HHS Public Access

Author manuscript

Environ Int. Author manuscript; available in PMC 2025 May 20.

Published in final edited form as:

Environ Int. 2025 May ; 199: 109516. doi:10.1016/j.envint.2025.109516.

Simultaneous profiling of mercapturic acids, glucuronic acids, and sulfates in human urine

Jin Y. Chen[a,b,c,d], **Saurin R. Sutaria**[a,b,c,d,e], **Zhengzhi Xie**[a,b,c,d], **Manjiri Kulkarni**[a,b,c,d], **Rachel J. Keith**[a,b,c,d], **Aruni Bhatnagar**[a,b,c,d], **Clara G. Sears**[a,b,c,d], **Pawel Lorkiewicz**[a,b,c,d,*], **Sanjay Srivastava**[a,b,c,d]

[a]Christina Lee Brown Envirome Institute, University of Louisville, Louisville, KY 40202, United States

[b]Superfund Research Center, University of Louisville, Louisville, KY 40202, United States

[c]American Heart Association-Tobacco Regulation and Addiction Center, University of Louisville, Louisville, KY 40202, United States

[d]Division of Environmental Medicine, Department of Medicine, University of Louisville, Louisville, KY 40202, United States

[e]Bellarmine University, Louisville, KY 40205, United States

Abstract

Humans are constantly exposed to both naturally-occurring and anthropogenic chemicals. Targeted mass spectrometry approaches are frequently used to measure a small panel of chemicals and their metabolites in environmental or biological matrices, but methods for comprehensive individual-level exposure assessment are limited. In this study, we applied an integrated library-guided analysis (ILGA) with ultraperformance liquid chromatography-quadrupole time-of-flight mass spectrometry (UPLC-QTOF/MS) to profile phase II metabolites, specifically mercapturic acids (MAs), glucuronic acids (GAs), and sulfates (SAs) in human urine samples (n = 844). We annotated 424 metabolites (146 MAs, 171 GAs, 107 SAs) by querying chromatographic features with in-house structural libraries, filtering against fragmentation patterns (common neutral loss and ion fragment), and comparing mass spectra with *in-silico* fragmentations and external spectral

*Corresponding author at: Christina Lee Brown Envirome Institute, University of Louisville, Louisville, Kentucky 40202, United States. pawel.lorkiewicz@louisville.edu (P. Lorkiewicz).

CRediT authorship contribution statement

Jin Y. Chen: Writing – review & editing, Writing – original draft, Validation, Software, Methodology, Formal analysis, Data curation. **Saurin R. Sutaria:** Validation, Software, Methodology, Data curation. **Zhengzhi Xie:** Validation, Software, Methodology, Conceptualization. **Manjiri Kulkarni:** Visualization, Software. **Rachel J. Keith:** Resources, Investigation. **Aruni Bhatnagar:** Resources, Funding acquisition. **Clara G. Sears:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology. **Pawel Lorkiewicz:** Writing – review & editing, Writing – original draft, Supervision, Resources, Project administration, Funding acquisition, Conceptualization. **Sanjay Srivastava:** Writing – review & editing, Writing – original draft, Supervision, Resources, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.envint.2025.109516.

libraries. These metabolites were derived from over 200 putative parent compounds of exogenous and endogenous sources, such as dietary compounds, benzene/monocyclic substituted aromatics, pharmaceuticals, polycyclic aromatic hydrocarbons, bile acids/bile salts, and 4-hydroxyalkenals associated with lipid peroxidation process. Further, we performed statistical analyses on 214 metabolites found in more than 75% of samples to examine the association between metabolites and demographic characteristics using integrated network analysis, principal component analysis (PCA), and multivariable linear regression models. The network analysis revealed four distinct communities of 37 positively correlated metabolites, and the PCA (using the 37 metabolites) presented 4 principal components that meaningfully explained at least 80% of the variance in the data. The multivariable linear regression models showed some positive and negative associations between metabolite profiles and certain demographic variables (e.g., age, sex, race, education, income, and tobacco use).

Keywords

Mercapturic acids; Glucuronic acids; Sulfates; Xenobiotics; Endobiotics; Integrated Library-Guided Analysis; LC-HRMS

1. Introduction

Humans are exposed to thousands of chemicals every day. These include naturally-occurring substances, anthropogenic compounds, and chemicals generated endogenously. Sources of xenobiotics include natural gas, industrial chemicals, petroleum products, household chemicals, dietary intake, agricultural products, tobacco smoke, automobile exhaust, wildfire emission, and pharmaceutical reagents (Gonzalez-Dominguez et al., 2020; Johnson et al., 2012; Scalbert et al., 2014; Xie et al., 2023). Some of these chemicals are innocuous or even beneficial, while others could be injurious to health (Crozier et al., 2009; Bhatnagar, 2006; Riggs et al., 2022; Sun et al., 2022). Moreover, metabolism can enhance the reactivity of several of these chemicals.

Xenobiotics are often oxidized, reduced, or hydrolyzed (phase I transformation), and subsequently these activated metabolites are conjugated with endogenous charged molecules such as glutathione (GSH), glucuronic acid (GA), sulfate (SA), and amino acids (phase II conjugation) (Testa, 2007; Croom, 2012). Electrophilic chemicals such as α,β-unsaturated carbonyls (e.g., *trans*-2-alkenals and *trans*-4-hydroxy-2-alkenals) readily react with cellular GSH to form Michael adducts across the α,β double bond, and the resultant electrophile-glutathione S-conjugates go through sequential metabolism, including hydrolysis, followed by N-acetylation in the kidney, to yield mercapturic acids (MAs), which are then excreted in the urine and feces (Hanna and Anders, 2019). Likely, glucuronidation is a major route of elimination for many environmental chemicals, pharmaceutical reagents, and low molecular weight endogenous compounds (Wells et al., 2004). Catalyzed by uridine 5′-diphospho-glucuronosyltransferases, a glucuronic acid is transferred from uridine-5′-diphosphate-α-D-glucuronic acid to a substrate (e.g., hydroxyl and carboxylic acid groups; primary, secondary, and tertiary amines; and thiols and sulfoxides) (Testa, 2007). Furthermore, sulfation, which transfers a sulfate from 3′-phosphoadenosine 5′-phosphosulfate to a

substrate under catalysis by sulfotransferases, also transforms hydrophobic endogenous compounds and xenobiotics into water-soluble sulfates and promotes their excretion from the body (Al-Horani and Desai, 2010). Together, MAs, GAs, and SAs account for ~70% of phase II metabolites (Jancova et al., 2010).

Several analytical techniques and assays are available to assess human exposure to environmental chemicals and associated metabolites at community and individual levels. Community-level exposure to parent organic environmental chemicals is often measured by gas chromatography-mass spectrometry (GC–MS) in indoor and outdoor air, whereas liquid chromatography-mass spectrometry (LC-MS), GC–MS, or nuclear magnetic resonance spectroscopy is employed to quantitate their metabolites in biofluids and tissues of individuals (Athersuch and Keun, 2015; Balcells et al., 2024). These assays are effective in measuring a select chemical or a small panel of analytes, but methods for comprehensive individual-level exposure assessment, especially phase II metabolites, are lacking. Moreover, assessment of xenobiotic metabolisms or phase II metabolites in human endogenous metabolome studies are limited. Therefore, many phase II metabolites remain unknown, uncharacterized, or overlooked in human chemical exposure assessments. We recently established an integrated library-guided analysis (ILGA) workflow using ultraperformance liquid chromatography with quadrupole time-of-flight mass spectrometry (UPLC-TOF/MS) with MS^E data-independent acquisition (DIA) for profiling MAs in human urine (Xie et al., 2023). Now, we have applied the ILGA platform for a comprehensive analysis of various classes of phase II metabolites (i.e., MAs, GAs, and SAs), originating from over 200 putative parent chemicals, in the urine samples from 844 study participants. Multivariable linear regression models, integrated network analysis, and principal component analysis were employed to examine associations between metabolite profiles and demographic characteristics of the study participants.

