solid Fuel Users in Xuanwei and Fuyuan, China

2.1 Introduction

About three billion people in the world rely on the household burning of solid fuels, which is associated with lung cancer risk.[30] Some of the highest rates of lung cancer in the world have been reported in Xuanwei and Fuyuan counties in Yunnan Province of China, where indoor combustion of smoky (bituminous) coal has been linked to an excess risk of lung cancer compared to the use smokeless (anthracite) coal or wood.[31-33] Lung cancer rates in Xuanwei and Fuyuan varied depending on the smoky coal subtypes used in different geologic locations and on stove type.[33, 34]

Polycyclic aromatic hydrocarbons (PAHs) are products of incomplete combustion and as a class are known to cause lung cancer in humans.[8] After exposure, PAHs undergo extensive metabolic activation and form adducts with DNA, which can drive mutations in protooncogenes or tumor suppresser genes.[8] PAHs are among the components of smoky coal emissions speculated to contribute to lung cancer in Xuanwei and Fuyuan. PAH levels in smoky coal emissions were comparable to those in occupational exposure in coke oven workers and higher than those in smokeless coal emissions.[32, 35] Replacing unventilated stoves with ventilated ones resulted in reductions in airborne PAH levels along with lower lung cancer rates in the region.[34] Polymorphisms in genes involved in PAH metabolic activation and clearance were observed to modulate lung cancer risks in Xuanwei.[36, 37] Moreover, the prevalence of G to T transversions in driver genes in lung cancer tissue of Xuanwei never-smokers were consistent with PAH-induced DNA damage.[38]

PAH-DNA adducts in bronchoalveolar lavage[39] and urinary biomarkers[40] have been explored to characterize PAH exposure in Xuanwei residents. Although metabolized PAHs are more abundant in urine, the levels of unmetabolized PAHs are presumably independent of PAH metabolism genotypes. Urinary PAH concentrations have been shown to correlate with airborne PAH levels and with occupational exposure status.[13, 14, 18, 41, 42] Mumford *et al.* compared urinary PAH levels in 16 female nonsmokers from Xuanwei who used smoky coal with those in 8 nonsmoking Kunming controls who used gas or electricity.[40] The average levels of 10 of 12 urinary PAHs in Xuanwei subjects were higher, albeit not significantly, than those in Kunming controls.[40] There has not been an extensive study of urinary PAH levels in Xuanwei and especially one comparing the levels between smoky coal users and users of other solid fuels.

In the current study, we report urinary PAH levels in 163 nonsmoking females from Xuanwei and Fuyuan who used smoky coal, smokeless coal, wood or plant material. We investigate the relationship between urinary PAH levels and matching exposure data, namely, airborne PAH levels and determinants of exposure such as fuel and stove types.

2.2 Materials and Methods

Urine and Air Samples

Urine and air samples were obtained with informed consent from subjects in China under protocols approved by the National Cancer Institute and local Chinese institutions. Urine

samples were collected from 163 nonsmoking female solid fuel users in Xuanwei and Fuyuan Counties in Yunnan, China during one to two visits in a study of indoor air pollution.[35, 43-46] Each visit included up to two successive 24-h periods of personal air sampling, at the end of which a spot urine sample was collected. Exposure characteristics, including fuel type, stove design, room size and season, were also recorded. Urine was stored at −80 °C for two to three years before analysis. Of the 198 unique urine samples, 21 were aliquoted in blinded duplicates and three as triplicates, resulting in a total of 225 aliquots for which urinary PAH levels were determined. Personal measurements of airborne PAHs have been reported.[35] Of the 198 unique urine samples, matching airborne PAH levels (via air sampling in the 24-h period prior to urine collection) were available for between 113 (TMN) and 179 (FLE) samples.

Urinary PAH and Creatinine Measurements

Sodium chloride and 2-mL amber-glass autosampler vials were heated at 160 °C in a vacuum oven overnight to minimize background PAH contamination. The 225 urine samples were processed in random batches of up to 10 samples. Upon thawing, urine was mixed by inversion and 750 μL was transferred to a microcentrifuge tube and centrifuged for 5 mins at 10 000× g to remove sediment. Twenty μL of the supernatant was diluted with 380 μL of water and stored at −20 °C prior to measurement of creatinine, as described in Chapter 1.2.

The levels of NAP, TMN, OMN, ACE, PHE, ANT, FLE, PYR, BAA and CHR were measured in urine in random batches of 10. A 0.7-mL aliquot of urine supernatant was transferred to an autosampler vial containing 0.3 g of NaCl. Then 2 μL of a mixture containing deuterium-labeled PAH internal standards (Sigma-Aldrich Co. LLC.) were added to give concentrations of 1 μg/L of ($^2H_8$)NAP, 200 ng/L of ($^2H_{10}$)PHE, 200 ng/L of ($^2H_{10}$)PYR and 200 ng/L of ($^2H_{12}$)BAP. The vial was capped immediately and the samples were subjected to automated headspace solid-phase microextraction (HS-SPME) with adsorption at 85 °C for one h and desorption at 275 °C for 20 min, followed by gas chromatography-mass spectrometry (GC-MS). The analytes were separated on a DB-5 (Agilent Technologies, Inc., Santa Clara, CA) fused silica column using 1.1 mL/min He carrier gas in a GC (Model 6890N, Agilent Technologies) ramped from 45 °C to 320 °C. The following ions were monitored in the selected-ion monitoring mode in MS (Model 5973N Agilent Technologies): *m/z* 128 (NAP), *m/z* 136 [($^2H_8$)NAP], *m/z* 142 (TMN and OMN), *m/z* 152 (ACY), *m/z* 178 (PHE and ANT), *m/z* 188 [($^2H_{10}$)PHE], *m/z* 202 (FLE and PYR), *m/z* 212 [($^2H_{10}$)PYR], and *m/z* 228 (BAA and CHR), and *m/z* 264 [($^2H_{12}$)BAP. The signal for BAP was not detected in the samples. PHE and ANT could not be chromatographically separated and were reported as PHE+ANT. Similarly, BAA and CHR were reported as BAA+CHR.

The calibration samples were generated by injecting 2 μL of acetonitrile-diluted PAH calibration mix 47 940-U (Sigma-Aldrich Co. LLC) and 2 μL of the internal standards mixture into autosampler vials containing 0.7 mL water (18.2 mΩ cm resistivity at 25 °C) and 0.3 g NaCl to give final concentrations of 0.1, 0.25, 0.5, 1, 2.5, 5, 10, 25, 100, 1000 or 5000 ng/L of each PAH in triplicates. The calibration curves ($R^2 > 0.94$) were generated by linear regression (intercepts set to zero) of calibration data from a range of concentrations that was similar to those of samples.

The internal-standard responses from the samples were smoothed to eliminate sporadic signals from co-eluting ions while retaining drifts in analytical sensitivity. After removing internal-standard responses greater than three times the median absolute deviation, a LOESS-smoothed fit over the order of analyses was computed with 7-fold cross-validation using the *bisoreg* package[47] in R[48]. Urinary PAH levels from the study were estimated by fitting ratios of

PAH responses to the smoothed responses of the corresponding (same ring number) internal standards.

Statistical Methods

Statistical analyses were performed using Stata (Stata Statistical Software: Release 14. College Station, TX). Of 225 urine samples, PAHs were detected in between 194 (86%, PYR) and 225 (100%, NAP, ACE, PHE+ANT and FLE) specimens. Coefficients of variation (CVs) were calculated from analysis of variance (ANOVA) of analyte levels of 24 urine aliquots that were prepared in duplicate or triplicate. Duplicate and triplicate measurements were summarized using their geometric means, resulting in up to 198 observations per urinary PAH (Table 2.1) for the urine samples. Urinary and airborne PAH levels and creatinine levels were transformed by taking natural logarithms, which were used for statistical analyses. Pairwise Pearson correlations between logged levels of urinary PAH were computed.

The fixed effects of airborne PAH levels, urinary creatinine levels and the age of the subject on urinary PAH levels were assessed by linear mixed-effects regression models with subject as a random effect, using an exchangeable covariance structure. In a given model for a urinary PAH, levels of the reference airborne PAH was used as a covariate. For a given model, age was omitted from the fixed effects if the preliminary model revealed that age was not a significant predictor (p-value > 0.1). Standardized residuals were calculated from the models, and final models were generated after omitting outlier observations with an absolute standardized residual of greater than three (one for NAP, one for OMN and two for BAA+CHR). Pearson correlations between (logged) levels of related urinary and airborne PAHs were computed after excluding outliers.

The fixed effects of exposure characteristics, creatinine levels and the age of subject on urinary PAH levels were assessed by linear mixed-effects regression with subject as a random effect. Identity covariance structure was selected, since not all models converged using the exchangeable covariance structure. Exposure characteristics were fuel type, stove design, room size and season that were coded as categorical variables. Standardized residuals were calculated from the models, and final models were generated after omitting outlier observations (one for each of NAP, TMN, OMN and BAA+CHR respectively).

Output from the mixed models were interpreted as follows. For predictors that had been log-transformed (*i.e.*, urinary PAH and creatinine levels), the exponent of the estimate (geometric mean change, GMC) represents the increase in a given urinary PAH in natural scale per unit increase in the predictor variable also in natural scale. For predictors that had not been log-transformed (*i.e.*, age and exposure characteristics), the exponent of the estimate (geometric mean ratio, GMR) represents the fold change in urinary PAH in natural scale per unit increase in the predictor in natural scale.

2.3 Results

Urinary PAH Levels and Pairwise Correlations

Urinary PAH levels are summarized in Table 2.1. Of 198 spot urine samples, PAHs were detected in 172 (87%, PYR) to 198 (100%, NAP, ACE, PHE+ANT and FLE) samples. Although the relatively high HS-SPME temperatures and long absorption time enabled the detection of four-ring PAHs, these high temperatures probably also contributed to imprecise measurements (median CV: 92%). Median levels of urinary PAHs ranged from 0.378 ng/L (BAA+CHR) to 152 ng/L (PHE+ANT). Two- or three-ring PAHs were generally more abundant than four-ring PAHs.

Pairs of all urinary PAHs were strongly correlated, except urinary ACE, which had small to moderate correlations with other urinary PAHs (Table 2.2). The largest correlation coefficients were 0.903 (PHE+ANT versus FLE), 0.870 (NAP versus TMN) and 0.863 (NAP versus OMN). Urinary PHE+ANT were highly correlated with all other urinary PAHs (rho > 0.7) except ACE.

Mixed Models of and Correlations between Urinary- versus Air-PAH Levels

To investigate whether PAH levels in air were linked to those in urine, we regressed logged urinary PAH levels on the covariates of logged airborne PAH levels, logged creatinine levels and the ages of subjects, via mixed effects models (Table 2.3). Airborne PAH levels were positively associated with urinary PAH levels ($p < 0.05$ for all PAHs except ACE). The models estimated that each ng/m^3 increase in the airborne PAH corresponded to a 1.05 to 1.24 ng/L increase in a given urinary PAH. Urinary creatinine was significantly and positively associated with urinary PAHs. Age had marginally significant ($p < 0.1$) and positive effects on urinary PAH levels in 4 of 8 models. It was estimated that each year increase in age gave rise to a one to two percent increase in urinary PAH levels.

Small but generally statistically significant ($p < 0.05$) Pearson correlations were found between logged levels of each pair of urinary vs airborne PAHs (Figure 2.1). The scatter of points in Figure 2.1 may be attributed to the large CVs due to imprecise urinary PAH data. Notwithstanding the largest CV of 199% for urinary PHE+ANT, the largest correlation was observed between urinary PHE+ANT and airborne PHE (rho = 0.28, $p = 0.0014$).

Mixed Models of Urinary PAHs versus Exposure Characteristics

The effects of exposure variables, recorded for the 24 h prior to urine collection, were investigated in mixed effects models (Table 2.4). Fuel type, stove design, room size and season were included as categorical variables to investigate their effects on urinary PAH levels, after adjustments for age and (logged) creatinine levels. The fuels were categorized as smokeless coal, 6 smoky coal subtypes, multiple coal types, multiple fuel types, plant products or wood. The smoky coal subtypes were coking coal from XW, coking coal from FY, 1/3 coking coal from FY, gas fat coal from FY, meagre lean coal from FY, and smoky coal of uncertain source (*i.e.* collected from a village in a smokeless coal producing area of FY).[35] With the exception of 'meager lean' and 'smoky coal of uncertain source', the use of the other 8 fuel types was typically associated with elevated urinary PAH levels compared to the use of smokeless coal from FY and XW. The use of coking coal from XW, coking coal from FY and multiple coal types were associated with significant increases (compared with smokeless coal from FY and XW) in the levels of three, five and four urinary PAHs respectively. Of note, some of the largest and most significant effects of fuel type were seen in association with urinary OMN, including an estimated 2.93 times and 3.76 times the urinary OMN levels in users of coking coal from XW and FY, respectively, compared to those in users of smokeless coal (FY and XW).

Across the models, one significant effect was found within each of stove design, room size, and season. Compared with the use of a ventilated stove, the use of an unventilated stove was consistently associated with an increase in urinary PAHs. In general, using a fire pit or mixed ventilation also led to increased urinary PAH levels. Compared with rooms smaller than 40 m^3, larger rooms were usually linked to lower PAH levels in urine. Fuel use during the spring or summer increased levels of all of the urinary PAHs compared to the fall.

2.4 Discussion

We report the levels of 10 PAHs in 198 urine samples from 163 Xuanwei women who were exposed to indoor solid fuel emissions. We explored the relationship between air and urine levels of PAHs. Several of our observations are in line with previous reports of urinary PAHs in human subjects. First, we observed generally strong pairwise correlations across urinary PAHs, consistent with [9, 16, 18] regardless of the type of solid fuel. Second, there were significant positive correlations between urinary and airborne PAH levels, as reported elsewhere in occupational exposure settings.[13, 42] Third, creatinine was consistently a strong, positive predictor of urinary PAH levels as previously observed by[18].

Using mixed effects models, Downward *et al.* found that fuel type, stove design, room size and season were significant determinants of PAH levels in 24-h personal air samples.[35] We used similar models to investigate the effects of those exposure characteristics on urinary PAH levels while adjusting for the effects of creatinine and age. Parallel to findings in Downward *et al.*, we observed that combustion of smoky coal, plant-based fuels and wood were generally associated with higher urinary PAH levels than combustion of smokeless coal. Of the urinary PAHs investigated, OMN had the largest GMRs for five out of six smoky coal subtypes. This suggests that OMN is a more suitable analyte than other urinary PAHs for discriminating between smoky and smokeless coal exposure. The use of unventilated stoves or fire pits was typically linked to higher urinary PAH levels than the use of ventilated stoves, analogous to what was observed for airborne PAHs.[35]

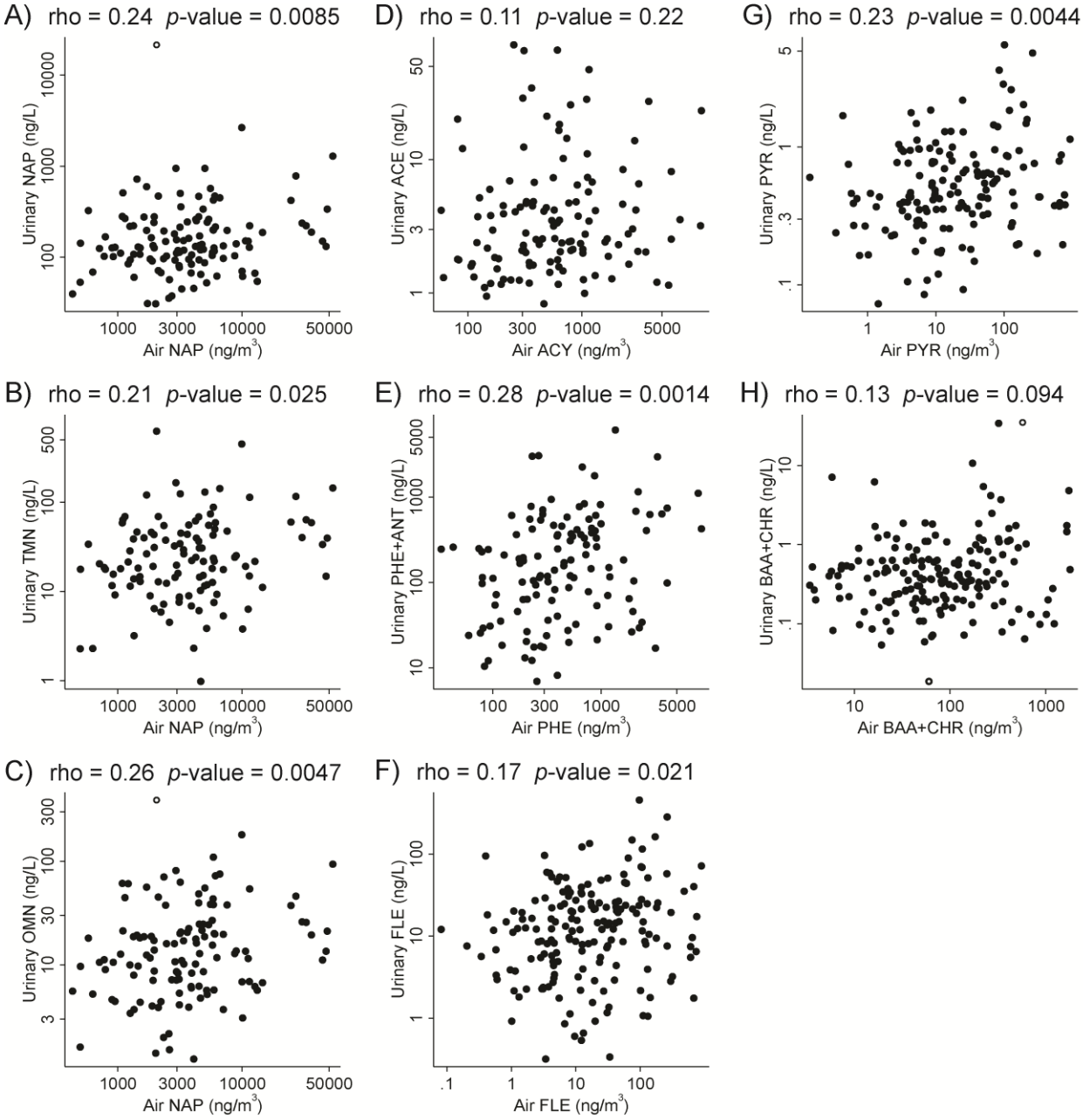
Living in rooms in the highest two quartiles of size led to lower urinary PAH levels compared to rooms in the first quartile. The opposite effect was noted in airborne levels of BAP and FLE.[35] There were non-significant increases in the levels of the two- to three-ring urinary PAHs in the winter, and non-significant decreases in the levels of four-ring urinary PAHs. By contrast, airborne PAH levels were predicted to be higher in the winter than in the fall.[35] Higher levels of PAHs in the 'spring or summer' compared to the autumn were observed in urine, but among the airborne PAHs this seasonal effect was only observed for NAP[35]. There are several possible reasons for the discrepancies between these earlier investigations of airborne PAHs and our study of urinary PAHs. First, we matched urinary PAHs to exposure variables on the previous day based on results from Sobus *et al.*, who reported levels of urinary NAP and PHE in workers exposed to asphalt emissions outdoors.[18] However, the patterns and duration of fuel usage and the resulting indoor exposures in our study could have differed from those of the asphalt workers, particularly considering the significant effects of home size and season on levels of urinary PAHs. Second, multiple routes of exposure, including inhalation, diet and dermal contact can contribute to urinary PAH levels, but we only considered airborne exposures in these analyses. Third, due to analytical limitations we were not able to detect urinary BAP and thus could not compare urinary levels to those of airborne BAP, which was a sentinel PAH in the earlier studies.

