Global Profiling of Urinary Mercapturic Acids Using Integrated Library-Guided Analysis

Zhengzhi Xie[1 a,b,c,f], Jin Y. Chen[1,a,b,c,f], Hong Gao[, a,b,c,f], Rachel J. Keith[a,b,c,f], Aruni Bhatnagar[a,b,c,f], Pawel Lorkiewicz*[a,b,c,d,e,f], and Sanjay Srivastava[a,b,c,f]

[1] Co-first author
[a]American Heart Association-Tobacco Regulation and Addiction Center, University of Louisville;
[b]Christina Lee Brown Envirome Institute, University of Louisville;
[c]Superfund Research Center, University of Louisville;
[d]Department Center for Cardiometabolic Science, University of Louisville;
[e]Department of Chemistry, University of Louisville; and,
[f] Division of Environmental Medicine, Department of Medicine, University of Louisville, Louisville, KY 40202, USA

Running Title: Untargeted profiling of urinary mercapturome

Emails: Zhengzhi Xie: zhengzhi.xie@louisville.edu; Jin Y. Chen: jin.chen@louisville.edu; Hong Gao: hong.gao@louisville.edu; Rachel J. Keith : rachel.keith@louisville.edu ; Aruni Bhatnagar: aruni@louisville.edu ;Pawel Lorkiewicz: pawel.lorkiewicz@louisville.edu; Sanjay Srivastava: sanjay.srivastava@louisville.edu

***Corresponding author:**

Pawel Lorkiewicz, Ph.D.
Center for Cardiometabolic Science
580 S. Preston Street, Delia Baxter Building, Rm. 421C,
University of Louisville, Louisville, KY, 40202
Tel: 502-852-5750
E-mail: pawel.lorkiewicz@louisville.edu

ABSTRACT

Urinary mercapturic acids (MAs) are often used as biomarkers for monitoring human exposures to occupational and environmental xenobiotics. Untargeted mass spectrometry-based approaches have been applied for the broad characterization of MAs, but the metabolite coverage was limited due to the lack of comprehensive MA databases. In this study, we developed an integrated library-guided analysis (ILGA) workflow using ultra-performance liquid chromatography-quadrupole time-of-flight mass spectrometry (UPLC-QTOF-MS). This method includes expanded assignment criteria and a curated library of ~240 MAs and addresses the shortcomings of previous untargeted approaches. We employed this workflow to profile MAs in the urine of 70 participants - 40 nonsmokers and 30 smokers. We identified approximately 500 MA candidates in each urine sample, and 118 MAs were putatively assigned with the ILGA approach. These include 29 previously unreported MAs derived mostly from alkenals and hydroxyalkenals. In addition, we observed that the levels of 70 MAs were comparable in nonsmokers and smokers, 2 MAs were higher in nonsmokers, and the levels of 46 MAs were elevated in the smokers. These included previously unreported mercapturates of polyaromatic hydrocarbons and hydroxyalkenals and well-documented MAs derived from toxicants present in cigarette smoke (e.g., acrolein, 1,3-butadiene, isoprene, acrylamide, benzene, and toluene). Our untargeted workflow led to the effective identification and discovery of known and unreported MAs derived from endogenous and environmental sources, and the levels of several of these MAs were increased by smoking. Our method can also be expanded and applied to other exposure-wide association studies.

Key Words: xenobiotic exposure, mercapturic acids, exposomics, mercapturomics, cigarettes, LC-MS, aliphatics, aromatics

Abbreviations: cardiovascular disease (CVD), volatile organic compound (VOC), glutathione (GSH), mercapturic acid (MA), integrated library-guided analysis (ILGA), ultra-performance liquid chromatography-quadrupole time-of-flight mass spectrometry (UPLC-QTOF-MS), data-independent acquisition (DIA), common neutral loss (CNL), harmful and potentially harmful constituents (HPHCs)

INTRODUCTION

Humans are likely to be exposed to 1-3 million chemicals in their lifetime[1]. Prominent sources of xenobiotic exposure include industrial chemicals, petroleum products, fossil fuels, household chemicals, tobacco smoke, plant products, pesticides, and pharmaceutical reagents[2, 3]. Several of these chemicals are highly reactive and injurious to health[4, 5]. Highly reactive chemicals are also generated endogenously, especially during the conditions of oxidative stress[6, 7]. Due to their high electrophilicity, unsaturated carbonyl compounds such as acrolein (abundant in cigarette smoke and automobile exhaust) and 4, hydroxy trans-2-nonenal (HNE, generated by lipid peroxidation) can react with cellular nucleophiles of proteins and DNA[6, 8-10]. Assessment of biomarkers of exposure to these toxicants will assist in developing appropriate remediation or detoxification mechanisms to limit the toxicity.

A common way to measure exposures to electrophilic xenobiotics is by analyzing their downstream N-acetyl-L-cysteine S-conjugate metabolites or mercapturic acids (MAs) in the urine[9, 11]. MAs are formed from xenobiotic-glutathione (GSH) S-conjugates through sequential transformation, including hydrolysis followed by N-acetylation in the kidney to form MAs, which are excreted in the urine[12]. The measurement of MAs is non-invasive and has been typically performed using liquid chromatography-mass spectrometry (LC-MS) targeted [13-15] and untargeted methods[16-18].

The untargeted approaches allow for discovery, identification, and semi-quantification of the detected species. Therefore, they have gained popularity for biomarker discovery and metabolite identification[16-19]. Several proof-of-concept studies

relying on untargeted approaches have been previously conducted by screening a common neutral loss (CNL) of 129 Da (a feature of N-acetyl-L-cysteine moiety) for MA detections[16, 18-20]. Newer methodologies that combined untargeted metabolomics, CNL scanning, and online database searching allowed for the identification of some MAs[17, 21]. Previous studies have analyzed urine samples from smokers and nonsmokers *via* an untargeted metabolomic approach, and putatively annotated 43 out of 91 MA candidates[17]. However, CNL-based screening may miss MA ions that do not undergo a loss of 129 Da. Similarly, CNL alone may lead to false-positive assignments. Moreover, small number of MAs listed in existing databases limits the range of annotated and identified species in previous mercapturomics studies. Due to a relatively small number of untargeted studies and insufficient MA coverage, the field of untargeted MA analysis is still in its infancy.

We developed an integrated library-guided analysis (ILGA) workflow using ultra-performance liquid chromatography-quadrupole time-of-flight mass spectrometry (UPLC-QTOF-MS) with MS^E data-independent acquisition (DIA) for MAs profiling in human urine. This method includes expanded assignment criteria, which yielded ~500 MA candidates in a typical urine sample – a sizable increase over previously reported workflows. In addition, we curated a library of ~240 structures, which markedly enhanced the range of identified MA species – 118 MAs in the urine of study participants.

MATERIALS AND METHODS

Chemicals and reagents

Thirteen authentic MA standards, including N-acetyl-S-(2,4-dimethylbenzene)-L-cysteine (MPhMA), N-acetyl-S-(3-hydroxypropyl)cysteine (3HPMA), N-acetyl-S-(2,3-dihydroxypropyl)-L-cysteine (23HPMA), N-acetyl-S-(3-amino-3-oxopropyl)-L-cysteine (2CaEMA), N-acetyl-S-(N-methylcarbamoyl)cysteine (MCaMA), benzylmercapturic acid (BzMA), propylmercapturic acid (1PMA), N-acetyl-S-(2-carboxyethyl)-L-cysteine (2CoEMA), N-acetyl-S-(2-cyanoethyl)-L-cysteine (2CyEMA), N-acetyl-S-(3,4-dihydroxybutyl)-L-cysteine (34HBMA), N-acetyl-S-[(2S)-3-amino-2-hydroxy-3-oxopropyl]-L-cysteine (2CaHEMA), N-acetyl-S-(3-hydroxy-1-methylpropyl)-L-cysteine (3HMPMA), and N-acetyl-S-[(1R)-2-hydroxy-1-phenylethyl]-L-cysteine (2HPhEMA) were purchased from Toronto Research Chemicals, Canada. UHPLC-MS grade water, UHPLC-MS grade acetonitrile, and LC-MS grade formic acid, and Infinity™ Creatinine Liquid Stable Reagent were purchased from Thermo Fisher Scientific Inc., Waltham MA.

Study population and sample collection

Spot urine specimens were collected from the participants (n=70) and stored at -80℃ until UPLC-QTOF MS^E DIA for MA profiling. **Figure 1** shows the general workflow developed for the analysis and assignment of MAs. Urinary cotinine levels were measured to determine the current smoking status of the study participants[22]. Demographic information (**Table 1**) including age, sex, race, smoking history, medical information, and other chronic health conditions, was obtained through the questionnaires. The study was approved by the University of Louisville Institutional Review Board (IRB 15.1260).

LC-MS analysis

The urine samples (50 µL) were thawed on ice and mixed with 0.1% formic acid (450 µL) in water. An aliquot of 7.5 µL of the mixture was analyzed using an Acquity I-Class UPLC system (Waters, MA). The separation was performed using a 2.1mm x 150mm Acquity Premier HSS T3 1.8 µm UPLC column (Waters, MA) maintained at 45℃ at a flow rate of 0.45 mL/min. The column was eluted with a gradient composed of 0.1% formic acid in water (solvent A) and 0.1% formic acid in acetonitrile (solvent B). The gradient profile started at 0% of B, increased to 23% B over 11 min, and then increased to 95% B over 3.6 min. The gradient was held at 95 %B for 2.4 min before returning to the initial conditions over 0.05 min, and re-equilibrated for 2.95 min before the next injection. One injection of QC sample, prepared by pooling equal amounts of each urine sample, was performed every ten injections of urine samples.

QTOF-MS data were collected using a Synapt XS HDMS (Waters, MA) with Masslynx 4.2 software and an electrospray ion source operated in negative mode. The capillary voltage was 2.25 kV, the source temperature was 120℃, the desolvation gas flow was 700L/h at a temperature of 650℃, and the cone gas flow was 150 L/h. The MS^E DIA acquisition was performed over the m/z range of 40−930 Da with low collision energy off (function 1) and high collision energy ramping from 10 to 40 V (function 2). Each of these functions employed a scan time of 0.2 s. Sodium formate was used for the mass calibration before the sample run, and leucine enkephalin with m/z ~ 554.2620 was used as the lock mass solution during the acquisition. The raw files were analyzed using UNIFI 1.9 software package (Waters, MA).

Identification and semi-quantification of urinary mercapturic acids

An ILGA workflow was established to profile MAs. This method started with curating a MA library in UNIFI which consists of structures of known and deduced MAs. Structures of 164 reported MAs were collected by searching structures and keywords (e.g., "mercapturic acid" and "L-cysteine, N-acetyl-S") in the SciFinder^{n} database, downloaded as .mol files, and imported into the library. Additional information, including retention times and MS/MS spectra, was added to the library after analyzing available MA standards. Furthermore, we deduced 75 prospective structures for MAs of harmful and potentially harmful constituents (HPHCs) found in tobacco smoke[23]. Proposed structures were deduced based on metabolic pathways that are known for major groups of reactive xenobiotics such as alkenals (e.g., acrolein),[11, 24] aromatic hydrocarbons (e.g., benzene),[25] and polycyclic aromatic hydrocarbons (PAHs; e.g., phenanthrene)[26]. Examples of known and proposed formation pathways of MAs from acrolein[11] and benzene[25] are shown in **Figure 1S**. These pathways were explored to propose MA structures of analogous aliphatic and aromatic compounds. Together, our library comprised ~240 MA structures. Chemical formula, CAS number IUPAC name and a specific acronym created according to the guidelines outlined by Tevis et al[27] were also provided for each MA structure in the library (**Table 1S)**.

Annotation and assignment criteria

Specific criteria were established for annotation and assignment of MS features. To be considered, selected features had to exhibit either a neutral loss m/z 129.043 Da ($-C_5H_7NO_3$), or at least one of the ion fragments specific to N-acetyl-L-cysteine moiety at

m/z 74.020 (C_3H_6S), 84.045 (C_4H_6NO), m/z 128.035 ($C_5H_6NO_3$), and m/z 162.023 ($C_5H_8NO_3S$) in the MS/MS spectra (function 2). **Figure 2S** shows representative MS and MS/MS spectra of N-acetyl-S-[1-(hydroxymethyl)-2-propenyl]-L-cysteine (1HMPeMA) that includes all specified features. Furthermore, confidence level criteria established by the Compound Identification workgroup of the Metabolomics Society (2017) were adapted and included in the ILGA analysis.[28] **Table 2S** describes confidence levels (level 0 to level 4) and criteria used for their assignment.

Typically, 500 out of 10,000 features in a single chromatogram matched the assignment criteria. These potential metabolites were inspected to remove duplicates or misassignments and screened against the curated library. Only features that matched monoisotopic mass with ±5 mDa accuracy in the MS domain (function 1), were putatively annotated and used for further analysis (level 4 features were not considered). Any additional information collected from authentic standard compounds was used to further confirm the identity of MAs. **Figure 3S** shows an example of such confirmation using the standard compound 3HPMA.

Statistical analysis

The characteristics of the total selected study subjects are expressed as mean ± standard deviation (SD) for continuous variables and frequency (%) for categorical variables. To examine the different characteristics between nonsmokers and smokers, the Student's *t*-test was performed for age, BMI, systolic blood pressure (SBP), diastolic blood pressure (DBP), and the Chi-square test was performed for sex, race, hypertension, and diabetes (**Table 1**).

All MAs were normalized to urinary creatinine levels to adjust for dilution. Intensity values for undetected MAs were imputed by dividing the detector response cutoff value (75 counts) by the square root of 2. Since the distributions of these normalized urinary MAs were right-skewed, they were log-transformed to improve the normality. The fold changes (FC) of the mean MAs were presented in the volcano plot, and 95% confidence limits (CL) between smokers and nonsmokers were presented in the forest plots.

Linear regression models were constructed to examine the associations between urinary cotinine and MAs. Both independent variables (cotinine) and dependent variables were log-transformed to improve the normality in the models. Since demographic and clinical characteristics were not significantly different between nonsmokers and smokers (**Table 1**), these potential confounders were not adjusted in the regression models. The statistical significance was set at the p-value <0.05.

The correlation heatmap of selected MAs and cotinine, and principal component analysis (PCA) of urinary mercapturome were conducted using the MetaboAnalyst 5.0 platform (https://www.metaboanalyst.ca/). Other statistical analyses were performed using SAS, version 9.4 (SAS Institute, Inc., NC), and the forest plots were produced in GraphPad Prism, version 9.1 (GraphPad Software, CA).

RESULTS

Characteristics of the Study Population

The demographic and clinical characteristics of the study participants are provided in **Table 1.** This study population comprised 69% females and 31% males, 83% White

and 13% Black. The mean age was 50 ± 12 years, and the mean BMI was 28.8 ± 6.9 kg/m^2. Forty-three percent of the participants were smokers (urinary cotinine level >40 µg/g creatinine),[29] 63% were hypertensive, and 17% were diabetics. No significant differences were observed in the demographics and clinical characteristics of smokers and nonsmokers.

Mercapturic acids levels in the study group

UPLC-MS analysis of the urine samples of study participants (n=70) revealed that out of ~500 annotated features that met assignment criteria, 118 MAs matched the curated library and were putatively assigned as metabolites of 64 parent xenobiotics. These xenobiotics account for 37 aliphatic (alkenals, hydroxyalkenals, halogenated aliphatics etc.) and 27 aromatic compounds (benzene and monocyclic substituted aromatic compounds, aromatic aldehydes, PAHs, halogenated aromatics, pharmaceutical reagents etc.). Full names, acronyms, structures, confidence levels of annotations, and CAS numbers of 118 MAs are provided in **Table 3S**. Notably, to our knowledge, 29 MAs (with no CAS number available) are reported here for the first time. As noted in **Table 3S**, these MAs were derived from 9 aliphatic (2-pentenal, 2-hexenal, 2-heptenal 2-octenal, 4-hydroxy-2-pentenal, 4-hydroxy-2-hexenal, 4-hydroxy-2-heptenal, 4-hydroxy-2-octenal, N-tert-butyl-acetamide) and 1 aromatic (phenanthrene) xenobiotic precursors.

Stratification of data between nonsmokers and smokers showed that 109 out of 118 MAs were detected in nonsmokers; however, as shown in **Table 2**, 55 MAs were above the instrument's detection limit in >25% samples, 25 MAs in >50% samples, and 9

MAs (primarily derived from aliphatic precursors) in >75% samples. Smokers' urine contained all 109 MAs found in nonsmokers and 9 additional MAs derived from aromatic (benzene, aniline, styrene, trimethylbenzene, naphthalene, coumarin) and aliphatic (4-hydroxy-2-hexenal, sulforaphane) compounds. Ninety-eight MAs were found in >25%, 55 MAs in >50%, and 37 MAs in >75% smokers' samples (**Table 2**). The levels of 46 MAs were significantly increased in smokers (**Table 3S**).

Forest plots illustrate significant changes in MA levels in each subgroup of parent compounds between smokers and nonsmokers (**Figure 3**). As shown in **Figure 3a,** 31 MAs derived from aliphatic precursors, were significantly different between smokers and nonsmokers. The levels of 2 MAs, N-acetyl-S-methylcysteine (MMA) and N-acetyl-S-ethylcysteine (EMA), were significantly higher in nonsmokers than smokers (**Table 3S**). Twenty nine MAs were significantly higher in smokers. These comprised of 9 MAs from alkenals, 6 MAs from hydroxyalkenals, 1 MA from halogenated aliphatic chemicals, and 13 MAs from other aliphatics. Levels of all 17 MAs derived from aromatic precursors were significantly higher in smokers (**Figure 3b**). The levels of 70 MAs derived from alkenals (acrolein, crotonaldehyde, 2-heptenal), hydroxyalkenals (4-hydroxy-2-pentenal, 4-hydroxy-2-hexenal, 4-hydroxy-2-heptenal, 4-hydroxy-2-octenal, 4-hydroxy-2-nonenal), halogenated aliphatics (halogenated propane, pentane, heptane, 2,3-dichloro-propene, ethyl chloroacetate), other aliphatics (ethylene oxide, allyl halide, 1,3-butadiene, acrylamide, acrylonitrile, methacrylonitrile, methylacrylate, butyl acrylate, 2-ethylhexyl acrylate, N-tert-butyl-acetamide, 4-methylthiobutyl isothiocyanate, allyl isothiocyanate, sulforaphane), and finally, aromatic compounds (benzene, anilline, styrene, xylene, trimethylbenzene, hydroquinone, trihydroxybenzene, orthocetamol, cinnamaldehyde, 2-

phenylpropenal, α-phenylacrylic acid, naphthalene, phenanthrene, bromocyclohexane, halogenated dinitrobenzene, acetaminophen, 1,4-benzoquinone, benzyl isothiocyanate, phenethyl isothiocyanate, coumarin), were comparable between nonsmokers and smokers.

Linear regression, and PCA analyses of all 118 MAs are provided in the Supplemental Materials (**Table 4S**, **Figure 4S**). However, because the reliability of statistical analyses deteriorates with increasing number of missing values, 55 MAs that were detected in > 50% of smokers were used for the bivariate and multivariate analysis. With the exception of 6HCycHeMA all the other 54 out of the 55 MAs were also detected in nonsmokers (**Table 2S**).

Linear regression analysis between urinary cotinine and 55 MAs detected in > 50% of smokers (**Table 3, Figure 3**) showed that 45 out of 55 MAs had positive associations ($\beta > 0$ and $p < 0.05$) with cotinine, and none showed significant negative association ($\beta < 0$ and $p < 0.05$). Notably, among the MAs with β values >0.4, metabolites of acrolein (2CoEMA, 3HPMA), crotonaldehyde (3HMPMA), 1,3-butadiene (N-acetyl-S-(2-hydroxy-3-buten-1-yl)-L-cysteine, 2HBeMA), acrylonitrile (2CyEMA), and isoprene (N-acetyl-S-(2-hydroxy-3-methyl-3-buten-1-yl)-L-cysteine, 2HMBeMA; N-acetyl-S-(4-hydroxy-2-methyl-2-buten-1-yl)-L-cysteine, 4HMBeMA), are well-known biomarkers of exposure to VOCs and HPHCs in cigarette smoke.[13, 22, 30]

A correlation heatmap was generated with Pearson correlation coefficients (r) between 55 MAs to continue to examine their relationship (**Figure 4**). Consistent with association results (**Table 3**), metabolites of 1,3-butadiene (2HBeMA), isoprene

(4HMBeMA) and acrylonitrile (2CyEMA) were highly correlated with cotinine (r >0.66; **Table 3** and **Figure 4**). Interestingly, metabolites of acrolein (2CoEMA, 3HPMA), crotonaldehyde (3HMPMA), 2-pentenal (N-acetyl-S-(1-ethyl-3-hydroxypropyl)-L-cysteine,1EHPMA), and ethyl acrylate (N-acetyl-S-(3-ethoxy-3-oxopropyl)-L-cysteine, 3EoOxPMA) were highly correlated with each other (r >0.67), suggesting co-exposure to the parent compounds of these MAs. We also observed that MAs from 2-octenal (2CoEPnMA) and 4-Hydroxy-2-octenal (3OxPHPnMA) were highly correlated to crotonaldehyde metabolite (3HMPMA), indicating the possible relationship between these parent compounds. Moreover, 3OxPMA, a proposed acrolein metabolite, showed a relatively poor correlation with the other two acrolein metabolites, 3HPMA and 2CoEMA.