2. Materials and methods

2.1. Study population and urine collection

Spot urine samples were collected throughout the day (outside first morning void) from 844 participants enrolled in the Health, Environment and Action in Louisville (HEAL) Study (Yeager et al., 2023) and stored at −80 °C. The study was approved by the University of Louisville Institutional Review Board (IRB 15.1260). Demographic information of the participants, including age, sex, race, and education, is provided in Supplemental Table S1.

2.2. Chemicals

LC-MS Ultra Chromasolv water was obtained from Honeywell (Muskegon, MI). Ultra-high-performance liquid chromatography (UHPLC)-MS grade acetonitrile, UHPLC-MS grade methanol, and LC-MS grade formic acid were purchased from Thermo Fisher Scientific (Waltham, MA). Product information of authentic chemical standards of 21 phase II metabolites and 6 deuterium-labeled analytical internal standards (IS) are detailed in Table S2. All the other chemicals were of the highest purity available.

2.3. Ultraperformance liquid chromatography-quadrupole time-of-flight mass spectrometry (UPLC-QTOF/MS)

Quality control (QC) stocks were prepared by pooling equal volumes of 60 individual urine samples. IS stock contained 6 deuterium-labeled compounds prepared in water. QC or urine samples (500 μL) were prepared by mixing 440 μL of 0.1% formic acid in water (solvent A), 10 μL of IS stock mixture, and 50 μL of urine sample or QC stock. IS samples were prepared with 490 μL of solvent A and 10 μL of IS stock mixture. Blanks consisted of 500 μL of solvent A. The urine samples (n = 844) were analyzed in 14 analytical batches. To monitor instrument stability, performance, and batch variations, QC and IS samples were included in each batch – one QC sample per10 urine samples and one IS sample per 30 urine samples.

The analytes were resolved, as described before (Xie et al., 2023), with slight modifications – lower sample volume (2 μL vs. 7.5 μL), shorter column (100 mm vs. 150 mm), and higher flow rate (0.55 mL/min vs. 0.45 mL/min). Briefly, samples (2 μL) were injected, and analytes were resolved on a Waters Acquity Premier HSS T3 VanGuard FIT column (2.1 × 100 mm, 1.8 μm) attached to a Waters Acquity I-Class UPLC system and maintained at 45 °C with a flow rate of 0.55 mL/min. The mobile phases were solvent A (0.1% formic acid in water) and solvent B (0.1% formic acid in acetonitrile). The gradient profile began at 0 % B, increased to 23 % B over 7.30 min, then to 95 % B over 2.40 min, held at 95% B for 1.60 min, returned to 0% B over 0.05 min, and held at 0% B for 1.95 min before the next injection. The total run time was 13.30 min.

The quadrupole time-of-flight mass spectrometric (QTOF/MS) data were collected using a Waters Synapt XS HDMS with MassLynx 4.2 software, operating in electrospray negative ion mode. The capillary voltage was set to 1.00 kV, source temperature to 112 °C, desolvation gas flow to 1100 L/h at 650 °C, and cone gas flow to 125 L/h. Sodium formate was used for mass calibration before the sample sequence run, and leucine enkephalin (m/z 554.2620) served as the lock mass during acquisition. The ion source sample cone was cleaned according to the vendor's guidelines before running each batch. MS^E data-independent acquisition was performed over an m/z range of 50–700 Da with a continuum scan time of 0.15 sec and acquisition time of 0–11.85 min. MS^E acquisition alternated between low-energy and high-energy channels, with collision energy off for the low-energy channel and a voltage ramp from 10 to 40 V for the high-energy channel. Data were collected and analyzed using the Waters UNIFI 1.9 software package.

Urinary creatinine was quantified using Ace Axcel® Clinical Chemistry System (Alfa Wassermann, West Caldwell, NJ) (Xie et al., 2023; Lorkiewicz et al., 2019; Xie et al., 2024), and urinary cotinine levels were measured by UPLC-MS/MS, as described before (Xie et al., 2023; Lorkiewicz et al., 2019). Participants with urinary cotinine level >40 μg/g creatinine were regarded as current tobacco users (Riggs et al., 2022; Lorkiewicz et al., 2019).

2.4. Curation of in-house metabolite libraries

Phase II metabolomics profiling was conducted using the ILGA workflow, as described previously (Xie et al., 2023). The process began with curating in-house metabolite

libraries that included reported and proposed metabolite molecular structures (.mol files). Reported structures were downloaded from SciFinder^{n} and the Human Metabolome Database (HMDB; http://hmdb.ca). Proposed metabolite structures were deduced based on known metabolic pathways of reactive compounds. For instance, acrolein, a 3-carbon α,β-unsaturated aldehyde, has been reported to yield MAs including N-acetyl-S-(3-hydroxypropyl)-L-cysteine, N-acetyl-S-(2-carboxyethyl)-L-cysteine and N-acetyl-S-(3-oxopropyl)-L-cysteine (Lorkiewicz et al., 2019; Hashmi et al., 1992). Based on this knowledge, we proposed MA structures for longer-chain α,β-unsaturated aldehydes such as 2-pentenal, 2-hexenal, 2-heptenal, and 2-octenal (see Supplemental Excel Tables sheet "Table 1 – mercapturic acids"). In addition, we used BioTransformer 3.0 (https://biotransformer.ca/) to predict phase I and phase II metabolites of parent compounds. We used ChemDraw (version 22.0.0) to sketch proposed structures and uploaded .mol files onto SciFinder^{n} to check if they have been reported or not. The.mol files were imported into a library repository created within the UNIFI software. The MA library contained 239 structures (200 reported and 39 proposed), the GA library had 431 structures (341 reported and 90 proposed), and the SA library had 148 reported structures. The MA library was created and expanded from the previous published work (Xie et al., 2023). The complete library in MS Excel format—including metabolite names and their corresponding International Chemical Identifiers (InChI), as well as downloadable files of metabolite structures in .mol and .sdf formats is available online at the following URL: https://osf.io/ch9tq/.