Interestingly, mixed effects models predicted that aging leads to increased urinary PAH concentrations. Upon repeating the models without adjusting for urinary creatinine, the positive effect of age on urinary PAH levels persisted (data not shown), indicating that the result was not due to the age-associated decline in urinary creatinine. This suggests that aging reduces capacity to metabolize PAHs, which would be consistent with previous investigations of benzene exposure.[49, 50] On the other hand McClean *et al.* reported a positive effect of age on urinary 1-hydroxypyrene (a metabolite of PYR) in workers exposed to asphalt emissions.[51]

In summary, our study showed that unmetabolized urinary PAHs have some utility as short-term biomarkers of household air pollution in residences employing indoor combustion of fossil fuels. Of the ten PAHs measured in subjects exposed to varied combustion sources, urinary OMN was the most discriminating biomarker. We also found that urinary PAH levels in Xuanwei and Fuyuan solid fuel users were linked to the corresponding airborne PAH levels in the same subjects. Finally, we observed that concentrations of urinary PAHs were elevated in subjects using smoky coal and unventilated stoves, both of which are risk factors for lung cancer.

Chapter 2 Figures

Figure 2.1. Log-scale scatter plots of urinary PAH levels versus airborne PAH levels. Open circles indicate outlier data points identified from the mixed effects regression (one in A, one in C and two in H) excluded from the Pearson correlations (rho) calculation. ACE, acenaphthene; ACY, acenaphthylene; BAA+CHR, benzo(*a*)anthracene plus chrysene; FLE, fluoranthene; NAP, naphthalene; OMN, 1-methylnaphthalene; PHE, phenanthrene; PHE+ANT, phenanthrene plus anthracene; PYR, pyrene; TMN, 2-methylnaphthalene.



Chapter 2 Tables

Table 2.1. Summary statistics of urinary PAHs (ng/L).

Polycyclic aromatic hydrocarbon (PAH)	Rings	N (k)	Minimum	25th percentile	Median	75th percentile	Maximum	CV (%)
Naphthalene (NAP)	2	198 (163)	30.8	94.0	139	242	21500	74.0
2-Methylnaphthalene (TMN)	2	178 (152)	0.980	11.9	21.8	40.4	780	95.1
1-Methylnaphthalene (OMN)	2	195 (162)	1.24	6.88	12.9	23.4	387	89.1
Acenaphthene (ACE)	3	198 (163)	0.832	1.74	2.55	4.85	72.5	82.6
Phenanthrene + Anthracene (PHE+ANT)	3	198 (163)	6.90	54.2	152	381	7440	199
Fluoranthene (FLE)	4	198 (163)	0.320	4.93	12.2	25.4	452	138
Pyrene (PYR)	4	172 (144)	0.0727	0.333	0.459	0.794	5.53	67.4
Benzo(*a*)anthracene + Chrysene (BAA+CHR)	4	195 (160)	0.0186	0.203	0.378	0.698	34.9	192

N, number of urine samples (maximum 198) in which the analyte was detected
k, number of subjects (maximum 163) represented by N
Analyte level for a replicated sample is represented by the geometric mean across replicates.
CV, sample preparation coefficient of variation calculated from ANOVA of analyte levels of 24 urine samples prepared in duplicate or triplicate.

Table 2.2. Pairwise Pearson correlations between log-scale levels of urinary PAHs (ng/L).

	NAP	TMN	OMN	ACE	PHE+ANT	FLE	PYR
TMN	0.863						
OMN	0.810	0.870					
ACE	0.424	0.409	0.360				
PHE+ANT	0.739	0.803	0.772	0.390			
FLE	0.682	0.728	0.738	0.377	0.903		
PYR	0.595	0.610	0.635	0.308	0.741	0.784	
BAA+CHR	0.619	0.662	0.632	0.240	0.750	0.726	0.615

P-value < 0.001 for all correlations.

ACE, acenaphthene; BAA+CHR benzo(*a*)anthracene plus chrysene; FLE, fluoranthene; NAP, naphthalene; OMN, 1-methylnaphthalene; PHE+ANT, phenanthrene plus anthracene; PYR, pyrene; TMN, 2-methylnaphthalene.

Table 2.3. Results of linear mixed effects models of log-scale levels of urinary PAHs as a function of airborne PAH measurements.

		Dependent variable, ln (urinary PAH, ng/L)							
		NAP	TMN	OMN	ACE	PHE+ANT	FLE	PYR	BAA+CHR
intercept	estimate	4.17	1.63	0.748	0.498	4.11	1.19	-1.44	-1.11
	(*P*-value)	(<0.0005)	(<0.0005)	(<0.0005)	(0.131)	(0.414)	(<0.0005)	(<0.0005)	(<0.0005)
ln (airborne	estimate,	0.128,	0.212,	0.180,	0.0477,	0.199,	0.113,	0.0827,	0.110,
PAH, ng/m3)	PAH	NAP	NAP	NAP	ACY	PHE	FLE	PYR	BAA+CHR
	GMC	1.14	1.24	1.20	1.05	1.22	1.12	1.09	1.12
	(*P*-value)	(0.015)	(<0.0005)	(<0.0005)	(0.542)	(0.024)	(0.003)	(0.001)	(0.029)
ln (creatinine, g/L)	estimate	0.808	1.26	0.973	0.488	1.47	1.45	0.735	0.977
	GMC	2.24	3.53	2.65	1.63	4.36	4.27	2.09	2.66
	(*P*-value)	(<0.0005)	(<0.0005)	(<0.0005)	(0.001)	(<0.0005)	(<0.0005)	(<0.0005)	(<0.0005)
age (years)	estimate	N.S.	N.S.	0.0104	0.0111	N.S.	0.0226	0.0120	N.S.
	GMR			1.01	1.01		1.02	1.01	
	(*P*-value)			(0.025)	(0.078)		(<0.0005)	(0.001)	

GMC, geometric mean change, or increase in urinary PAH (in natural scale) per unit increase in predictor (in natural scale)

GMR, geometric mean ratio, or fold change in urinary PAH (in natural scale) per unit increase in predictor (in natural scale)

ACY, airborne acenaphthylene levels were used since majority of the air levels of ACE were below limit of detection[35]

N.S., not significant, *i.e.* age has a *P*-value > 0.1 in the full model and was excluded from the final model.

Table 2.4. Results of linear mixed effects models of log-scale levels of urinary PAHs as a function of exposure characteristics.

	NAP		TMN		OMN		ACE		PHE+ANT		FLE		PYR		BAA+CHR	
intercept	**4.78*** **		**2.17*** **		**1.41*** **		0.471		**4.15*** **		**1.12***		**-1.52*** **		**-1.20***	
	estimate	GMC	estimate	GMC	estimate	GMC	estimate	GMC	estimate	GMC	estimate	GMC	estimate	GMC	estimate	GMC
ln (creatinine, g/L)	**0.829*** **	**2.29**	**1.11*** **	**3.02**	**0.953*** **	**2.59**	**0.520*** **	**1.68**	**1.38*** **	**3.99**	**1.46*** **	**4.3**	**0.762*** **	**2.14**	**1.01*** **	**2.75**
	estimate	GMR	estimate	GMR	estimate	GMR	estimate	GMR	estimate	GMR	estimate	GMR	estimate	GMR	estimate	GMR
age (years)	0.00414	1.00	0.00813	1.01	**0.0126** **	**1.01**	**0.0122***	**1.01**	**0.0137***	**1.01**	**0.0232*** **	**1.02**	**0.0137*** **	**1.01**	0.00801	1.01
Fuel Type																
smokeless coal (FY and XW)	ref	1	ref	1	ref	1	ref	1	ref	1	ref	1	ref	1	ref	1
coking coal from XW	**0.558***	**1.75**	**0.929** **	**2.53**	**1.07*** **	**2.93**	0.547	1.73	0.368	1.44	0.324	1.38	0.241	1.27	0.526	1.69
coking coal from FY	**0.587***	**1.8**	**0.910***	**2.48**	**1.33*** **	**3.76**	0.0603	1.06	**1.03***	**2.8**	0.814	2.26	0.386	1.47	**0.967***	**2.63**
1/3 coking FY	0.383	1.47	0.378	1.46	**0.771***	**2.16**	-0.164	0.849	0.838	2.31	0.837	2.31	0.435	1.54	**0.883***	**2.42**
gas fat coal FY	0.358	1.43	0.487	1.63	0.529	1.7	0.0316	1.03	0.125	1.13	0.207	1.23	0.117	1.12	-0.0877	0.916
meager lean	-0.29	0.748	-0.0207	0.98	0.307	1.36	-0.435	0.647	-0.00766	0.992	-0.215	0.807	-0.228	0.796	-0.0366	0.964
multiple coal types	0.354	1.43	**0.769***	**2.16**	**0.815***	**2.26**	0.227	1.25	0.455	1.58	0.415	1.51	**0.506***	**1.66**	**0.739***	**2.09**
multiple fuel types	0.411	1.51	0.541	1.72	**0.813** **	**2.25**	0.0953	1.1	0.117	1.12	0.355	1.43	0.0197	1.02	0.384	1.47
smoky coal of uncertain source	-0.202	0.817	-0.232	0.793	0.543	1.72	-0.928	0.395	-0.245	0.783	0.114	1.12	-0.488	0.614	0.152	1.16
plant products	**0.801***	**2.23**	0.275	1.32	**1.04***	**2.83**	0.179	1.2	0.535	1.71	0.887	2.43	0.502	1.65	0.526	1.69
wood	0.529	1.7	0.205	1.23	0.723	2.06	0.0634	1.07	0.266	1.3	0.703	2.02	0.258	1.29	0.0228	1.02
Stove Design																
ventilated stove	ref	1	ref	1	ref	1	ref	1	ref	1	ref	1	ref	1	ref	1
unventilated stove	0.271	1.31	**0.653***	**1.92**	0.444	1.56	0.00814	1.01	0.183	1.2	0.286	1.33	0.202	1.22	0.139	1.15
fire pit	-0.0238	0.976	0.404	1.5	0.142	1.15	-0.197	0.821	0.0416	1.04	0.344	1.41	0.124	1.13	0.309	1.36
portable stove	-0.0975	0.907	0.137	1.15	0.062	1.06	-0.118	0.889	-0.0205	0.98	-0.167	0.846	-0.249	0.78	-0.151	0.86
mixed ventilation	0.0569	1.06	0.0876	1.09	0.151	1.16	-0.0811	0.922	0.113	1.12	0.334	1.4	0.267	1.31	-0.195	0.822
unknown ventilation	-0.318	0.727	-0.0472	0.954	-0.146	0.864	-0.495	0.609	-0.524	0.592	-0.33	0.719	-0.18	0.835	-0.119	0.888
Room Size																
< 40 m^3	ref	1	ref	1	ref	1	ref	1	ref	1	ref	1	ref	1	ref	1
40 m^3 to 49 m^3	-0.15	0.861	-0.186	0.83	-0.341	0.711	0.0293	1.03	-0.0114	0.989	0.114	1.12	0.227	1.25	-0.151	0.86
50 m^3 to 67 m^3	-0.355	0.702	-0.257	0.774	**-0.445***	**0.641**	-0.0678	0.934	-0.151	0.86	-0.207	0.813	-0.13	0.878	-0.429	0.651
> 67 m^3	-0.222	0.801	-0.0829	0.92	-0.219	0.803	-0.155	0.857	-0.177	0.838	-0.227	0.797	-0.118	0.889	-0.204	0.816
unknown	-0.301	0.74	-0.113	0.893	-0.398	0.671	0.00983	1.01	-0.0221	0.978	-0.268	0.765	0.0639	1.07	-0.14	0.869

Table 2.4. *(continued)*

	NAP		TMN		OMN		ACE		PHE+ANT		FLE		PYR		BAA+CHR	
Season																
autumn	ref	1	ref	1	ref	1	ref	1	ref	1	ref	1	ref	1	ref	1
spring or summer	0.124	1.13	0.268	1.31	0.239	1.27	**0.446***	**1.56**	0.461	1.59	0.284	1.33	0.00294	1	0.0783	1.08
winter	0.0234	1.02	0.128	1.14	0.0962	1.1	0.0872	1.09	0.133	1.14	-0.105	0.9	-0.0298	0.971	-0.201	0.818

GMC, geometric mean change, or increase in urinary PAH (in natural scale) per unit increase in predictor (in natural scale)
GMR, geometric mean ratio, or fold change in urinary PAH (in natural scale) per unit increase in predictor (in natural scale)
Boldface indicates estimates with *P*-value < 0.05 and corresponding GMCs or GMRs
* 0.005 <= *P*-value < 0.05
** 0.0005 <= *P*-value < 0.005
*** *P*-value < 0.0005

Chapter 3. Profiling the Serum Albumin Cys34 Adductome of Solid Fuel Users in Xuanwei and Fuyuan, China[a]

3.1 Introduction

Lung cancer is the leading cause of cancer mortality worldwide.[52] While most lung cancers can be attributed to cigarette smoking, in East Asia an estimated 61% of female lung cancers are observed in never-smokers,[53] especially those exposed to household air pollution from coal combustion.[54] Domestic fuel combustion has been recognized as a major source of exposure to carcinogens that affects about 3 billion people worldwide.[30] Indeed, Xuanwei and Fuyuan Counties in China, where smoky (bituminous) coal is used for domestic cooking and heating, have the highest lung cancer incidence and mortality in the world.[31, 32]

Since women from Xuanwei and Fuyuan rarely smoke, the high incidence of lung cancer has motivated scrutiny of possible risk factors. Nonsmoking Xuanwei women, who use smoky coal, have a 30-fold greater risk of lung cancer than those who use smokeless (anthracite) coal or wood.[33] Compared to smokeless coal, smoky coal emits significantly more particulate matter (PM), polycyclic aromatic hydrocarbons (PAHs), and silica, all of which are known lung carcinogens.[32, 35, 43, 44] Of these potentially causal exposures in Xuanwei, PAHs have been scrutinized, on the basis of the detection of PAH–DNA adducts,[39] characteristic mutational spectra in lung tumors,[38] and risk modulation by genes involved in PAH metabolism.[36, 37] However, the heterogeneity of emissions of PAHs and other combustion products, even across subtypes of smoky coal, has complicated the analysis of exposure–response relationships.[33, 35, 43, 44]

Many environmental toxicants that emanate from combustion of solid fuels either are reactive electrophiles or are metabolized to such species in the body. Reactive electrophiles can produce DNA mutations and modify functional proteins[55, 56] and can alter the redox proteome.[57] Since reactive electrophiles have short lifetimes, investigators have studied their dispositions in vivo by measuring adducts from reactions with abundant proteins in the blood, mainly hemoglobin and human serum albumin (HSA).[58] HSA is particularly interesting because it contains a nucleophilic hotspot, Cys34, that efficiently scavenges reactive oxygen species (ROS) and other small electrophiles in serum, where it represents about 80% of the antioxidant capacity.[59] Oxidation of Cys34 to the reactive sulfenic acid (Cys34-SOH) can lead to formation of mixed Cys34-disulfides from reactions between Cys34-SOH and circulating low-molecular-weight thiols.[60] These Cys34 disulfides represent potential biomarkers of the redox state of the serum over the 1 month residence time of HSA.[61, 62]

Our laboratory has recently developed an *untargeted* assay for characterizing modifications at the Cys34 locus of HSA that we refer to as "Cys34 adductomics".[63] The scheme focuses on the third largest tryptic peptide of HSA ("T3") with a sequence of ALVLIAFAQYLQQC[34]PFEDHVK and a mass of 2432 Da. Adducts of this hydrophobic peptide

[a]This chapter is based on a publication. Reprinted with permission from Lu SS, Grigoryan H, Edmands WMB, Hu W, Iavarone AT, Hubbard A, Rothman N, Vermeulen R, Lan Q, Rappaport SM. Profiling the Serum Albumin Cys34 Adductome of Solid Fuel Users in Xuanwei and Fuyuan, China. *Environ Sci Technol* **2017,** *51* (1), 46-57. DOI: 10.1021/acs.est.6b03955. Copyright © 2016 American Chemical Society.

are separated by nanoflow liquid chromatography (nLC) and detected by high-resolution mass spectrometry (HRMS). A bioinformatic pipeline is used to locate T3 modifications from tandem MS2 spectra, to annotate modifications based on accurate masses, and to quantitate and normalize peak areas.