The PCA analysis performed using the same 55 MAs found in > 50% of smokers (**Figure 5)** shows that this unsupervised technique produced only partial separation between the two groups. In the PCA score plot (**Figure 5a)**, smokers were located at the high end of the axis of principal component 1 (PC1). The PC1, which accounts for the highest share of the total variance among all PCs, represents MAs abundant in smokers' urine, such as the ones from acrolein (2CoEMA, 3HPMA), crotonaldehyde (3HMPMA), 1,3-butadiene (2HBeMA), isoprene (4HMBeMA) and acrylonitrile (2CyEMA) (**Figure 5b**). Table 3 shows these MAs were significantly higher in smokers' urine, thus driving high PC1 scores. However, instead of clustering at the low end of the PC1, nonsmokers distributed widely throughout the PC1 range (**Figure 5a).** Some were at the high end of the PC1, overlapping with smokers.

DISCUSSION

We developed a UPLC-QTOF-MS with MS^E DIA workflow that utilizes the ILGA method to profile MAs in the urine of 70 participants. Our workflow expands on and adds to the previous untargeted profiling approaches[16, 18-20]. To increase confidence, in addition to utilizing the CNL 129 Da, we expanded assignment criteria to include one or more common MS/MS fragments (162.0229, 128.0349, 84.0454, or 74.0244) characteristic to N-acetyl-L-cysteine moiety. Furthermore, the ILGA approach establishes the connection between MAs and their precursors through the generation of an inclusive and expandable library that combines known and imputed structures of ~240 MAs. The structural information of the metabolites and the metabolite-parent connection greatly facilitated the interpretation of the results and improved assignment confidence.

The population characteristics in this study were appropriately distributed, which allowed the comparison of MA levels without adjusting for demographic variables such as age, gender, race, and health conditions.

Our approach allowed for uncovering nearly 500 metabolite candidates and putatively assigning 118 MAs derived from 67 parent xenobiotics. The ILGA workflow uncovered 29 previously unreported MAs derived from 10 xenobiotic precursors mostly derived from alkenals and hydroxyalkenals. Interestingly, two MAs, MMA and EMA, derived from methyl and ethyl halides, respectively, were significantly higher in nonsmokers. Ethyl chloride and halogenated methane are industrial chemicals, used as alkylating agents, and are classified as potential carcinogens and mutagens[31]. Our studies also showed that the levels of 70 MAs were comparable in nonsmokers and

smokers, suggesting that these MAs are derived from organic pollutants present in the ambient air.

Endogenous lipid peroxidation and dietary sources are the other potential sources of MAs. Hydroxyalkenals are predominantly derived from the oxidation of ω-6 polyunsaturated fatty acids. Due to their high electrophilicity, these unsaturated aldehydes are highly reactive and can crosslink with DNAs and proteins. Increased accumulation of α,β- unsaturated aldehydes has been observed in several pathological conditions including atherosclerosis, Alzheimer's, and Parkinson diseases[32-34] Lipid peroxidation also generates high concentrations of other alkanal and alkenals. Myeloperoxidase-derived reactions generate alkenals, such as acrolein during inflammatory conditions, metabolism of drugs such as cyclophosphamide and industrial chemicals such as allylamine[32-35]. Longer chain alkenals - 2-Hexenal and 2-octenal are natural food constituents used widely as flavoring agents[36, 37].

Increased abundance of MAs of several hydroxalkenals and alkenals in smokers is consistent with increased oxidative stress. Our analyses are also in agreement with previous studies demonstrating that metabolites of acrolein (2CoEMA, 3HPMA), crotonaldehyde (3HMPMA), 1,3-butadiene (2HBeMA), N,-N-dimethylformamide (N-acetyl-S-(N-methylcarbamoyl)cysteine, MCaMA), and acrylamide (N-acetyl-S-(2-carbamoylethyl)-L-cysteine, 2CaEMA; N-acetyl-S-(2-carbamoyl-2-hydroxyethyl)-L-cysteine, 2CaHEMA) have been reported to elevate in smokers and routinely used as biomarkers of exposure to tobacco smoke[13, 15, 38]. Also, metabolites of glycidol (N-acetyl-S-(2,3-dihydroxypropyl)-L-cysteine, 23HPMA) and isoprene (N-acetyl-S-[(2E)-4-hydroxy-

2-methyl-2-buten-1-yl]-L-cysteine, 4MBeMA; 2HMBeMA, 4HMBeMA) were previously found to be elevated in the urine of smokers,[30, 39, 40] but are not specific biomarkers of cigarette use. Moreover, well-established biomarkers of cigarette exposure: aromatics-derived MAs were also elevated in smokers. These included MAs from benzene (N-acetyl-S-(6-hydroxy-2,4-cyclohexadien-1-yl)-L-cysteine, 6HCycHeMA),[41] toluene (benzylmercapturic acid, BzMA),[13] and xylene (N-acetyl-S-[(2-methylphenyl)methyl]-L-cysteine, 2MPhMMA)[42].

Similar to several aliphatic compounds, we also observed significantly higher levels of MAs derived from aromatic compounds including naphthalene (N-acetyl-S-(3,4-dihydroxy-1-naphthalenyl)-L-cysteine, 34HNeMA), and phenanthrene (2-phenanthrenemercapturic acid, 2PaMA; 4-phenanthrenemercapturic acid, 4PaMA) in the urine of smokers. These polyaromatic hydrocarbons (PAHs) have been found in combustible tobacco products[43], and, to the best of our knowledge, this is the first time to report mercapturic acid metabolites of these PAH in human urine. Furthermore, substantially higher levels of PAH MAs in the urine of smokers than nonsmokers suggest these molecules are predominantly derived from cigarette smoking. Additionally, in smokers, we discovered a significantly higher level of MAs derived from precursors present in different environmental settings. For instance, exposures to halogenated alkanes, ethyl acrylate, isobutyl acrylates, trimethylbenzene, 2-fluorobenzaldehyde, 4-tert-butylbenzaldehyde, and 4-chloronitrobenzene are primarily industrial and occupational[31, 44-47]. Detection of MAs derived from 4-methoxybenzaldehyde, acetaminophen, and 6-chloropurine indicate exposures to biogenic and pharmaceutical parent compounds.[48-50] Interestingly, the source of 3-(2-furanyl)-3-hydroxy-1-(4-

methylphenyl)-1-propanone (FHMPP), the precursor of N-acetyl-S-[1-(2-furanyl)-3-(4-methylphenyl)-3-oxopropyl]-L-cysteine (2FuMPhOxPMA), is not well known.

Interestingly, some of the MAs derived from the same parent compounds were not necessarily correlated with each other. For example, among the 3 MAs derived from acrolein, 3OxPMA was not correlated with 3HPMA and 2CoEMA. This could possibly be be due to the differences in the metabolism and pharmacokinetics of these metabolites.

CONCLUSIONS

We developed a UPLC-QTOF-MS with MS^E DIA workflow that utilizes the ILGA approach and applied it for profiling MAs in urine samples collected from 70 participants. Our studies identified 29 new MAs, and detected 109 out of 118 MAs in both nonsmokers and smokers, suggesting that these MAs are derived, at least in parts, from environmental exposure. Significantly higher levels of 46 MAs in smokers reveal that smoking increases the levels of these metabolites. The ILGA method used in this study, could be expanded to profile a wider set of metabolites for biomonitoring human exposure to VOCs and xenobiotics.

Acknowledgements: This work was supported by National Institutes of Health Grants U54HL120163, P42ES023716 and S10OD026840.

Table 1: Characteristics of the participants by nonsmokers and smokers.

Variable	Total (n=70)	Nonsmokers (n = 40)	Smokers (n = 30)	p value
Sex, Male	22 (31%)	14 (35%)	8 (27%)	0.457
Race				0.241
White	58 (83%)	33 (83%)	25 (83%)	
Black	9 (13%)	4 (10%)	5 (17%)	
Other	3 (4%)	3 (8%)	0 (0%)	
Age (years)	50 ± 12.3	49.2 ± 13.2	51 ± 11.1	0.543
Body mass index, BMI (kg/m2)	28.8 ± 6.9	28.8 ± 7.2	28.8 ± 6.5	1.000
systolic blood pressure, SBP (mmHg)	129 ± 21.3	126.8 ± 23.6	132 ± 17.7	0.311
diastolic blood pressure, DBP (mmHg)	79.5 ± 13.9	77.8 ± 14.4	81.7 ± 13.1	0.251
Hypertension	44 (63%)	27 (68%)	17 (57%)	0.353
Diabetes	12 (17%)	7 (18%)	5 (17%)	0.927

Values represent mean ± standard deviation (SD) for continuous variables, and frequency (%) for categorical variables.

Statistical tests for the comparison between nonsmokers and smokers: the Student's *t*-test for age, BMI, SBP, and DBP, the Chi square test for sex, race, hypertension and diabetes.

Table 2: Number of putative mercapturic acids detected in urine samples categorized by their parent compound subgroups.

Main group of the parent compound	Subgroup of the parent compound	Nonsmokers				Smokers			
		At least in one sample	> 25% samples	> 50% samples	> 75% samples	At least in one sample	> 25% samples	> 50% samples	> 75% samples
Aliphatics	Alkenals	13	11	8	3	13	14	11	11
	Hydroxyalkenals	21	8	3	1	22	19	8	5
	Halogenated aliphatics	8	4	1	N.D.	8	8	N.D.	N.D.
	Other aliphatics	28	12	4	3	29	22	15	11
	Aliphatics total	70	35	16	7	72	63	34	27
Aromatics	Benzene and monocyclic substituted aromatics	9	4	3	1	13	12	6	3
	Polycyclic aromatic hydrocarbons	8	5	1	1	8	9	5	3
	Aromatic aldehydes	5	5	4	N.D.	5	7	5	3
	Halogenated aromatics	2	1	1	N.D.	2	3	2	1
	Pharmaceuticals	11	2	N.D.	N.D.	13	2	2	N.D.
	Other aromatics	4	3	N.D.	N.D.	5	2	1	N.D.
	Aromatics total	39	20	9	2	46	35	21	10
Grand Total		109	55	25	9	118	98	55	37

N.D.: Not detected

Table 3. Associations between urinary cotinine and 55 MAs detected in more than 50% of smokers. β is linear regression coefficient; SE is standard error of β. Before the linear regression, the creatinine normalized levels of MAs were log transformed.

Main group of parent	Subgroup	Harmonized acronym	Full name	β	SE	p value
Aliphatics	Alkenals	3HPPnMA	3-Hydroxypropyl-1-pentylmercapturic acid	0.131	0.038	**0.001**
		2CoEPnMA	2-Carboxyethyl-1-pentylmercapturic acid	0.325	0.051	**<0.001**
		2HEBMA	N-Acetyl-S-[1-(2-hydroxyethyl)butyl]-L-cysteine	0.120	0.057	**0.035**
		2CoEEMA	2-Carboxyethyl-1-ethylmercapturic acid	0.187	0.044	**<0.001**
		3HPMA	N-Acetyl-S-(3-hydroxypropyl)cysteine	0.418	0.062	**<0.001**
		2CoEMA	N-acetyl-S-(2-carboxyethyl)-L-cysteine	0.406	0.052	**<0.001**
		3HMPMA	N-Acetyl-S-(3-hydroxy-1-methylpropyl)-L-cysteine	0.434	0.053	**<0.001**
		2CoMEMA	2-Carboxy-1-methylethylmercapturic acid	0.081	0.085	0.338
		1EHPMA	N-Acetyl-S-(1-ethyl-3-hydroxypropyl)-L-cysteine	0.365	0.047	**<0.001**
		2CoEPMA	2-Carboxyethyl-1-propylmercapturic acid	0.242	0.073	**0.001**
		3OxPMA	N-Acetyl-S-(3-oxopropyl)-L-cysteine	-0.081	0.084	0.337
	Hydroxyalkenals	5HPnFuMA	N-Acetyl-S-(tetrahydro-5-hydroxy-2-pentyl-3-furanyl)-L-cysteine	0.104	0.050	**0.036**
		5OxPnFuMA	N-Acetyl-S-(tetrahydro-5-oxo-2-pentyl-3-furanyl)-L-cysteine	0.037	0.067	0.584
		3OxPHPMA	3-Oxopropyl-1-(2-hydroxypentyl)mercapturic acid	0.274	0.074	**<0.001**
		3OxPHPMA2	3-Oxopropyl-1-(2-hydroxypentyl)mercapturic acid	0.303	0.039	**<0.001**

Main group of parent	Subgroup	Harmonized acronym	Full name	β	SE	p value
		3OxPHPnMA	3-Oxopropyl-1-(2-hydroxypentyl)mercapturic acid	0.197	0.076	**0.009**
		3OxPHEMA	3-Oxopropyl-1-(2-hydroxyethyl)mercapturic acid	0.233	0.070	**0.001**
		5MDfMA	4-(5-Methyl-dihydrofuryl)mercapturic acid	0.227	0.040	**<0.001**
		3HPHEMA	3-Hydroxypropyl-1-(2-hydroxyethyl)mercapturic acid	-0.068	0.074	0.359
	Other aliphatics	4HMBeMA	N-Acetyl-S-(4-hydroxy-2-methyl-2-buten-1-yl)-L-cysteine	0.543	0.045	**<0.001**
		23HHMPMA2	N-Acetyl-S-[2,3-dihydroxy-1-(hydroxymethyl)propyl]-L-cysteine M1	0.264	0.047	**<0.001**
		2MPoOxPMA	N-Acetyl-S-[3-(2-methylpropoxy)-3-oxopropyl]-L-cysteine	0.151	0.047	**0.001**
		2CaEMA	N-Acetyl-S-(3-amino-3-oxopropyl)-L-cysteine	0.320	0.057	**<0.001**
		MCaMA	N-Acetyl-S-(N-methylcarbamoyl)cysteine	0.366	0.047	**<0.001**
		2CyEMA	N-Acetyl-S-(2-cyanoethyl)-L-cysteine	0.509	0.057	**<0.001**
		34HBMA	N-Acetyl-S-(3,4-dihydroxybutyl)-L-cysteine	0.075	0.066	0.256
		2CaHEMA	N-Acetyl-S-[(2S)-3-amino-2-hydroxy-3-oxopropyl]-L-cysteine	0.207	0.032	**<0.001**
		2HMBeMA	N-Acetyl-S-(2-hydroxy-3-methyl-3-buten-1-yl)-L-cysteine	0.446	0.062	**<0.001**
		4MBeMA	N-Acetyl-S-[(2E)-4-hydroxy-2-methyl-2-buten-1-yl]-L-cysteine	0.206	0.071	**0.003**
		2HBeMA	N-Acetyl-S-(2-hydroxy-3-buten-1-yl)-L-cysteine	0.440	0.040	**<0.001**
		1CyHEMA	N-Acetyl-S-(1-cyano-2-hydroxyethyl)cysteine	0.137	0.060	**0.023**

Main group of parent	Subgroup	Harmonized acronym	Full name	β	SE	p value
		23HPMA	N-Acetyl-S-(2,3-dihydroxypropyl)-L-cysteine	0.162	0.040	**<0.001**
		2CoPMA	N-Acetyl-S-(2-carboxypropyl)-L-cysteine	0.096	0.053	0.069
		3EoOxPMA	N-Acetyl-S-(3-ethoxy-3-oxopropyl)-L-cysteine	0.230	0.040	**<0.001**
Aromatics	Benzene and monocyclic substituted aromatics	4NPhMA	N-Acetyl-S-(4-nitrophenyl)-L-cysteine	0.173	0.046	**<0.001**
		6HCycHeMA	N-Acetyl-S-(6-hydroxy-2,4-cyclohexadien-1-yl)-L-cysteine	0.208	0.041	**<0.001**
		2MPhMMA	N-Acetyl-S-[(2-methylphenyl)methyl]-L-cysteine	0.264	0.068	**<0.001**
		235HPhMA	N-Acetyl-S-(2,3,5-trihydroxyphenyl)-L-cysteine	-0.014	0.066	0.827
		245HPhMA	N-Acetyl-S-(2,4,5-trihydroxyphenyl)-L-cysteine	-0.033	0.081	0.682
		25HPhMA	N-Acetyl-S-(2,5-dihydroxyphenyl)-L-cysteine	0.093	0.046	**0.041**
	Polycyclic aromatic hydrocarbons	10HPaMA	N-Acetyl-S-(9,10-dihydro-10-hydroxy-9-phenanthrenyl)-L-cysteine	0.232	0.090	**0.010**
		3PaMA	3-Phenanthrenemercapturic acid	0.069	0.084	0.412
		2ANeMA	N-acetyl-3-[(2-amino-1-naphthyl)thio]-alanine	0.205	0.087	**0.018**
		2NeMA	N-Acetyl-S-(2-naphthalenylmethyl)-L-cysteine	0.277	0.047	**<0.001**
		34HNeMA	N-Acetyl-S-(3,4-dihydroxy-1-naphthalenyl)-L-cysteine	0.080	0.023	**<0.001**
	Aromatic aldehydes	2FPhMMA	N-Acetyl-S-[(2-fluorophenyl)methyl]-L-cysteine	0.219	0.087	**0.012**
		4MoPhMMA	N-Acetyl-S-[(4-methoxyphenyl)methyl]-L-cysteine	0.176	0.058	**0.002**

Main group of parent	Subgroup	Harmonized acronym	Full name	β	SE	p value
		11MEPhMMA	N-Acetyl-S-[[4-(1,1-dimethylethyl)phenyl]methyl]-L-cysteine	0.178	0.046	**<0.001**
		BzMA	Benzylmercapturic acid	0.157	0.037	**<0.001**
		3HPhPMA	N-Acetyl-S-(3-hydroxy-1-phenylpropyl)-L-cysteine	0.062	0.072	0.392
	Halogenated aromatics	6PuMA	S-(6-Purinyl)-N-acetyl-L-cysteine	0.210	0.049	**<0.001**
		2HCycHxMA	N-Acetyl-S-(2-hydroxycyclohexyl)-L-cysteine	0.199	0.058	**0.001**
	Pharmaceuticals	4AcAPhMA	N-Acetyl-S-[4-(acetylamino)phenyl]-L-cysteine	0.161	0.032	**<0.001**
		5AcAHPhMA	N-acetyl-S-[5-(acetylamino)-2-hydroxyphenyl]-L-cysteine	0.265	0.145	0.067
	Other aromatics	2FuMPhOxPMA	N-Acetyl-S-[1-(2-furanyl)-3-(4-methylphenyl)-3-oxopropyl]-L-cysteine	0.168	0.046	**<0.001**

Figure Legends

Figure 1. Workflow displaying general steps of MA analysis and profiling.

Figure 2. Volcano plot showing differences in levels of 118 MAs between smokers and nonsmokers. The creatinine normalized levels of MAs were log-transformed, and shown as a point in the plot. The Student's t-test was conducted to examine the mean difference of the natural log-transformed 118 MAs between smokers and nonsmokers. Red points represent fold change (FC) >2 and $p < 0.05$. Blue points represent FC <0.5 and $p < 0.05$. Grey points represent $p > 0.05$ or $0.5 < FC < 2$.

Figure 3. Forest plots showing differences in levels of 118 MAs derived from **a**) aliphatic, and **b**) aromatic precursors between smokers and nonsmokers. The mean differences (round points) and 95% confidence limits (horizontal bars) were presented for each MA. The dotted lines represent no change.

Figure 4. Correlation heatmap of urinary cotinine and 55 MAs in the total selected subjects (n = 70) including smokers and nonsmokers. The dendrogram was generated by hierarchical cluster analysis based on Pearson correlation coefficient (r). The MAs and cotinine are in both columns and rows with their r values represented by color (blue is low and red is high).

Figure 5. PCA analysis of 55 MAs in the total selected subjects (n = 70) including smokers and nonsmokers: **a**) PCA score plot of PC1 and PC2 with symbols denoting nonsmoker (blue circles) and smoker (red triangles) samples, and **b**) loading plot of PC1 and PC2 with each dot representing individual MA.