2.5. Metabolite annotation

The raw UPLC-QTOF/MS data files were exported into and processed using UNIFI, and each metabolite group was analyzed separately. We used a trial sample set (consisted of 15 QCs and 60 urine samples) for the metabolite discovery step, in which *in-silico* fragmentation was employed to generate predicted fragments from library structures. After data processing and peak picking, the detected chromatographic features (peaks with *m*/*z* and retention time information) were queried against structures in the in-house libraries. Filtered peaks had to meet the following criteria to be considered for annotation: (1) response threshold >3000, (2) mass error of ±10 mDa, (3) observed *m*/*z* range between 100 and 700, and (4) observed retention time (RT) between 0.4 and 9.0 min. To facilitate annotation, fragmentation signatures like common neutral loss (CNL) and common ion fragment (CIF) specific to MA, GA, and SA were employed (Xie et al., 2023; Levsen et al., 2005; Wagner et al., 2006; Huber et al., 2022; Yan et al., 2003; Evich et al., 2022; Lafaye et al., 2004). As shown in Table 1, MA has one CNL (129.0426 Da) and four CIFs (*m*/*z* 162.0230, 128.0353, 84.0455 and 74.0196); GA has two CNLs (194.0427 and 176.0321 Da) and 10 CIFs (*m*/*z* 193.0354, 175.0248, 117.0193, 113.0244, 103.0401, 99.0088, 95.0139, 87.0088, 85.0295, and 75.0088); SA has two CNLs (97.9674 and 79.9568 Da) and four CIFs (*m*/*z* 96.9601, 95.9523, 80.9652 and 79.9574). Thus, for a filtered peak to be considered as a potential MA, GA, or SA candidate, it needed to have at least one CNL and one CIF, or two CIFs present in the MS/MS spectra of one sample. Examples of characteristic fragmentation patterns present in the experimental MS/MS spectra of a MA (N-acetyl-S-(2-cyanoethyl)-L-cysteine (metabolite of acrylonitrile), a GA (benzoyl glucuronide, metabolite of benzoic acid), and a SA (phenyl sulfate, metabolite of phenol) are provided in Fig. 1.

Before applying the filtering criteria mentioned above, ~18,000 candidate peaks were detected in the raw data. After applying the filtering criteria, ~3,000 MA, ~4,700 GA, and ~3,700 SA candidates were detected. For candidate features that had a match with a library structure, we manually examined their MS/MS spectra against *in-silico* fragmentation and external spectral libraries (i.e., ChemBank, DrugBank, HMDB, MassBank, NIST Spectra, PubMed, and Urine Metabolome Database), and putatively assigned 146 MAs, 171 GAs, and 107 SAs. Next, detection results such as observed RTs and ion fragments of annotated metabolites were sent to in-house libraries to use for metabolite screening step. For the screening step, we used expected RT and ion fragments to annotate analytes in all sample batches. In this phase, we applied a RT error window of <0.10 min, mass error of ±10 mDa and response >100, and excluded the CNL and CIF filtering criteria. In scenarios where multiple candidate peaks matched one metabolite compound due to the presence of possible isomers, the retention time is added to the compound's name (e.g., 4-aminophenyl sulfate_0.73 vs 4-aminophenyl sulfate_1.16) as shown in Supplemental Excel Tables sheet "Table 3 – sulfates".

We assessed true positive rate (TPR), false negative rate (FNR), false discovery rate (FDR) and precision to determine the performance of our metabolite annotation. We analyzed a mixture sample containing only 21 reagent standards (listed in Table S2) by screening against a library containing all 424 annotated metabolites detected in this study. Chemicals spiked in the mixture sample and detected were marked as true positives (TPs), whereas chemicals spiked into the sample but not detected were denoted as false negatives (FNs). Analytes which were detected but were not spiked in the mixture sample were regarded as false positives (FPs). Because our analysis was restricted to 424 annotated metabolites, true negatives (TNs) are not expected in this investigation. Our analysis of reagent mixture identified all 21 TPs (as expected), 6 FPs, and 0 FN (Table S3). Accordingly, our TPR was 1.0 [TP/(TP + FN)] and FNR was 0 [FN/(TP + FN)]. Our FDR was 0.22 [FP/(TP + FP)] and precision was 0.78 [TP/(TP + FP)] (Table S3).

Furthermore, to evaluate the performance of the phase II metabolite filtering process based on CNL and CIF criteria, we compared detections of 21 reagent standard in a mixture sample vs a solvent blank. Before applying the criteria, all 21 standards were detected in the mixture sample, and 5 were detected in the solvent blank with no CNL and CIF found (Table S4). After applying the CNL and CIF criteria, no standard analytes were found in the solvent blank. This shows the applied CNL and CIF filtering criteria was able to remove 100% false positives in the solvent blank.

2.6. Confidence levels of annotation

The annotated metabolites were classified into different confidence levels based on profiling descriptions and data requirements (detailed in Table 2) adapted from Blazenovic et al. (2018). The highest level of confidence is Level 1, followed by Levels 2A, 2B and 3. Level 1 annotations were confidently identified metabolites whose MS/MS spectra and retention times were confirmed with authentic standards. In this study, we divided Level 2 annotations into two sublevels to differentiate a probable structure matched to both an external database

and in-house library (Level 2A) or to in-house library only (Level 2B). Level 3 annotations had a match a proposed structure in the curated libraries.

2.7. Missing value imputation and batch normalization analysis

All annotated LC-MS peaks and peak response values were exported to Microsoft Excel 365 spreadsheets. For values below the response threshold of 100, or instances where an analyte was not detected in a specific sample, imputation was carried out by dividing the response cutoff value (100) by the square root of 2 per established protocols (Richardson and Ciampi, 2003; Succop et al., 2004). As illustrated in Table S5, the imputed values were <25% for 214 metabolites, 25%–50% for 88 metabolites, 50%–75% for 65 metabolites, and >75% for 57 metabolites. Subsequently, the Norm ISWSVR (https://github.com/Dingxian666/Norm-ISWSVR) algorithm developed by Ding et al. (2022) was used to perform batch normalization analysis. Briefly, Norm ISWSVR calculates the peak's relative standard deviation (RSD) across QC samples, and applies internal standards to normalize peak response. If normalized response has smaller RSD in QC samples, then the internal standard that reduces RSD the most is selected for data normalization. Detailed procedure of Norm ISWSVR is described by Ding et al. (2022). Fig. S1 compares RSD of MAs, GAs, and SAs found in all QC samples before and after batch normalization. Prior to normalization, distribution of RSD ≤20% accounted for 3% MAs, 0% GAs, and 4% SAs. After normalization, distribution of RSD ≤20% improved to 87% MAs, 91% GAs, and 92% SAs.

2.8. Statistical analysis

After batch correction, response values were normalized to creatinine to account for urine dilution. For metabolites with <25% imputation (or found >75% of samples), ln-transformation and standardization of values were performed. Descriptive statistics were calculated for each metabolite, and bivariate correlations among metabolites within each group (MA, GA, SA) were examined using Spearman's rank correlation coefficients. Further correlations among metabolites were examined through an integrated network analysis using the R package xMWAS, with data from n = 844 participants. Pairwise data integration and community detection were conducted by this framework, and the eigenvector centrality measure was used to evaluate the importance of each node (metabolite) in the network. Higher eigenvector values indicated nodes with greater influence. Only associations significant at $p < 0.05$ and meeting a correlation threshold of 0.6 were included in the network analysis. Cytoscape software version 3.10.2 was used to visualize the network and associations between metabolites. Based on these criteria, 37 metabolites were selected for principal component (PC) analysis. Four PCs were identified, explaining at least 80% of the variance in the data, and the factor loadings were rotated using a varimax rotation. To examine associations between metabolite profiles and demographic characteristics, multivariable linear regression models were used, including age (continuous, in years), sex (categorical), race (categorical), education (categorical), income (categorical), and tobacco use based on urine cotinine values (categorical, cotinine <40 μg/g creatinine). Of the 844 participants, 57 were excluded from this analysis due to missing demographic data.