Given the constellation of electrophiles generated during combustion of fossil fuels, it is difficult to hypothesize about particular adducts or classes of protein modifications that might be observed in blood from Xuanwei and Fuyuan subjects. Cys34 adductomics offers a data-driven approach for comparing adduct features across populations differentially exposed to combustion effluents and thereby for discovering potential biomarkers of relevance to human health. Discriminating adduct features can be identified and targeted for follow-up studies to investigate effects of exposure and to develop mechanistic understanding. Here, we describe application of our methodology to characterize Cys34 adducts in plasma from 29 healthy nonsmoking women from Xuanwei and Fuyuan, China, who used smoky coal, smokeless coal, or wood and 10 local controls who used electricity/gas. We detected 50 T3-derived peptides in these women and explored relationships between adduct levels and the types of solid fuel as well as personal measurements of airborne PM and a carcinogenic PAH (benzo(*a*)pyrene, BaP). Despite the small sample sizes, we detected several highly significant associations between adduct levels and covariates that should generate hypotheses for follow-up studies.

3.2 Materials and Methods

Reagents

Acetonitrile (LC/MS grade), dimethyl sulfoxide, ethylenediaminetetraacetic acid, triethylammonium bicarbonate buffer (1 M), and trypsin (from porcine pancreas, catalog number T0303) were from Sigma-Aldrich (St. Louis, MO). Formic acid and methanol were from Fisher Scientific (Optima LC/MS, Fair Lawn, NJ). Water (18.2 mΩ cm resistivity at 25 °C) was purified by a PureLab Classic system (ELGA LabWater, Woodridge, IL). Isotopically labeled T3 peptide (iT3) with sequence AL-[^{15}N,^{13}C-Val]-LIAFAQYLQQCPFEDH-[^{15}N,^{13}C-Val]-K was custom-made (>95%, BioMer Technology, Pleasanton, CA). The carbamidomethylated iT3 peptide (IAA-iT3) was used as an internal standard for monitoring mass and retention time (RT) stabilities and was prepared as reported previously.[64]

Plasma Samples and Air Measurements

Plasma samples were obtained with informed consent from subjects in China under protocols approved by the National Cancer Institute and local Chinese institutions. Plasma from 29 nonsmoking female subjects using smoky coal, smokeless coal, or wood (hereafter, "exposed subjects") was collected in 2008 and 2009 as part of a cross-sectional study in Xuanwei and Fuyuan counties, China. Details of this study, including the demographic characteristics of the subjects, and collection of air and biological samples have been described.[35, 43-46] Blinded duplicate aliquots from four exposed subjects were also included to assess sample-processing variability and quality assurance, resulting in a total of 33 plasma samples from exposed subjects. Archived plasma from 10 nonsmoking female electricity or gas users (hereafter, "control subjects") were collected between 2007 and 2010 in nearby hospitals in Fuyuan, Qujing, and Xuanwei Counties. Control subjects were being treated for conditions unrelated to tobacco smoking, smoky coal use, and lung disease and had the same age distribution as the exposed subjects. Plasma samples were stored at −80 °C for 4–8 years before analysis.

Personal $PM_{2.5}$ samples (*i.e.*, PM with aerodynamic diameter less than 2.5 μm) were collected on Teflon filters for all exposed subjects in the 24 h period prior to the blood draw.[35, 44] Nineteen of these samples were extracted with dichloromethane and analyzed for BaP by gas chromatography–mass spectrometry.[35] Exposed subjects for whom BaP exposures were not measured were imputed the median BaP levels estimated among other subjects with the same type of fuel. Three categories of BaP and $PM_{2.5}$ exposures were established as follows: controls, low exposure (below the median value), or high exposure (at or above the median value) (median values: BaP = 36.7 ng/m^3, $PM_{2.5}$ = 145 μg/m^3). Personal BaP and $PM_{2.5}$ levels for the 10 control subjects were imputed the minimum values observed in any exposed subject divided by √2.

Sample Processing

The 43 plasma samples were processed in four random batches of 10 or 11 samples. Samples were analyzed as previously described.[63] Briefly, 5 μL of plasma was mixed with 60 μL of 50% methanol for 15 min and centrifuged. Fifty microliters of the supernatant was mixed with 200 μL of digestion buffer (50 mM triethylammonium bicarbonate, 1 mM ethylenediaminetetraacetic acid, pH 8.0) and stored at −80 °C prior to digestion. One hundred thirty-eight microliters of the solution was transferred to a MicroTube (MT-96, Pressure Biosciences Inc., South Easton, MA) to which 2 μL of 10 μg/μL trypsin was added (~1:10 ratio of trypsin/protein, w/w). The tube was capped (MC150-96, Pressure Biosciences Inc.), vortexed briefly, and placed in a pressurized system (Barocycler NEP2320, Pressure Biosciences Inc.) that cycled between 1380 bar (45 s) and ambient pressure (15 s) for 30 min at 37 °C. We had previously shown that pressure cycling of serum/plasma extracts containing 10–20% methanol promoted rapid tryptic digestion, even without prior reduction of disulfide bonds in HSA.[63] After digestion, 3 μL of 10% formic acid was added to stop digestion, and the digest was briefly vortexed and centrifuged to remove particles. Twenty microliters of the digest and 1 μL of a 20 pmol/μL solution of internal standard (IAA-iT3) were transferred to a silanized autosampler vial containing 79 μL of an aqueous solution of 2% acetonitrile and 0.1% formic acid. The diluted digest was stored at −80 °C and/or queued at 4 °C for up to 36 h prior to analysis by nLC-HRMS.

Nanoflow Liquid Chromatography−Mass Spectrometry

Digests were analyzed by nLC-HRMS with an LTQ Orbitrap XL Hybrid mass spectrometer coupled to a Dionex UltiMate 3000 nLC system via a Flex Ion nanoelectrospray ionization source (Thermo Fisher Scientific, Waltham, MA), operated in positive ion mode, as described previously.[63] Briefly, duplicate 1 μL portions of each diluted digest were injected into the nLC and separated on a Dionex PepSwift monolithic column (100 μm i.d. × 25 cm) (Thermo Scientific, Sunnyvale, CA, USA). Full scan MS spectra (*m/z* 350–1200) were acquired with a resolution of 60 000 at *m/z* 400 in the Orbitrap. In data-dependent mode, up to six intense triply charged precursor ions from each MS1 scan were fragmented by collision-induced dissociation and tandem mass spectra (MS2) were acquired in the linear ion trap. The column was washed after every pair of duplicate injections with 1 μL of a solution containing 80% acetonitrile, 10% acetic acid, 5% dimethyl sulfoxide, and 5% water.

Locating T3-Related Peptides with MS2 Spectra

As described previously,[63] RAW data files were converted to mzXML format using the ProteoWizard msConvert tool (3.06387, 64-bit)[65] without filters. All MS2 spectra collected in the elution window between 20 and 35 min were screened for putative adducts using in-house software written in R.[48] Briefly, the screening algorithm focused on signature ions from the T3

peptide and required the presence of at least five unmodified b^+-series fragment ions with signal-to-noise ratios >3 (b_3^+–b_6^+ and b_{11}^+–b_{13}^+) plus a set of at least four fragment ions indicative of the prominent y_{14}^{2+} through y_{18}^{2+} ions with relative intensities ≥20% of the base peak. Spectra that passed the screening algorithm were considered to represent T3-related peptides. These T3 peptides were then clustered with each nearest neighbor having a monoisotopic mass (MIM) within 0.003 *m/z* and a RT within 0.4 min. For each group, an isotope distribution consistent with its respective triply charged precursor MIM was verified, and the means of MIMs and RTs were calculated.

Annotation of Putative Adducts

Putative T3 adducts were annotated as described previously.[63] Briefly, the masses added to the thiolate form of the T3 peptide (Cys34-S^-) were calculated and plausible elemental compositions were probed or confirmed using ChemCalc Molecular Formula finder,[66] Molecular Weight Calculator (version 6.50, https://omics.pnl.gov/software/molecular-weight-calculator/), UNIMOD (http://www.unimod.org/) and MetFrag.[67] Mass accuracy of the assigned elemental composition was assessed in terms of the difference (<3 ppm) between theoretical and observed MIMs. A modification at Cys34 is indicated by MS2 spectra displaying unmodified y_7^+ or y_7^{2+} (*i.e.*, from Pro35 to the C-terminus) plus mass-shifted b_{14}^+ (*i.e.*, from the N-terminus to Cys34), y_8^+, or y_8^{2+} (*i.e.*, from Cys34 to the C-terminus). Conversely, the presence of unmodified b_{14}^+, y_8^+, and y_8^{2+} indicates that modification(s) were not at Cys34.[63] Adducts lacking unambiguous diagnostic ions were annotated as having unclear modification sites.

Quantitation of T3-Related Peptides

Automated peak integration of T3-related peptides was performed using Processing Methods in Xcalibur software (version 2.0.7 SP1, Thermo Fisher Scientific, Inc., Waltham, MA) based on MIMs and RTs with 5 ppm mass accuracy using the Genesis algorithm without smoothing and with >3 signal-to-noise ratio. Each low-abundance peak of a putative T3 peptide was verified by comparing the observed isotopic pattern against the expected pattern. To quantitate and adjust adduct levels for the amounts of HSA in individual digests, peak areas were divided by the corresponding peak areas of a "housekeeping peptide" (HK), with sequence LVNEVTEFA, that appears as a doubly charged peptide (MIM = 575.31113 *m/z*; average RT = 13.5 min). The peak area ratio (PAR, adduct peak area/HK peak area) was previously shown to be a robust linear predictor of adduct concentrations over at least a 500-fold range (0.01–5 μM).[63] Approximate adduct concentrations with units of pmol adduct/mg HSA were estimated as previously described.[63]

Batch Adjustment

Peak-area ratios were log-transformed and adjusted for batch effects with a mixed-effects model similar to that described previously,[63] using Stata software (Stata Statistical Software: Release 13. College Station, TX). Since data included four blinded sample replicates as well as injection replicates for all samples, the following model was used:

$$\ln\left(\frac{Adduct}{HK\ peptide}\right) = \beta_0 + \beta_{1_i} + \mu_{0_j} + \mu_{1_{jk}} + \varepsilon_{ijkh} \quad (1)$$

where β_0 is the fixed overall mean value of the logged peak-area ratio (intercept), β_{1i} is the fixed effect for the i^{th} batch, μ_{0j} is the random effect for the j^{th} subject, μ_{1jk} is the random effect for the k^{th} replicate sample from the j^{th} subject (duplicate samples from four subjects), and ε_{ijkh} is the residual error for the h^{th} injection for a given sample (duplicate injections for all subjects).

Restricted maximum likelihood (REML) estimation was used to fit the models. Coefficients of variation (CVs), representing sample replicates and duplicate injections, were estimated as $\sqrt{e^{(\widehat{\sigma_P}^2-1)}}$ or $\sqrt{e^{(\widehat{\sigma_M}^2-1)}}$, respectively, where $\widehat{\sigma_P}^2$ is the estimated variance component for replicate samples and $\widehat{\sigma_M}^2$ is the estimated variance component for replicate injections. Intraclass correlation coefficients (*ICC*s) were estimated as $ICC = \frac{\widehat{\sigma_B}^2}{\widehat{\sigma_B}^2+\widehat{\sigma_P}^2+\widehat{\sigma_M}^2}$, where $\widehat{\sigma_B}^2$ is the estimated between-subject variance component. Adducts whose models failed to fit (n = 2) or with *ICC*s < 0.1 (n = 12) were eliminated from statistical testing; however, each of their median levels was estimated across subjects with PARs.

After batch adjustment with model (1), subject-specific PARs were predicted as ln(PAR) = $\beta_0 + \mu_{0j}$ for each adduct[63] and these values were used for statistical tests. When a given adduct was not detected in all replicates from a given subject, the ln(PAR) was imputed a value of *minimum* – ln(√2) where *minimum* is the smallest ln(PAR) of a given adduct observed in any subject. Among 36 adducts with *ICC*s ≥ 0.1, 11 had between 1 and 39 nondetected values (median = 15).

Four sets of structurally related adducts were collapsed into clusters, namely, two peaks of *S*-homocysteine (*S*-hCys), and the respective sodium and/or potassium adducts of unadducted T3, *S*-cysteine (*S*-Cys), and *S*-cysteinylglycine (*S*-CysGly). For each cluster, predicted subject-specific logged adduct levels were exponentiated, summed, and log-transformed. After clustering, 32 adducts and clusters were subjected to statistical analyses.

Statistical Analyses

Statistical analyses were performed with Stata software using predicted logged PARs from model (1) (×10 000 for scaling) for each of the 32 adducts and clusters. Three permutation Kruskal–Wallis tests were performed using the *permute* command of Stata with 100 000 replications, under the null hypotheses that fuel types, BaP categories, or $PM_{2.5}$ categories had the same median adduct levels. Significance levels were corrected for multiple testing at a 5% uncorrected false discovery rate (FDR) using the *simes* option (Benjamini-Hochberg method)[68] of the *multproc* program.[69] For each significant Kruskal–Wallis test, a Wilcoxon rank sum test was performed between all pairs of exposure categories with the *dunntest* package[70] with significance levels corrected at 5% FDR using the *simes* option. Adducts with absolute values of Spearman correlation coefficients (r_S) greater than 0.5 were organized into a network generated with Cytoscape.[71] Multivariable linear regression was used to model each adduct or cluster as a function of the fuel types (as dichotomous variables) plus age and log-transformed levels of BaP as covariates. Exploratory analyses, using backward stepwise elimination, revealed that BMI and $PM_{2.5}$ were weaker predictors than fuel groups, BaP, and age and thus were not included in the multivariable models for power considerations.

3.3 Results

Annotation of Adducts

The adductomics workflow identified 50 distinct T3-related peptides (numbered M1 through M50) as summarized in Table 3.1. Median adduct levels spanned a 19 500-fold range with PARs ranging from 0.19 to 3640, corresponding to approximate adduct concentrations of 0.080 to 1590 pmol/mg HSA. For 40 of the T3 peptides, the observed MIM was within 3 ppm of the theoretical value of a modification having a plausible elemental composition. Previously reported modifications[63] include truncations (M1–M4), a labile adduct (M5), unmodified T3

(M6), the T3 dimer (M7, 6+ charge state), T3 methylation at a site other than Cys34 (M9), and Cys34 oxidation products (M8, M11, and M14). The largest class of modifications consisted of 22 mixed disulfides of Cys34, most of which have been reported,[63] including two isomeric modifications of *S*-hCys (M28 and M29), four Na and K adducts of *S*-Cys and *S*-CysGly (M31, M33, M40, and M41), and two apparent modifications of *S*-hCys or *S*-CysGly (M32 and M38). Other putative adducts that have not been reported previously include: *S*-methylthiolation (M13), a Cys34 adduct of pyruvate or malonate semialdehyde (M16), a variant of the *S*-Cys adduct (M19, possibly $NH_2 \rightarrow OH$, $-H_2O$), a Cys34 adduct of oxindole (M27), and a Cys34 trisulfide, *i.e. S*-*S*-hCys (M36).

Summary Statistics and Global Comparisons

Median PARs (×10 000) and CVs are shown in Table 3.2 for all adducts or clusters. The levels of Cys34 oxidation products for a given subject always followed the order: sulfinic acid (dioxidation, M11) > Cys34-Gln cross-link (mono-oxidation, M8) ≫ sulfonic acid (trioxidation, M14), as previously reported for healthy volunteers.[63] Among 36 adducts with *ICC*s ≥ 0.1 (*ICC*: median = 0.73; range: 0.19–0.98), CVs for replicate injections (CV_M: median = 21%; range: 7.7–70%) tended to be greater than those for replicate samples (CV_P: median = 5.5%; range: 0–78%).

Table 3.2 also shows median adduct levels aggregated by fuel type and categories of BaP and $PM_{2.5}$ exposures, along with results of Kruskal–Wallis tests that investigated global associations for 32 adducts. After multiple testing correction ($\alpha = 0.0078$), five adducts had significant differences across fuel groups, *i.e*., the T3 labile adduct (M5), the *S*-hCys cluster (M28 + M29), a Cys34 adduct with unknown annotation (M30, likely composition: $+C_4H_9O_3S$), *S*-γ-glutamylcysteine (*S*-γ-GluCys, M43), and *S*-glutathione (*S*-GSH, M44). The latter three adducts (M30, M43, and M44) also differed significantly across categories of exposures to both BaP and $PM_{2.5}$ ($\alpha = 0.0047$).