References

(1) Idle, J. R.; Gonzalez, F. J. Metabolomics. *Cell Metab* **2007**, *6* (5), 348-351. DOI: 10.1016/j.cmet.2007.10.005 From NLM Medline.
(2) Montero-Montoya, R.; Lopez-Vargas, R.; Arellano-Aguilar, O. Volatile Organic Compounds in Air: Sources, Distribution, Exposure and Associated Illnesses in Children. *Ann Glob Health* **2018**, *84* (2), 225-238. DOI: 10.29024/aogh.910 From NLM Medline.
(3) Stefanac, T.; Grgas, D.; Landeka Dragicevic, T. Xenobiotics-Division and Methods of Detection: A Review. *J Xenobiot* **2021**, *11* (4), 130-141. DOI: 10.3390/jox11040009 From NLM PubMed-not-MEDLINE.
(4) Organization, W. H. Agents classified by the IARC monographs. *World Health Organization, International Agency for Research on Cancer. http://monographs. iarc. fr/ENG/Classification. Last Accessed* **2016**.
(5) U.S. Food & Drug Administration. *Harmful and Potentially Harmful Constituents in Tobacco Products and Tobacco Smoke: Established List*. 2012. www.fda.gov (accessed 2017 November 6,).
(6) Esterbauer, H.; Schaur, R. J.; Zollner, H. Chemistry and biochemistry of 4-hydroxynonenal, malonaldehyde and related aldehydes. *Free Radic Biol Med* **1991**, *11* (1), 81-128. DOI: 10.1016/0891-5849(91)90192-6 From NLM Medline.
(7) Bhatnagar, A. Environmental cardiology: studying mechanistic links between pollution and heart disease. *Circ.Res.* **2006**, *99* (7), 692-705.
(8) Uchida, K.; Kanematsu, M.; Morimitsu, Y.; Osawa, T.; Noguchi, N.; Niki, E. Acrolein is a product of lipid peroxidation reaction - Formation of free acrolein and its conjugate with lysine residues in oxidized low density lipoproteins. *J Biol Chem* **1998**, *273* (26), 16058-16066. DOI: 10.1074/jbc.273.26.16058.
(9) van Welie, R. T.; van Dijck, R. G.; Vermeulen, N. P.; van Sittert, N. J. Mercapturic acids, protein adducts, and DNA adducts as biomarkers of electrophilic chemicals. *Critical reviews in toxicology* **1992**, *22* (5-6), 271-306. DOI: 10.3109/10408449209146310 From NLM Medline.
(10) Farmer, P. B.; Bailey, E.; Naylor, S.; Anderson, D.; Brooks, A.; Cushnir, J.; Lamb, J. H.; Sepai, O.; Tang, Y. S. Identification of endogenous electrophiles by means of mass spectrometric determination of protein and DNA adducts. *Environ Health Perspect* **1993**, *99*, 19-24. DOI: 10.1289/ehp.939919 From NLM Medline.
(11) M. DE ROOIJ JAN N. M. COMMANDEUR NICO P. E. VERMEULEN, B. Mercapturic acids as biomarkers of exposure to electrophilic chemicals: applications to environmental and industrial chemicals. *Biomarkers* **1998**, *3* (4-5), 239-303.
(12) Hanna, P. E.; Anders, M. W. The mercapturic acid pathway. *Critical reviews in toxicology* **2019**, *49* (10), 819-929. DOI: 10.1080/10408444.2019.1692191 From NLM Medline.
(13) Alwis, K. U.; Blount, B. C.; Britt, A. S.; Patel, D.; Ashley, D. L. Simultaneous analysis of 28 urinary VOC metabolites using ultra high performance liquid chromatography coupled with electrospray ionization tandem mass spectrometry (UPLC-ESI/MSMS). *Anal Chim Acta* **2012**, *750*, 152-160. DOI: 10.1016/j.aca.2012.04.009 READCUBE.

(14) Pluym, N.; Gilch, G.; Scherer, G.; Scherer, M. Analysis of 18 urinary mercapturic acids by two high-throughput multiplex-LC-MS/MS methods. *Anal Bioanal Chem* **2015**, *407* (18), 5463-5476. DOI: 10.1007/s00216-015-8719-x.
(15) Lorkiewicz, P.; Riggs, D. W.; Keith, R. J.; Conklin, D. J.; Xie, Z.; Sutaria, S.; Lynch, B.; Srivastava, S.; Bhatnagar, A. Comparison of Urinary Biomarkers of Exposure in Humans Using Electronic Cigarettes, Combustible Cigarettes, and Smokeless Tobacco. *Nicotine & tobacco research : official journal of the Society for Research on Nicotine and Tobacco* **2019**, *21* (9), 1228-1238. DOI: 10.1093/ntr/nty089.
(16) Wagner, S.; Scholz, K.; Sieber, M.; Kellert, M.; Voelkel, W. Tools in metabonomics: an integrated validation approach for LC-MS metabolic profiling of mercapturic acids in human urine. *Analytical chemistry* **2007**, *79* (7), 2918-2926. DOI: 10.1021/ac062153w READCUBE.
(17) Frigerio, G.; Mercadante, R.; Campo, L.; Polledri, E.; Boniardi, L.; Olgiati, L.; Missineo, P.; Fustinoni, S. Urinary biomonitoring of subjects with different smoking habits. Part I: Profiling mercapturic acids. *Toxicol Lett* **2020**, *327*, 48-57. DOI: 10.1016/j.toxlet.2020.03.010.
(18) Scholz, K.; Dekant, W.; Volkel, W.; Pahler, A. Rapid detection and identification of N-acetyl-L-cysteine thioethers using constant neutral loss and theoretical multiple reaction monitoring combined with enhanced product-ion scans on a linear ion trap mass spectrometer. *Journal of the American Society for Mass Spectrometry* **2005**, *16* (12), 1976-1984. DOI: 10.1016/j.jasms.2005.08.003.
(19) Jamin, E. L.; Costantino, R.; Mervant, L.; Martin, J.-F.; Jouanin, I.; Blas-Y-Estrada, F.; Guéraud, F.; Debrauwer, L. Global Profiling of Toxicologically Relevant Metabolites in Urine: Case Study of Reactive Aldehydes. *Analytical chemistry* **2020**, *92* (2), 1746-1754. DOI: 10.1021/acs.analchem.9b03146.
(20) Wagner, S.; Scholz, K.; Donegan, M.; Burton, L.; Wingate, J.; Volkel, W. Metabonomics and biomarker discovery: LC-MS metabolic profiling and constant neutral loss scanning combined with multivariate data analysis for mercapturic acid analysis. *Analytical chemistry* **2006**, *78* (4), 1296-1305. DOI: 10.1021/ac051705s.
(21) Bloch, R.; Schuetze, S. E.; Mueller, E.; Roeder, S.; Lehmann, I.; Brack, W.; Krauss, M. Non-targeted mercapturic acid screening in urine using LC-MS/MS with matrix effect compensation by postcolumn infusion of internal standard (PCI-IS). *Analytical and Bioanalytical Chemistry* **2019**, *411* (29), 7771-7781. DOI: 10.1007/s00216-019-02166-6.
(22) Lorkiewicz, P.; Riggs, D. W.; Keith, R. J.; Conklin, D. J.; Xie, Z. Z.; Sutaria, S.; Lynch, B.; Srivastava, S.; Bhatnagar, A. Comparison of Urinary Biomarkers of Exposure in Humans Using Electronic Cigarettes, Combustible Cigarettes, and Smokeless Tobacco. *Nicotine & Tobacco Research* **2019**, *21* (9), 1228-1238. DOI: 10.1093/ntr/nty089.
(23) US Food Drug Administration. Harmful and potentially harmful constituents in tobacco products and tobacco smoke; established list. *Fed Regist* **2012**, *77* (64), 20034-20037.
(24) Stevens, J. F.; Maier, C. S. Acrolein: sources, metabolism, and biomolecular interactions relevant to human health and disease. *Mol Nutr Food Res* **2008**, *52* (1), 7-25. DOI: 10.1002/mnfr.200700412 From NLM Medline.
(25) Waidyanatha, S.; Rothman, N.; Li, G.; Smith, M. T.; Yin, S.; Rappaport, S. M. Rapid determination of six urinary benzene metabolites in occupationally exposed and

unexposed subjects. *Anal Biochem* **2004**, *327* (2), 184-199. DOI: 10.1016/j.ab.2004.01.008.
(26) Hecht, S. S.; Villalta, P. W.; Hochalter, J. B. Analysis of phenanthrene diol epoxide mercapturic acid detoxification products in human urine: relevance to molecular epidemiology studies of glutathione S-transferase polymorphisms. *Carcinogenesis* **2008**, *29* (5), 937-943. DOI: 10.1093/carcin/bgn015 From NLM Medline.
(27) Tevis, D. S.; Flores, S. R.; Kenwood, B. M.; Bhandari, D.; Jacob, P., 3rd; Liu, J.; Lorkiewicz, P. K.; Conklin, D. J.; Hecht, S. S.; Goniewicz, M. L.; et al. Harmonization of acronyms for volatile organic compound metabolites using a standardized naming system. *International journal of hygiene and environmental health* **2021**, *235*, 113749. DOI: 10.1016/j.ijheh.2021.113749 From NLM Medline.
(28) Blaženović, I.; Kind, T.; Ji, J.; Fiehn, O. Software tools and approaches for compound identification of LC-MS/MS data in metabolomics. *Metabolites* **2018**, *8* (2), 31.
(29) Riggs, D. W.; Malovichko, M. V.; Gao, H.; McGraw, K. E.; Taylor, B. S.; Krivokhizhina, T.; Rai, S. N.; Keith, R. J.; Bhatnagar, A.; Srivastava, S. Environmental exposure to volatile organic compounds is associated with endothelial injury. *Toxicol Appl Pharmacol* **2022**, *437*, 115877. DOI: 10.1016/j.taap.2022.115877 From NLM Medline.
(30) Etemadi, A.; Poustchi, H.; Chang, C. M.; Blount, B. C.; Calafat, A. M.; Wang, L. Q.; De Jesus, V. R.; Pourshams, A.; Shakeri, R.; Shiels, M. S.; et al. Urinary Biomarkers of Carcinogenic Exposure among Cigarette, Waterpipe, and Smokeless Tobacco Users and Never Users of Tobacco in the Golestan Cohort Study. *Cancer Epidem Biomar* **2019**, *28* (2), 337-347. DOI: 10.1158/1055-9965.Epi-18-0743.
(31) Fishbein, L. Potential halogenated industrial carcinogenic and mutagenic chemicals. III. Alkane halides, alkanols and ethers. *Sci Total Environ* **1979**, *11* (3), 223-257. DOI: 10.1016/0048-9697(79)90076-7 From NLM Medline.
(32) Srivastava, S.; Chandra, A.; Wang, L. F.; Seifert, W. E.; DaGue, B. B.; Ansari, N. H.; Srivastava, S. K.; Bhatnagar, A. Metabolism of the lipid peroxidation product, 4-hydroxy-trans-2-nonenal, in isolated perfused rat heart. *J Biol Chem* **1998**, *273* (18), 10893-10900. DOI: 10.1074/jbc.273.18.10893.
(33) Srivastava, S.; Watowich, S. J.; Petrash, J. M.; Srivastava, S. K.; Bhatnagar, A. Structural and kinetic determinants of aldehyde reduction by aldose reductase. *Biochemistry-Us* **1999**, *38* (1), 42-54. DOI: 10.1021/bi981794l.
(34) Vladykovskaya, E.; Sithu, S. D.; Haberzettl, P.; Wickramasinghe, N. S.; Merchant, M. L.; Hill, B. G.; McCracken, J.; Agarwal, A.; Dougherty, S.; Gordon, S. A.; et al. Lipid peroxidation product 4-hydroxy-trans-2-nonenal causes endothelial activation by inducing endoplasmic reticulum stress. *J Biol Chem* **2012**, *287* (14), 11398-11409. DOI: 10.1074/jbc.M111.320416 From NLM.
(35) Srivastava, S.; Sithu, S. D.; Vladykovskaya, E.; Haberzettl, P.; Hoetker, D. J.; Siddiqui, M. A.; Conklin, D. J.; D'Souza, S. E.; Bhatnagar, A. Oral exposure to acrolein exacerbates atherosclerosis in apoE-null mice. *Atherosclerosis* **2011**, *215* (2), 301-308. DOI: 10.1016/j.atherosclerosis.2011.01.001.
(36) Major, R. T.; Collins, O. D.; Marchini, P.; Schnabel, H. W. Formation of 2-hexenal by leaves. *Phytochemistry* **1972**, *11* (2), 607-610.

(37) Venkateshwarlu, C.; Chandravadana, M. V.; Tewari, R. P. Volatile flavour components of some edible mushrooms (Basidiomycetes). *Flavour and Fragrance Journal* **1999**, *14* (3), 191-194.
(38) Goniewicz, M. L.; Smith, D. M.; Edwards, K. C.; Blount, B. C.; Caldwell, K. L.; Feng, J.; Wang, L.; Christensen, C.; Ambrose, B.; Borek, N.; et al. Comparison of Nicotine and Toxicant Exposure in Users of Electronic Cigarettes and Combustible Cigarettes. *JAMA Network Open* **2018**, *1* (8). DOI: 10.1001/jamanetworkopen.2018.5937.
(39) Alwis, K. U.; Bailey, T. L.; Patel, D.; Wang, L. Q.; Blount, B. C. Measuring urinary N-acetyl-S-(4-hydroxy-2-methyl-2-buten-1-yl)-L-cysteine (IPMA3) as a potential biomarker of isoprene exposure. *Anal Chim Acta* **2016**, *941*, 61-66. DOI: 10.1016/j.aca.2016.08.023.
(40) Eckert, E.; Gries, W.; Goen, T.; Leng, G. Evidence for S-(3,4-dihydroxybutyl) mercapturic acid as the main metabolite of 2-chloroprene in humans. *N-S Arch Pharmacol* **2011**, *383*, 92-92.
(41) Carrieri, M.; Tranfo, G.; Pigini, D.; Paci, E.; Salamon, F.; Scapellato, M. L.; Fracasso, M. E.; Manno, M.; Bartolucci, G. B. Correlation between environmental and biological monitoring of exposure to benzene in petrochemical industry operators. *Toxicol Lett* **2010**, *192* (1), 17-21. DOI: 10.1016/j.toxlet.2009.07.015.
(42) Sabatini, L.; Barbieri, A.; Indiveri, P.; Mattioli, S.; Violante, F. S. Validation of an HPLC-MS/MS method for the simultaneous determination of phenylmercapturic acid, benzylmercapturic acid and o-methylbenzyl mercapturic acid in urine as biomarkers of exposure to benzene, toluene and xylenes. *J Chromatogr B Analyt Technol Biomed Life Sci* **2008**, *863* (1), 115-122. DOI: 10.1016/j.jchromb.2008.01.022.
(43) Rodgman, A.; Smith, C. J.; Perfetti, T. A. The composition of cigarette smoke: a retrospective, with emphasis on polycyclic components. *Hum Exp Toxicol* **2000**, *19* (10), 573-595. DOI: 10.1191/096032700701546514 From NLM Medline.
(44) Walker, A. M.; Cohen, A. J.; Loughlin, J. E.; Rothman, K. J.; DeFonso, L. R. Mortality from cancer of the colon or rectum among workers exposed to ethyl acrylate and methyl methacrylate. *Scand J Work Environ Health* **1991**, *17* (1), 7-19. DOI: 10.5271/sjweh.1731 From NLM Medline.
(45) Kostrzewski, P.; Wiaderna-Brycht, A.; Czerski, B. Biological monitoring of experimental human exposure to trimethylbenzene. *Sci Total Environ* **1997**, *199* (1-2), 73-81. DOI: 10.1016/s0048-9697(97)05504-6 From NLM Medline.
(46) Yu, W. H.; Zhang, Z.; Wang, H.; Ge, Z. H.; Pinnavala, T. J. Peroxide oxidation of 4-tert-butyltoluene to 4-tert-butylbenzaldehyde over titanium(IV)-functionalized mesostructured silica. *Micropor Mesopor Mat* **2007**, *104* (1-3), 151-158. DOI: 10.1016/j.micromeso.2007.01.022.
(47) Jones, C. R.; Liu, Y. Y.; Sepai, O.; Yan, H.; Sabbioni, G. Internal exposure, health effects, and cancer risk of humans exposed to chloronitrobenzene. *Environ Sci Technol* **2006**, *40* (1), 387-394. DOI: 10.1021/es050693p From NLM Medline.
(48) Ghalla, H.; Issaoui, N.; Bardak, F.; Atac, A. Intermolecular interactions and molecular docking investigations on 4-methoxybenzaldehyde. *Computational Materials Science* **2018**, *149*, 291-300. DOI: 10.1016/j.commatsci.2018.03.042.
(49) Taylor, R.; Pergolizzi, J. V.; Raffa, R. B. Acetaminophen (paracetamol): properties, clinical uses, and adverse effects. In *Acetaminophen: Properties, Clinical Uses and Adverse Effects*, Nova Science Publishers, Inc., 2012; pp 1-24.

(50) Bendich, A.; Russell Jr, P. J.; Fox, J. J. The Synthesis and Properties of 6-Chloropurine and Purine1. *J. Am. Chem. Soc.* **1954**, *76* (23), 6073-6077.

Figure 1.

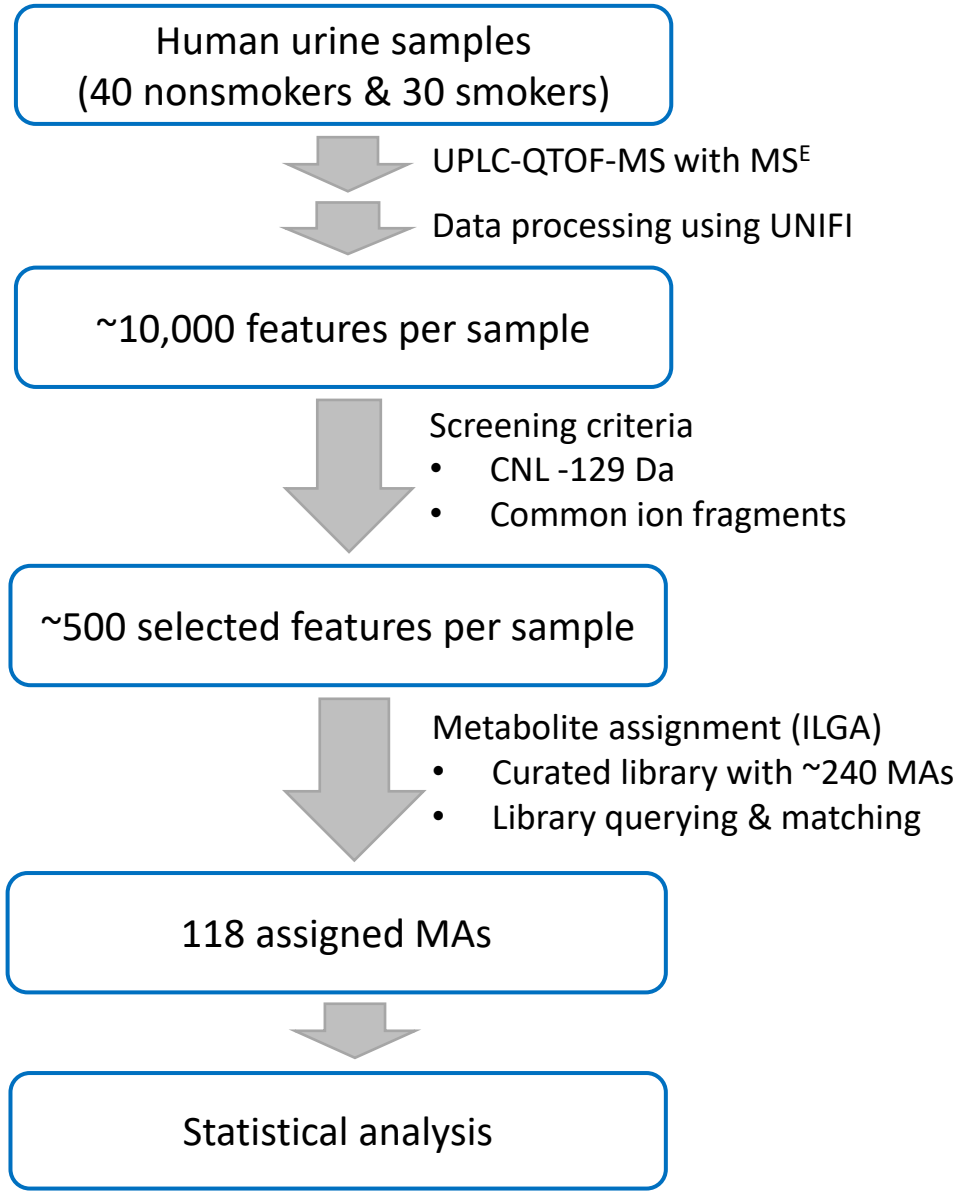

Human urine samples
(40 nonsmokers & 30 smokers)
UPLC-QTOF-MS with MS^E
Data processing using UNIFI
~10,000 features per sample
Screening criteria
• CNL -129 Da
• Common ion fragments
~500 selected features per sample
Metabolite assignment (ILGA)
• Curated library with ~240 MAs
• Library querying & matching
118 assigned MAs
Statistical analysis

Figure 2.