3. Results

3.1. Annotation, parent compounds, and confidence levels of phase II metabolites

An overview of analytical experiments, metabolite annotations, and statistical analyses is illustrated in Fig. 2. Briefly, we analyzed 844 urine samples by UPLC-QTOF/MS and processed raw data for peak picking and metabolite annotation. Then, we imputed missing response values, corrected for batch variation, and normalized responses to creatinine for urine dilution. Subsequently, we performed integrated network analysis and principal component analysis and examined the association between demographic characteristics and metabolite principal components.

3.2. Annotated metabolites and their parent compounds

In all participant samples, we annotated 424 metabolites (146 MAs, 171 GAs, and 107 SAs) as shown in Table 2 and Fig. 3. These metabolites are derived from putative parent compounds of various exogenous, endogenous, anthropogenic, and biogenic sources. As shown in Table S6, the parent compounds belong to three distinct structural classes-the aliphatics, the aromatics, and the alicyclics. Each class is divided into sub-groups based on chemical functionality or sources of parent compounds. The aliphatics class consists of three groups: aliphatic aldehydes, halogenated aliphatics, and other aliphatics. The aromatics class is divided into seven groups: aromatic aldehydes, benzene/monocyclic substituted aromatics, dietary phenolics and flavonoids, halogenated aromatics, other aromatics, pharmaceuticals, and polycyclic aromatic hydrocarbons. The alicyclics class comprises six groups: bile acids/bile salts, epoxides, halogenated alicyclics, monoterpenes and derivatives, other alicyclics, and steroids.

Particularly, the annotated MAs were derived from 83 putative parent compounds comprising of 43 aliphatics, 29 aromatics, and 11 alicyclics; the GAs originated from 124 putative parent compounds, including 23 aliphatics, 81 aromatics, and 23 alicyclics; and the SAs derived from 43 putative parent compounds, including 3 aliphatics, 37 aromatics, and 3 alicyclics (Table S6).

3.3. Parent compounds shared among metabolites

We observed that 131 phase II metabolites (40 MAs, 44 GAs, and 47 SAs) shared a common parent compound with at least one other metabolite group (Table 3). Four parent compounds, including 3 environmental chemicals (benzo[a]pyrene, naphthalene, and trihydroxybenzene) and 1 pharmaceutical agent (acetaminophen), had at least one MA, GA, and SA annotated. Naphthalene yielded 13 metabolites (9 MAs, 3 GAs, and 1 SA), trihydroxybenzene (2 MAs, 3 GAs, and 4 SAs) and acetaminophen (2 MAs, 2 GAs, and 5 SAs) afforded 9 metabolites each, and benzo[a]pyrene formed 4 metabolites (1 MA, 2GAs and 1 SA). Moreover, 8 parent molecules, including aliphatic aldehydes (4-hydroxy-2-heptenal, 4-hydroxy-2-nonenal, 4-hydroxy-2-octenal, and cinnamaldehyde), aromatic chemicals (hydroquinone, phenanthrene, and styrene), and a monoterpene (limonene), had at least one MA and GA. Furthermore, 12 parent chemicals including aromatic compounds (benzoic acid, catechol, cresol, indole, and phenol), dietary phenolics (caffeic acid, ferulic acid, and guaiacol), neuromodulators or associated metabolites

(dopamine, 3-methoxy-4-hydroxyphenylglycol and homovanillic acid), and a bile salt (glycochenodeoxycholic acid) generated at least one GA and SA.

3.4. Confidence levels of annotated metabolites

Among all annotated metabolites, 21 Level 1 (10 MAs, 4 GAs, and 7 SAs) derived from 17 unique parent compounds (see Table 4) were confidently identified by matching RTs and mass spectra against authentic chemical standards analyzed under the same analytical conditions (Figs. S2–S22). Furthermore, 46 Level 2A metabolites (3 MAs, 13 GAs, and 30 SAs) were annotated by matching parent ions ($[M-H]^-$) and MS/MS spectra with the external spectral libraries (Table 2). The Level 2A metabolites had a range of 3–85% matching spectral intensity. Moreover, 288 Level 2B and 69 Level 3 metabolites were annotated by matching reported and proposed structures, respectively, in our libraries. Specifically, the Level 3 metabolites consisted of 40 MAs and 29 GAs. Among Level 3 MAs, 23 of them were first reported by us (Xie et al., 2023). For Level 3 GAs, 4 of them (perillic acid-8,9-diol-glucuronide, limonene-1,2-diol-glucuronide_7.64, limonene-1,2-diol-glucuronide_8.76, and 2-hydroxy-p-menth-8-en-7-oic acid-glucuronide) were derived from limonene and its phase I metabolites. We previously identified them using the ILGA workflow (Xie et al., 2024). Complete lists of all metabolites and their descriptions (e.g., *m*/*z*, RT, CAS #, annotation confidence level, canonical SMILES, proposed parent, etc.) are provided in the Supplemental Excel Table S1–S3.

3.5. Metabolites found in >75% of participant samples

As the reliability of statistical analyses weakens with increasing number of missing values, the subsequent statistical analyses were performed for the 214 metabolites detected above response threshold in >75% of samples (see Fig. 4 and Table S5). These included 50 MAs of 36 unique parent compounds (20 aliphatics, 13 aromatics, and 3 alicyclics), 83 GAs of 68 unique parents (14 aliphatics, 36 aromatics, and 18 alicyclics), and 81 SAs of 36 unique parents (2 aliphatics, 33 aromatics, and 1 alicyclics). For these metabolites, we observed modest to strong positive and negative bivariate correlations among each type of metabolite. The correlation coefficients ranged from −0.21 to 1.00 for MAs, −0.24 to 0.91 for GAs, and −0.15 to 0.95 for SAs (see Supplementary Excel – “Correlation Coefficients”).

3.6. Integrated network analysis

We identified four distinct communities of 37 positively correlated metabolites (Fig. 5). Community 1 had 9 metabolites (4 GAs and 5 SAs) whose parent compounds belong to benzene/monocyclic substituted aromatics and dietary phenols and flavonoids of exogenous sources. The node (or metabolite) most central to this community was phenylpropanoic acid metabolite (1-benzenepropanoate β-D-glucopyranuronic acid_6.89, GA17, eigenvector centrality measure = 0.77), followed by methylparaben metabolite (1-methyl-4-(sulfooxy)benzoate_3.98, SA41, 0.74); see Supplementary Excel – “Network_PCA”. Community 2 had 11 metabolites (3 GAs and 8 SAs). The node most central to this community was catechol metabolite (2-hydroxyphenyl β-D-glucopyranosiduronic acid, GA43, eigenvector centrality measure = 1.0), followed by p-Methylguaiacol metabolite (2-methoxy-4-methylphenyl β-D-glucopyranosiduronic acid_6.71, GA49, 0.92). Most metabolites in Community 2 consisted of dietary phenols and flavonoids. Community

3 consisted of 7 metabolites (3 GAs, 3 SAs, and 1 MA). The node most central to the community was 4-aminophenol metabolite (4-aminophenyl sulfate_1.16, SA108, eigenvector centrality measure = 0.87), followed by vanillyl alcohol metabolite (vanillyl alcohol sulfate_2.81, SA52, 0.82). Community 4 had 10 metabolites (5 SAs, 4 MAs, and 1 GA). The nodes most central to this community were 4-oxo-2-heptenoic acid metabolite (4-oxo-2-heptenoic acid glucuronide, GA95, eigenvector centrality measure = 0.96), and 2-fluorobenzaldehyde metabolite (N-acetyl-S-[(2-fluorophenyl)methyl]-L-cysteine, MA137, 0.91).