Pairwise Differences between Exposure Categories

The sources of global differences across exposure categories (Table 3.2) were investigated pairwise with Wilcoxon rank sum tests, several of which had *P*-values that remained significant after corrections for multiple testing. Subjects using electric/gas fuel or smoky coal had significantly lower levels of the T3 labile adduct (M5) than those using wood or smokeless coal (Figure 3.1A); those using electric/gas fuel or smoky coal had significantly lower levels of *S*-hCys (M28 + M29) than those using smokeless coal, while those using electric/gas fuel also had significantly lower levels of *S*-hCys than those using wood (Figure 3.1B); those using either type of coal had significantly lower levels of *S*-γ-GluCys (M43) and *S*-GSH (M44) than those using electric/gas fuel (Figure 3.1D), and those using each solid fuel had significantly lower levels of a Cys34 adduct with unknown annotation (M30) and *S*-GSH (M44) than those using electric/gas fuel (Figure 3.1C,E).

Extending pairwise comparisons to subjects classified by exposures to BaP and $PM_{2.5}$, the global differences observed in Table 3.2 for a Cys34 adduct with unknown annotation (M30), *S*-γ-GluCys (M43), and *S*-GSH (M44) reflect significantly higher adduct levels in controls compared to either low- or high-exposed subjects for both BaP and $PM_{2.5}$ (Figure 3.2).

Correlation of Adduct Levels

Figure 3.3 shows a correlation map of the 25 adducts having at least one $|r_S|$ greater than 0.5 with another adduct. Many of the moderate to strong correlations were between structurally or biochemically related adducts. For example, *S*-GSH (M44) was correlated with *S*-γ-GluCys

(M43), which in turn was correlated with *S*-Cys (M24). Unadducted T3 (M6) and the earlier-eluting *S*-hCys disulfide (M28) were strongly correlated with their methylated counterparts (M9 and M32, respectively). Unadducted T3 (M6), *S*-Cys (M24), and *S*-CysGly (M37) were correlated with their potassium adducts (M12, M33, and M41, respectively). Oxidation products (M8, M11, and M14) were very strongly correlated with each other, as well as with Cys34 truncations (M1 and M3). In fact, the Cys34-Gln cross-link (M8) and sulfinic acid (M11) had the strongest overall correlation (r_S = 0.95). Grigoryan *et al.*[64] proposed two pathways of cross-link formation between Cys34 and Gln33, from the Cys34 sulfenic acid (−SOH) or from the sulfinic acid ($-SO_2H$), the latter of which is corroborated by our data.

Multivariable Models

We regressed the log-scale estimates of levels of each of the 32 adducts and clusters with sufficient data on the covariates of fuel type, ln(BaP), and age to identify significant covariate effects and reduce possible confounding (Table 3.3). Models for 9 adducts (M5, M17, M19, M25, M30, M32, M34, M43, and M44) contained at least one significant effect with *a P*-value < 0.05 (18 in all). All of the 12 significant effects for fuel group were negative, implying that, after adjusting for BaP and age, the use of each solid fuel was typically associated with lower adduct levels than those in controls. On the other hand, all four of the significant BaP effects (M5, M19, M34, and M44) were positive, indicating that exposure to BaP increased levels of these adducts after adjusting for fuel type and age. Also, both of the significant effects of age (M17 and M32) were positive, suggesting that levels of these two adducts increased significantly with age. Only two of the 18 covariate effects with *P*-values < 0.05 were significant after FDR adjustment (α = 0.0017), namely, *S*-GSH (M44) and *S*-γ-GluCys (M43), consistent with the univariate analyses. Interestingly, the *S*-hCys cluster (M28 + M29), which had been strongly associated with fuel type in univariate analyses (Table 3.2 and Figure 3.1B), did not detect the same associations after adjustment for BaP and age, both of which were marginally associated with *S*-hCys levels (Table 3.3). Also, the strong effects of fuel type and BaP on levels of the unannotated adduct, M30, (Table 3.2) were greatly reduced in the multivariable model, where only smokeless coal showed evidence of an association (*P*-value = 0.036).

3.4 Discussion

This is the first application of Cys34 adductomics to investigate populations exposed to high levels of combustion products that are known to contribute to lung disease. Indeed, nonsmoking women exposed to indoor emissions from smoky coal have among the highest lung cancer incidence and mortality in the world.[31, 32] Constituents of smoky coal and its emissions have been explored in an attempt to pinpoint those that account for lung cancer risk.[33, 35, 43-46] Here, we integrated untargeted adductomics with external exposure measurements to investigate the influence of fuel type and external exposures on downstream biological processes that are reflected by Cys34 adducts.

The 50 T3-peptides detected in this study of Chinese women are similar to the 43 T3-peptides reported by Grigoryan *et al.*,[63] who applied the same methodology to plasma from healthy smokers and nonsmokers in the U.S. The Venn diagram in Figure 3.5 compares the features reported from these two studies, 31 of which were detected in both. Grigoryan *et al.*[63] reported that cigarette smokers had significantly higher levels of adducts representing Cys34 addition of ethylene oxide and acrylonitrile (two constituents of cigarette smoke) as well as the T3-methylation product, and also had decreased levels of the Cys34 sulfinic acid and Cys34-*S*-

Cys adduct. All of the women in the current study were nonsmokers, and a different set of adducts was identified that distinguished solid fuel users from controls.

The most prominent class of Cys34 adducts detected in our investigation were the Cys34 mixed disulfides (22 of 50 T3 features in Table 3.1) that reflect reactions with low-molecular-weight thiols.[72] The median contributions of the five most abundant Cys34 disulfides are compared in Table 3.4 with those from targeted analysis of the same species in another investigation by Lepedda *et al.*[73] The similar percentages derived from sets of independent data indicate that our Cys34 adductomics pipeline is quantitatively reliable.

Several adducts detected in our study were significantly associated with the fuel type and BaP exposures. The strongest associations between adduct levels and fuel type involved the disulfides *S*-GSH (M44) and *S*-γ-GluCys (M43) (Tables 3.2 and 3.3). Intracellular GSH plays a principle role in the elimination of reactive electrophiles, including ROS, and is depleted under oxidative stress.[74, 75] The *S*-GSH adduct represents the reaction between Cys34 and GSH that can involve the unstable Cys34 sulfenic acid (−SOH) as an intermediate.[60] Thus, the lower levels of *S*-GSH observed in the solid fuel groups relative to controls could reflect depletion of intracellular GSH that is mediated by exposures to reactive electrophiles generated by combustion products from solid fuels.

Using the estimated regression coefficients from multivariable models (Table 3.3), the fold change (control: exposed) for smoky coal = 1/exp(−1.4522) = 4.27 compared to 3.30 for wood and 2.49 for smokeless coal. This indicates that the *S*-GSH adduct was present at lower concentrations in smoky-coal users compared to smokeless-coal and wood-fuel users after adjustment for BaP exposure and age and suggests that smoky coal may be a more potent cause of GSH depletion than either smokeless coal or wood.

Decreased levels of circulating GSH have been observed in various diseases and cancers,[72] and the null genotype of glutathione *S*-transferase M1 was associated with increased lung-cancer risk in Asian populations exposed to indoor combustion of coal,[76] and smoky coal in Xuanwei County.[37] Similar decreases in the *S*-γ-GluCys adduct (Table 3.3) probably reflect the fact that γ-GluCys is a dipeptide precursor for the GSH tripeptide.[77] Membrane-bound γ-glutamyltranspeptidase catabolizes conversion of extracellular GSH to CysGly, stimulating the production of pro-oxidant species, and is upregulated in various cancer cells and by depletion of intracellular GSH.[75, 77-79] It is interesting that the ratio of Cys34-*S*-CysGly (M34) to Cys34-*S*-GSH (M44) was elevated in all exposed groups relative to controls (Figure 3.4), especially for smoky coal which showed a much stronger effect (*P*-value = 0.00042) than for wood-fuel (*P*-value = 0.031) or smokeless coal (*P*-value = 0.034). This suggests that γ-glutamyltranspeptidase activity may have contributed to the decrease in circulating GSH via catabolism to CysGly, especially for subjects using smoky coal.

One adduct that differed substantially across fuel types was the T3 labile adduct (M5) which has been reported previously.[63] We suspect that this labile adduct disaggregates in the nanoelectrospray source because it has an accurate mass and MS2 spectrum identical to those of the unadducted T3 peptide but has a distinct retention time (eluting about 30 s earlier than the T3 peptide). Levels of this labile adduct were significantly lower in users of smoky coal compared to other fuel groups after adjustment for exposure to BaP and age but increased with exposure to BaP after adjusting for fuel type and age (Table 3.3). Although the identity of this adduct has not been ascertained, its levels were correlated with several Cys34 disulfides, particularly M17 (*S*-mercaptoacetic acid) and the two isoforms of *S*-hCys (M28 and M32). However, in our previous

adductomic analysis, we observed that this labile T3 adduct was not affected by tris(2-carboxyethyl)phosphine treatment, suggesting that it is not a Cys34 disulfide.[63]

Further research is required to annotate a number of adducts that were associated with exposure to combustion products, particularly the labile T3 adduct (M5) and M30 (likely composition, $+C_4H_9O_3S$). Also, the relationship between levels of *S*-hCys (M28 + M29) and solid fuel, which was highly significant in univariate analyses (Table 3.2) but not in the multivariable model (Table 3.3), requires additional investigation.

In summary, our study detected a host of HSA adducts in plasma from 39 nonsmoking Chinese women. Several of these adducts were significantly influenced by solid fuel use and pollutant exposures, particularly *S*-GSH and *S*-γ-GluCys, which were both present at lower levels in subjects using solid fuels than in controls (Tables 3.2 and 3.3). We realize that this study is small and will require validation with larger samples sizes. Another limitation is the lack of measurements of $PM_{2.5}$ and BaP exposures among control subjects, although it is reasonable to expect that nonsmoking controls who used electricity/gas had lower exposures to $PM_{2.5}$ and BaP than the solid fuel users.

Chapter 3 Figures

Figure 3.1. Pairwise comparisons of adducts showing significant global differences across fuel groups by Kruskal–Wallis tests (Table 3.2): **(A)** the labile T3 adduct (M5), **(B)** the *S*-hCys cluster, **(C)** a Cys34 adduct with unknown annotation (M30), **(D)** *S*-γ-GluCys (M43) and **(E)** *S*-GSH (M44). *P*-values indicate significant Wilcoxon rank sum tests after correction for multiple testing.

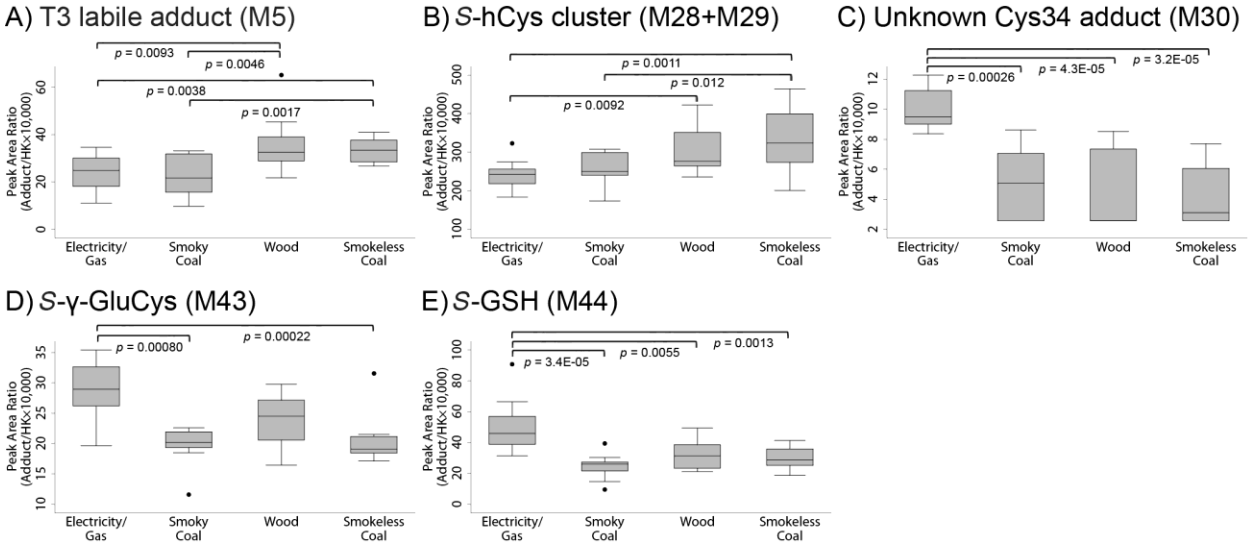


Figure 3.2. Pairwise comparisons of the following adducts showing significant global differences across exposure categories by Kruskal-Wallis tests (Table 3.2): **(A&D)** Cys34 adduct with unknown annotation (M30), **(B&E)** *S*-γ-GluCys (M43) and **(C&F)** *S*-GSH (M44). *P*-values indicate significant Wilcoxon rank sum tests after multiple testing correction.

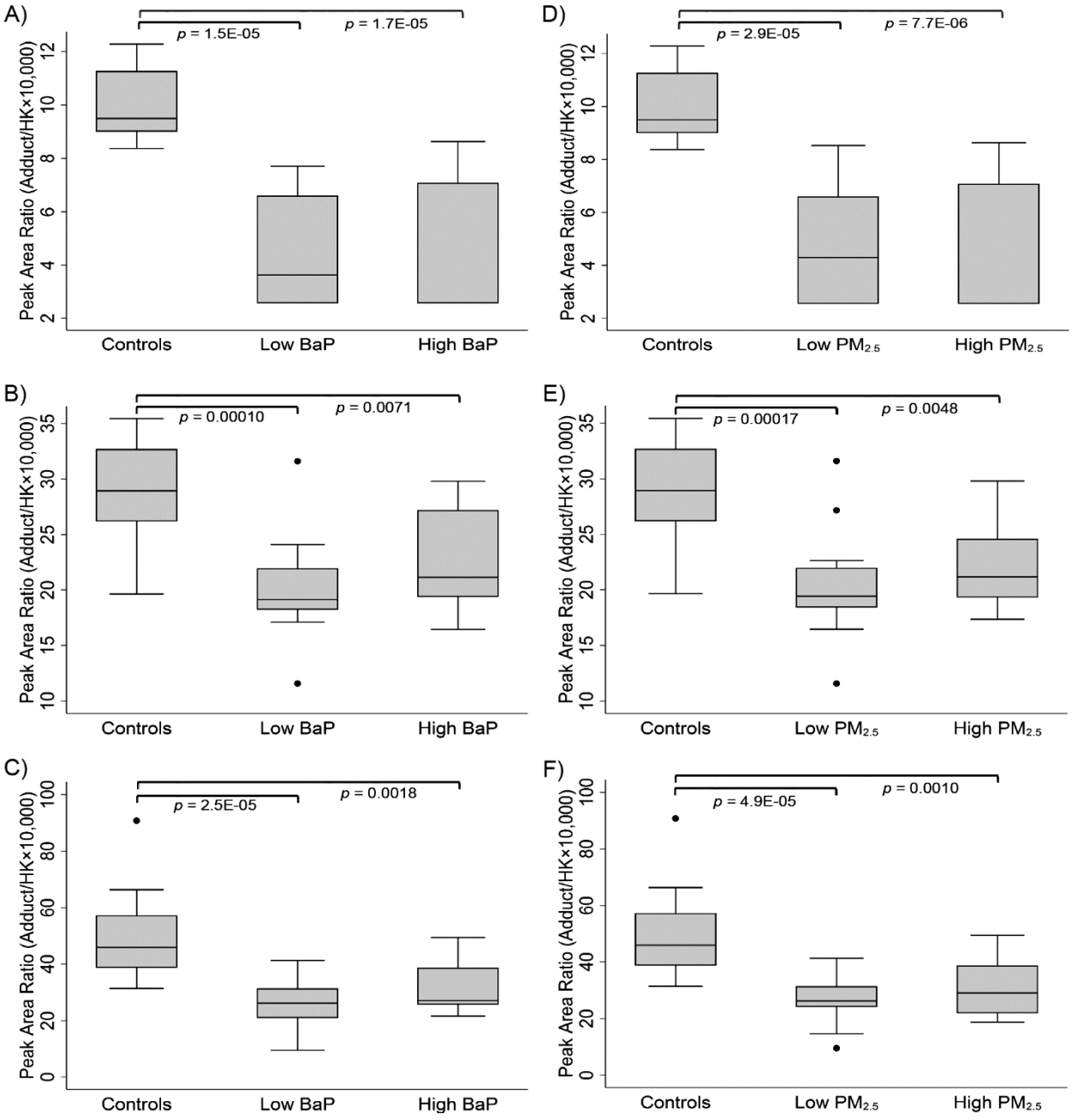


Figure 3.3. Map displaying adducts with moderate to very strong Spearman correlations ($|r_S| > 0.5$). Each adduct is represented by a circle, whose area is linearly related to the median logged adduct level. Each correlation is represented by a line, whose darkness corresponds to the strength of correlation.