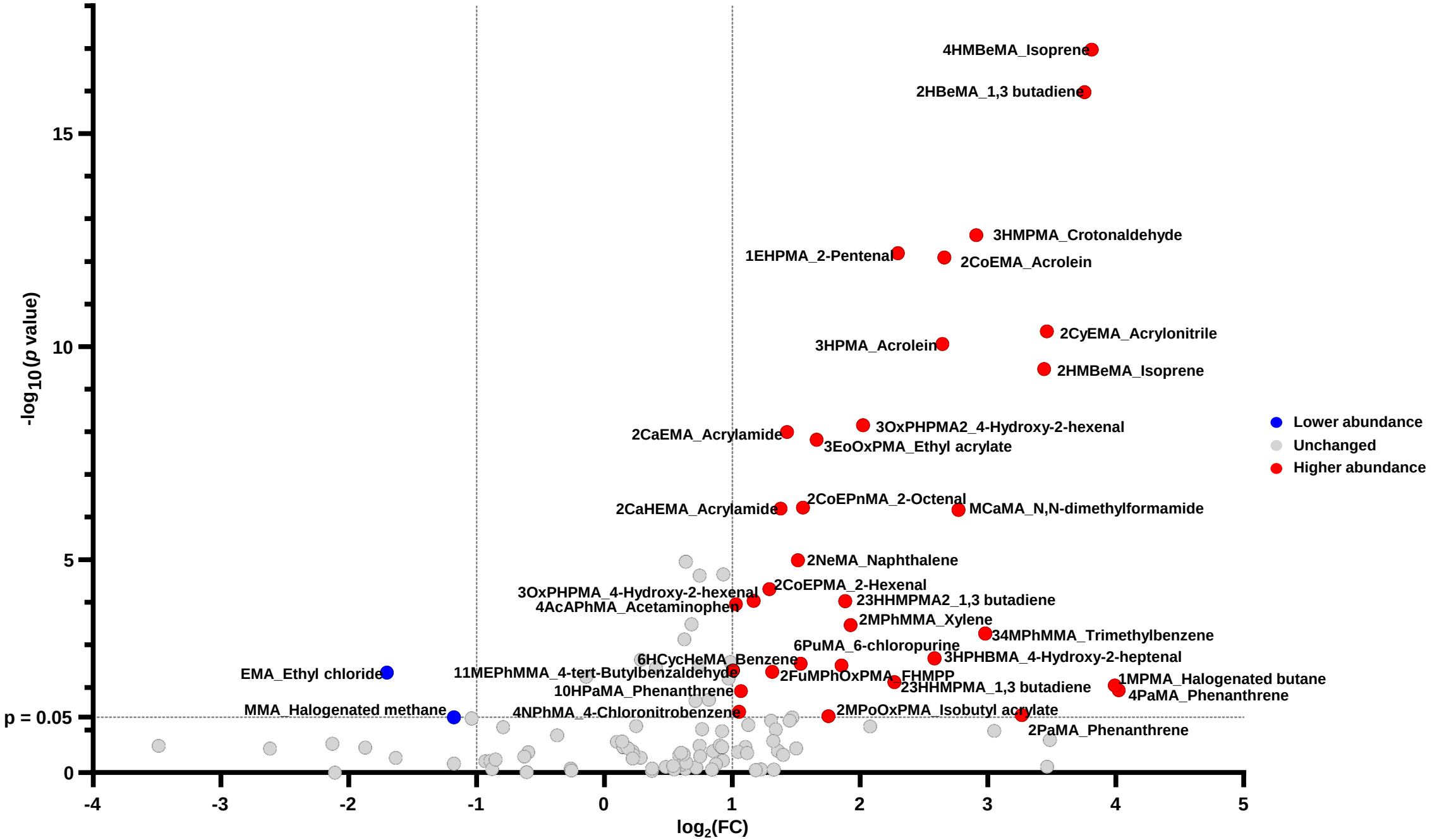

4HMBeMA_Isoprene
2HBeMA_1,3 butadiene
3HMPMA_Crotonaldehyde
1EHPMA_2-Pentenal
2CoEMA_Acrolein
2CyEMA_Acrylonitrile
3HPMA_Acrolein
2HMBeMA_Isoprene
2CaEMA_Acrylamide
3OxPHPMA2_4-Hydroxy-2-hexenal
3EoOxPMA_Ethyl acrylate
2CaHEMA_Acrylamide
2CoEPnMA_2-Octenal
MCaMA_N,N-dimethylformamide
2NeMA_Naphthalene
3OxPHPMA_4-Hydroxy-2-hexenal
2CoEPMA_2-Hexenal
4AcAPhMA_Acetaminophen
23HHMPMA2_1,3 butadiene
2MPhMMA_Xylene
34MPhMMA_Trimethylbenzene
6PuMA_6-chloropurine
3HPHBMA_4-Hydroxy-2-heptenal
6HCycHeMA_Benzene
EMA_Ethyl chloride
11MEPhMMA_4-tert-Butylbenzaldehyde
2FuMPhOxPMA_FHMPP
1MPMA_Halogenated butane
23HHMPMA_1,3 butadiene
10HPaMA_Phenanthrene
4PaMA_Phenanthrene
MMA_Halogenated methane
4NPhMA_4-Chloronitrobenzene
2MPoOxPMA_Isobutyl acrylate
2PaMA_Phenanthrene
Lower abundance
Unchanged
Higher abundance
-log10(p value)
p = 0.05
log2(FC)
15
10
5
0
-4
-3
-2
-1
0
1
2
3
4
5

Figure 3.

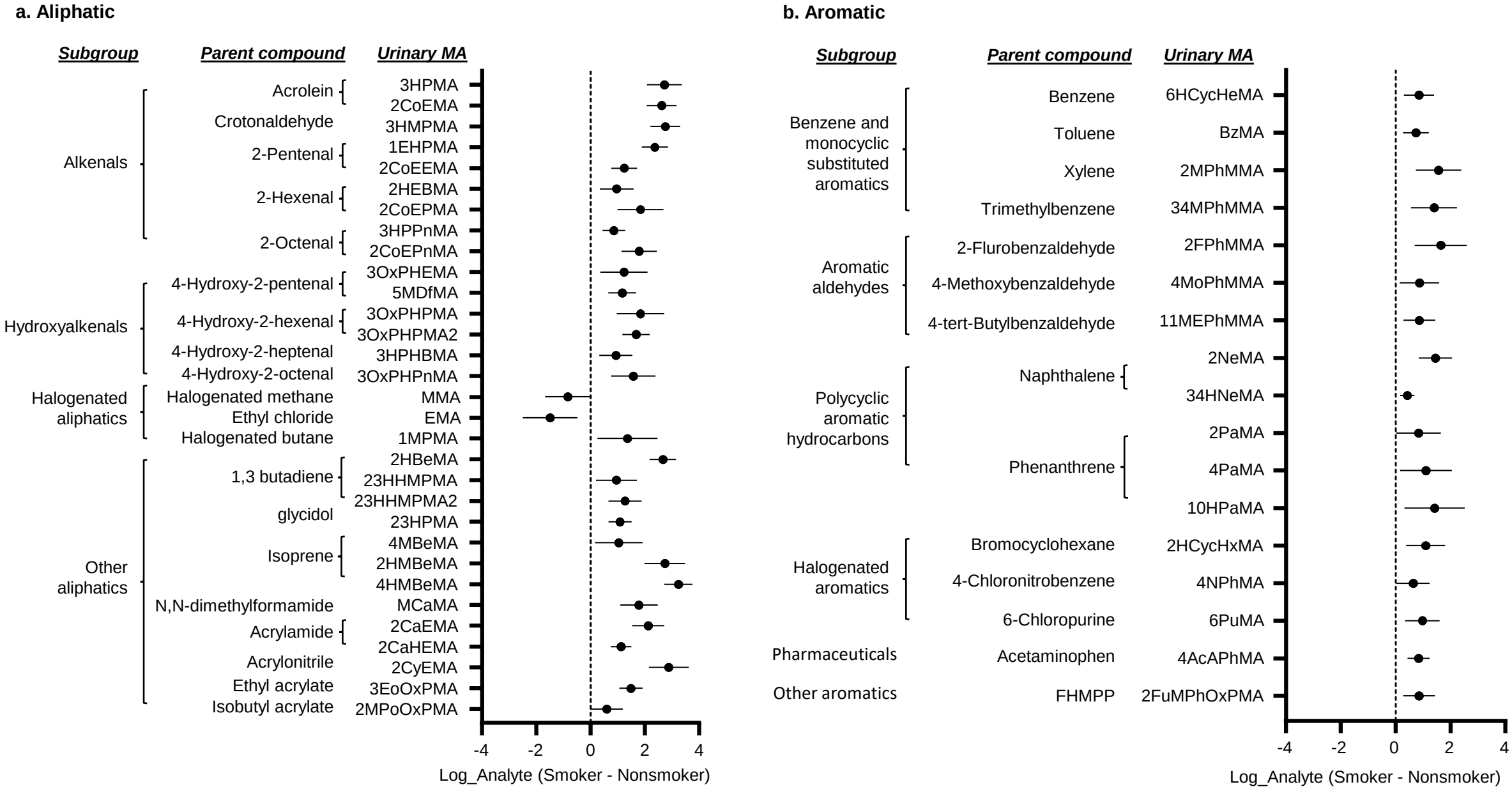
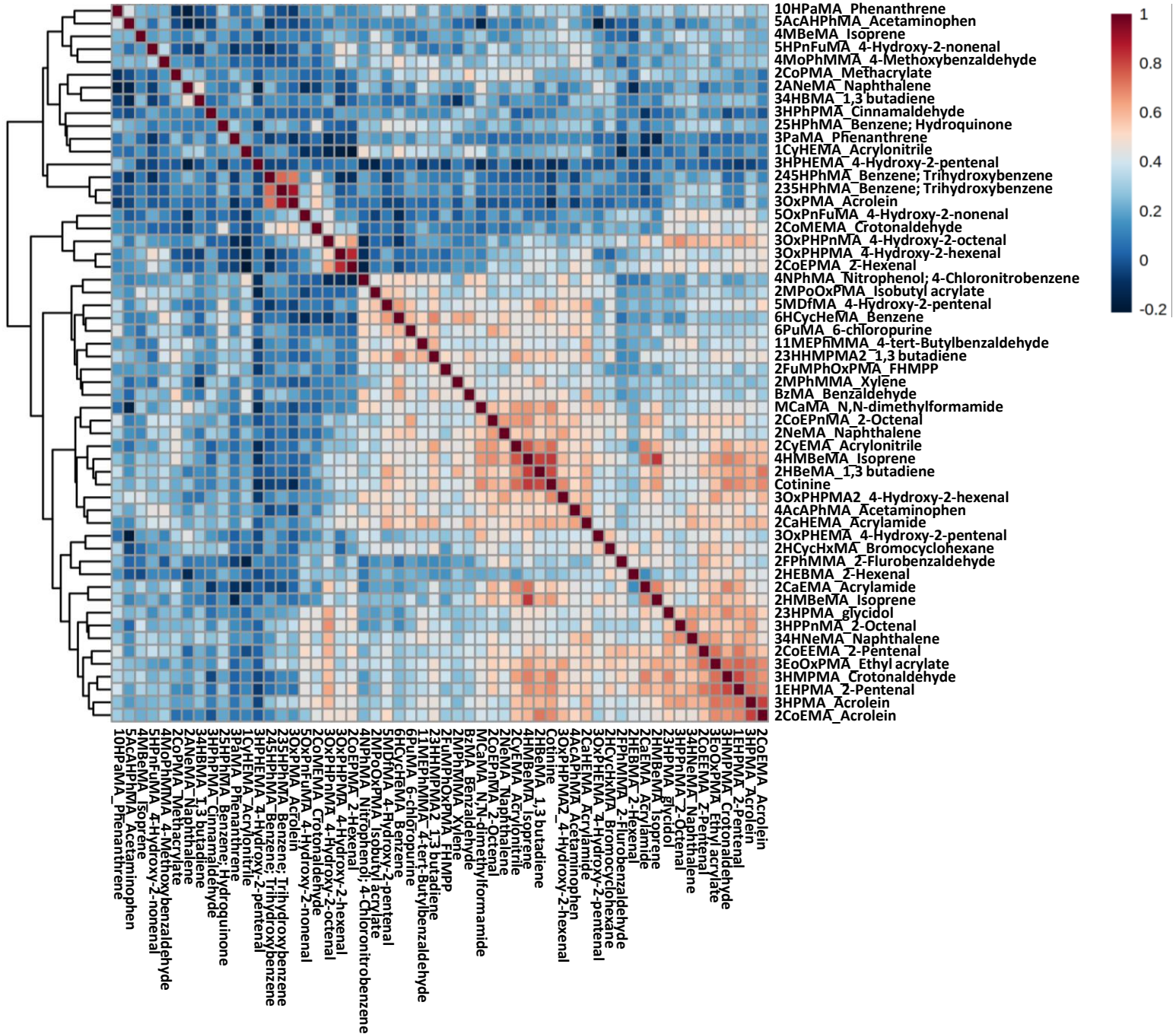
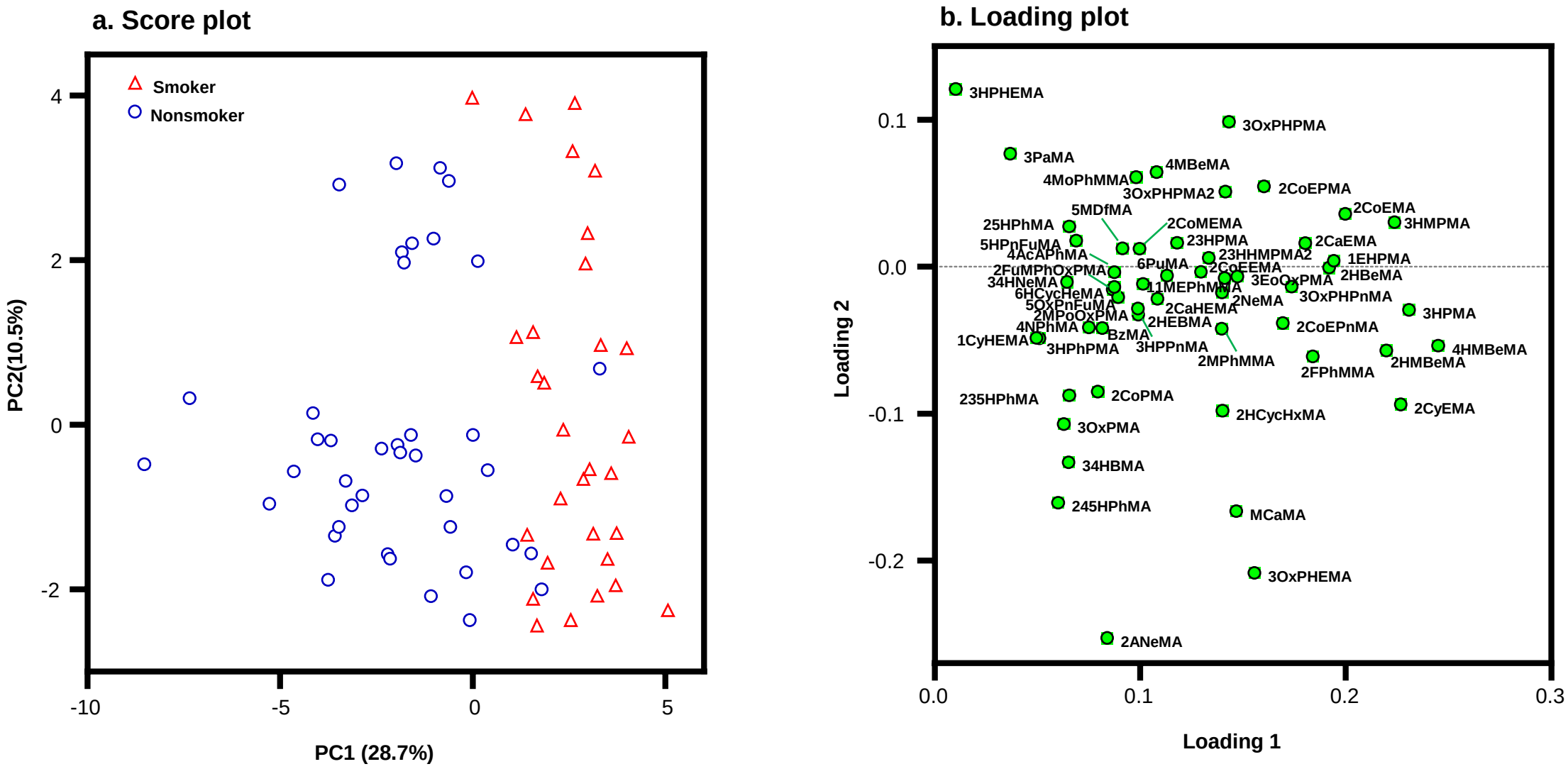

a. Aliphatic
Subgroup
Parent compound
Urinary MA
Alkenals
Hydroxyalkenals
Halogenated aliphatics
Other aliphatics
Acrolein
Crotonaldehyde
2-Pentenal
2-Hexenal
2-Octenal
4-Hydroxy-2-pentenal
4-Hydroxy-2-hexenal
4-Hydroxy-2-heptenal
4-Hydroxy-2-octenal
Halogenated methane
Ethyl chloride
Halogenated butane
1,3 butadiene
glycidol
Isoprene
N,N-dimethylformamide
Acrylamide
Acrylonitrile
Ethyl acrylate
Isobutyl acrylate
3HPMA
2CoEMA
3HMPMA
1EHPMA
2CoEEMA
2HEBMA
2CoEPMA
3HPPnMA
2CoEPnMA
3OxPHEMA
5MDfMA
3OxPHPMA
3OxPHPMA2
3HPHBMA
3OxPHPnMA
MMA
EMA
1MPMA
2HBeMA
23HHMPMA
23HHMPMA2
23HPMA
4MBeMA
2HMBeMA
4HMBeMA
MCaMA
2CaEMA
2CaHEMA
2CyEMA
3EoOxPMA
2MPoOxPMA
-4
-2
0
2
4
Log_Analyte (Smoker - Nonsmoker)
b. Aromatic
Subgroup
Parent compound
Urinary MA
Benzene and monocyclic substituted aromatics
Aromatic aldehydes
Polycyclic aromatic hydrocarbons
Halogenated aromatics
Pharmaceuticals
Other aromatics
Benzene
Toluene
Xylene
Trimethylbenzene
2-Flurobenzaldehyde
4-Methoxybenzaldehyde
4-tert-Butylbenzaldehyde
Naphthalene
Phenanthrene
Bromocyclohexane
4-Chloronitrobenzene
6-Chloropurine
Acetaminophen
FHMPP
6HCycHeMA
BzMA
2MPhMMA
34MPhMMA
2FPhMMA
4MoPhMMA
11MEPhMMA
2NeMA
34HNeMA
2PaMA
4PaMA
10HPaMA
2HCycHxMA
4NPhMA
6PuMA
4AcAPhMA
2FuMPhOxPMA
-4
-2
0
2
4
Log_Analyte (Smoker - Nonsmoker)

Figure 4.



55 MAs were detected in more than 50% of smokers

Figure 5.


a. Score plot
Smoker
Nonsmoker
PC2(10.5%)
PC1 (28.7%)
4
2
0
-2
-10
-5
0
5
b. Loading plot
Loading 2
Loading 1
0.1
0.0
-0.1
-0.2
0.0
0.1
0.2
0.3
3HPHEMA
3OxPHPMA
3PaMA
4MBeMA
4MoPhMMA
3OxPHPMA2
2CoEPMA
2CoEMA
3HMPMA
5MDfMA
25HPhMA
2CoMEMA
23HPMA
2CaEMA
5HPnFuMA
4AcAPhMA
23HHMPMA2
1EHPMA
6PuMA
2FuMPhOxPMA
2CoEEMA
2HBeMA
34HNeMA
3EoOxPMA
11MEPhMMA
6HCycHeMA
2NeMA
3OxPHPnMA
5OxPnFuMA
2CaHEMA
3HPMA
2MPoOxPMA
2HEBMA
2CoEPnMA
4NPhMA
BzMA
1CyHEMA
3HPhPMA
3HPPnMA
2MPhMMA
4HMBeMA
2HMBeMA
2FPhMMA
235HPhMA
2CoPMA
2CyEMA
2HCycHxMA
3OxPMA
34HBMA
245HPhMA
MCaMA
3OxPHEMA
2ANeMA

Global Profiling of Urinary Mercapturic Acids Using Integrated Library-Guided Analysis

Zhengzhi Xie[1 a,b,c,f], Jin Y. Chen[1,a,b,c,f], Hong Gao[, a,b,c,f], Rachel J. Keith[a,b,c,f], Aruni Bhatnagar[a,b,c,f], Pawel Lorkiewicz*[a,b,c,d,e,f], and Sanjay Srivastava[a,b,c,f]

[1] Co-first author
[a]American Heart Association-Tobacco Regulation and Addiction Center, University of Louisville;
[b]Christina Lee Brown Envirome Institute, University of Louisville;
[c]Superfund Research Center, University of Louisville;
[d]Department Center for Cardiometabolic Science, University of Louisville;
[e]Department of Chemistry, University of Louisville; and,
[f]Division of Environmental Medicine, Department of Medicine, University of Louisville, Louisville, KY 40202, USA

Supplementary Materials

Figure Legends

Figure 1S. MS/MS fragments and features specific to mercapturic acids. Representative MS and MS/MS spectra are shown for metabolite N-acetyl-S-[1-(hydroxymethyl)-2-propenyl]-L-cysteine (1HMPeMA).

Figure 2S. Metabolic routes leading to MA formation from acrolein. Dashed arrow represents the proposed MA structure. S-NAc = S-N-acetyl-L-cysteine.

Figure 3S. Validation of putative assignment using an authentic standard. Example of 3HPMA. Upper insert represents chromatograms and MS, and MS/MS spectra of 3HPMA standard. Lower panel represents the corresponding results for the putatively assigned 3HPMA in a urine sample.

Figure 4S. Correlation heatmap of urinary cotinine and 118 MAs in the total selected subjects (n = 70) including smokers and nonsmokers. The dendrogram was generated by hierarchical cluster analysis based on Pearson correlation coefficient (r). The MAs and cotinine are in both columns and rows with their r values represented by color (blue is low and red is high).

Table 1S: Substituent abbreviations for MAs.