3.7. Principal component analysis

In principal component analysis (PCA) using the 37 metabolites from the network analysis, we identified four PCs that meaningfully explained at least 80% of the variance in the data (see Fig. 6 and Supplementary Excel – "Network_PCA"). PC1 explained 64% of the variance in the data, with 6.4%, 3.9%, and 3.3% of the variance explained by PC2, PC3, and PC4, respectively. Individuals with higher scores (0.505 to 0.763) for PC1 tended to have higher levels of metabolites of dietary phenols and flavonoids (e.g., ferulic acid 4-sulfate_5.35 (SA94) and vanillyl alcohol sulfate_2.81 (SA52)), benzene/monocyclic substituted aromatics (e.g., 2-hydroxyphenyl β-D-glucopyranosiduronic acid (GA43) and 4-aminophenyl sulfate_1.16 (SA108))), other aromatics (e.g., 3-hydroxypyridine sulfate (SA85)), and aliphatics (e.g., N-Acetyl-S-(2-chloro-2-propen-1-yl)-L-cysteine (MA60)). For PC2, higher scores (0.511 to 0.899) tended to be correlated with higher levels of metabolites such as MAs of isoprene (N-acetyl-S-(4-hydroxy-2-methyl-2-buten-1-yl)-L-cysteine (MA70)), 4-hydroxy-2-pentenal (N-acetyl-S-[1-(2-oxoethyl)-2-hydroxypropyl]-L-cysteine (MA25)), halogenated compounds (N-acetyl-S-(2-hydroxycyclohexyl)-L-cysteine (MA127), and N-acetyl-S-[(2-fluorophenyl)methyl]-L-cysteine (MA137)); SAs of 4-ethoxybenzyl alcohol (benzenemethanol, 4-ethoxy-, 1-(hydrogen sulfate)_7.01(SA46)) and acetylphenol (4-acetylphenol sulfate_5.68 (SA20) and 3-acetylphenol sulfate (SA19)); and GAs of phenylpropanoic acid (1-benzenepropanoate β-D-glucopyranuronic acid_6.89 (GA17)) and guaiacol (2-methoxyphenyl β-D-glucopyranosiduronic acid (GA53)). For PC3, higher scores (0.638 to 0.834) were most strongly correlated with higher levels of SAs and MAs exclusively derived from dietary compounds. These included metabolites of ferulic acid (SA92, SA93, GA75 and GA76), 4-vinylguaiacol (SA32), caffeic acid (SA71), and isovanillic acid (SA67). For PC4, higher scores (0.520 to 0.803) were correlated with metabolites of 4-oxo-2-hexenoic acid (GA96), 4-hydroxy-2-heptenal (GA38), trihydroxybenzene (GA164 and SA13), and dopamine (SA50).

3.8. Association between demographic characteristics and metabolites

To illustrate the relationship between metabolites and certain demographic characteristics (including age, sex, race, education, income, and tobacco use; detailed in Table 5), multivariable linear regression models were constructed. To begin, older age was associated with higher scores on PC1 (β = 0.12 per year; 95% CI = [0.05, 0.19]), PC2 (β = 0.13 per year; 95% CI = [0.07, 0.19]), and PC3 (β = 0.1 per year; 95% CI = [0.03, 0.17]). Next, self-reporting race as non-Hispanic Black was associated with lower scores on PC1 (β = −0.38; 95% CI = [−0.58, −0.20]) and PC4 (β = −0.35; 95% CI = [−0.54, −0.15]) when compared with non-Hispanic White participants. Comparing to participants with high school

education or less, higher PC1 scores were associated with some college or 2-year degree (β = 0.30; 95% CI = [0.13, 0.47]) and a 4-year college degree or more (β = 0.58; 95% CI = [0.39, 0.77]). Moreover, those who used tobacco tended to have higher scores on PC2 (β = 1.24; 95% CI = [1.12, 1.36]) and lower scores on PC3 (β = −0.24; 95% CI = [−0.35, −0.05]) when compared with non-tobacco use participants. For sex and income variables, the scores were not significant in relation to respective references.

4. Discussion

The phase II metabolism renders the lipophilic organic molecules into hydrophilic species and facilitates their excretion from the body. We utilized our UPLC-QTOF/MS with MS^E DIA workflow and ILGA approach to profile 424 phase II metabolites – GAs, SAs, and MAs in the urine of 844 study participants. To the best of our knowledge, this is the most comprehensive phase II exposomics study, both in the number of analytes measured and the sample size. The phase II metabolite libraries curated by this study comprises structures of 239 MAs, 431 GAs, and 148 SAs, which serve as a unique database for exposomics analyses. The tiered peak annotation confidence approach ascribes precision and the degree of confidence in metabolite assignment. Our rigorous peak annotation criteria helped to minimize spurious peak assignment and ensures appropriate peak identity. Two hundred and fourteen of these phase II metabolites originating from various aliphatic aldehydes, benzene and monocyclic substituted aromatic chemicals, dietary phenols and flavonoids, halogenated alicyclic compounds, halogenated aliphatic and halogenated aromatic substances, were present in the urine of >75% participants. One hundred and thirty-one phase II metabolites (40 MAs, 44 GAs, and 47 SAs) shared a common parent compound with at least one other metabolite group. The network analysis identified four distinct communities (1 to 4) of 37 positively correlated metabolites, and the PCA recognized four PCs that meaningfully explained at least 80% of the variance in the data.

Multiple approaches, depending on instrument and software used for data acquisition and processing, are employed for metabolite discovery and characterization. These include constant neutral loss/precursor ion monitoring, fragmentation filtering, in-source collision-induced dissociation, mass defect filtering, and isotope pattern filtering (Murray et al., 2023). To enhance compound discovery and confidence in annotation, we used a fragmentation filtering technique post data acquisition. Compounds with related structural moieties, such as phase II conjugates, share patterns in fragmentation (Levsen et al., 2005; Murray et al., 2023). Utilization of both CNL and CIF for metabolite screening increased specificity and decreased false positive hits. Moreover, unlike previous studies which used external databases for structural annotation (Chen et al., 2024; Frigerio et al., 2020; Correia et al., 2019), we built in-house libraries comprising not only published structures but also the putative structures of various metabolites (Table 2). This facilitated a more comprehensive profiling of phase II metabolites in this study. Furthermore, the dynamic nature of our libraries enables their expansion as more structures are reported or deduced.

Previous studies, including ours (Xie et al., 2023) profiled only a specific class of phase II metabolites such as MA or GA or SA conjugates (Chen et al., 2024; Frigerio et al., 2020; Fitzgerald et al., 2022), which limited the assessment of biomarkers of exposure

to a select metabolite. Also in the past, many metabolite discovery, phenotyping, and exposomics studies used a small sample size (n = <100) (Frigerio et al., 2020; Lu et al., 2019; Jin et al., 2020; Palermo et al., 2017). This limits the evaluation of biological and sample-to-sample variation and precise environmental exposure assessment at the individual level. In this study, we analyzed various MA, GA, and SA metabolites in a single chromatographic run. Our dilute and shoot approach eliminated the cumbersome sample processing (e.g., derivatization of functional groups, hydrolysis, extraction, etc.) and facilitated the sample analyses in a large cohort (n = 844 participants). Profiling of multiple phase II metabolites, representing different classes (Table 3), of a given chemical strengthens exposure assessment. This is very important because polymorphism of a select gene in a given metabolic pathway may affect analyte level or their potential biological activity. Measurement of multiple metabolites originating by distinctly different pathways not only provides comprehensive and rigorous exposure assessment but will also be useful in examining the biological activities and their association with health outcomes and potential toxicity in prospective studies.