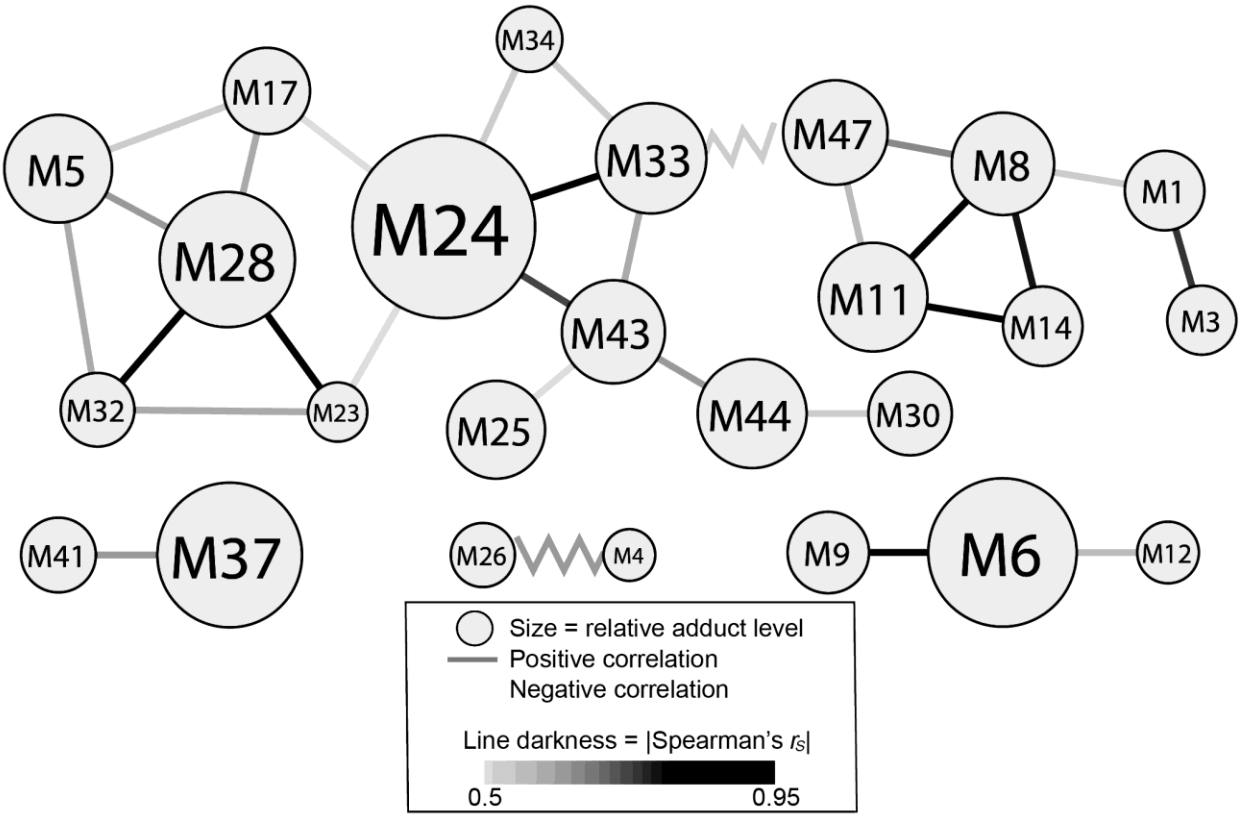


Figure 3.4. Pairwise comparisons of ratios of *S*-CysGly (M37+M40+M41) to *S*-GSH (M44) across fuel groups (Kruskal–Wallis, *P*-value = 0.0107).

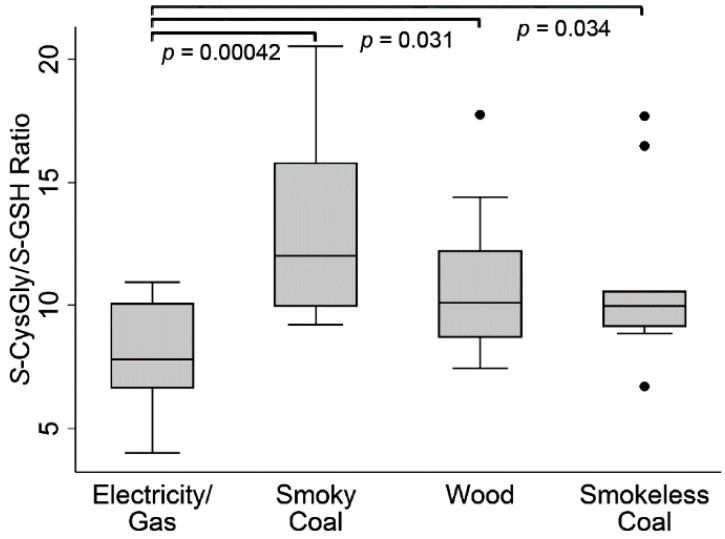
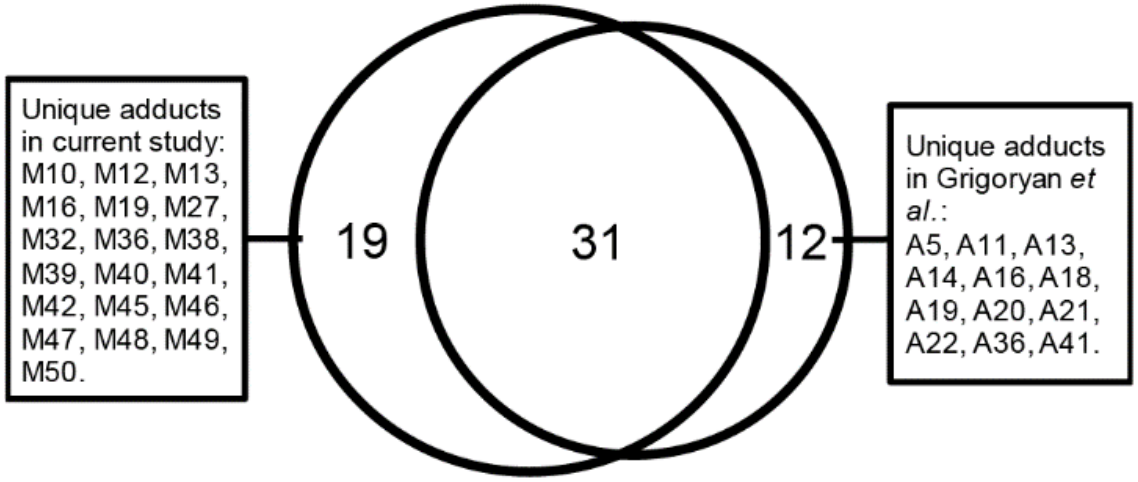


Figure 3.5. Venn diagram comparing the numbers of common and unique adducts detected in the current study with those reported by Grigoryan *et al.*[63]



Chapter 3 Tables

Table 3.1. Putative T3 peptides in the current study.

Adduct	Retention time (min)	PAR[a] (×10 000)	Concn[b] (pmol/mg HSA)	*m/z*, 3+, observed	Mass (Da) added to T3 (Cys-S$^-$)	Added mass composition	*m/z*, 3+, theoretical	Δmass (ppm)	Putative annotation
M1	27.79	4.56	2.00	796.43091	−45.98870	$-CH_2S$	796.43009	−1.03	Cys34 → Gly
M2	28.83	0.476	0.208	800.43213	−33.98505	$-H_2S$	800.43009	−2.54	Cys34 → Dehydroalanine
M3	27.40	2.33	1.02	805.76263	−17.99356	$-H_2S$, +O	805.76173	−1.11	Cys34 → Oxoalanine
M4	27.95	0.781	0.341	808.73005	−9.09130				Not Cys34 adduct
M5	27.88	26.7	11.7	811.76011	1.0072		811.75933	−0.95	T3 labile adduct
M6	28.49	399	174	811.76048	1.00831		811.75933	−1.41	Unadducted T3[d]
M7	30.35	26.5[c]	11.6	811.42514	2431.24780	$+C_{114}H_{172}N_{27}O_{30}S$	811.42394	−1.47	T3 dimer[d]
M8	27.30	19.5	8.53	816.42006	13.97875	$-H_2$, +O	816.41909	−1.19	Cys34-Gln crosslink[d]
M9	28.92	5.53	2.42	816.43203	15.02192	$+CH_3$	816.43122	−0.99	Methylation, not Cys34
M10	28.51	2.69	1.18	819.08685	22.98639	M6−H+Na	819.08665	−0.24	Na adduct of M6
M11	27.30	30.4	13.3	822.42354	32.99646	$+HO_2$	822.42261	−1.13	Sulfinic acid[d]
M12	28.51	1.47	0.644	824.41032	38.95681	M6−H+K	824.41130	1.18	K adduct of M6
M13	29.51	1.28	0.558	827.09009	46.99610	$+CH_3S$	827.08858	−1.83	*S*-Methylthiolation
M14	27.63	4.54	1.99	827.75482	48.99030	$+HO_3$	827.75425	−0.69	Sulfonic acid[d]
M15	28.50	1.43	0.624	829.39659	53.91560				Unclear modification site
M16	27.40	4.38	1.91	841.09883	89.02234	$+C_3H_5O_3$	841.09802	−0.97	Pyruvate or malonate semialdehyde
M17	28.25	7.35	3.21	841.75250	90.98332	$+C_2H_3O_2S$	841.75185	−0.76	*S*-Mercaptoacetic acid
M18	27.27	14.9	6.52	845.42505	102.00098	$+C_3H_4NOS$	845.42385	−1.42	*S*-Cys ($-H_2O$)
M19	28.55	0.457	0.200	845.75278	102.98417	$+C_3H_3O_2S$	845.75185	−1.09	*S*-Cys (possibly $NH_2 \rightarrow OH$, $-H_2O$)
M20	29.53	1.98	0.867	847.10752	107.04839	$+C_7H_7O$	847.10662	−1.06	Benzaldehyde
M21	27.29	0.497	0.217	847.76593	109.02363				Cys34 adduct with unknown annotation
M22	28.31	0.454	0.199	849.06903	112.93293	$+HO_3S_2$	849.06896	−0.08	*S*-*S*-Sulfonic acid trisulfide
M23	28.12	1.28	0.559	850.09692	116.01661	$+C_4H_6NOS$	850.09573	−1.41	*S*-hCys ($-H_2O$)
M24	26.36	3640	1590	851.42850	120.01134	$+C_3H_6NO_2S$	851.42737	−1.33	*S*-Cys[d]
M25	27.70	15.1	6.60	851.75712	120.99719	$+C_3H_5O_3S$	851.75537	−2.05	*S*-Cys ($NH_2 \rightarrow OH$)
M26	27.58	1.72	0.753	853.78324	127.07557				Cys34 adduct with unknown annotation
M27	29.02	0.186	0.0815	855.43732	132.03779	$+C_8H_6NO$	855.43837	1.23	Oxindole
M28	26.61	170	74.3	856.10012	134.02621	$+C_4H_8NO_2S$	856.09925	−1.02	*S*-hCys[d]
M29	26.95	104	45.3	856.09983	134.02533	$+C_4H_8NO_2S$	856.09925	−0.68	*S*-hCys[d]
M30	27.70	7.56	3.31	857.09973	137.02503	$+C_4H_9O_3S$	857.09914	−0.69	Cys34 adduct with unknown annotation
M31	26.39	30.9	13.5	858.75401	141.98788	M24−H+Na	858.75468	0.78	Na adduct of M24
M32	27.06	3.30	1.44	860.77178	148.04118	$+C_5H_{10}NO_2S$	860.77113	−0.75	*S*-hCys, plus methylation not at Cys34
M33	26.38	33.6	14.7	864.07707	157.95704	M24−H+K	864.07933	2.62	K adduct of M24
M34	27.05	1.85	0.811	864.43189	159.02152	$+C_5H_7N_2O_2S$	864.43100	−1.03	*S*-CysGly ($-H_2O$)
M35	27.53	1.04	0.457	865.43150	162.02033	$+C_5H_8NO_3S$	865.43089	−0.70	*S*-(*N*-acetyl)Cys
M36	27.01	1.45	0.634	866.75716	165.99731	$+C_4H_8NO_2S_2$	866.75661	−0.63	*S*-*S*-hCys trisulfide
M37	26.10	316	138	870.43565	177.03279	$+C_5H_9N_2O_3S$	870.43452	−1.30	*S*-CysGly[d]
M38	26.32	4.58	2.00	875.10623	191.04454	M37+CH_2	875.10640	0.20	*S*-CysGly, plus methylation not at Cys34
M39	27.66	1.18	0.516	875.42305	191.99498				Cys34 adduct with unknown annotation
M40	26.10	3.12	1.37	877.76149	199.01031	M37−H+Na	877.76184	0.40	Na adduct of M37
M41	26.14	3.55	1.55	883.08462	214.97971	M37−H+K	883.08648	2.11	K adduct of M37
M42	24.99	1.37	0.600	894.12694	248.10667				Cys34 adduct with unknown annotation
M43	26.67	22.6	9.89	894.44219	249.05241	$+C_8H_{13}N_2O_5S$	894.44156	−0.70	*S*-γ-GluCys
M44	26.55	31.6	13.8	913.44928	306.07367	$+C_{10}H_{16}N_3O_6S$	913.44872	−0.61	*S*-GSH[d]
M45	25.37	5.00	2.19	931.82122	361.18949				Cys34 adduct with unknown annotation
M46	25.30	1.03	0.450	941.15696	389.19672				Unclear modification site
M47	25.40	22.9	10.0	965.49160	462.20064	$+C_{18}H_{32}N_5O_7S$	965.49080	−0.83	Cys34 adduct with unknown annotation
M48	25.33	1.92	0.840	974.50721	489.24747				Cys34 adduct with unknown annotation
M49	26.96	2.82	1.23	976.82030	496.18675				Cys34 adduct with unknown annotation
M50	25.25	1.58	0.692	981.49559	510.21261				Cys34 adduct with unknown annotation

[a]Median peak-area ratio (adduct/housekeeping peptide) in natural scale before imputing nondetects.

[b]Approximate adduct concentration.

[c]6+ charge state, peak areas obtained by extracting the second heavy isotope ion (*m/z* 811.76, more readily detected than MIM).

[d]Annotation confirmed with synthetic standard.

Table 3.2. Median peak-area ratios (adduct/HK × 10 000) in natural scale.[a]