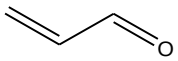
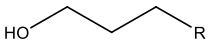
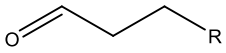
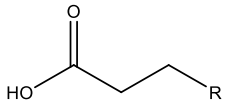
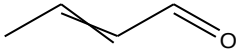
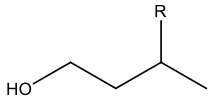
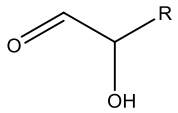
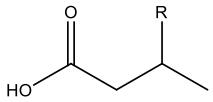
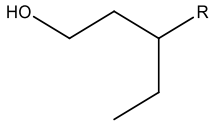
Substituent	Abbreviation
Hydroxy	H
Methyl	M
Ethyl	E
Propyl	P
Butyl	B
Pentyl	Pn
Hexyl	Hx
Heptyl	Hp
Propenyl	Pe
Butenyl	Be
Propynyl	Py
Butynyl	By
Hexadienyl	He
Carboxy	Co
Methoxy	Mo
Ethoxy	Eo
Propoxy	Po
Butoxy	Bo
Hexoxy	Ho
Cyano	Cy
Fluoro	F
Chloro	C
Bromo	Br
Iodo	I
Phenyl	Ph
Benzyl	Bz
Thiazoline	T
Thiazolidine	Tl
Thioxo	To
Vinyl	V
Nitro	N
Amino	A
Carbamoyl	Ca
Naphthyl	Np

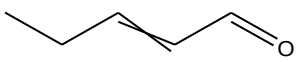
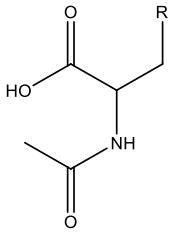
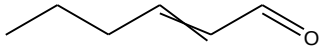
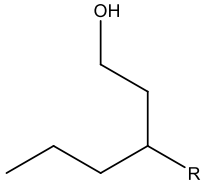
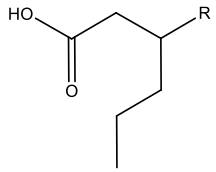
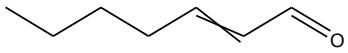
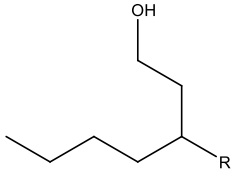
Naphthalenyl	Ne
Phenanthrenyl	Pa
Furanyl	Fu
Acetyl	Ac
Purinyl	Pu
Sulfinyl	Sf
Oxo	Ox
Cyclo	Cyc
Benzo	Bzo
Pyranyl	Pr
Thio	Th
Cysteine	Cs
Dihydrofuryl	Df
Benzoquinone	Bq
Coumarinyl	Cou

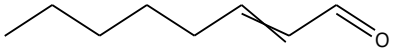
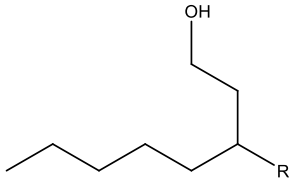
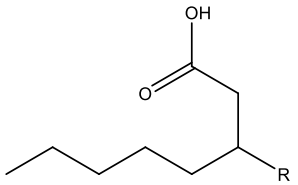
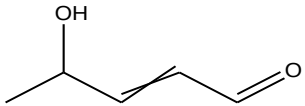
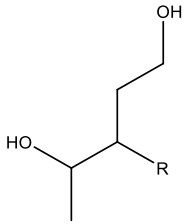
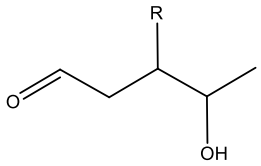
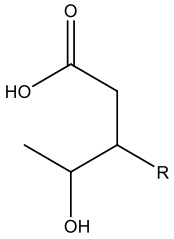
Table 2S: Confidence levels of compound annotations adapted from Blazenovic et al., 2018.

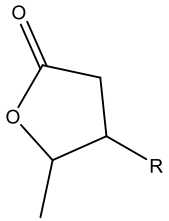
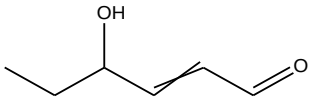
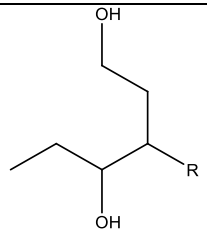
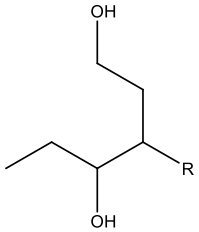
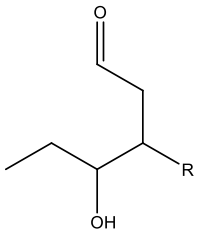
Confidence Level	Description	Data requirements
Level 0	Unambiguous 3D structure. Isolated pure compomd, including full stereochemistry.	Following natural product guidelines for 3D structure determination.
Level 1	Confident 2D structure. Matches reference standard or full structure elucidation.	At least MS/MS and RT match.
Level 2	Probable structure. Matches curated library and/or other databases, and literature.	At least two orthogonal pieces of information, including evidence excluding other candidates.
Level 3	Possible MA. Matches one or more selection criteria (neutral loss m/z 129.043 Da (-$C_5H_7NO_3$), or at least one of the ion fragments specific to N-acetyl-L-cysteine moiety at m/z 74.020 (C_3H_6S), 84.045 (C_4H_6NO), m/z 128.035 ($C_5H_6NO_3$), and m/z 162.023 ($C_5H_8NO_3S$) in MS/MS spectra (function 2)	One or more candidates are possible. Requires at least one piece of information that supports the candidate.
Level 4	Unknown feature	Found in sample

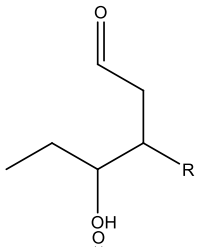
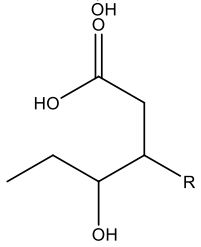
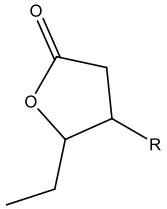
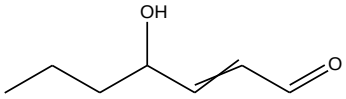
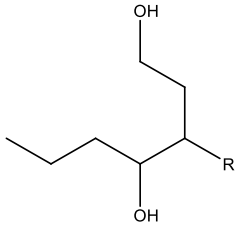
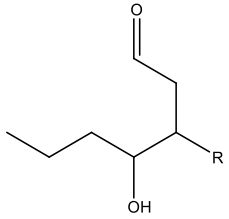
Table 3S: Names and structures of proposed parent compounds and 118 putative identified MAs. Bolded MAs indicated significant changes in levels between smokers and nonsmokers. Underlined MAs were validated using authentic standards.

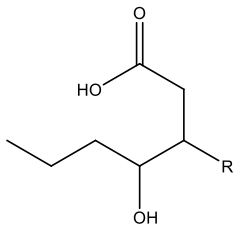
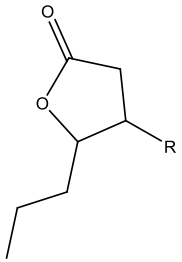
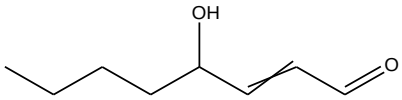
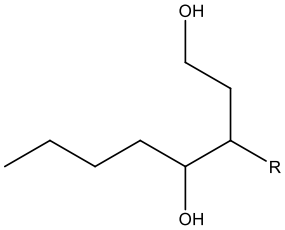
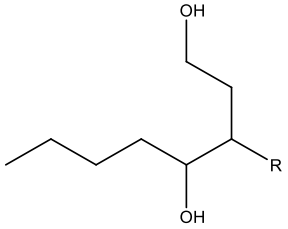
Proposed parent & structure	MA full name & acronym	Structure	CAS #	Confidence Level	Detected in
		Alkenals			
Acrolein 	**N-Acetyl-S-(3-hydroxypropyl)cysteine; 3HPMA**		23127-40-4	1	Smokers & Nonsmokers
	N-Acetyl-S-(3-oxopropyl)-L-cysteine; 3OxPMA		140226-30-8	2	Smokers & Nonsmokers
	N-acetyl-S-(2-carboxyethyl)-L-cysteine; 2CoEMA		51868-61-2	1	Smokers & Nonsmokers
Crotonaldehyde 	**N-Acetyl-S-(3-hydroxy-1-methylpropyl)-L-cysteine; 3HMPMA**		33164-64-6	1	Smokers & Nonsmokers
	N-Acetyl-S-(1-hydroxy-2-oxoethyl)-L-cysteine; 1HOxEMA		1456776-73-0	2	Smokers & Nonsmokers
	2-Carboxy-1-methylethylmercapturic acid; 2CoMEMA		33164-65-7	2	Smokers & Nonsmokers
	N-Acetyl-S-(1-ethyl-3-hydroxypropyl)-L-cysteine; 1EHPMA		207226-97-9	2	Smokers & Nonsmokers

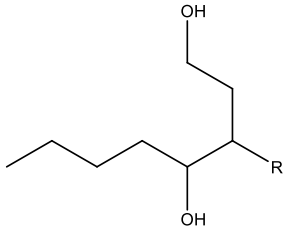
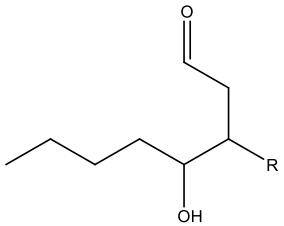
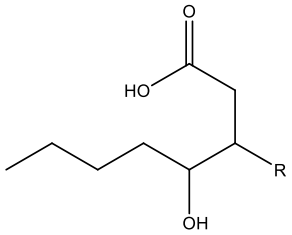
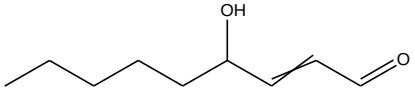
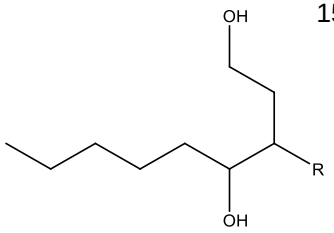
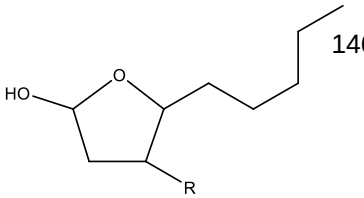
2-Pentenal 	**2-Carboxyethyl-1-ethylmercapturic acid; 2CoEEMA***		N.A.	2	Smokers & Nonsmokers
2-Hexenal 	**N-Acetyl-S-[1-(2-hydroxyethyl)butyl]-L-cysteine; 2HEBMA**		879555-83-6	2	Smokers & Nonsmokers
	2-Carboxyethyl-1-propylmercapturic acid; 2CoEPMA*		N.A.	2	Smokers & Nonsmokers
2-Heptenal 	3-Hydroxypropyl-1-butylmercapturic acid; 3HPBMA*		N.A.	2	Smokers & Nonsmokers

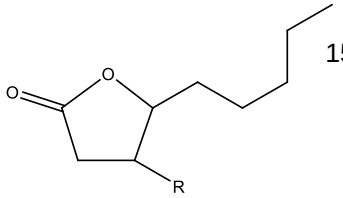
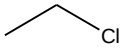
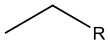
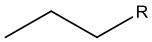
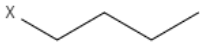
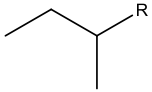
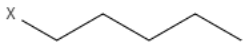
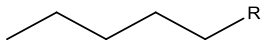
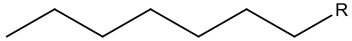
2-Octenal 	**3-Hydroxypropyl-1-pentylmercapturic acid; 3HPPnMA***		N.A.	2	Smokers & Nonsmokers
	2-Carboxyethyl-1-pentylmercapturic acid; 2CoEPnMA*		N.A.	2	Smokers & Nonsmokers
		Hydroxyalkenals			
4-Hydroxy-2-pentenal 	3-Hydroxypropyl-1-(2-hydroxyethyl)mercapturic acid; 3HPHEMA*		N.A.	2	Smokers & Nonsmokers
	3-Oxopropyl-1-(2-hydroxyethyl)mercapturic acid; 3OxPHEMA*		N.A.	2	Smokers & Nonsmokers
	Carboxyethyl-1-(2-hydroxyethyl)mercapturic acid; CoEHEMA*		N.A.	2	Smokers & Nonsmokers

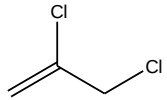
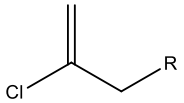
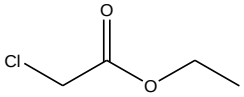
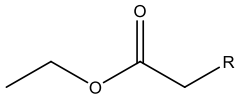
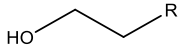
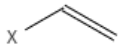
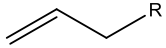
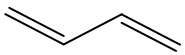
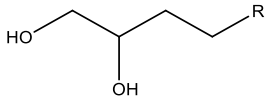
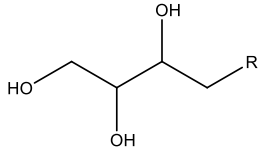
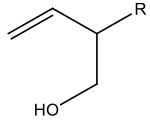
	4-(5-Methyl-dihydrofuryl)mercapturic acid; 5MDfMA*		N.A.	2	Smokers & Nonsmokers
4-Hydroxy-2-hexenal 	3-Hydroxypropyl-1-(2-hydroxypropyl)mercapturic acid; 3HPHPMA*		N.A.	2	Smokers & Nonsmokers
	3-Hydroxypropyl-1-(2-hydroxypropyl)mercapturic acid; 3HPHPMA2*		N.A.	2	Smokers
	3-Oxopropyl-1-(2-hydroxypropyl)mercapturic acid; 3OxPHPMA*		N.A.	2	Smokers & Nonsmokers

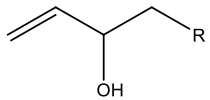
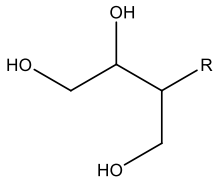
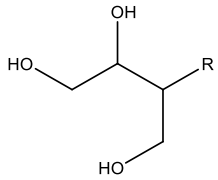
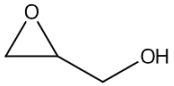
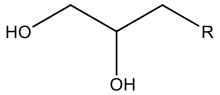
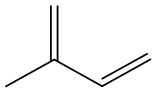
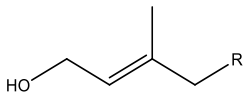
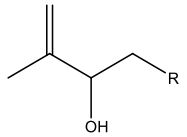
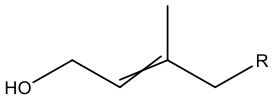
	3-Oxopropyl-1-(2-hydroxypropyl)mercapturic acid; 3OxPHPMA2* 		N.A.	2	Smokers & Nonsmokers
	Carboxyethyl-1-(2-hydroxypropyl)mercapturic acid; CoEHPMA* 		N.A.	2	Smokers & Nonsmokers
	4-(5-Ethyl-Dihydrofuryl)mercapturic acid; 5EDfMA* 		N.A.	2	Smokers & Nonsmokers
4-Hydroxy-2-heptenal 	**3-Hydroxypropyl-1-(2-hydroxybutyl)mercapturic acid; 3HPHBMA*** 		N.A.	2	Smokers & Nonsmokers
	3-Oxopropyl-1-(2-hydroxybutyl)mercapturic acid; 3OxPHBMA* 		N.A.	2	Smokers & Nonsmokers

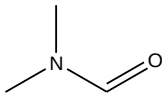
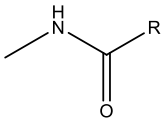
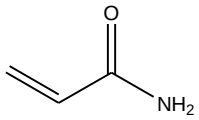
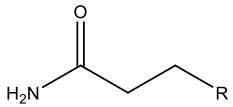
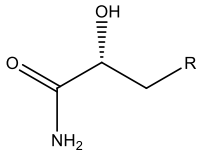
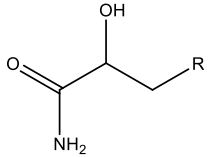
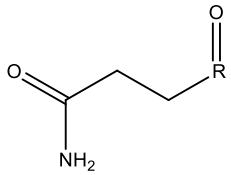
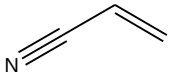
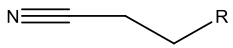
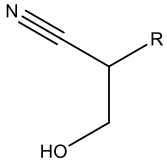
	Carboxyethyl-1-(2-hydroxybutyl)mercapturic acid; CoEHBMA*		N.A.	2	Smokers & Nonsmokers
	4-(5-Propyl-Dihydrofuryl)mercapturic acid; 5PDfMA*		N.A.	2	Smokers & Nonsmokers
4-Hydroxy-2-octenal 	3-Hydroxypropyl-1-(2-hydroxypentyl)mercapturic acid; 3HPHPnMA*		N.A.	2	Smokers & Nonsmokers
	3-Hydroxypropyl-1-(2-hydroxypentyl)mercapturic acid; 3HPHPnMA2*		N.A.	2	Smokers & Nonsmokers

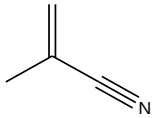
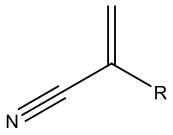
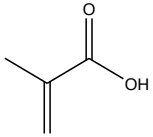
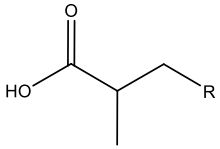
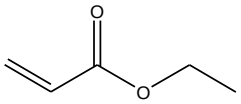
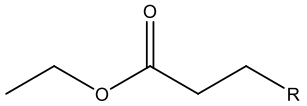
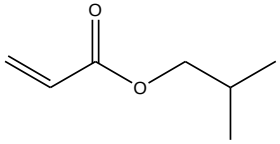
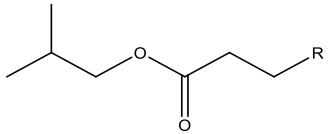
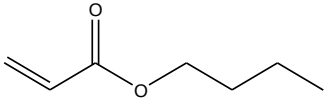
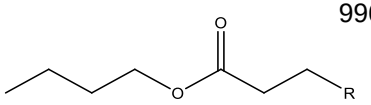
	3-Hydroxypropyl-1-(2-hydroxypentyl)mercapturic acid; 3HPHPnMA3*		N.A.	2	Smokers & Nonsmokers
	3-Oxopropyl-1-(2-hydroxypentyl)mercapturic acid; 3OxPHPnMA*		N.A.	2	Smokers & Nonsmokers
	Carboxyethyl-1-(2-hydroxypentyl)mercapturic acid; CoEHPnMA*		N.A.	2	Smokers & Nonsmokers
4-Hydroxy-2-nonenal 	N-Acetyl-S-[2-hydroxy-1-(2-hydroxyethyl)heptyl]-L-cysteine; 2HHEHpMA		158822-69-6	2	Smokers & Nonsmokers
	N-Acetyl-S-(tetrahydro-5-hydroxy-2-pentyl-3-furanyl)-L-cysteine; 5HPnFuMA		146764-24-1	2	Smokers & Nonsmokers

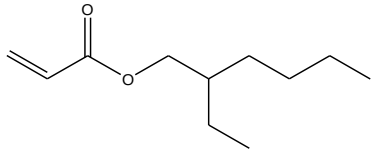
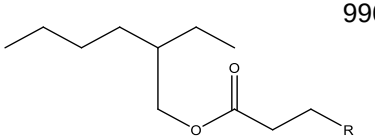
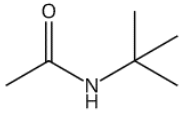
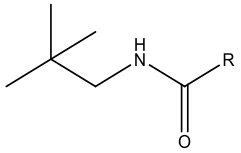
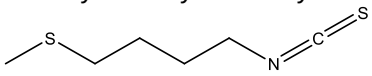
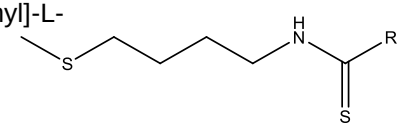
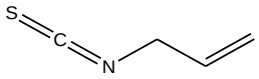
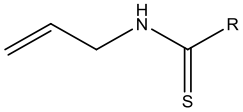
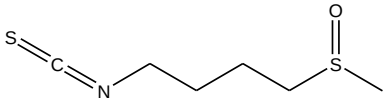
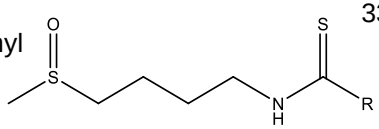
	N-Acetyl-S-(tetrahydro-5-oxo-2-pentyl-3-furanyl)-L-cysteine; 5OxPnFuMA		158822-71-0	2	Smokers & Nonsmokers
	Halogenated aliphatics				
Halogenated methane 	**N-Acetyl-S-methylcysteine; MMA**		24947-73-7	2	Smokers & Nonsmokers
Ethyl chloride 	**N-Acetyl-S-ethylcysteine; EMA**		31386-36-4	2	Smokers & Nonsmokers
Halogenated propane 	N-Acetyl-S-(n-propyl)-l-cysteine; 1PMA		14402-54-1	1	Smokers & Nonsmokers
Halogenated butane 	**N-Acetyl-S-(1-methylpropyl)-L-cysteine; 1MPMA**		101871-35-6	2	Smokers & Nonsmokers
Halogenated pentane 	N-Acetyl-S-pentyl-L-cysteine; PnMA		35985-42-3	2	Smokers & Nonsmokers
Halogenated heptane 	N-Acetyl-S-heptyl-L-cysteine; HpMA		35874-34-1	2	Smokers & Nonsmokers