Among the phase II urinary metabolites, GAs were the most prominent species. Approximately 70% of GAs (171 metabolites from 124 parent chemicals) were derived from aromatic compounds, predominantly from dietary phenolics and flavonoids (e.g., biochanin A, caffeic acid, p-methylguaiacol, vanillin) (Sarfraz et al., 2020; Magnani et al., 2014; Singh et al., 2011; Yu et al., 2013), benzene/monocyclic substituted aromatic compounds and industrial chemicals (e.g., benzoic acid, hydroquinone, phenol) (IARC, 2000), and pollutants generated from incomplete combustion of fossil fuels (e.g., styrene) (Miller et al., 1994), and biomass burning (e.g., catechol) (Finewax et al., 2018). Additionally, 17% of GAs were metabolites of aliphatic chemicals (e.g., hydroxyalkenals, hydroxyalkanoic acids, ethanol, glycerol, and methanol), and 13% were of cyclic parent compounds (e.g., menthol, limonene and perillic acid). Several of the GAs assayed by our study were also profiled by Chen et al. (2024) in the urine of 20 colorectal cancer patients by a chemical isotope labeling and dual-filtering strategy. They used N,N-Dimethylethylenediamine (DMED-d_0) and its deuterated counterpart DMED-d_6 to derivatize carboxyl-containing urinary metabolites and analyzed samples by UPLC-HRMS/MS in ESI positive mode, followed by a dual-filtering strategy to screen for DMED-labeled glucuronide metabolites. In another liquid chromatography-high resolution mass spectrometric study of the urine samples of 10 non-smokers and 10 combustible cigarette users, Kachhadia et al. (2024) observed that 171 biomarkers of exposure (peaks) were increased in combustible cigarette users, and 30 of these “features” were attributed to GA.

Apart from glucuronidation, sulfation is another major pathway of phase II metabolism. Several endogenous anabolic androgenic steroids, pharmaceutical reagents, dietary compounds, and microbiome are excreted from the body as sulfates (Correia et al., 2020; D’Agostino et al., 2024; Liu et al., 2021). In this study, we annotated 107 species derived from 43 proposed parent compounds. Approximately 93% of SAs were metabolites of aromatic xenobiotics, such as benzene/monocyclic substituted aromatics (e.g., 3-methoxy-4-hydroxyphenylglycol, 4-aminophenol, 4-ethoxybenzyl alcohol, benzoic acid, catechol, ethylbenzene, homovanillic acid, and trihydroxybenzene) and dietary phenols and flavonoids (e.g., 4-vinylguaiacol, caffeic acid, coumaric acid, eugenol, ferulic acid, and isovanillic

acid). The remaining 7% SAs were derived from three other aliphatic parents (2-buten-1-ol, glucose, and N-acetylglucosamine), one bile salt (glycochenodeoxycholic acid), one steroid (i.e., epiandrosterone), and two other cyclic compounds (ascorbic acid and pyridine). Some of these urinary organic sulfates (e.g., catechol derivatives) were reported to be elevated in 10 combustible cigarette users by Kachhadia et al. (2024).

Profiling of urinary MAs revealed that 111 out of 116 MAs reported by us recently (Xie et al., 2023) were also detected in the present study and depicted the reproducibility of our assay. Now we have expanded our library by adding more reported and proposed structures, which lead to annotating 35 new MA features (derived from compounds including limonene, 2-hexenal, cinnamaldehyde, halogenated butene, benzene, and phenanthrene) when compared to our previous work. Several MAs, especially those are metabolites of acrolein, acrylamide, 1,3-butadiene, crotonaldehyde, trans 2-hexenal, N,N-dimethylformamide, glycidol, and 4-hydroxy-2-pentenal were also identified by Kachhadia et al. (2024). We observed that the urinary levels of several of these MA, especially those originating from acrolein, crotonaldehyde, 1,3-butadiene, N,N-dimethylformamide, acrylamide, naphthalene, and phenanthrene, were markedly higher in smokers than non-smokers (Xie et al., 2023). Sources of MAs of halogenated alkanes, ethyl acrylate, isobutyl acrylates, trimethylbenzene, 2-fluorobenzaldehyde, 4-*tert*-butylbenzaldehyde, and 4-chloronitrobenzene are likely to be industrial and occupational (Fishbein, 1979; Walker et al., 1991; Kostrzewski et al., 1997; Yu et al., 2007; Jones et al., 2006), whereas probable sources of MAs of 4-methoxybenzaldehyde, acetaminophen, and 6-chloropurine include biogenic and pharmaceutical parent compounds (Ghalla et al., 2018; Taylor et al., 2012; Bendich et al., 1954). A potential source for MAs of hydroxyalkenals is endogenous lipid peroxidation, especially of ω-6 polyunsaturated fatty acids. Additionally, lipid oxidation also produces other alkanals and alkenals. Moreover, alkenals such as acrolein are also generated as a byproduct of myeloperoxidase-driven inflammation and by the metabolism of the chemotherapeutic agent cyclophosphamide and the industrial chemical allylamine (Srivastava et al., 1998; Vladykovskaya et al., 2012; Srivastava et al., 1999). Due to their high electrophilicity, the alkenals hydroxyalkenals readily crosslink with DNA and proteins. Increased accumulation of protein adducts of these molecules has been detected in pathological conditions such as atherosclerosis, Alzheimer's, and Parkinson's diseases (Srivastava et al., 1998; Vladykovskaya et al., 2012; Srivastava et al., 1999). We observed that MAs of acrolein and 1,3-butadience were associated with endothelial dysfunction and may increase the risk for hypertension (McGraw et al., 2021; McGraw et al., 2023). MAs of ethylbenzene/styrene, benzene, and xylene allied with endothelial injury (Riggs et al., 2022), and MAs of acrolein; acrylamide, acrylonitrile, butadiene, crotonaldehyde, and styrene were positively correlated with alkaline phosphatase, a biomarker for cholestatic injury (Wahlang et al., 2023).

Leveraging these data in network and principal component analyses allowed us to detect distinct clusters of metabolites that could be related to common co-exposures and demographic characteristics in this population. For example, the most influential metabolites in cluster 1 were phenylpropanoic acid and methylparaben. These compounds are both benzene/monocyclic substituted aromatics that can be found in processed foods and cosmetics as preservatives. Thus, higher co-exposure may occur from common sources,

and the combined effect of these exposures may need to be considered in studies of health effects. Analyses examining the association between principal components and demographic characteristics could also provide insights about factors that are associated with specific exposure profiles. For example, higher PC1 scores indicated higher levels of metabolites for dietary phenols and flavonoids, which may be related with higher consumption of fruits and vegetables. Participants with higher metabolites corresponding with these types of exposures tended to be non-Hispanic White, as well as have a higher education and income. Future research linking these metabolite data with diet and behavioral information, as well as health measures, in these types of analyses could identify common sources contributing to co-exposures and disease.