Adduct or cluster	CV_P (%)	CV_M (%)	Global median (n = 39)	Medians for fuel type: Control (n = 10)	Smoky (n = 10)	Wood (n = 9)	Smokeless (n = 10)	KW P-value (95% CI)	Medians for BaP exposure category: Control (n = 10)	Low (n = 15)	High (n = 14)	KW P-value (95% CI)	Medians for $PM_{2.5}$ exposure category: Control (n = 10)	Low (n = 15)	High (n = 14)	KW P-value (95% CI)
M1	0	14.7	4.70	4.66	4.78	5.23	4.33	0.54 (0.54, 0.54)	4.66	4.30	5.04	0.23 (0.23, 0.23)	4.66	4.65	4.92	0.54 (0.54, 0.54)
M2	3×10^{-5}	40.2	0.431	0.441	0.478	0.397	0.438	n/a[c]	0.441	0.420	0.499	n/a[c]	0.441	0.420	0.449	n/a[c]
M3	0	25.7	2.40	2.86	2.37	2.65	2.16	0.29 (0.29, 0.29)	2.86	2.05	2.62	0.043 (0.042, 0.045)	2.86	2.42	2.20	0.29 (0.29, 0.29)
M4	28.4	33.6	0.784	1.38	0.678	0.635	0.789	0.035 (0.034, 0.036)	1.38	0.784	0.631	0.018 (0.017, 0.018)	1.38	0.635	0.748	0.035 (0.034, 0.036)
M5	12.6	16.4	28.9	24.9	21.7	32.5	33.4	**1.1×10^{-3} (9.2×10^{-4}, 1.3×10^{-3})**[b]	24.9	32.4	26.9	0.070 (0.069, 0.072)	24.9	28.9	32.4	0.63 (0.63, 0.63)
M6+M10+M12	n/a	n/a	402	366	425	402	449	0.12 (0.12, 0.12)	366	424	432	0.072 (0.071, 0.074)	366	424	442	0.037 (0.036, 0.038)
M7	37.7	48.9	26.9	27.9	29.6	23.3	29.6	n/a[c]	27.9	27.8	25.8	n/a[c]	27.9	31.0	24.5	n/a[c]
M8	0	14.3	20.5	19.2	20.3	18.2	24.0	0.25 (0.25, 0.25)	19.2	23.2	20.2	0.54 (0.54, 0.55)	19.2	20.2	20.6	0.65 (0.65, 0.65)
M9	0	14.9	5.31	5.11	6.34	5.31	5.66	0.34 (0.33, 0.34)	5.11	5.13	6.34	0.14 (0.14, 0.15)	5.11	5.13	6.49	0.13 (0.13, 0.13)
M11	0	14.0	30.8	29.1	30.5	28.8	35.2	0.28 (0.27, 0.28)	29.1	34.7	30.3	0.47 (0.46, 0.47)	29.1	30.9	32.7	0.67 (0.67, 0.67)
M13	0	70.3	1.27	1.28	1.02	0.858	1.47	0.28 (0.28, 0.28)	1.28	1.01	1.16	0.56 (0.56, 0.57)	1.28	1.01	1.16	0.67 (0.66, 0.67)
M14	20.0	21.7	4.68	4.26	4.49	4.70	4.98	0.29 (0.29, 0.29)	4.26	4.71	4.70	0.48 (0.48, 0.48)	4.26	4.70	4.75	0.54 (0.54, 0.55)
M15	29.4	37.0	1.48	1.40	1.32	1.59	1.77	n/a[c]	1.40	1.69	1.45	n/a[c]	1.40	1.41	1.65	n/a[c]
M16	0	29.6	<0.720	<0.720	<0.720	<0.720	<0.720	0.053 (0.052, 0.055)	<0.720	<0.720	<0.720	0.083 (0.082, 0.085)	<0.720	<0.720	<0.720	0.092 (0.09, 0.094)
M17	17.9	13.5	7.29	6.62	7.22	8.80	7.36	0.17 (0.16, 0.17)	6.62	7.99	8.08	0.14 (0.14, 0.14)	6.62	6.72	8.41	0.11 (0.11, 0.11)
M18	38.0	11.7	15.7	19.1	15.3	18.8	13.8	n/a[c]	19.1	15.7	15.3	n/a[c]	19.1	15.0	18.3	n/a[c]
M19	77.9	41.9	0.316	<0.276	0.341	0.604	<0.276	0.028 (0.027, 0.029)	<0.276	0.313	0.517	0.028 (0.027, 0.029)	<0.276	0.313	0.535	0.024 (0.023, 0.025)
M20	0	43.3	1.13	1.44	1.07	0.942	0.989	0.64 (0.64, 0.64)	1.44	1.13	0.929	0.50 (0.49, 0.50)	1.44	0.761	1.09	0.32 (0.32, 0.33)
M21	0	38.1	<0.310	<0.310	<0.310	<0.310	<0.310	n/a[c]	<0.310	<0.310	<0.310	n/a[c]	<0.310	<0.310	<0.310	n/a[c]
M22	2.39	31.6	<0.386	0.570	<0.386	<0.386	0.475	n/a[c]	0.570	<0.386	<0.386	n/a[c]	0.570	<0.386	<0.386	n/a[c]
M23	0	25.1	1.27	1.10	1.16	1.44	1.71	0.062 (0.061, 0.064)	1.10	1.47	1.25	0.35 (0.35, 0.35)	1.10	1.33	1.36	0.50 (0.50, 0.50)
M24+M31+M33	n/a	n/a	3690	3840	3460	3690	3630	0.37 (0.36, 0.37)	3840	3560	3650	0.29 (0.29, 0.29)	3840	3560	3650	0.25 (0.26, 0.26)
M25	0	15.1	15.4	16.6	12.6	22.9	12.4	0.44 (0.43, 0.44)	16.6	12.5	14.3	0.84 (0.84, 0.84)	16.6	12.3	15.8	0.42 (0.42, 0.43)
M26	5.49	18.5	1.74	1.62	1.90	1.91	1.75	0.29 (0.29, 0.29)	1.62	1.70	2.03	0.051 (0.050, 0.053)	1.62	1.89	1.80	0.21 (0.20, 0.21)
M27	n/a[d]	n/a[d]	<0.164	<0.164	<0.164	<0.164	<0.164	n/a[d]	<0.164	<0.164	<0.164	n/a[d]	<0.164	<0.164	<0.164	n/a[d]
M28+M29	n/a	n/a	274	242	250	276	323	**0.0050 (0.0046, 0.0055)**[b]	242	303	280	0.034 (0.033, 0.035)	242	287	280	0.027 (0.026, 0.028)
M30	49.9	19.2	6.06	9.50	5.05	<3.63	<3.63	**1×10^{-5} (2.5×10^{-7}, 5.6×10^{-5})**[b]	9.50	3.63	<3.63	**0 (0, 3.7×10^{-5})**[b]	9.50	4.30	<3.63	**0 (0, 3.7×10^{-5})**[b]
M32	22.8	20.0	3.22	2.85	3.10	3.64	3.43	0.026 (0.025, 0.026)	2.85	3.42	3.34	0.054 (0.053, 0.056)	2.85	3.31	3.42	0.042 (0.040, 0.043)
M34	46.2	33.0	1.89	1.96	1.84	2.07	1.64	0.050 (0.049, 0.051)	1.96	1.67	2.07	0.099 (0.097, 0.10)	1.96	1.67	2.02	0.31 (0.31, 0.31)
M35	23.0	39.4	1.06	1.12	0.944	1.24	1.07	0.026 (0.025, 0.027)	1.12	1.04	1.05	0.81 (0.81, 0.81)	1.12	1.03	1.12	0.43 (0.43, 0.43)
M36	n/a[d]	n/a[d]	<0.453	<0.453	<0.453	<0.453	<0.453	n/a[d]	<0.453	<0.453	<0.453	n/a[d]	<0.453	<0.453	<0.453	n/a[d]
M37+M40+M41	n/a	n/a	317	346	283	337	302	0.10 (0.10, 0.11)	346	279	323	0.035 (0.034, 0.036)	346	273	323	0.016 (0.016, 0.017)
M38	30.0	35.1	4.46	5.20	4.04	5.17	3.62	n/a[c]	5.20	3.91	5.02	n/a[c]	5.20	3.59	4.67	n/a[c]
M39	0	35.4	0.970	1.26	0.831	0.443	1.01	0.065 (0.063, 0.066)	1.26	0.970	0.662	0.049 (0.048, 0.050)	1.26	0.970	0.509	0.031 (0.030, 0.032)
M42	0	36.3	1.43	1.20	1.37	1.64	1.59	0.82 (0.82, 0.83)	1.20	1.51	1.48	0.74 (0.74, 0.74)	1.20	1.12	1.80	0.017 (0.016, 0.018)
M43	17.9	12.3	21.5	28.9	20.2	24.5	19.1	**5.6×10^{-4} (4.2×10^{-4}, 7.3×10^{-4})**[b]	28.9	19.1	21.1	**3.8×10^{-4} (2.7×10^{-4}, 5.2×10^{-4})**[b]	28.9	19.4	21.1	**6.2×10^{-4} (4.8×10^{-4}, 7.9×10^{-4})**[b]
M44	5.41	13.7	31.3	45.9	25.9	31.3	28.8	**1.5×10^{-4} (8.4×10^{-5}, 2.5×10^{-4})**[b]	45.9	26.3	27.1	**6.0×10^{-5} (2.2×10^{-5}, 1.3×10^{-4})**[b]	45.9	26.3	29.0	**7.0×10^{-5} (2.8×10^{-5}, 1.4×10^{-4})**[b]
M45	0	22.3	5.34	4.79	5.08	5.63	5.09	0.44 (0.43, 0.44)	4.79	5.30	5.43	0.60 (0.60, 0.60)	4.79	4.82	5.44	0.57 (0.56, 0.57)
M46	4×10^{-6}	24.6	0.783	<0.778	<0.778	0.934	0.921	0.25 (0.25, 0.25)	<0.778	0.783	0.898	0.31 (0.31, 0.32)	<0.778	<0.778	0.946	0.13 (0.13, 0.14)
M47	17.8	13.6	23.5	18.7	23.5	21.4	28.0	0.39 (0.38, 0.39)	18.7	25.4	23.4	0.35 (0.35, 0.35)	18.7	24.0	23.7	0.41 (0.41, 0.41)
M48	35.2	32.3	2.05	2.06	2.26	2.30	1.67	0.12 (0.12, 0.12)	2.06	1.74	2.36	0.012 (0.011, 0.013)	2.06	2.02	2.05	0.57 (0.57, 0.58)
M49	49.0	35.5	2.73	3.02	2.61	3.76	2.40	n/a[c]	3.02	2.73	2.65	n/a[c]	3.02	2.59	3.60	n/a[c]
M50	35.6	47.6	0.975	1.12	1.11	1.05	0.863	n/a[c]	1.12	0.875	1.39	n/a[c]	1.12	0.876	1.16	n/a[c]

[a]HK: housekeeping peptide; CV_P: coefficient of variation for replicate samples; CV_M: coefficient of variation for replicate injections; P-values and confidence interval (CI) from permuted Kruskal–Wallis (KW) tests. Left-censored (nondetect) median PARs are represented by "<" minimum (subject-specific PAR estimate).

[b]Values in boldface represent significant P-values after adjustment for multiple testing using the Benjamini-Hochberg method.

[c]Dropped from statistical analyses because the intraclass correlation coefficient (ICC) < 0.1.

[d]Dropped from statistical analyses due to insufficient observations for fitting the mixed-effects model.

Table 3.3. Results of multivariable linear regression models with ln(peak area ratio × 10 000) for each adduct or cluster as the dependent variable.[a]

Adduct or cluster	*intercept*	*smokeless*	*smoky*	*wood*	*ln(BaP)*	*age*	*Adj. R^2 (%)*
Cys34 → Gly, M1	1.7	−0.50 (0.11)	−0.68 (0.14)	−0.56 (0.27)	0.22 (0.11)	−0.0049 (0.32)	3.6
Cys34 → Oxoalanine, M3	1.2	−0.49 (0.12)	−0.76 (0.10)	−0.51 (0.31)	0.17 (0.20)	−0.0062 (0.20)	5.4
Not Cys34 adduct, M4	0.45	−0.093 (0.86)	−0.15 (0.84)	−0.021 (0.98)	−0.20 (0.37)	−0.0028 (0.73)	14
T3 labile adduct, M5	2.7	−0.062 (0.81)	**−0.80 (0.043)**	−0.43 (0.31)	**0.23 (0.045)**	0.0061 (0.13)	38
Unadducted T3 cluster, M6+M10+M12	6.0	−0.044 (0.81)	−0.26 (0.33)	−0.31 (0.29)	0.12 (0.14)	−0.0032 (0.26)	3.9
Cys34-Gln crosslink, M8	2.9	−0.10 (0.72)	−0.57 (0.19)	−0.69 (0.15)	0.21 (0.099)	−0.0031 (0.50)	5.6
Methylation, not Cys34, M9	1.7	−0.26 (0.36)	−0.60 (0.16)	−0.72 (0.12)	0.23 (0.062)	−0.0037 (0.41)	2.4
Sulfinic acid, M11	3.4	−0.095 (0.72)	−0.49 (0.22)	−0.53 (0.22)	0.17 (0.15)	−0.0022 (0.60)	0.073
S-Methylthiolation, M13	0.13	−0.26 (0.59)	−0.71 (0.32)	−0.96 (0.23)	0.17 (0.43)	0.00027 (0.97)	0
Sulfonic acid, M14	1.4	0.070 (0.56)	−0.032 (0.86)	−0.067 (0.73)	0.022 (0.67)	0.00076 (0.69)	0
Pyruvate or malonate semialdehyde, M16	1.8	−0.91 (0.32)	−1.6 (0.24)	−1.4 (0.35)	0.061 (0.88)	−0.021 (0.15)	17
S-Mercaptoacetic acid, M17	1.4	−0.090 (0.70)	−0.18 (0.61)	−0.12 (0.75)	0.091 (0.36)	**0.0074 (0.049)**	11
S-Cys (possibly NH_2 → OH, $-H_2O$), M19	−1.6	**−0.76 (0.045)**	−1.0 (0.071)	−0.68 (0.27)	**0.38 (0.022)**	0.0028 (0.64)	30
Benzaldehyde, M20	−0.24	0.30 (0.75)	1.6 (0.26)	1.3 (0.41)	−0.50 (0.23)	0.020 (0.19)	0
S-hCys ($-H_2O$), M23	−0.25	0.050 (0.89)	−0.71 (0.21)	−0.55 (0.37)	0.20 (0.22)	0.0046 (0.44)	12
S-Cys cluster, M24+M31+M33	8.2	−0.19 (0.074)	−0.24 (0.12)	−0.22 (0.20)	0.041 (0.35)	0.0019 (0.25)	4.2
S-Cys (NH_2 → OH), M25	2.3	−0.58 (0.099)	**−1.1 (0.042)**	−0.92 (0.11)	0.26 (0.085)	0.0060 (0.27)	5.2
Cys34 adduct with unknown annotation, M26	0.37	−0.16 (0.45)	−0.29 (0.34)	−0.26 (0.44)	0.13 (0.14)	−0.000041 (0.99)	2.3
S-hCys cluster, M28+M29	5.2	0.059 (0.70)	−0.29 (0.20)	−0.20 (0.41)	0.11 (0.11)	0.0048 (0.051)	33
Cys34 adduct with unknown annotation, M30	2.7	**−0.80 (0.036)**	−0.63 (0.26)	−0.75 (0.22)	−0.044 (0.78)	−0.0069 (0.24)	42
S-hCys, plus methylation not at Cys34, M32	0.67	0.052 (0.78)	−0.32 (0.24)	−0.24 (0.43)	0.12 (0.13)	**0.0059 (0.045)**	29
S-CysGly ($-H_2O$), M34	0.72	**−0.41 (0.010)**	**−0.50 (0.033)**	−0.33 (0.19)	**0.13 (0.050)**	−0.0025 (0.31)	25
S-(*N*-acetyl)Cys, M35	−0.10	−0.25 (0.34)	−0.62 (0.12)	−0.34 (0.44)	0.14 (0.24)	0.0016 (0.70)	5.5
S-CysGly cluster, M37+M40+M41	6.0	−0.16 (0.35)	−0.20 (0.44)	−0.068 (0.81)	−0.00075 (0.99)	−0.0028 (0.31)	5.2
Cys34 adduct with unknown annotation, M39	1.1	−0.14 (0.81)	−0.070 (0.94)	−0.29 (0.76)	−0.19 (0.45)	−0.0097 (0.30)	17
Cys34 adduct with unknown annotation, M42	−0.50	−0.035 (0.95)	0.21 (0.79)	0.35 (0.69)	−0.096 (0.67)	0.015 (0.073)	0
S-γ-GluCys, M43	3.2	**−0.52 (0.0016)**	**−0.66 (0.0066)**	**−0.51 (0.048)**	0.087 (0.20)	0.0017 (0.49)	38
S-GSH, M44	3.9	**−0.91 (0.0010)**	**−1.5 (0.00056)**	**−1.2 (0.0070)**	**0.22 (0.049)**	−0.0033 (0.41)	44
Cys34 adduct with unknown annotation, M45	1.9	0.13 (0.62)	0.15 (0.70)	0.39 (0.37)	−0.037 (0.75)	−0.0067 (0.12)	0
Unclear modification site, M46	−0.80	0.11 (0.70)	−0.11 (0.78)	−0.010 (0.98)	0.049 (0.68)	0.0075 (0.090)	6.4
Cys34 adduct with unknown annotation, M47	3.0	0.36 (0.31)	0.29 (0.58)	0.20 (0.73)	−0.033 (0.83)	−0.00069 (0.90)	0
Cys34 adduct with unknown annotation, M48	0.61	−0.21 (0.38)	−0.032 (0.93)	0.14 (0.72)	−0.017 (0.87)	0.0022 (0.57)	5.2

[a]Each row shows the modeled adduct (or cluster) and regression coefficients (with *P*-values in parentheses) for each covariate. (The three fuel types, i.e., smoky and smokeless coal and wood, were dichotomous variables with electricity/gas as the reference group, and ln(BaP) (ng/m^3) and age were continuous variables). Values in boldface represent *P*-values <0.05.

Table 3.4. Relative percentages of 5 plasma HSA-Cys34 mixed disulfides detected in the current study and in an independent study by Lepedda *et al.*[73]

Adduct or cluster	Lepedda *et al.* (*n* = 27)	Current study (*n* = 39)
S-Cys cluster, M24+M31+M33	85.84 ± 5.15	84.72 ± 1.90
S-hCys cluster, M28+M29	5.10 ± 1.69	6.49 ± 1.49
S-CysGly cluster, M37+M40+M41	8.01 ± 4.22	7.49 ± 1.39
S-γ-GluCys, M43	0.32 ± 0.15	0.52 ± 0.089
S-GSH, M44	0.73 ± 0.33	0.78 ± 0.31

Chapter 4. Conclusions and Future Directions

4.1 Conclusions

Given the established classification of diesel exhaust (DE) as a Group 1 carcinogen[1] and the challenges in quantifying exposure-response relationships due to the lack of quantitative data[4-6], Chapter 1 investigated whether urinary levels of polycyclic aromatic hydrocarbons (PAHs) increase following short-term, controlled exposure to DE. Previous studies detected unmetabolized urinary PAHs in occupational exposures to emissions containing PAHs[9, 11-18] and postulated that urinary PAHs may reflect the levels of exposure to PAH mixtures in workers exposed to coke-oven emissions, asphalt fumes and DE.[9] Using headspace solid-phase microextraction (HS-SPME) and gas chromatography-mass spectrometry (GC-MS), the levels of two- to five-ring PAHs were measured in urine samples from 28 non-smokers exposed to DE in controlled chamber studies conducted at the US Environmental Protection Agency and Lund University. Analysis using mixed-effects models did not find a significant association between DE exposure and increased urinary PAH levels immediately after nor at 24 hours following exposure, after adjusting for other covariates. The unexpected null result suggests that other unaccounted-for exposure sources such as diet and background environmental PAHs, may have masked the contribution of controlled DE exposure. Another surprising result was that median levels of urinary PAHs, probably reflecting background PAH exposures, and mean levels of chamber airborne PAHs were strongly correlated. The significant variability in baseline urinary PAH concentrations across different populations in these and other[9, 11-14, 18] studies underscores the necessity of having geographically matched controls in biomarker studies. Because urinary PAHs are not promising biomarkers of short-term DE exposure, this finding has limited generalizability to their potential utility in real-world environmental, indoor and occupational exposure settings.

Whereas Chapter 1 highlights the limitations of using urinary PAH levels as biomarkers for short-term DE exposure, Chapter 2 complements Chapter 1 by focusing on levels of urinary PAHs in subjects chronically exposed to emissions from household solid-fuel combustion in geographical regions with some of the highest lung cancer rates in the world. Prior research implicated massive air concentrations of PAHs from indoor combustion of smoky (bituminous) coal with the elevated lung cancer risks in Xuanwei and Fuyuan counties in Yunnan Province of China.[31-33, 35] In Chapter 2, urinary PAH concentrations measured via HS-SPME and GC-MS were reported in 163 nonsmoking females from Xuanwei and Fuyuan who used solid fuels including smoky coal, smokeless coal, wood, or plant material. The study sought to establish the relationship between urinary PAH levels and airborne PAH levels and determinants of exposure such as fuel and stove types. Despite analytical imprecision likely attributable to an effort to detect four-ring PAHs, several salient insights emerged. First, the significant pairwise correlations between urinary PAHs regardless of fuel type were consistent with earlier findings.[9, 16, 18] Second, mixed models uncovered significant positive correlations between matched levels of urinary PAHs and airborne PAHs, indicating that urinary PAHs are candidate biomarkers of exposure to household air pollution from indoor combustion of solid fuels. Third, mixed models showed that among other determinants, smoky coal and unventilated stove use, which are known risk factors for lung cancer,[31-34] were associated with increased urinary PAH levels. Finally, age had a marginal positive effect on levels of some urinary PAHs, warranting further inquiry into the implication of aging on PAH metabolism and clearance. Overall, Chapter 2 validates the

merit of considering various exposure characteristics when assessing the dynamics of PAH exposure.