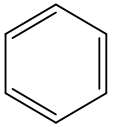
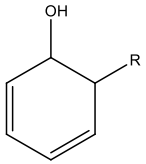
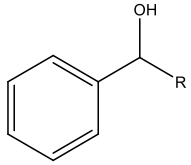
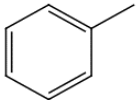
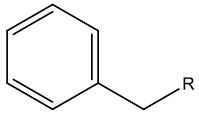
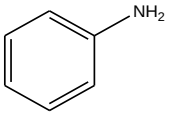
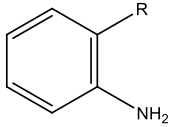
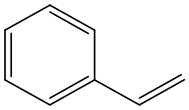
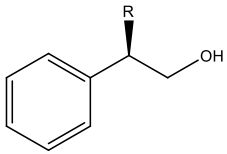
2,3-Dichloro-1-propene 	N-Acetyl-S-(2-chloro-2-propen-1-yl)-L-cysteine; CPeMA		109702-06-9	2	Smokers & Nonsmokers
Ethyl chloroacetate 	N-Acetyl-S-(2-ethoxy-2-oxoethyl)-L-cysteine; 3EoOxEMA		77549-13-4	2	Smokers & Nonsmokers
	Other aliphatics				
Ethylene oxide 	N-Acetyl-S-(2-hydroxyethyl)-L-cysteine; 2HEMA		15060-26-1	1	Smokers & Nonsmokers
Allyl halide 	N-Acetyl-S-2-propen-1-yl-L-cysteine; 2PeMA		23127-41-5	2	Smokers & Nonsmokers
1,3-Butadiene 	N-Acetyl-S-(3,4-dihydroxybutyl)-L-cysteine; 34HBMA		144889-50-9	1	Smokers & Nonsmokers
	N-Acetyl-S-(2,3,4-trihydroxybutyl)-L-cysteine; 234HBMA		219965-90-9	1	Smokers & Nonsmokers
	N-acetyl-S-[1-(hydroxymethyl)-2-propenyl]-L-cysteine; 1HMPeMA		144889-51-0	2	Smokers & Nonsmokers

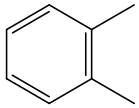
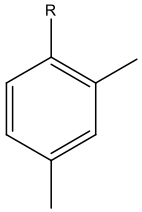
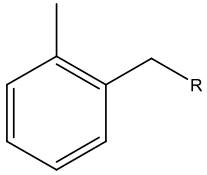
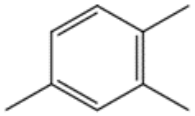
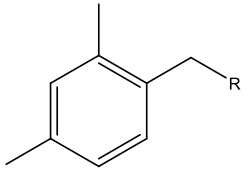
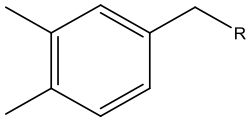
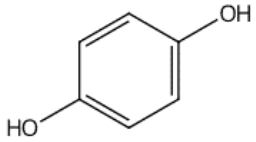
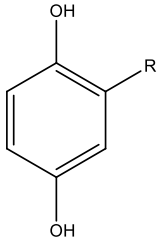
	N-Acetyl-S-(2-hydroxy-3-buten-1-yl)-L-cysteine; 2HBeMA		159092-64-5	2	Smokers & Nonsmokers
	N-Acetyl-S-[2,3-dihydroxy-1-(hydroxymethyl)propyl]-L-cysteine; 23HHMPMA		176914-98-0	1	Smokers & Nonsmokers
	N-Acetyl-S-[2,3-dihydroxy-1-(hydroxymethyl)propyl]-L-cysteine; 23HHMPMA2		176914-98-0	2	Smokers & Nonsmokers
Glycidol 	**N-Acetyl-S-(2,3-dihydroxypropyl)-L-cysteine; 23HPMA**		23255-33-6	1	Smokers & Nonsmokers
Isoprene 	**L-Cysteine, N-acetyl-S-[(2E)-4-hydroxy-2-methyl-2-buten-1-yl]-; 4MBeMA**		2165415-10-9	2	Smokers & Nonsmokers
	L-Cysteine, N-acetyl-S-(2-hydroxy-3-methyl-3-buten-1-yl)-; 2HMBeMA		2002427-60-1	2	Smokers & Nonsmokers
	N-Acetyl-S-(4-hydroxy-2-methyl-2-buten-1-yl)-L-cysteine; 4HMBeMA		2002427-58-7	2	Smokers & Nonsmokers

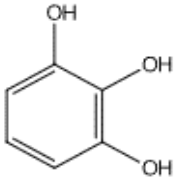
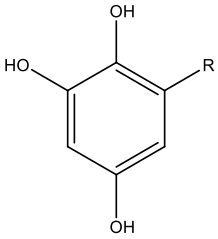
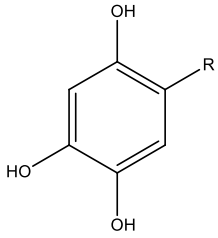
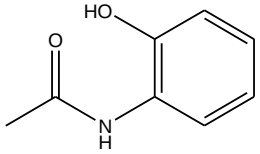
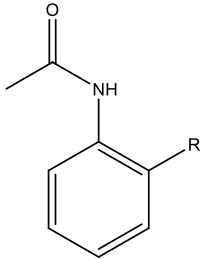
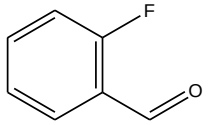
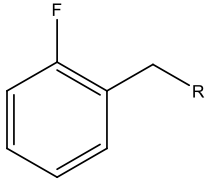
N,N-dimethylformamide 	**N-Acetyl-S-(N-methylcarbamoyl)cysteine; MCaMA**		103974-29-4	1	Smokers & Nonsmokers
Acrylamide 	**N-Acetyl-S-(2-carbamoylethyl)-L-cysteine; 2CaEMA**		81690-92-8	1	Smokers & Nonsmokers
	N-Acetyl-S-(2-carbamoyl-2-hydroxyethyl)-L-cysteine; 2CaHEMA		196620-45-8	1	Smokers & Nonsmokers
	N-Acetyl-S-(2-carbamoyl-2-hydroxyethyl)-L-cysteine; 2CaHEMA2		137698-08-9	2	Smokers & Nonsmokers
	N-Acetyl-3-[(3-amino-3-oxopropyl)sulfinyl]-L-alanine; 3AOxPSfMA		861959-88-8	2	Smokers & Nonsmokers
Acrylonitrile 	**N-Acetyl-S-(2-cyanoethyl)-L-cysteine; 2CyEMA**		74514-75-3	1	Smokers & Nonsmokers
	N-Acetyl-S-(1-cyano-2-hydroxyethyl)cysteine; 1CyHEMA		116477-44-2	2	Smokers & Nonsmokers

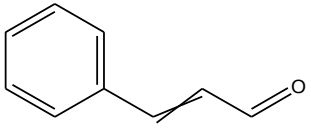
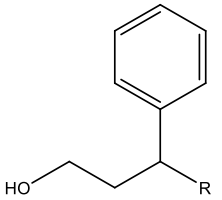
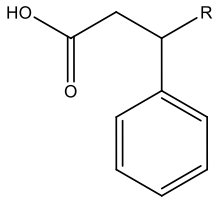
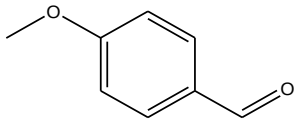
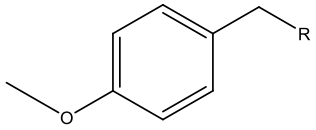
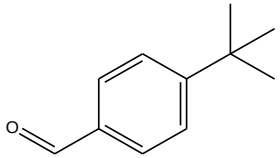
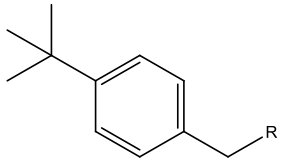
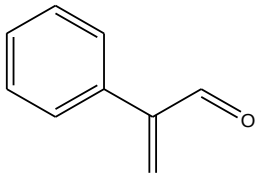
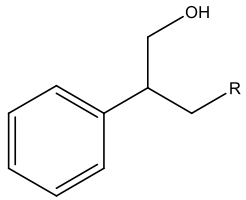
Methacrylonitrile 	N-Acetyl-S-(1-cyanovinyl)-L-cysteine; 1CyVMA		116477-45-3	2	Smokers & Nonsmokers
Methacrylate 	N-Acetyl-S-(2-carboxypropyl)-L-cysteine; 2CoPMA		73614-35-4	2	Smokers & Nonsmokers
Ethyl acrylate 	**N-Acetyl-S-(3-ethoxy-3-oxopropyl)-L-cysteine; 3EoOxPMA**		99651-61-3	2	Smokers & Nonsmokers
Isobutyl acrylate 	**N-Acetyl-S-[3-(2-methylpropoxy)-3-oxopropyl]-L-cysteine; 2MPoOxPMA**		99651-63-5	2	Smokers & Nonsmokers
Butyl acrylate 	N-Acetyl-S-(3-butoxy-3-oxopropyl)-L-cysteine; 3BoOxPMA		99651-62-4	2	Smokers & Nonsmokers

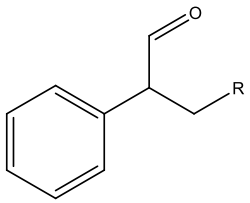
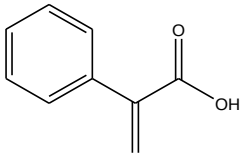
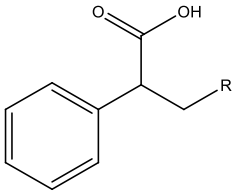
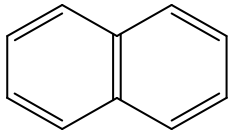
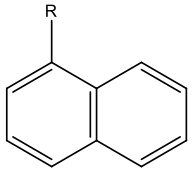
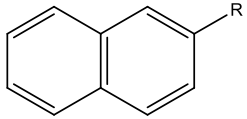
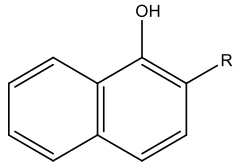
2-Ethylhexyl acrylate 	N-Acetyl-S-[3-[(2-ethylhexyl)oxy]-3-oxopropyl]-L-cysteine; 2EHoOxPMA		99651-64-6	2	Smokers & Nonsmokers
N-tert-Butyl-acetamide 	N-Acetyl-S-(isobutylcarbamoyl)-L-cysteine; 2BCaMA*		N.A.	2	Smokers & Nonsmokers
4-Methylthiobutyl isothiocyanate 	N-Acetyl-S-[[[4-(methylthio)butyl]amino]thioxomethyl]-L-cysteine; MTBAToMMA		547755-87-3	2	Smokers & Nonsmokers
Allyl isothiocyanate 	N-Acetyl-S-[(2-propen-1-ylamino)thioxomethyl]-L-cysteine; 2PeAToMMA		87321-45-7	2	Smokers & Nonsmokers
Sulforaphane 	N-Acetyl-S-[[[4-(methylsulfinyl)butyl]amino]thioxomethyl]-L-cysteine; 4MSfBAToMMA		334829-66-2	2	Smokers
Benzene and monocyclic substituted aromatics					

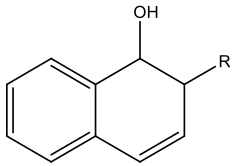
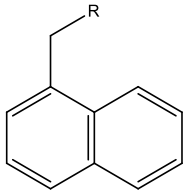
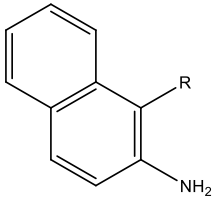
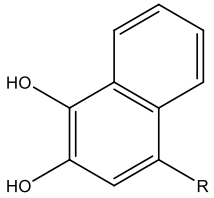
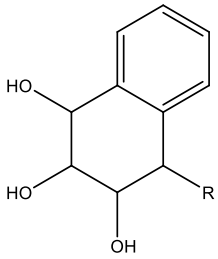
Benzene 	**N-Acetyl-S-(6-hydroxy-2,4-cyclohexadien-1-yl)-L-cysteine; 6HCycHeMA**		122083-16-3	2	Smokers
	N-Acetyl-S-(hydroxyphenylmethyl)-L-cysteine; HPhMMA		321835-24-9	2	Smokers & Nonsmokers
Toluene 	**<u>Benzylmercapturic acid; BzMA</u>**		19538-71-7	1	Smokers & Nonsmokers
Aniline 	N-Acetyl-S-(2-aminophenyl)cysteine; 2APhMA		96346-79-1	2	Smokers
Styrene 	<u>N-Acetyl-S-[(1R)-2-hydroxy-1-phenylethyl]-L-cysteine; 2HPhEMA</u>		81522-00-1	1	Smokers

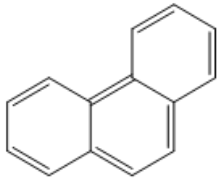
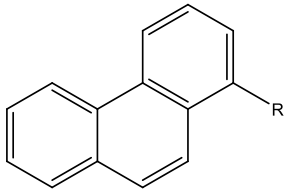
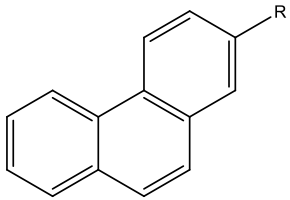
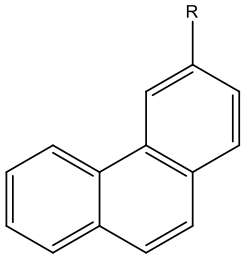
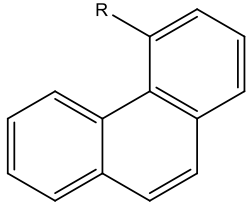
Xylene 	N-Acetyl-S-(2,4-dimethylbenzene)-L-cysteine; MPhMA		581076-69-9	1	Smokers & Nonsmokers
	N-Acetyl-S-[(2-methylphenyl)methyl]-L-cysteine; 2MPhMMA		73898-18-7	2	Smokers & Nonsmokers
Trimethylbenzene 	N-Acetyl-S-[(2,4-dimethylphenyl)methyl]-L-cysteine; 24MPhMMA		263746-56-1	2	Smokers
	N-Acetyl-S-[(3,4-dimethylphenyl)methyl]-L-cysteine; 34MPhMMA		263746-58-3	2	Smokers & Nonsmokers
Hydroquinone 	N-Acetyl-S-(2,5-dihydroxyphenyl)-L-cysteine; 25HPhMA		39484-09-8	2	Smokers & Nonsmokers

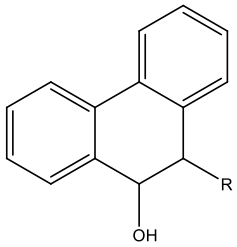
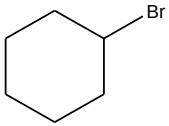
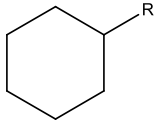
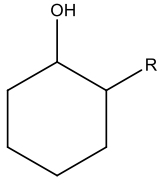
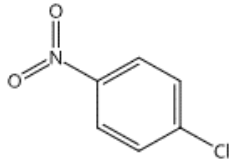
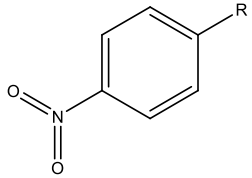
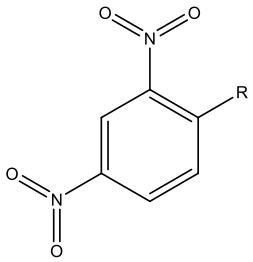
Trihydroxybenzene 	N-Acetyl-S-(2,3,5-trihydroxyphenyl)-L-cysteine; 235HPhMA		215814-13-4	2	Smokers & Nonsmokers
	N-Acetyl-S-(2,4,5-trihydroxyphenyl)-L-cysteine; 245HPhMA		215814-12-3	2	Smokers & Nonsmokers
Orthocetamol 	N-Acetyl-S-[2-(acetylamino)phenyl]-L-cysteine; 2AcAPhMA		321835-23-8	2	Smokers & Nonsmokers
	Aromatic aldehydes				
2-Flurobenzaldehyde 	**N-Acetyl-S-[(2-fluorophenyl)methyl]-L-cysteine; 2FPhMMA**		85680-00-8	2	Smokers & Nonsmokers

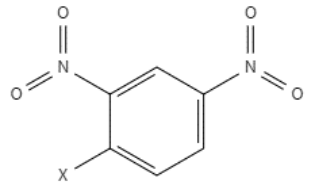
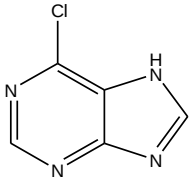
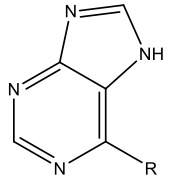
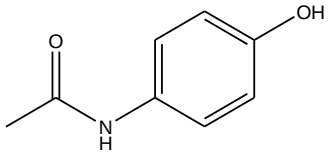
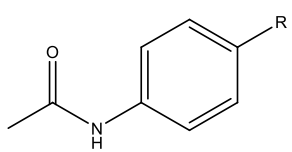
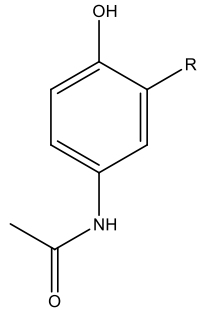
Cinnamaldehyde 	N-Acetyl-S-(3-hydroxy-1-phenylpropyl)-L-cysteine; 3HPhPMA		74032-83-0	2	Smokers & Nonsmokers
	β-[[2-(Acetylamino)-2-carboxyethyl]thio]benzenepropanoic acid; 1CoEMBzMA		74032-84-1	2	Smokers & Nonsmokers
4-Methoxybenzaldehyde 	**N-Acetyl-S-[(4-methoxyphenyl)methyl]-L-cysteine; 4MoPhMMA**		126252-88-8	2	Smokers & Nonsmokers
4-tert-Butylbenzaldehyde 	**N-Acetyl-S-[[4-(1,1-dimethylethyl)phenyl]methyl]-L-cysteine; 11MEPhMMA**		871515-50-3	2	Smokers & Nonsmokers
2-Phenylpropenal 	N-Acetyl-S-(3-hydroxy-2-phenylpropyl)-L-cysteine; 3HPhPMA2		188907-25-7	2	Smokers & Nonsmokers

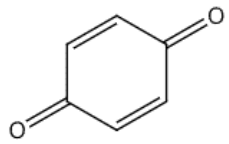
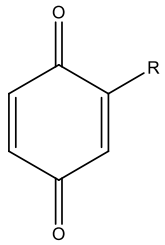
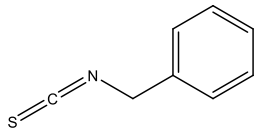
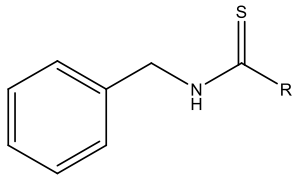
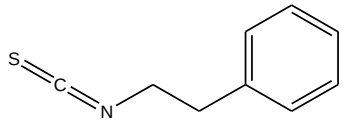
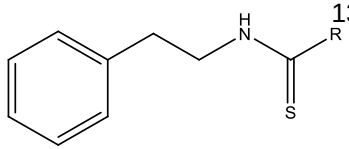
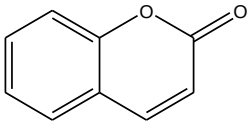
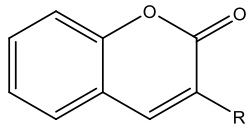
	N-Acetyl-S-(3-oxo-2-phenylpropyl)-L-cysteine; 3OxPhPMA		188907-27-9		Smokers & Nonsmokers
a-Phenylacrylic acid 	α-[[[(2R)-2-(Acetylamino)-2-carboxyethyl]thio]methyl]benzeneacetic acid; 2CoEMBzMA		188907-26-8	2	Smokers & Nonsmokers
	Polycyclic aromatic hydrocarbons (PAHs)				
Naphthalene 	N-Acetyl-S-1-naphthalenyl-L-cysteine; 1NeMA		51325-34-9	2	Smokers & Nonsmokers
	N-Acetyl-S-(2-naphthalenyl)-L-cysteine; 2NeMA		159878-15-6	2	Smokers & Nonsmokers
	N-Acetyl-S-(1-hydroxy-2-naphthalenyl)-L-cysteine; 1HNeMA		321835-22-7	2	Smokers

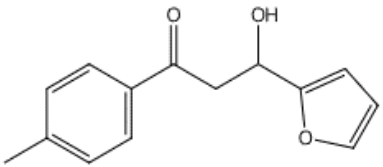
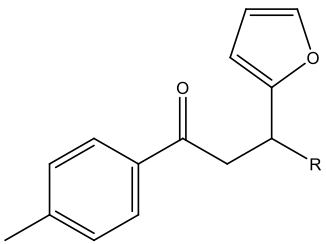
N-Acetyl-S-(1,2-dihydro-1-hydroxy-2-naphthalenyl)-L-cysteine; 12HHNeMA		84608-90-2	2	Smokers
N-Acetyl-S-(1-naphthalenylmethyl)-L-cysteine; 1NeMMA		17238-35-6	2	Smokers & Nonsmokers
Alanine, N-acetyl-3-[(2-amino-1-naphthyl)thio]-; 2ANeMA		3432-30-2	2	Smokers & Nonsmokers
N-Acetyl-S-(3,4-dihydroxy-1-naphthalenyl)-L-cysteine; 34HNeMA		132493-16-4	2	Smokers & Nonsmokers
L-Cysteine, N-acetyl-S-(1,2,3,4-tetrahydro-2,3,4-trihydroxy-1-naphthalenyl)-; 234HNeMA		160567-13-5	2	Smokers & Nonsmokers

Phenanthrene 	Phenanthrene mercapturic acid 1; 1PaMA*		N.A.	2	Smokers & Nonsmokers
	Phenanthrene mercapturic acid 2; 2PaMA*		N.A.	2	Smokers & Nonsmokers
	Phenanthrene mercapturic acid; 3PaMA*		N.A.	2	Smokers & Nonsmokers
	Phenanthrene mercapturic acid 4; 4PaMA*		N.A.	2	Smokers & Nonsmokers

	N-Acetyl-S-(9,10-dihydro-10-hydroxy-9-phenanthrenyl)-L-cysteine; 10HPaMA		572-46-3	2	Smokers & Nonsmokers
		Halogenated aromatics			
Bromocyclohexane 	N-Acetyl-S-cyclohexyl-L-cysteine; CycHxMA		13392-33-1	2	Smokers & Nonsmokers
	N-Acetyl-S-(2-hydroxycyclohexyl)-L-cysteine; 2HCycHxMA		13392-34-2	2	Smokers & Nonsmokers
4-Chloronitrobenzene 	**N-Acetyl-S-(4-nitrophenyl)-L-cysteine; 4NPhMA**		91088-55-0	2	Smokers & Nonsmokers
	N-Acetyl-S-(2,4-dinitrophenyl)-L-cysteine; 24NPhMA		35897-25-7	2	Smokers & Nonsmokers

Halogenated dinitrobenzene 					
6-Chloropurine 	**S-(6-Purinyl)-N-acetyl-L-cysteine; 6PuMA**		136039-99-1	2	Smokers & Nonsmokers
		Pharmaceuticals			
Acetaminophen 	**N-Acetyl-S-[4-(acetylamino)phenyl]-L-cysteine; 4AcAPhMA**		64014-07-9	2	Smokers & Nonsmokers
	Acetaminophen mercapturate; L-Cysteine, N-acetyl-S-[5-(acetylamino)-2-hydroxyphenyl]-; 5AcAHPhMA		52372-86-8	2	Smokers & Nonsmokers
		Other aromatics			

1,4-Benzoquinone 	2-Benzoquinone mercapturic acid; 2BqMA		39484-08-7	2	Smokers & Nonsmokers
Benzyl isothiocyanate 	N-Acetyl-S-[[(phenylmethyl)amino]thioxomethyl]-L-cysteine; PhMAToMMA		62959-32-4	2	Smokers & Nonsmokers
Phenethyl isothiocyanate 	N-Acetyl-S-[[(2-phenylethyl)amino]thioxomethyl]-L-cysteine; 2PhEAToMMA		131918-97-3	2	Smokers & Nonsmokers
Coumarin 	3-Coumaryl mercapturic acid; 3CouMA		135505-65-6	2	Smokers

3-(2-furanyl)-3-hydroxy-1-(4-methylphenyl)-1-propanone (FHMPP) 	**N-Acetyl-S-[1-(2-furanyl)-3-(4-methylphenyl)-3-oxopropyl]-L-cysteine; 2FuMPhOxPMA**		1095010-04-0	2	Smokers & Nonsmokers

X: a halogen in parent structures

R: the N-acetyl-S-L-cysteine moiety in MA structures

*: MAs first reported in our study

Table 4S. Associations between urinary cotinine and 118 MAs. β is linear regression coefficient; SE is standard error of β. Before the linear regression, the creatinine normalized levels of MAs were log transformed.