5. Limitations

While our assay is robust and comprehensive to profile phase II urinary metabolites, there are certain limitations to the approach. *One*, we have used spot urine for phase II metabolite profiling. Therefore, the data provide only a snapshot of exposures. Further studies with a 24 hr urine collection or multiple urine samples per individual per day are required to account for cumulative short-term (24 hr) exposure. *Two*, we diluted the urine samples and directly injected them into the UPLC-QTOF/MS system for analysis. It is possible that intensities of some analytes might be suppressed or enhanced by coeluting molecules in the samples (Zhou et al., 2017). *Three*, our libraries contain both reported and proposed structures of compounds and metabolites. Because it is a dynamic and expandable process, data analyses are limited to published and putative metabolites. As the field grows and new metabolites are discovered and published, they will be included in the library. *Four*, only Level 1 metabolites were identified with authentic standards, and annotations of Levels 2A, 2B, and 3 are putative. Thus, additional studies are required to confirm these putative/potential metabolites, especially metabolites that were found prevalently and/or at higher levels in samples.

6. Conclusion

This study has established a new robust and comprehensive LC-MS assay for the profiling of a plethora of phase II metabolites derived from environmental exposures, dietary compounds, tobacco products, and pharmaceutical agents and generated endogenously by oxidative stress and inflammation. The dynamic nature of our library facilitates constant updates to the database as more metabolites are uncovered.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

This work was supported in parts by NIH grants P42ES023716, R01ES029846, R01ES033531, U54HL120163 and S10OD026840. J.Y. C. and S.R.S. were supported by NIH T32 training grant (T32ES011564).

Data availability

Data will be made available on request.

References

Al-Horani RA, Desai UR, 2010. Chemical sulfation of small molecules-advances and challenges. Tetrahedron 66 (16), 2907–2918. 10.1016/j.tet.2010.02.015. [PubMed: 20689724]

Athersuch TJ, Keun HC, 2015. Metabolic profiling in human exposome studies. Mutagenesis 30 (6), 755–762. 10.1093/mutage/gev060. [PubMed: 26290610]

Balcells C, Xu Y, Gil-Solsona R, Maitre L, Gago-Ferrero P, Keun HC, 2024. Blurred lines: Crossing the boundaries between the chemical exposome and the metabolome. Curr. Opin. Chem. Biol 78, 102407. [PubMed: 38086287]

Bendich A, Russell PJ Jr, Fox JJ, 1954. The synthesis and properties of 6-chloropurine and purine1. J. Am. Chem. Soc 76 (23), 6073–6077.

Bhatnagar A, 2006. Environmental cardiology - studying mechanistic links between pollution and heart disease. Circ. Res 99 (7), 692–705. 10.1161/01.RES.0000243586.99701.cf. [PubMed: 17008598]

Blazenovic I, Kind T, Ji J, Fiehn O, 2018. Software tools and approaches for compound identification of LC-MS/MS data in metabolomics. Metabolites 8 (2), 31. 10.3390/metabo8020031. [PubMed: 29748461]

Chen ZQ, Yang RJ, Zhu CW, Li Y, Yan R, Wan JB, 2024. Chemical isotope labeling and dual-filtering strategy for comprehensive profiling of urinary glucuronide conjugates. Anal. Chem 96 (33), 13576–13587. 10.1021/acs.analchem.4c02339. [PubMed: 39102235]

Correia MSP, Jain A, Alotaibi W, Young Tie Yang P, Rodriguez-Mateos A, Globisch D, 2020. Comparative dietary sulfated metabolome analysis reveals unknown metabolic interactions of the gut microbiome and the human host. Free Radic. Biol. Med 160, 745–754. 10.1016/j.freeradbiomed.2020.09.006. [PubMed: 32927015]

Correia MS, Rao M, Ballet C, Globisch D, 2019. Coupled enzymatic treatment and mass spectrometric analysis for identification of glucuronidated metabolites in human samples. Chembiochem 20 (13), 1678–1683. [PubMed: 30803115]

Croom E, 2012. Metabolism of xenobiotics of human environments. Prog. Mol. Biol. Transl. Sci 112, 31–88. 10.1016/B978-0-12-415813-9.00003-9. [PubMed: 22974737]

Crozier A, Jaganath IB, Clifford MN, 2009. Dietary phenolics: chemistry, bioavailability and effects on health. Nat. Prod. Rep 26 (8), 1001–1043. 10.1039/b802662a. [PubMed: 19636448]

D'Agostino GD, Chaudhari SN, Devlin AS, 2024. Host-microbiome orchestration of the sulfated metabolome. Nat. Chem. Biol 20 (4), 410–421. 10.1038/s41589-023-01526-9. [PubMed: 38347214]

Ding X, Yang F, Chen Y, Xu J, He J, Zhang R, Abliz Z, 2022. Norm ISWSVR: a data integration and normalization approach for large-scale metabolomics. Anal. Chem 94 (21), 7500–7509. 10.1021/acs.analchem.1c05502. [PubMed: 35584098]

Evich MG, Mosley JD, Ntai I, Cavallin JE, Villeneuve DL, Ankley GT, Collette TW, Ekman DR, 2022. Untargeted MS(n)-based monitoring of glucuronides in fish: screening complex mixtures for contaminants with biological relevance. ACS ES T Water 2 (12), 2481–2490. 10.1021/acsestwater.2c00310. [PubMed: 37288388]

Finewax Z, de Gouw JA, Ziemann PJ, 2018. Identification and quantification of 4-nitrocatechol formed from OH and NO(3) radical-initiated reactions of catechol in air in the presence of NO(x): implications for secondary organic aerosol formation from biomass burning. Environ. Sci. Technol 52 (4), 1981–1989. 10.1021/acs.est.7b05864. [PubMed: 29353485]

Fishbein L, 1979. Potential halogenated industrial carcinogenic and mutagenic chemicals. III. Alkane halides, alkanols and ethers. Sci. Total Environ 11 (3), 223–257. 10.1016/0048-9697(79)90076-7. [PubMed: 155878]

Fitzgerald CC, Hedman R, Uduwela DR, Paszerbovics B, Carroll AJ, Neeman T, Cawley A, Brooker L, McLeod MD, 2022. Profiling urinary sulfate metabolites with mass spectrometry. Front. Mol. Biosci 9, 829511. [PubMed: 35281273]

Frigerio G, Mercadante R, Campo L, Polledri E, Boniardi L, Olgiati L, Missineo P, Nash WJ, Dunn WB, Fustinoni S, 2020. Urinary biomonitoring of subjects with different smoking habits. Part II: an untargeted metabolomic approach and the comparison with the targeted measurement of mercapturic acids. Toxicol. Lett 329, 56–66. 10.1016/j.toxlet.2020.03.020. [PubMed: 32380120]

Ghalla H, Issaoui N, Bardak F, Atac A, 2018. Intermolecular interactions and molecular docking investigations on 4-methoxybenzaldehyde. Comput. Mater. Sci 149, 291–300. 10.1016/j.commatsci.2018.03.042.