Chapter 3 takes a novel approach by investigating modifications at the Cys34 locus of human serum albumin (HSA), referred to as "Cys34 adductomics" in 29 non-smoking, female solid-fuel users from Xuanwei and Fuyuan and 10 local control subjects who used electricity or gas. Reactive electrophiles, such as toxicants generated from solid-fuel combustion or their metabolites, can produce DNA mutations and modify functional proteins[55, 56] and can alter the redox proteome.[57] The Cys34 sulfhydryl group is a nucleophilic hotspot on the 'T3' tryptic peptide of HSA that reacts with reactive oxygen species (ROS), accounting for 80% of the antioxidant capacity in serum.[59] The one-month residence time of HSA enables the discovery of Cys34 adducts as potential biomarkers of the serum redox state in longer-term exposures.[61, 62] In Xuanwei and Fuyuan, the heterogeneity of PM2.5, PAHs and other emissions from different fuel types has complicated the analysis of exposure-response relationships.[33, 35, 43, 44] Employing an untargeted assay using nanoflow liquid chromatography (nLC) and high-resolution mass spectrometry (HRMS),[63] this study sought to characterize Cys34 adducts on the T3 peptide of HSA to facilitate the identification of biomarkers across populations with differential exposure to combustion effluents. Fifty T3 adducts were quantified, including 40 modifications annotated with plausible elemental compositions, including 31 that were previously reported.[63] Strong correlations were observed among structurally or biochemically related adducts, hinting at potential common pathways of adduct formation. Significant differences in the levels of adducts were observed across different fuel groups and categories of exposure to benzo(*a*)pyrene (BaP) and PM2.5. Concentrations of several adducts, notably Cys34-*S*-glutathione and Cys34-*S*-γ-glutamylcysteine, were lower in solid-fuel users compared with controls, suggesting that solid fuel use may lead to depletion of the antioxidant glutathione (GSH) mediated by exposures to reactive electrophiles generated by PM2.5 and combustion byproducts. Smoky coal appeared to induce more GSH depletion compared to other solid fuels, as evidenced by lower Cys34-*S*-GSH levels in smoky-coal users compared to smokeless-coal and wood-fuel users, even after adjusting for airborne BaP levels and age using multivariate regression. Moreover, the disparity in ratios of Cys34-*S*-cysteinylglycine to Cys34-*S*-GSH generated the hypothesis that decline in circulating GSH, especially in smoky-coal users, might be facilitated by the catabolism of GSH to cysteinylglycine (CysGly) by γ-glutamyltranspeptidase. This small pilot study adds focus toward the burgeoning landscape of Cys34 adductomics as an agnostic approach for uncovering biomarkers of relevance to human health that has led to subsequent applications to cancer epidemiology.[80-82] It becomes crucial to validate the observed connections between adduct levels, fuel type, exposures, and pertinent biological mechanisms with larger sample sizes.

4.2 Future Directions

Future research should aim to address limitations identified in this dissertation and expand upon its findings. The following avenues can be explored to better characterize exposures to DE and solid fuels:

1. Methodological developments for measuring urinary PAHs:
    a. To address the technical imprecision in measuring urinary unmetabolized PAHs, more sensitive methods can be utilized. For instance, solid-phase microextraction of PAHs may be performed by direct immersion of the SPME fiber in the urine sample (SPME)[15, 41, 83-85] rather than adsorption from the

headspace as described in Chapters 1 and 2 (HS-SPME). The SPME method facilitates the extraction and detection of four- to six-ring PAHs which, in addition to being generally less abundant, are more difficult to volatilize compared to two- to three-ring PAHs. Certain four- to six-ring PAHs are known carcinogens and contain a bay region akin to BaP, making them suitable biomarkers. Incorporating the use of tandem mass spectrometry (GC-MS/MS) can further increase the percentage of urine samples in which PAHs are quantifiable, and with low coefficients of variation.[84, 86]

b. Using enhanced methods, future studies should investigate whether urinary PAHs can serve as biomarkers for long-term ambient DE exposure. A study in a population living near a solid waste incinerator found that vehicular traffic, one of the potential confounding factors, had significant independent associations with some urinary PAHs,[84] indicating that urinary PAHs could reflect ambient DE exposure. Additionally, studies should account for other background sources of PAH exposure, such as diet and smoking, which also influence urinary PAH levels.

c. Researchers should also consider methods for measuring unmetabolized PAHs in other biological matrices, such as blood[87] and serum[88]. Although more invasive, blood sampling avoids the need to account for variability in diuresis, which is necessary with urine sampling via urinary creatinine levels.

2. Validation studies:

a. With improved analytical precision, repeating measurements of PAHs in urine samples from Xuanwei and Fuyuan would confirm correlations between urinary and airborne PAHs and validate the exposure characteristics influencing urinary PAH levels.

b. To corroborate the preliminary findings in Chapter 3, Cys34 adducts should be measured in a larger subset of plasma samples from the case-control, cross-sectional study in Xuanwei and Fuyuan. Future studies on using Cys34 adducts as biomarkers of air pollution can include longitudinal data[89] and involve diverse populations with geographical controls. The development of adductomics methods for dried blood spots[90-93] and cord blood[94] presents opportunities to retrospectively study the whether adducts linked to in-utero exposure to air pollution predict health risks later in life.

c. The identities of new putative adducts should be verified with synthetic standards of modified T3.[63]

d. It is important to annotate unknown adducts associated with exposure. These include a T3 labile adduct likely caused by electrospray dissociation. Intriguingly, the levels of the labile adduct were significantly lower in users of smoky coal compared to other fuel groups after adjustment for exposure to BaP and age, yet increased with exposure to BaP after adjusting for fuel type and age (Chapter 3). This labile adduct was significantly elevated in benzene-exposed workers[80] and in colorectal cancer.[82] Approaches to prevent electrospray dissociation should be explored, including adjusting mobile phase compositions and tuning electrospray ionization parameters.

Subsequently, this relatively abundant labile adduct can be identified by its estimated elution order and concentration.

e. An untargeted adductomics approach favors detecting high-abundance adducts. To enhance the quantitation of low-abundance adducts, specific adducts of interest can be chosen for targeted investigation.[89] Targeted adductomics can also be performed upon compiling a library of known adducts reported in previous studies.[94]

3. Studies on biological mechanisms:
    a. Based on the findings described in Chapter 3 and other articles, research should elucidate the biological mechanism through which exposure to combustion effluents affects the formation of Cys34 adducts. Paradoxical findings between adductomics studies necessitate a nuanced understanding. For example, the Cys34-*S*-GSH adduct was negatively associated with smoky coal (Chapter 3) and with chronic obstructive pulmonary disease and ischemic heart disease,[95] but positively associated with DE in exposed workers.[96] Investigating to what extent the Cys34-*S*-GSH adduct reflects circulating GSH levels or serum oxidation status would be insightful. Also, it would be interesting to test the hypothesis linking the Cys34-*S*-CysGly to Cys34-*S*-GSH ratio with GSH depletion and γ-glutamyltranspeptidase upregulation, especially in groups with elevated air pollution exposure.
    b. Studies measuring both unmetabolized and metabolized urinary PAHs will enable inferences about determinants of metabolism.[18, 84, 86] For instance, it would be possible to verify whether the increased urinary unmetabolized PAHs in aging observed in Chapter 2 can be attributed to a decrease in PAH metabolism and clearance with age.
    c. When sufficient data, analyzing the relationship between adducts and urinary PAHs (instead of airborne BaP as in Chapter 3) in Xuanwei and Fuyuan subjects would provide insight into exposure-response relationships in this region. Specifically, examining whether urinary PAHs, serving as internal exposure indicators, exhibit dose-dependent effects on the levels of particular adducts is warranted. Additionally, testing whether subtypes of smoky coal or other solid fuels from distinct geographic regions affect adduct levels, while adjusting for urinary PAH levels and other covariates, could reveal unique adduct patterns. These differences may hint at or bolster the hypothetical biological mechanisms underlying geographic variations in lung cancer rates.

References

1. Benbrahim-Tallaa L, Baan RA, Grosse Y, Lauby-Secretan B, El Ghissassi F, Bouvard V, Guha N, Loomis D, Straif K. Carcinogenicity of diesel-engine and gasoline-engine exhausts and some nitroarenes. *Lancet Oncol* **2012,** *13* (7), 663-4.
2. Wichmann HE. Diesel Exhaust Particles. *Inhalation Toxicol* **2007,** *19* (s1), 241-4.
3. Sydbom A, Blomberg A, Parnia S, Stenfors N, Sandstrom T, Dahlen SE. Health effects of diesel exhaust emissions. *Eur Respir J* **2001,** *17* (4), 733-46.
4. Pronk A, Coble J, Stewart PA. Occupational exposure to diesel engine exhaust: A literature review. *J Expo Sci Environ Epidemiol* **2009,** *19* (5), 443-57.
5. Laumbach RJ, Kipen HM. Does Diesel Exhaust Cause Lung Cancer (Yet)? *Am J Respir Crit Care Med* **2011,** *183* (7), 843-4.
6. Steenland K, Deddens J, Stayner L. Diesel exhaust and lung cancer in the trucking industry: exposure-response analyses and risk assessment. *Am J Ind Med* **1998,** *34* (3), 220-8.
7. Sobus JR, Pleil JD, Madden MC, Funk WE, Hubbard HF, Rappaport SM. Identification of surrogate measures of diesel exhaust exposure in a controlled chamber study. *Environ Sci Technol* **2008,** *42* (23), 8822-8.
8. IARC Working Group on the Evaluation of Carcinogenic Risks to Humans. Some non-heterocyclic polycyclic aromatic hydrocarbons and some related exposures. *IARC Monogr Eval Carcinog Risks Hum* **2010,** *92*, 1-853.
9. Sobus JR, Waidyanatha S, McClean MD, Herrick RF, Smith TJ, Garshick E, Laden F, Hart JE, Zheng Y, Rappaport SM. Urinary naphthalene and phenanthrene as biomarkers of occupational exposure to polycyclic aromatic hydrocarbons. *Occup Environ Med* **2008,** *66* (2), 99-104.
10. Lin YS. Air samples versus biomarkers for epidemiology. *Occup Environ Med* **2005,** *62* (11), 750-60.
11. 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 Env Hyg* **2007,** *4* (sup1), 100-10.
12. Campo L, Mercadante R, Rossella F, Fustinoni S. Quantification of 13 priority polycyclic aromatic hydrocarbons in human urine by headspace solid-phase microextraction gas chromatography–isotope dilution mass spectrometry. *Anal Chim Acta* **2009,** *631* (2), 196-205.
13. Serdar B, Egeghy PP, Waidyanatha S, Gibson R, Rappaport SM. Urinary Biomarkers of Exposure to Jet Fuel (JP-8). *Environ Health Perspect* **2003,** *111* (14), 1760-4.
14. Waidyanatha S, Zheng Y, Rappaport SM. Determination of polycyclic aromatic hydrocarbons in urine of coke oven workers by headspace solid phase microextraction and gas chromatography–mass spectrometry. *Chem-Biol Interact* **2003,** *145* (2), 165-74.
15. Campo L, Fustinoni S, Bertazzi P. Quantification of carcinogenic 4- to 6-ring polycyclic aromatic hydrocarbons in human urine by solid-phase microextraction gas chromatography–isotope dilution mass spectrometry. *Anal Bioanal Chem* **2011,** *401* (2), 625-34.

16. Campo L, Rossella F, Pavanello S, Mielzynska D, Siwinska E, Kapka L, Bertazzi PA, Fustinoni S. Urinary profiles to assess polycyclic aromatic hydrocarbons exposure in coke-oven workers. *Toxicol Lett* **2010,** *192* (1), 72-8.
17. 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* (8), 509-16.
18. Sobus JR, McClean MD, Herrick RF, Waidyanatha S, Onyemauwa F, Kupper LL, Rappaport SM. Investigation of PAH Biomarkers in the Urine of Workers Exposed to Hot Asphalt. *Ann Occup Hyg* **2009,** *53* (6), 551-60.
19. Wierzbicka A, Nilsson PT, Rissler J, Sallsten G, Xu Y, Pagels JH, Albin M, Österberg K, Strandberg B, Eriksson A, Bohgard M, Bergemalm-Rynell K, Gudmundsson A. Detailed diesel exhaust characteristics including particle surface area and lung deposited dose for better understanding of health effects in human chamber exposure studies. *Atmos Environ* **2014,** *86* (0), 212-9.
20. Hubbard HF, Sobus JR, Pleil JD, Madden MC, Tabucchi S. Application of novel method to measure endogenous VOCs in exhaled breath condensate before and after exposure to diesel exhaust. *J Chromatogr B* **2009,** *877* (29), 3652-8.
21. Pleil JD, Stiegel MA, Madden MC, Sobus JR. Heat map visualization of complex environmental and biomarker measurements. *Chemosphere* **2011,** *84* (5), 716-23.
22. Sällsten G, Gustafson P, Johansson L, Johannesson S, Molnár P, Strandberg B, Tullin C, Barregard L. Experimental Wood Smoke Exposure in Humans. *Inhalation Toxicol* **2006,** *18* (11), 855-64.
23. Heinegard D, Tiderstrom G. Determination of serum creatinine by a direct colorimetric method. *Clin Chim Acta* **1973,** *43* (3), 305-10.
24. Sobus JR, McClean MD, Herrick RF, Waidyanatha S, Nylander-French LA, Kupper LL, Rappaport SM. Comparing Urinary Biomarkers of Airborne and Dermal Exposure to Polycyclic Aromatic Compounds in Asphalt-Exposed Workers. *Ann Occup Hyg* **2009,** *53* (6), 561-71.
25. DeMarini DM. Genotoxicity biomarkers associated with exposure to traffic and near-road atmospheres: a review. *Mutagenesis* **2013,** *28* (5), 485-505.
26. Adonis M. Susceptibility and exposure biomarkers in people exposed to PAHs from diesel exhaust. *Toxicol Lett* **2003,** *144* (1), 3-15.
27. Laumbach R, Tong J, Zhang L, Ohman-Strickland P, Stern A, Fiedler N, Kipen H, Kelly-McNeil K, Lioy P, Zhang J. Quantification of 1-aminopyrene in human urine after a controlled exposure to diesel exhaust. *J Environ Monit* **2009,** *11* (1), 153.
28. Huyck S, Ohman-Strickland P, Zhang L, Tong J, Xu XU, Zhang J. Determining times to maximum urine excretion of 1-aminopyrene after diesel exhaust exposure. *J Expo Sci Environ Epidemiol* **2010,** *20* (7), 650-5.
29. Bamford HA, Bezabeh DZ, Schantz S, Wise SA, Baker JE. Determination and comparison of nitrated-polycyclic aromatic hydrocarbons measured in air and diesel particulate reference materials. *Chemosphere* **2003,** *50* (5), 575-87.
30. Smith KR, Bruce N, Balakrishnan K, Adair-Rohani H, Balmes J, Chafe Z, Dherani M, Hosgood HD, Mehta S, Pope D, Rehfuess E, Group HCRE. Millions dead: how do we know and what does it mean? Methods used in the comparative risk assessment of household air pollution. *Annu Rev Public Health* **2014,** *35*, 185-206.