Main group of parent	Subgroup	Harmonized acronym	Full name	β	SE	p value
Aliphatics	Alkenals	3HPPnMA	3-Hydroxypropyl-1-pentyl MA	0.131	0.038	**0.001**
		2CoEPnMA	2-Carboxyethyl-1-Pentyl MA	0.325	0.051	**0.000**
		2HEBMA	N-Acetyl-S-[1-(2-hydroxyethyl)butyl]-L-cysteine	0.120	0.057	**0.035**
		2CoEEMA	2-Carboxyethyl-1-ethyl MA	0.187	0.044	**0.000**
		3HPMA	N-Acetyl-S-(3-hydroxypropyl)cysteine	0.418	0.062	**0.000**
		2CoEMA	N-acetyl-S-(2-carboxyethyl)-L-cysteine	0.406	0.052	**0.000**
		3HMPMA	N-Acetyl-S-(3-hydroxy-1-methylpropyl)-L-cysteine	0.434	0.053	**0.000**
		2CoMEMA	2-Carboxy-1-methylethylmercapturic acid	0.081	0.085	0.338
		1EHPMA	N-Acetyl-S-(1-ethyl-3-hydroxypropyl)-L-cysteine	0.365	0.047	**0.000**
		3HPBMA	3-Hydroxypropyl-1-butyl MA	0.163	0.056	**0.004**
		1HOxEMA	N-Acetyl-S-(1-hydroxy-2-oxoethyl)-L-cysteine	0.096	0.042	**0.022**
		2CoEPMA	2-Carboxyethyl-1-propyl MA	0.242	0.073	**0.001**
		3OxPMA	N-Acetyl-S-(3-oxopropyl)-L-cysteine	-0.081	0.084	0.337
	Hydroxyalkenals	5HPnFuMA	N-Acetyl-S-(tetrahydro-5-hydroxy-2-pentyl-3-furanyl)-L-cysteine	0.104	0.050	**0.036**
		5OxPnFuMA	N-Acetyl-S-(tetrahydro-5-oxo-2-pentyl-3-furanyl)-L-cysteine	0.037	0.067	0.584
		2HHEHpMA	N-Acetyl-S-[2-hydroxy-1-(2-hydroxyethyl)heptyl]-L-cysteine	0.035	0.064	0.586
		3OxPHBMA	3-Oxopropyl-1-(2-hydroxybutyl) MA	0.098	0.049	**0.048**
		5PDfMA	4-(5-Propyl-Dihydrofuryl) MA	0.099	0.053	0.060
		CoEHBMA	Carboxyethyl-1-(2-hydroxybutyl) MA	0.002	0.048	0.972
		3HPHBMA	3-Hydroxypropyl-1-(2-hydroxybutyl) MA	0.200	0.045	**0.000**
		3OxPHPMA	3-Oxopropyl-1-(2-hydroxypropyl) MA	0.274	0.074	**0.000**
		3OxPHPMA2	3-Oxopropyl-1-(2-hydroxypropyl) MA	0.303	0.039	**0.000**
		5EDfMA	4-(5-Ethyl-Dihydrofuryl) MA	0.043	0.059	0.460
		CoEHPMA	Carboxyethyl-1-(2-hydroxypropyl) MA	0.107	0.046	**0.019**
		3HPHPMA	3-Hydroxypropyl-1-(2-hydroxypropyl) MA	0.069	0.044	0.114
		3HPHPMA2	3-Hydroxypropyl-1-(2-hydroxypropyl) MA	0.110	0.041	**0.007**
		3OxPHPnMA	3-Oxopropyl-1-(2-hydroxypentyl) MA	0.197	0.076	**0.009**
		CoEHPnMA	Carboxyethyl-1-(2-hydroxypentyl) MA	0.111	0.043	**0.011**
		3HPHPnMA	3-Hydroxypropyl-1-(2-hydroxypentyl) MA	-0.032	0.062	0.601

Main group of parent	Subgroup	Harmonized acronym	Full name	β	SE	p value
		3HPHPnMA2	3-Hydroxypropyl-1-(2-hydroxypentyl) MA	0.081	0.040	**0.044**
		3HPHPnMA3	3-Hydroxypropyl-1-(2-hydroxypentyl) MA	0.088	0.051	0.083
		3OxPHEMA	3-Oxopropyl-1-(2-hydroxyethyl) MA	0.233	0.070	**0.001**
		5MDfMA	4-(5-Methyl-dihydrofuryl) MA	0.227	0.040	**0.000**
		CoEHEMA	Carboxyethyl-1-(2-hydroxyethyl) MA	0.163	0.048	**0.001**
		3HPHEMA	3-Hydroxypropyl-1-(2-hydroxyethyl) MA	-0.068	0.074	0.359
	Other aliphatics	4HMBeMA	N-Acetyl-S-(4-hydroxy-2-methyl-2-buten-1-yl)-L-cysteine	0.543	0.045	**0.000**
		2BCaMA	N-Acetyl-S-(isobutylcarbamoyl)-L-cysteine	0.001	0.055	0.982
		2PeAToMMA	N-Acetyl-S-[(2-propen-1-ylamino)thioxomethyl]-L-cysteine	0.012	0.074	0.871
		MTBAToMMA	N-Acetyl-S-[[[4-(methylthio)butyl]amino]thioxomethyl]-L-cysteine	0.136	0.051	**0.008**
		23HHMPMA	N-Acetyl-S-[2,3-dihydroxy-1-(hydroxymethyl)propyl]-L-cysteine	0.201	0.055	**0.000**
		23HHMPMA2	N-Acetyl-S-[2,3-dihydroxy-1-(hydroxymethyl)propyl]-L-cysteine M1	0.264	0.047	**0.000**
		2MPoOxPMA	N-Acetyl-S-[3-(2-methylpropoxy)-3-oxopropyl]-L-cysteine	0.151	0.047	**0.001**
		2EHoOxPMA	N-Acetyl-S-[3-[(2-ethylhexyl)oxy]-3-oxopropyl]-L-cysteine	0.069	0.059	0.241
		2PeMA	N-Acetyl-S-2-propen-1-yl-L-cysteine	-0.027	0.093	0.770
		2CaEMA	N-Acetyl-S-(3-amino-3-oxopropyl)-L-cysteine	0.320	0.057	**0.000**
		MCaMA	N-Acetyl-S-(N-methylcarbamoyl)cysteine	0.366	0.047	**0.000**
		2CyEMA	N-Acetyl-S-(2-cyanoethyl)-L-cysteine	0.509	0.057	**0.000**
		34HBMA	N-Acetyl-S-(3,4-dihydroxybutyl)-L-cysteine	0.074	0.066	0.256
		4MSfBAToMMA	N-Acetyl-S-[[[4-(methylsulfinyl)butyl]amino]thioxomethyl]-L-cysteine	0.072	0.044	0.105
		2CaHEMA	N-Acetyl-S-[(2S)-3-amino-2-hydroxy-3-oxopropyl]-L-cysteine	0.207	0.032	**0.000**
		2HEMA	N-Acetyl-S-(2-hydroxyethyl)-L-cysteine	0.045	0.044	0.316
		2CaHEMA2	isoGAMA N-Acetyl-S-(3-amino-2-hydroxy-3-oxopropyl)-L-cysteine	0.136	0.046	**0.003**
		2HMBeMA	L-Cysteine, N-acetyl-S-(2-hydroxy-3-methyl-3-buten-1-yl)-	0.446	0.062	**0.000**

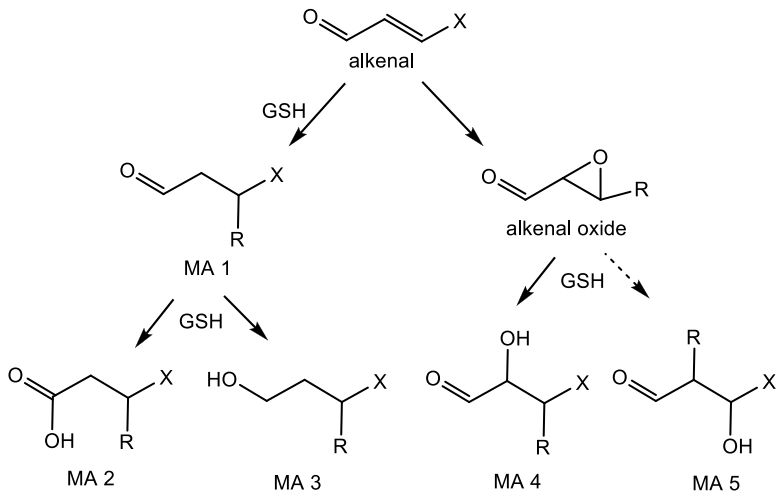
Main group of parent	Subgroup	Harmonized acronym	Full name	β	SE	p value
		4MBeMA	L-Cysteine, N-acetyl-S-[(2E)-4-hydroxy-2-methyl-2-buten-1-yl]-	0.206	0.071	**0.003**
		1HMPeMA	N-acetyl-S-[1-(hydroxymethyl)-2-propenyl]-L-cysteine	0.092	0.043	**0.032**
		2HBeMA	N-Acetyl-S-(2-hydroxy-3-buten-1-yl)-L-cysteine	0.440	0.040	**0.000**
		3AOxPSfMA	N-Acetyl-3-[(3-amino-3-oxopropyl)sulfinyl]-L-alanine	0.111	0.050	**0.028**
		1CyHEMA	N-Acetyl-S-(1-cyano-2-hydroxyethyl)cysteine	0.137	0.060	**0.023**
		1CyVMA	N-Acetyl-S-(1-cyanovinyl)-L-cysteine	0.155	0.042	**0.000**
		234HBMA	N-Acetyl-S-(2,3,4-trihydroxybutyl)-L-cysteine	0.063	0.073	0.387
		23HPMA	N-Acetyl-S-(2,3-dihydroxypropyl)-L-cysteine	0.161	0.040	**0.000**
		2CoPMA	N-Acetyl-S-(2-carboxypropyl)-L-cysteine	0.096	0.053	0.069
		3BoOxPMA	N-Acetyl-S-(3-butoxy-3-oxopropyl)-L-cysteine	0.147	0.052	**0.004**
		3EoOxPMA	N-Acetyl-S-(3-ethoxy-3-oxopropyl)-L-cysteine	0.230	0.040	**0.000**
Aromatics	Benzene and monocyclic substituted aromatics	2HPhEMA	12PHEMA_SR N-Acetyl-S-[(1R)-2-hydroxy-1-phenylethyl]-L-cysteine	0.131	0.038	**0.001**
		6HCycHeMA	N-Acetyl-S-(6-hydroxy-2,4-cyclohexadien-1-yl)-L-cysteine	0.208	0.041	**0.000**
		HPhMMA	N-Acetyl-S-(hydroxyphenylmethyl)-L-cysteine	0.080	0.043	0.067
		24MPhMMA	N-Acetyl-S-[(2,4-dimethylphenyl)methyl]-L-cysteine	0.055	0.043	0.199
		2MPhMMA	N-Acetyl-S-[(2-methylphenyl)methyl]-L-cysteine	0.263	0.068	**0.000**
		34MPhMMA	N-Acetyl-S-[(3,4-dimethylphenyl)methyl]-L-cysteine	0.278	0.060	**0.000**
		2AcAPhMA	N-Acetyl-S-[2-(acetylamino)phenyl]-L-cysteine	0.089	0.046	0.054
		MPhMA	N-Acetyl-S-(2,4-dimethylbenzene)-L-cysteine	0.127	0.047	**0.007**
		BzMA	Benzylmercapturic acid	0.157	0.037	**0.000**
		235HPhMA	N-Acetyl-S-(2,3,5-trihydroxyphenyl)-L-cysteine	-0.014	0.066	0.827
		245HPhMA	N-Acetyl-S-(2,4,5-trihydroxyphenyl)-L-cysteine	-0.033	0.081	0.682

Main group of parent	Subgroup	Harmonized acronym	Full name	β	SE	p value
		25HPhMA	N-Acetyl-S-(2,5-dihydroxyphenyl)-L-cysteine	0.093	0.046	**0.041**
		2APhMA	N-Acetyl-S-(2-aminophenyl)cysteine	0.117	0.043	**0.007**
	Polycyclic aromatic hydrocarbons	10HPaMA	N-Acetyl-S-(9,10-dihydro-10-hydroxy-9-phenanthrenyl)-L-cysteine	0.232	0.090	**0.010**
		1NeMA	N-Acetyl-S-1-naphthalenyl-L-cysteine	0.077	0.054	0.154
		1PaMA	Phenanthrene MA1	0.046	0.053	0.388
		2PaMA	Phenanthrene MA2	0.137	0.067	**0.040**
		3PaMA	Phenanthrene MA3	0.069	0.084	0.412
		4PaMA	Phenanthrene MA4	0.199	0.069	**0.004**
		2ANeMA	Alanine, N-acetyl-3-[(2-amino-1-naphthyl)thio]-	0.205	0.087	**0.018**
		234HNeMA	L-Cysteine, N-acetyl-S-(1,2,3,4-tetrahydro-2,3,4-trihydroxy-1-naphthalenyl)-	-0.023	0.066	0.731
		12HHNeMA	N-Acetyl-S-(1,2-dihydro-1-hydroxy-2-naphthalenyl)-L-cysteine	0.101	0.046	**0.029**
		1HNeMA	N-Acetyl-S-(1-hydroxy-2-naphthalenyl)-L-cysteine	0.056	0.045	0.219
		1NeMMA	N-Acetyl-S-(1-naphthalenylmethyl)-L-cysteine	0.044	0.043	0.304
		2NeMA	N-Acetyl-S-(2-naphthalenyl)-L-cysteine	0.277	0.047	**0.000**
		34HNeMA	N-Acetyl-S-(3,4-dihydroxy-1-naphthalenyl)-L-cysteine	0.080	0.023	**0.000**
	Aromatic aldehydes	2FPhMMA	N-Acetyl-S-[(2-fluorophenyl)methyl]-L-cysteine	0.219	0.087	**0.012**
		4MoPhMMA	N-Acetyl-S-[(4-methoxyphenyl)methyl]-L-cysteine	0.176	0.058	**0.002**
		11MEPhMMA	N-Acetyl-S-[[4-(1,1-dimethylethyl)phenyl]methyl]-L-cysteine	0.178	0.046	**0.000**
		2CoEMBzMA	α-[[[(2R)-2-(Acetylamino)-2-carboxyethyl]thio]methyl]benzeneacetic acid	0.050	0.047	0.284
		1CoEMBzMA	β-[[2-(Acetylamino)-2-carboxyethyl]thio]benzenepropanoic acid	0.143	0.049	**0.003**
		3HPhPMA	N-Acetyl-S-(3-hydroxy-1-phenylpropyl)-L-cysteine	0.062	0.072	0.392
		3HPhPMA2	N-Acetyl-S-(3-hydroxy-2-phenylpropyl)-L-cysteine	0.001	0.075	0.989
		3OxPhPMA	N-Acetyl-S-(3-oxo-2-phenylpropyl)-L-cysteine	0.046	0.048	0.340
		EMA	N-Acetyl-S-ethylcysteine	-0.241	0.083	**0.004**

Main group of parent	Subgroup	Harmonized acronym	Full name	β	SE	p value
	Halogenated aliphatics	HpMA	N-Acetyl-S-heptyl-L-cysteine	0.118	0.045	**0.008**
		MMA	N-Acetyl-S-methylcysteine	-0.075	0.070	0.284
		PnMA	N-Acetyl-S-pentyl-L-cysteine	0.007	0.059	0.908
		1PMA	Propylmercapturic acid	0.015	0.055	0.788
		1MPMA	N-Acetyl-S-(1-methylpropyl)-L-cysteine	0.244	0.082	**0.003**
		CPeMA	N-Acetyl-S-(2-chloro-2-propen-1-yl)-L-cysteine	-0.044	0.114	0.700
		3EoOxEMA	N-Acetyl-S-(2-ethoxy-2-oxoethyl)-L-cysteine	0.147	0.057	**0.009**
		4NPhMA	N-Acetyl-S-(4-nitrophenyl)-L-cysteine	0.173	0.046	**0.000**
		CycHxMA	N-Acetyl-S-cyclohexyl-L-cysteine	0.094	0.045	**0.039**
		6PuMA	S-(6-Purinyl)-N-acetyl-L-cysteine	0.210	0.049	**0.000**
		24NPhMA	N-Acetyl-S-(2,4-dinitrophenyl)-L-cysteine	0.035	0.043	0.419
		2HCycHxMA	N-Acetyl-S-(2-hydroxycyclohexyl)-L-cysteine	0.198	0.058	**0.001**
	Pharmaceuticals	4AcAPhMA	N-Acetyl-S-[4-(acetylamino)phenyl]-L-cysteine	0.161	0.032	**0.000**
		5AcAHPhMA	L-Cysteine, N-acetyl-S-[5-(acetylamino)-2-hydroxyphenyl]-	0.265	0.145	0.067
	Other aromatics	2PhEAToMMA	N-Acetyl-S-[[(2-phenylethyl)amino]thioxomethyl]-L-cysteine	-0.097	0.081	0.232
		PhMAToMMA	N-Acetyl-S-[[(phenylmethyl)amino]thioxomethyl]-L-cysteine	-0.036	0.129	0.780
		2FuMPhOxPMA	N-Acetyl-S-[1-(2-furanyl)-3-(4-methylphenyl)-3-oxopropyl]-L-cysteine	0.168	0.046	**0.000**
		3CouMA	N-Acetyl-S-(2-oxo-2H-1-benzopyran-3-yl)-L-cysteine	0.056	0.052	0.285
		2BqMA	N-Acetyl-S-(3,6-dioxo-1,4-cyclohexadien-1-yl)-L-cysteine	0.145	0.067	**0.031**

Figure 1S.