31. Barone-Adesi F, Chapman RS, Silverman DT, He X, Hu W, Vermeulen R, Ning B, Fraumeni JF, Jr., Rothman N, Lan Q. Risk of lung cancer associated with domestic use of coal in Xuanwei, China: retrospective cohort study. *BMJ* **2012,** *345*, e5414.
32. Mumford JL, He XZ, Chapman RS, Cao SR, Harris DB, Li XM, Xian YL, Jiang WZ, Xu CW, Chuang JC, et al. Lung cancer and indoor air pollution in Xuan Wei, China. *Science* **1987,** *235* (4785), 217-20.
33. Lan Q, He X, Shen M, Tian L, Liu LZ, Lai H, Chen W, Berndt SI, Hosgood HD, Lee KM, Zheng T, Blair A, Chapman RS. Variation in lung cancer risk by smoky coal subtype in Xuanwei, China. *Int J Cancer* **2008,** *123* (9), 2164-9.
34. Lan Q, Chapman RS, Schreinemachers DM, Tian L, He X. Household stove improvement and risk of lung cancer in Xuanwei, China. *J Natl Cancer Inst* **2002,** *94* (11), 826-35.
35. Downward GS, Hu W, Rothman N, Reiss B, Wu G, Wei F, Chapman RS, Portengen L, Qing L, Vermeulen R. Polycyclic aromatic hydrocarbon exposure in household air pollution from solid fuel combustion among the female population of Xuanwei and Fuyuan counties, China. *Environ Sci Technol* **2014,** *48* (24), 14632-41.
36. Lan Q, Mumford JL, Shen M, Demarini DM, Bonner MR, He X, Yeager M, Welch R, Chanock S, Tian L, Chapman RS, Zheng T, Keohavong P, Caporaso N, Rothman N. Oxidative damage-related genes AKR1C3 and OGG1 modulate risks for lung cancer due to exposure to PAH-rich coal combustion emissions. *Carcinogenesis* **2004,** *25* (11), 2177-81.
37. Lan Q, He X, Costa DJ, Tian L, Rothman N, Hu G, Mumford JL. Indoor coal combustion emissions, GSTM1 and GSTT1 genotypes, and lung cancer risk: a case-control study in Xuan Wei, China. *Cancer Epidemiol Biomarkers Prev* **2000,** *9* (6), 605-8.
38. Hosgood HD, 3rd, Pao W, Rothman N, Hu W, Pan YH, Kuchinsky K, Jones KD, Xu J, Vermeulen R, Simko J, Lan Q. Driver mutations among never smoking female lung cancer tissues in China identify unique EGFR and KRAS mutation pattern associated with household coal burning. *Respir Med* **2013,** *107* (11), 1755-62.
39. Mumford JL, Lee X, Lewtas J, Young TL, Santella RM. DNA adducts as biomarkers for assessing exposure to polycyclic aromatic hydrocarbons in tissues from Xuan Wei women with high exposure to coal combustion emissions and high lung cancer mortality. *Environ Health Perspect* **1993,** *99*, 83-7.
40. Mumford JL, Li X, Hu F, Lu XB, Chuang JC. Human exposure and dosimetry of polycyclic aromatic hydrocarbons in urine from Xuan Wei, China with high lung cancer mortality associated with exposure to unvented coal smoke. *Carcinogenesis* **1995,** *16* (12), 3031-6.
41. Campo L, Fustinoni S, Consonni D, Pavanello S, Kapka L, Siwinska E, Mielzynska D, Bertazzi P. Urinary carcinogenic 4-6 ring polycyclic aromatic hydrocarbons in coke oven workers and in subjects belonging to the general population: role of occupational and environmental exposure. *Int J Hyg Environ Health* **2014,** *217* (2-3), 231-8.
42. Campo L, Buratti M, Fustinoni S, Cirla PE, Martinotti I, Longhi O, Cavallo D, Foa V. Evaluation of exposure to PAHs in asphalt workers by environmental and biological monitoring. *Ann N Y Acad Sci* **2006,** *1076*, 405-20.
43. Downward GS, Hu W, Large D, Veld H, Xu J, Reiss B, Wu G, Wei F, Chapman RS, Rothman N, Qing L, Vermeulen R. Heterogeneity in coal composition and implications

for lung cancer risk in Xuanwei and Fuyuan counties, China. *Environ Int* **2014,** *68*, 94-104.
44. Hu W, Downward GS, Reiss B, Xu J, Bassig BA, Hosgood HD, 3rd, Zhang L, Seow WJ, Wu G, Chapman RS, Tian L, Wei F, Vermeulen R, Lan Q. Personal and indoor PM2.5 exposure from burning solid fuels in vented and unvented stoves in a rural region of China with a high incidence of lung cancer. *Environ Sci Technol* **2014,** *48* (15), 8456-64.
45. Hosgood HD, Lan Q, Vermeulen R, Wei H, Reiss B, Coble J, Wei F, Jun X, Wu G, Rothman N. Combustion-derived nanoparticle exposure and household solid fuel use in Xuanwei and Fuyuan, China. *Int J Environ Health Res* **2012,** *22* (6), 571-81.
46. Seow WJ, Hu W, Vermeulen R, Hosgood Iii HD, Downward GS, Chapman RS, He X, Bassig BA, Kim C, Wen C, Rothman N, Lan Q. Household air pollution and lung cancer in China: a review of studies in Xuanwei. *Chin J Cancer* **2014,** *33* (10), 471-5.
47. Curtis SM. *bisoreg: Bayesian Isotonic Regression with Bernstein Polynomials*, 2015; https://cran.r-project.org/src/contrib/Archive/bisoreg/.
48. The R Foundation. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing: Vienna, Austria, 2015.
49. Rappaport SM, Waidyanatha S, Qu Q, Shore R, Jin X, Cohen B, Chen LC, Melikian AA, Li G, Yin S, Yan H, Xu B, Mu R, Li Y, Zhang X, Li K. Albumin adducts of benzene oxide and 1,4-benzoquinone as measures of human benzene metabolism. *Cancer Res* **2002,** *62* (5), 1330-7.
50. Kim S, Vermeulen R, Waidyanatha S, Johnson BA, Lan Q, Rothman N, Smith MT, Zhang L, Li G, Shen M, Yin S, Rappaport SM. Using urinary biomarkers to elucidate dose-related patterns of human benzene metabolism. *Carcinogenesis* **2006,** *27* (4), 772-81.
51. Mcclean MD, Rinehart RD, Ngo L, Eisen EA, Kelsey KT, Wiencke JK, Herrick RF. Urinary 1-hydroxypyrene and polycyclic aromatic hydrocarbon exposure among asphalt paving workers. *Ann Occup Hyg* **2004,** *48* (6), 565-78.
52. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. *CA: Cancer J Clin* **2015,** *65* (2), 87-108.
53. Sun S, Schiller JH, Gazdar AF. Lung cancer in never smokers--a different disease. *Nat Rev Cancer* **2007,** *7* (10), 778-90.
54. Hosgood HD, 3rd, Boffetta P, Greenland S, Lee YC, McLaughlin J, Seow A, Duell EJ, Andrew AS, Zaridze D, Szeszenia-Dabrowska N, Rudnai P, Lissowska J, Fabianova E, Mates D, Bencko V, Foretova L, Janout V, Morgenstern H, Rothman N, Hung RJ, Brennan P, Lan Q. In-home coal and wood use and lung cancer risk: a pooled analysis of the International Lung Cancer Consortium. *Environ Health Perspect* **2010,** *118* (12), 1743-7.
55. Miller EC, Miller JA. Mechanisms of chemical carcinogenesis: nature of proximate carcinogens and interactions with macromolecules. *Pharmacol Rev* **1966,** *18* (1), 805-38.
56. Brodie BB, Reid WD, Cho AK, Sipes G, Krishna G, Gillette JR. Possible mechanism of liver necrosis caused by aromatic organic compounds. *Proc Natl Acad Sci USA* **1971,** *68* (1), 160-4.
57. Go YM, Jones DP. The redox proteome. *J Biol Chem* **2013,** *288* (37), 26512-20.
58. Rubino FM, Pitton M, Di Fabio D, Colombi A. Toward an "omic" physiopathology of reactive chemicals: thirty years of mass spectrometric study of the protein adducts with endogenous and xenobiotic compounds. *Mass Spectrom Rev* **2009,** *28* (5), 725-84.

59. Aldini G, Vistoli G, Regazzoni L, Gamberoni L, Facino RM, Yamaguchi S, Uchida K, Carini M. Albumin is the main nucleophilic target of human plasma: a protective role against pro-atherogenic electrophilic reactive carbonyl species? *Chem Res Toxicol* **2008,** *21* (4), 824-35.
60. Carballal S, Alvarez B, Turell L, Botti H, Freeman BA, Radi R. Sulfenic acid in human serum albumin. *Amino Acids* **2007,** *32* (4), 543-51.
61. Aldini G, Yeum K-J, Vistoli G. Covalent Modifications of Albumin Cys34 as a Biomarker of Mild Oxidative Stress. In *Biomarkers for Antioxidant Defense and Oxidative Damage: Principles and Practical Applications*, Aldini G, Yeum K-J, Niki E, Russell RM, Eds. Wiley-Blackwell: Oxford, UK, 2010; pp 229-41.
62. Nagumo K, Tanaka M, Chuang VT, Setoyama H, Watanabe H, Yamada N, Kubota K, Tanaka M, Matsushita K, Yoshida A, Jinnouchi H, Anraku M, Kadowaki D, Ishima Y, Sasaki Y, Otagiri M, Maruyama T. Cys34-cysteinylated human serum albumin is a sensitive plasma marker in oxidative stress-related chronic diseases. *PLoS One* **2014,** *9* (1), e85216.
63. Grigoryan H, Edmands W, Lu SS, Yano Y, Regazzoni L, Iavarone AT, Williams ER, Rappaport SM. Adductomics Pipeline for Untargeted Analysis of Modifications to Cys34 of Human Serum Albumin. *Anal Chem* **2016,** *88* (21), 10504-12.
64. Grigoryan H, Li H, Iavarone AT, Williams ER, Rappaport SM. Cys34 adducts of reactive oxygen species in human serum albumin. *Chem Res Toxicol* **2012,** *25* (8), 1633-42.
65. Chambers MC, Maclean B, Burke R, Amodei D, Ruderman DL, Neumann S, Gatto L, Fischer B, Pratt B, Egertson J, Hoff K, Kessner D, Tasman N, Shulman N, Frewen B, Baker TA, Brusniak MY, Paulse C, Creasy D, Flashner L, Kani K, Moulding C, Seymour SL, Nuwaysir LM, Lefebvre B, Kuhlmann F, Roark J, Rainer P, Detlev S, Hemenway T, Huhmer A, Langridge J, Connolly B, Chadick T, Holly K, Eckels J, Deutsch EW, Moritz RL, Katz JE, Agus DB, MacCoss M, Tabb DL, Mallick P. A cross-platform toolkit for mass spectrometry and proteomics. *Nat Biotechnol* **2012,** *30* (10), 918-20.
66. Patiny L, Borel A. ChemCalc: a building block for tomorrow's chemical infrastructure. *J Chem Inf Model* **2013,** *53* (5), 1223-8.
67. Wolf S, Schmidt S, Müller-Hannemann M, Neumann S. In silico fragmentation for computer assisted identification of metabolite mass spectra. *BMC Bioinform* **2010,** *11* (1), 1-12.
68. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J Roy Stat Soc Ser B (Stat Method)* **1995,** *57* (1), 289-300.
69. Newson R. *SMILEPLOT: Stata module to create plots for use with multiple significance tests*. Boston College: Chestnut Hill, MA, 2012.
70. Dinno A. *dunntest: Dunn's Test of Multiple Comparisons Using Rank Sums*, 2014; http://www.alexisdinno.com/stata/dunntest.html.
71. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **2003,** *13* (11), 2498-504.
72. Isokawa M, Kanamori T, Funatsu T, Tsunoda M. Analytical methods involving separation techniques for determination of low-molecular-weight biothiols in human plasma and blood. *J Chromatogr B Analyt Technol Biomed Life Sci* **2014,** *964*, 103-15.

73. Lepedda AJ, Zinellu A, Nieddu G, De Muro P, Carru C, Spirito R, Guarino A, Piredda F, Formato M. Human serum albumin Cys34 oxidative modifications following infiltration in the carotid atherosclerotic plaque. *Oxid Med Cell Longev* **2014,** *2014*, 690953.
74. Nel A. Atmosphere. Air pollution-related illness: effects of particles. *Science* **2005,** *308* (5723), 804-6.
75. Pompella A, Visvikis A, Paolicchi A, Tata VD, Casini AF. The changing faces of glutathione, a cellular protagonist. *Biochem Pharmacol* **2003,** *66* (8), 1499-503.
76. Hosgood HD, 3rd, Berndt SI, Lan Q. GST genotypes and lung cancer susceptibility in Asian populations with indoor air pollution exposures: a meta-analysis. *Mutat Res* **2007,** *636* (1-3), 134-43.
77. Lu SC. Glutathione synthesis. *Biochim Biophys Acta* **2013,** *1830* (5), 3143-53.
78. Pompella A, De Tata V, Paolicchi A, Zunino F. Expression of gamma-glutamyltransferase in cancer cells and its significance in drug resistance. *Biochem Pharmacol* **2006,** *71* (3), 231-8.
79. Dominici S, Paolicchi A, Lorenzini E, Maellaro E, Comporti M, Pieri L, Minotti G, Pompella A. Gamma-glutamyltransferase-dependent prooxidant reactions: a factor in multiple processes. *BioFactors* **2003,** *17* (1-4), 187-98.
80. Grigoryan H, Edmands WMB, Lan Q, Carlsson H, Vermeulen R, Zhang L, Yin SN, Li GL, Smith MT, Rothman N, Rappaport SM. Adductomic signatures of benzene exposure provide insights into cancer induction. *Carcinogenesis* **2018,** *39* (5), 661-8.
81. Grigoryan H, Imani P, Sacerdote C, Masala G, Grioni S, Tumino R, Chiodini P, Dudoit S, Vineis P, Rappaport SM. HSA Adductomics Reveals Sex Differences in NHL Incidence and Possible Involvement of Microbial Translocation. *Cancer Epidemiol Biomarkers Prev* **2023,** *32* (9), 1217-26.
82. Grigoryan H, Schiffman C, Gunter MJ, Naccarati A, Polidoro S, Dagnino S, Dudoit S, Vineis P, Rappaport SM. Cys34 Adductomics Links Colorectal Cancer with the Gut Microbiota and Redox Biology. *Cancer Res* **2019,** *79* (23), 6024-31.
83. De Craemer S, Croes K, van Larebeke N, Sioen I, Schoeters G, Loots I, Nawrot T, Nelen V, Campo L, Fustinoni S, Baeyens W. Investigating unmetabolized polycyclic aromatic hydrocarbons in adolescents' urine as biomarkers of environmental exposure. *Chemosphere* **2016,** *155*, 48-56.
84. Gatti MG, Bechtold P, Campo L, Barbieri G, Quattrini G, Ranzi A, Sucato S, Olgiati L, Polledri E, Romolo M, Iacuzio L, Carrozzi G, Lauriola P, Goldoni CA, Fustinoni S. Human biomonitoring of polycyclic aromatic hydrocarbonsand metals in the general population residing near the municipal solid waste incinerator of Modena, Italy. *Chemosphere* **2017,** *186*, 546-57.
85. Ranzi A, Fustinoni S, Erspamer L, Campo L, Gatti MG, Bechtold P, Bonassi S, Trenti T, Goldoni CA, Bertazzi PA, Lauriola P. Biomonitoring of the general population living near a modern solid waste incinerator: a pilot study in Modena, Italy. *Environ Int* **2013,** *61*, 88-97.
86. Campo L, Hanchi M, Olgiati L, Polledri E, Consonni D, Zrafi I, Saidane-Mosbahi D, Fustinoni S. Biological Monitoring of Occupational Exposure to Polycyclic Aromatic Hydrocarbons at an Electric Steel Foundry in Tunisia. *Ann Occup Hyg* **2016,** *60* (6), 700-16.
87. Gruber B, Schneider J, Fohlinger M, Buters J, Zimmermann R, Matuschek G. A minimal-invasive method for systemic bio-monitoring of the environmental pollutant

phenanthrene in humans: Thermal extraction and gas chromatography - mass spectrometry from 1 mL capillary blood. *J Chromatogr A* **2017,** *1487*, 254-7.
88. Zhou S, Zhang L, Guo C, Zhong Y, Luo X, Pan X, Yang Z, Tan L. Comparing liquid-liquid, solid-phase, and supported-liquid extraction for the determination of polycyclic aromatic hydrocarbons in serum samples and their application for human biomonitoring. *Microchemical Journal* **2022,** *181*, 107812.
89. Smith JW, O'Meally RN, Ng DK, Chen JG, Kensler TW, Cole RN, Groopman JD. Biomonitoring of Ambient Outdoor Air Pollutant Exposure in Humans Using Targeted Serum Albumin Adductomics. *Chem Res Toxicol* **2021,** *34* (4), 1183-96.
90. Yano Y, Grigoryan H, Schiffman C, Edmands W, Petrick L, Hall K, Whitehead T, Metayer C, Dudoit S, Rappaport S. Untargeted adductomics of Cys34 modifications to human serum albumin in newborn dried blood spots. *Anal Bioanal Chem* **2019,** *411* (11), 2351-62.
91. Yano Y, Schiffman C, Grigoryan H, Hayes J, Edmands W, Petrick L, Whitehead T, Metayer C, Dudoit S, Rappaport S. Untargeted adductomics of newborn dried blood spots identifies modifications to human serum albumin associated with childhood leukemia. *Leuk Res* **2020,** *88*, 106268.
92. Funk WE, Montgomery N, Bae Y, Chen J, Chow T, Martinez MP, Lurmann F, Eckel SP, McConnell R, Xiang AH. Human Serum Albumin Cys34 Adducts in Newborn Dried Blood Spots: Associations With Air Pollution Exposure During Pregnancy. *Front Public Health* **2021,** *9*, 730369.
93. Petrick LM, Uppal K, Funk WE. Metabolomics and adductomics of newborn bloodspots to retrospectively assess the early-life exposome. *Curr Opin Pediatr* **2020,** *32* (2), 300-7.
94. Lin ET, Bae Y, Birkett R, Sharma AM, Zhang R, Fisch KM, Funk W, Mestan KK. Cord Blood Adductomics Reveals Oxidative Stress Exposure Pathways of Bronchopulmonary Dysplasia. *Antioxidants (Basel)* **2024,** *13* (4).
95. Liu S, Grigoryan H, Edmands WMB, Dagnino S, Sinharay R, Cullinan P, Collins P, Chung KF, Barratt B, Kelly FJ, Vineis P, Rappaport SM. Cys34 Adductomes Differ between Patients with Chronic Lung or Heart Disease and Healthy Controls in Central London. *Environ Sci Technol* **2018,** *52* (4), 2307-13.
96. Wong JYY, Imani P, Grigoryan H, Bassig BA, Dai Y, Hu W, Blechter B, Rahman ML, Ji BT, Duan H, Niu Y, Ye M, Jia X, Meng T, Bin P, Downward G, Meliefste K, Leng S, Fu W, Yang J, Ren D, Xu J, Zhou B, Hosgood HD, Vermeulen R, Zheng Y, Silverman DT, Rothman N, Rappaport SM, Lan Q. Exposure to diesel engine exhaust and alterations to the Cys34/Lys525 adductome of human serum albumin. *Environ Toxicol Pharmacol* **2022,** *95*, 103966.