a. Alkenals



b. Benzene

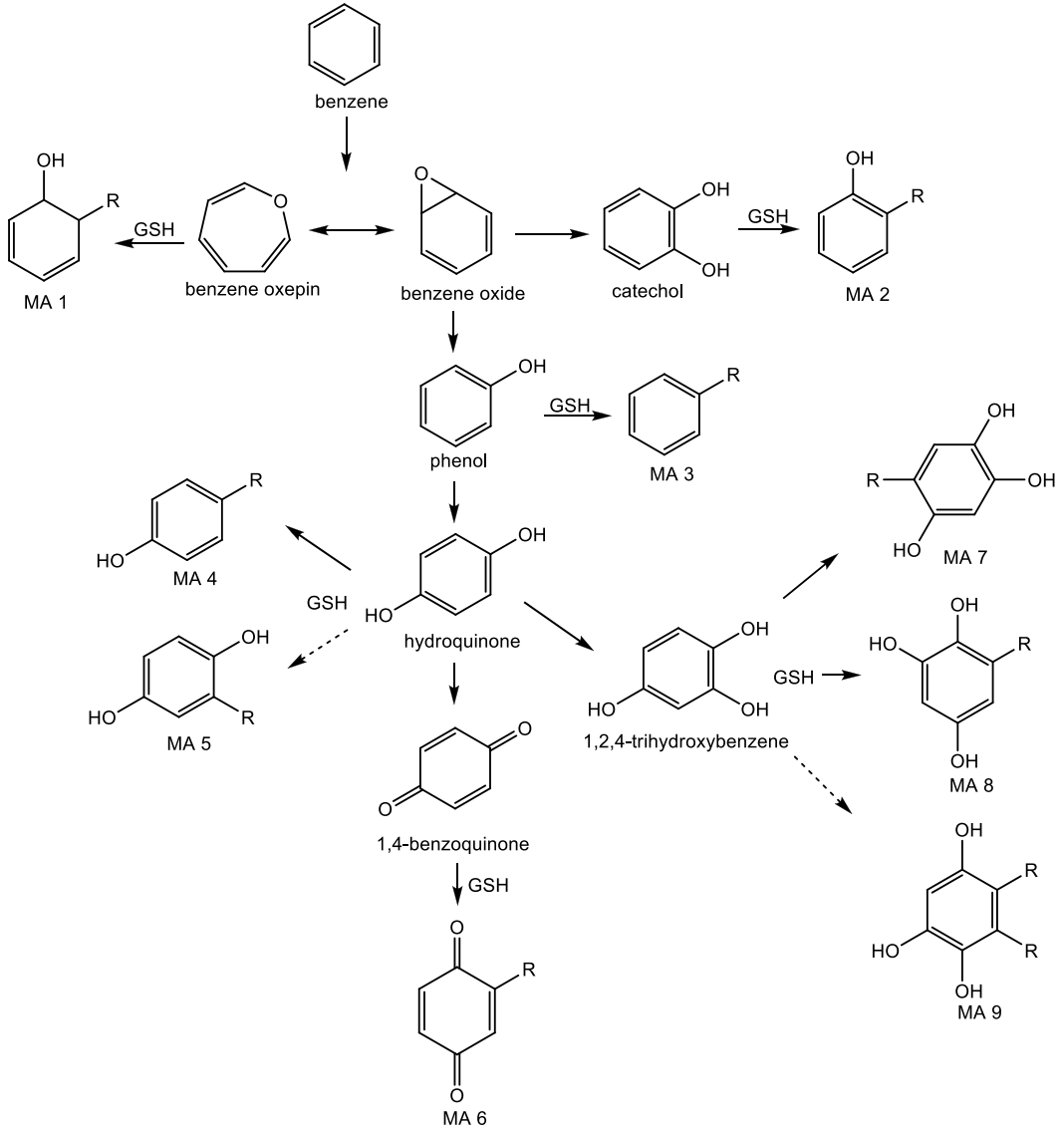
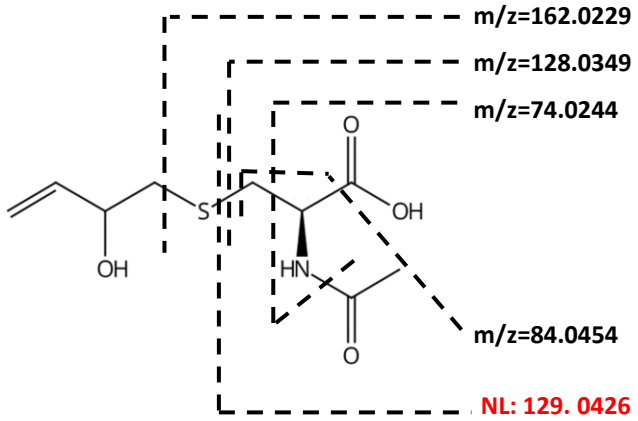


Figure 2S.

a. Characteristic fragmentation patterns



b. MS and MS/MS spectra

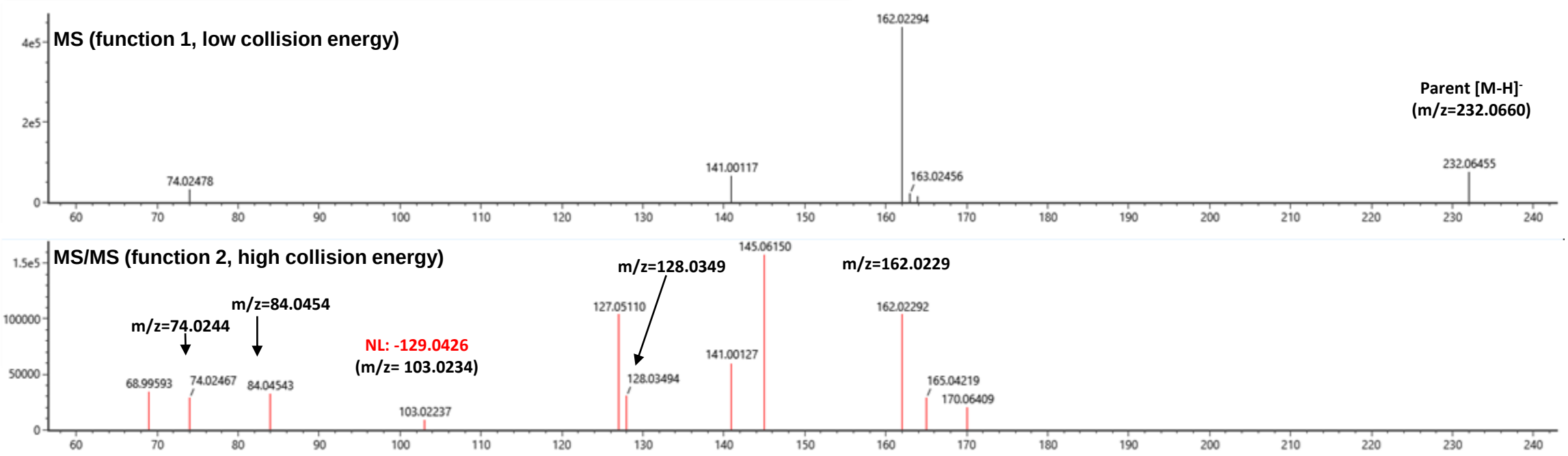


Figure 3S.

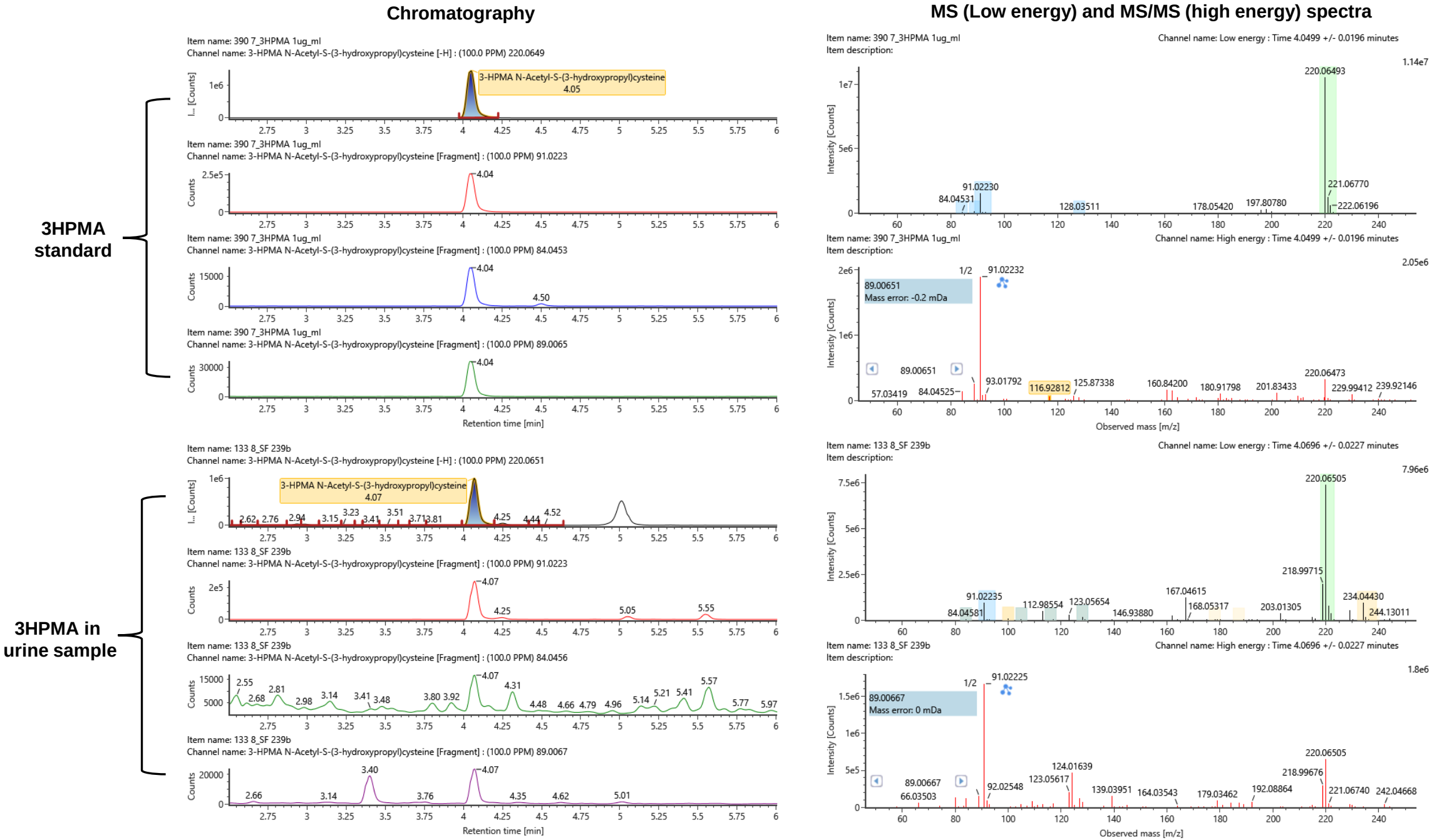
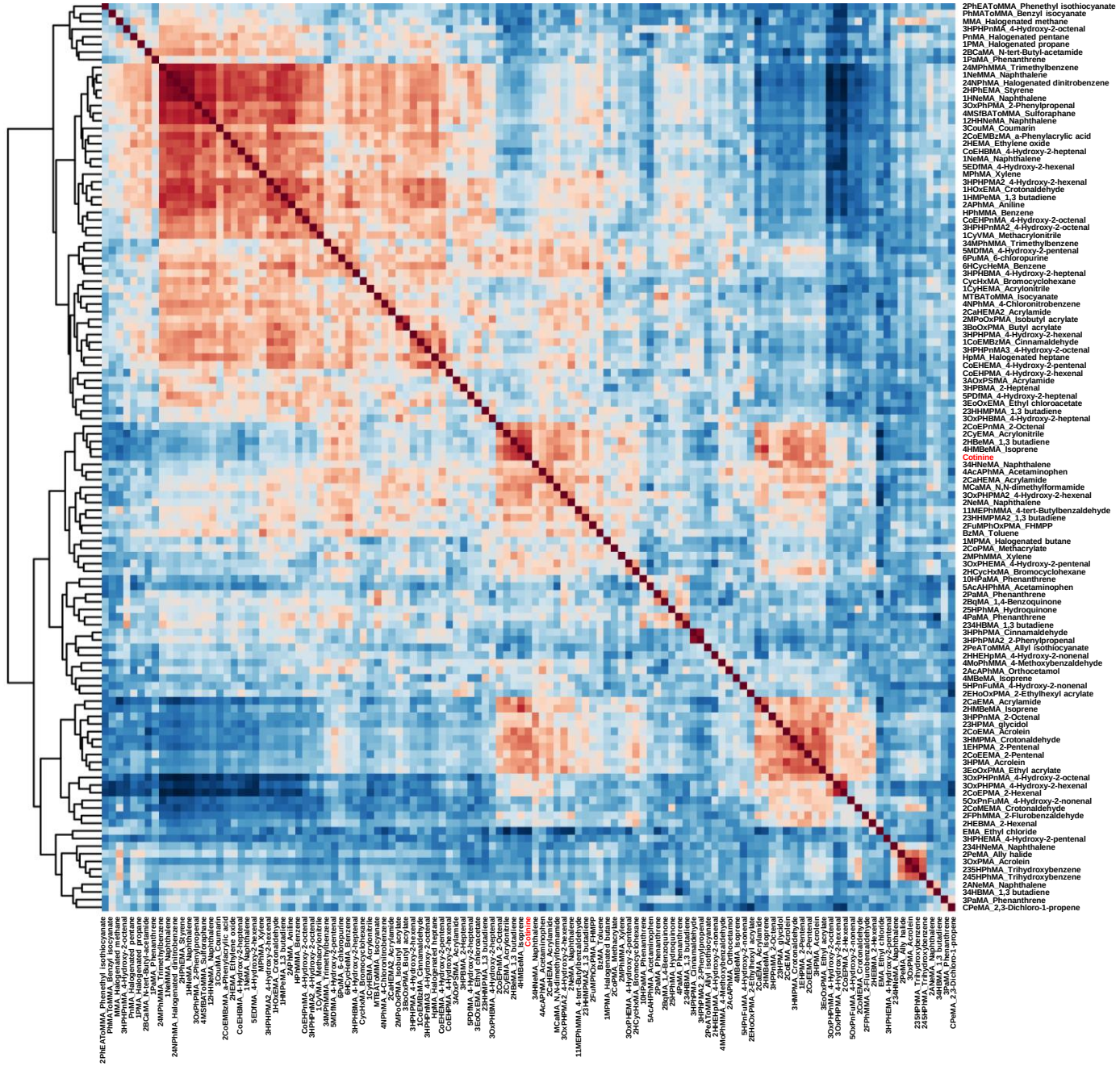

Chromatography
MS (Low energy) and MS/MS (high energy) spectra
3HPMA standard
3HPMA in urine sample
Item name: 390 7_3HPMA 1ug_ml
Channel name: 3-HPMA N-Acetyl-S-(3-hydroxypropyl)cysteine [-H] : (100.0 PPM) 220.0649
3-HPMA N-Acetyl-S-(3-hydroxypropyl)cysteine
4.05
Channel name: 3-HPMA N-Acetyl-S-(3-hydroxypropyl)cysteine [Fragment] : (100.0 PPM) 91.0223
Channel name: 3-HPMA N-Acetyl-S-(3-hydroxypropyl)cysteine [Fragment] : (100.0 PPM) 84.0453
Channel name: 3-HPMA N-Acetyl-S-(3-hydroxypropyl)cysteine [Fragment] : (100.0 PPM) 89.0065
Item name: 133 8_SF 239b
Channel name: 3-HPMA N-Acetyl-S-(3-hydroxypropyl)cysteine [-H] : (100.0 PPM) 220.0651
3-HPMA N-Acetyl-S-(3-hydroxypropyl)cysteine
4.07
Channel name: 3-HPMA N-Acetyl-S-(3-hydroxypropyl)cysteine [Fragment] : (100.0 PPM) 91.0223
Channel name: 3-HPMA N-Acetyl-S-(3-hydroxypropyl)cysteine [Fragment] : (100.0 PPM) 84.0456
Channel name: 3-HPMA N-Acetyl-S-(3-hydroxypropyl)cysteine [Fragment] : (100.0 PPM) 89.0067
Retention time [min]
Item description:
Channel name: Low energy : Time 4.0499 +/- 0.0196 minutes
Channel name: High energy : Time 4.0499 +/- 0.0196 minutes
89.00651
Mass error: -0.2 mda
Channel name: Low energy : Time 4.0696 +/- 0.0227 minutes
Channel name: High energy : Time 4.0696 +/- 0.0227 minutes
89.00667
Mass error: 0 mDa
Intensity [Counts]
Observed mass [m/z]

Figure 4S.


2PhEAToMMA_Phenethyl isothiocyanate
PhMAToMMA_Benzyl isocyanate
MMA_Halogenated methane
3HPHPnMA_4-Hydroxy-2-octenal
PnMA_Halogenated pentane
1PMA_Halogenated propane
2BCaMA_N-tert-Butyl-acetamide
1PaMA_Phenanthrene
24MPhMMA_Trimethylbenzene
1NeMMA_Naphthalene
24NPhMA_Halogenated dinitrobenzene
2HPhEMA_Styrene
1HNeMA_Naphthalene
3OxPhPhMA_2-Phenylpropenal
4MSfBAToMMA_Sulforaphane
12HHNeMA_Naphthalene
3CouMA_Coumarin
2CoEMBzMA_a-Phenylacrylic acid
2HEMA_Ethylene oxide
CoEHBMA_4-Hydroxy-2-heptenal
1NeMA_Naphthalene
5EDfMA_4-Hydroxy-2-hexenal
MPhMA_Xylene
3HPHPMA2_4-Hydroxy-2-hexenal
1HOxEMA_Crotonaldehyde
1HMPeMA_1,3 butadiene
2APhMA_Aniline
HPhMMA_Benzene
CoEHPnMA_4-Hydroxy-2-octenal
3HPHPnMA2_4-Hydroxy-2-octenal
1CyVMA_Methacrylonitrile
34MPhMMA_Trimethylbenzene
5MDfMA_4-Hydroxy-2-pentenal
6PuMA_6-chloropurine
6HCycHeMA_Benzene
3HPHBMA_4-Hydroxy-2-heptenal
CycHxMA_Bromocyclohexane
1CyHEMA_Acrylonitrile
MTBATcMMA_Isocyanate
4NPhMA_4-Chloronitrobenzene
2CaHEMA2_Acrylamide
2MPoOxPMA_Isobutyl acrylate
3BoOxPMA_Butyl acrylate
3HPHPMA_4-Hydroxy-2-hexenal
1CoEMBzMA_Cinnamaldehyde
3HPHPnMA3_4-Hydroxy-2-octenal
HpMA_Halogenated heptane
CoEHEMA_4-Hydroxy-2-pentenal
CoEHPMA_4-Hydroxy-2-hexenal
3AOxPSfMA_Acrylamide
3HPBMA_2-Heptenal
5PDfMA_4-Hydroxy-2-heptenal
3EoOxEMA_Ethyl chloroacetate
23HHMPMA_1,3 butadiene
3OxPHBMA_4-Hydroxy-2-heptenal
2CoEPnMA_2-Octenal
2CyEMA_Acrylonitrile
2HBeMA_1,3 butadiene
4HMBeMA_Isoprene
Cotinine
34HNeMA_Naphthalene
4AcAPhMA_Acetaminophen
2CaHEMA_Acrylamide
MCaMA_N,N-dimethylformamide
3OxPHPMA2_4-Hydroxy-2-hexenal
2NeMA_Naphthalene
11MEPhMMA_4-tert-Butylbenzaldehyde
23HHMPMA2_1,3 butadiene
2FuMPhOxPMA_FHMPP
BzMA_Toluene
1MPMA_Halogenated butane
2CoPMA_Methacrylate
2MPhMMA_Xylene
3OxPHEMA_4-Hydroxy-2-pentenal
2HCycHxMA_Bromocyclohexane
10HPaMA_Phenanthrene
5AcAHPhMA_Acetaminophen
2PaMA_Phenanthrene
2BqMA_1,4-Benzoquinone
25HPhMA_Hydroquinone
4PaMA_Phenanthrene
234HBMA_1,3 butadiene
3HPhPMA_Cinnamaldehyde
3HPhPMA2_2-Phenylpropenal
2PeATcMMA_Allyl isothiocyanate
2HHEHpMA_4-Hydroxy-2-nonenal
4MoPhMMA_4-Methoxybenzaldehyde
2AcAPhMA_Orthocetamol
4MBeMA_Isoprene
5HPnFuMA_4-Hydroxy-2-nonenal
2EHoOxPMA_2-Ethylhexyl acrylate
2CaEMA_Acrylamide
2HMBeMA_Isoprene
3HPPnMA_2-Octenal
23HPMA_glycidol
2CoEMA_Acrolein
3HMPMA_Crotonaldehyde
1EHPMA_2-Pentenal
2CoEEMA_2-Pentenal
3HPMA_Acrolein
3EoOxPMA_Ethyl acrylate
3OxPHPnMA_4-Hydroxy-2-octenal
3OxPHPMA_4-Hydroxy-2-hexenal
2CoEPMA_2-Hexenal
5OxPnFuMA_4-Hydroxy-2-nonenal
2CoMEMA_Crotonaldehyde
2FPhMMA_2-Fluorobenzaldehyde
2HEBMA_2-Hexenal
EMA_Ethyl chloride
3HPHEMA_4-Hydroxy-2-pentenal
234HNeMA_Naphthalene
2PeMA_Allyl halide
3OxPMA_Acrolein
235HPhMA_Trihydroxybenzene
245HPhMA_Trihydroxybenzene
2ANeMA_Naphthalene
34HBMA_1,3 butadiene
3PaMA_Phenanthrene
CPeMA_2,3-Dichloro-1-propene
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4