Gonzalez-Dominguez R, Jauregui O, Queipo-Ortuno MI, Andres-Lacueva C, 2020. Characterization of the human exposome by a comprehensive and quantitative large-scale multianalyte metabolomics platform. Anal. Chem 92 (20), 13767–13775. 10.1021/acs.analchem.0c02008. [PubMed: 32966057]

Hanna PE, Anders MW, 2019. The mercapturic acid pathway. Crit. Rev. Toxicol 49 (10), 819–929. 10.1080/10408444.2019.1692191. [PubMed: 31944156]

Hashmi M, Vamvakas S, Anders M, 1992. Bioactivation mechanism of S-(3-oxopropyl)-N-acetyl-L-cysteine, the mercapturic acid of acrolein. Chem. Res. Toxicol 5 (3), 360–365. [PubMed: 1504259]

Huber C, Krauss M, Reinstadler V, Denicolò S, Mayer G, Schulze T, Brack W, Oberacher H, 2022. In silico deconjugation of glucuronide conjugates enhances tandem mass spectra library annotation of human samples. Anal. Bioanal. Chem 414 (8), 2629–2640. [PubMed: 35080654]

IARC, 2000. IARC Monographs on the Evaluation of Carcinogenic Risks to Humans: Some Industrial Chemicals; IARC Working Group on the Evaluation of Carcinogenic Risks to Humans.

Jancova P, Anzenbacher P, Anzenbacherova E, 2010. Phase II drug metabolizing enzymes. Biomed. Pap. Med. Fac. Univ. Palacky Olomouc Czech Repub 154 (2), 103–116. [PubMed: 20668491]

Jin R, McConnell R, Catherine C, Xu S, Walker DI, Stratakis N, Jones DP, Miller GW, Peng C, Conti DV, et al. , 2020. Perfluoroalkyl substances and severity of nonalcoholic fatty liver in children: an untargeted metabolomics approach. Environ. Int 134, 105220. 10.1016/j.envint.2019.105220. [PubMed: 31744629]

Johnson CH, Patterson AD, Idle JR, Gonzalez FJ, 2012. Xenobiotic metabolomics: major impact on the metabolome. Annu. Rev. Pharmacol. Toxicol 52, 37–56. 10.1146/annurev-pharmtox-010611-134748. [PubMed: 21819238]

Jones CR, Liu YY, Sepai O, Yan H, Sabbioni G, 2006. Internal exposure, health effects, and cancer risk of humans exposed to chloronitrobenzene. Environ. Sci. Technol 40 (1), 387–394. 10.1021/es050693p. [PubMed: 16433376]

Kachhadia A, Burkhardt T, Scherer G, Scherer M, Pluym N, 2024. Development of an LC-HRMS non-targeted method for comprehensive profiling of the exposome of nicotine and tobacco product users - a showcase for cigarette smokers. J. Chromatogr. B Anal. Technol. Biomed. Life Sci 1247, 124330. 10.1016/j.jchromb.2024.124330.

Kostrzewski P, Wiaderna-Brycht A, Czerski B, 1997. Biological monitoring of experimental human exposure to trimethylbenzene. Sci. Total Environ 199 (1–2), 73–81. 10.1016/s0048-9697(97)05504-6. [PubMed: 9200849]

Lafaye A, Junot C, Ramounet-Le Gall B, Fritsch P, Ezan E, Tabet JC, 2004. Profiling of sulfoconjugates in urine by using precursor ion and neutral loss scans in tandem mass spectrometry. Application to the investigation of heavy metal toxicity in rats. J. Mass Spectrom 39 (6), 655–664. 10.1002/jms.635. [PubMed: 15236304]

Levsen K, Schiebel HM, Behnke B, Dotzer R, Dreher W, Elend M, Thiele H, 2005. Structure elucidation of phase II metabolites by tandem mass spectrometry: an overview. J. Chromatogr. A 1067 (1–2), 55–72. 10.1016/j.chroma.2004.08.165. [PubMed: 15844510]

Liu KH, Lee CM, Singer G, Bais P, Castellanos F, Woodworth MH, Ziegler TR, Kraft CS, Miller GW, Li S, et al. , 2021. Large scale enzyme based xenobiotic identification for exposomics. Nat. Commun 12 (1), 5418. 10.1038/s41467-021-25698-x. [PubMed: 34521839]

Lorkiewicz P, Riggs DW, Keith RJ, Conklin DJ, Xie Z, Sutaria S, Lynch B, Srivastava S, Bhatnagar A, 2019. Comparison of urinary biomarkers of exposure in humans using electronic cigarettes,

combustible cigarettes, and smokeless tobacco. Nicotine Tobacco Res. : Off. J. Soc. Res. Nicotine Tobacco 21 (9), 1228–1238. 10.1093/ntr/nty089.

Lu Y, Gao K, Li X, Tang Z, Xiang L, Zhao H, Fu J, Wang L, Zhu N, Cai Z, et al. , 2019. Mass spectrometry-based metabolomics reveals occupational exposure to per- and polyfluoroalkyl substances relates to oxidative stress, fatty acid beta-oxidation disorder, and kidney injury in a manufactory in China. Environ. Sci. Technol 53 (16), 9800–9809. 10.1021/acs.est.9b01608. [PubMed: 31246438]

Magnani C, Isaac VLB, Correa MA, Salgado HRN, 2014. Caffeic acid: a review of its potential use in medications and cosmetics. Anal. Methods-UK 6 (10), 3203–3210. 10.1039/c3ay41807c.

McGraw KE, Riggs DW, Rai S, Navas-Acien A, Xie Z, Lorkiewicz P, Lynch J, Zafar N, Krishnasamy S, Taylor KC, et al. , 2021. Exposure to volatile organic compounds - acrolein, 1,3-butadiene, and crotonaldehyde - is associated with vascular dysfunction. Environ. Res 196, 110903. 10.1016/j.envres.2021.110903. [PubMed: 33636185]

McGraw KE, Konkle SL, Riggs DW, Rai SN, DeJarnett N, Xie Z, Keith RJ, Oshunbade A, Hall ME, Shimbo D, et al. , 2023. Exposure to volatile organic compounds is associated with hypertension in black adults: the Jackson heart study. Environ. Res 223, 115384. 10.1016/j.envres.2023.115384. [PubMed: 36796615]

Miller RR, Newhook R, Poole A, 1994. Styrene production, use, and human exposure. Crit. Rev. Toxicol 24 (Sup1), S1–10. 10.3109/10408449409020137. [PubMed: 7818766]

Murray KJ, Villalta PW, Griffin TJ, Balbo S, 2023. Discovery of modified metabolites, secondary metabolites, and xenobiotics by structure-oriented LC-MS/MS. Chem. Res. Toxicol 36 (11), 1666–1682. 10.1021/acs.chemrestox.3c00209. [PubMed: 37862059]

Palermo A, Botre F, de la Torre X, Zamboni N, 2017. Non-targeted LC-MS based metabolomics analysis of the urinary steroidal profile. Anal. Chim. Acta 964, 112–122. 10.1016/j.aca.2017.01.055. [PubMed: 28351627]

Richardson DB, Ciampi A, 2003. Effects of exposure measurement error when an exposure variable is constrained by a lower limit. Am. J. Epidemiol 157 (4), 355–363. 10.1093/aje/kwf217. [PubMed: 12578806]

Riggs DW, Malovichko MV, Gao H, McGraw KE, Taylor BS, Krivokhizhina T, Rai SN, Keith RJ, Bhatnagar A, Srivastava S, 2022. Environmental exposure to volatile organic compounds is associated with endothelial injury. Toxicol. Appl. Pharmacol 437, 115877. 10.1016/j.taap.2022.115877. [PubMed: 35045333]

Sarfraz A, Javeed M, Shah MA, Hussain G, Shafiq N, Sarfraz I, Riaz A, Sadiqa A, Zara R, Zafar S, et al. , 2020. Biochanin A: a novel bioactive multifunctional compound from nature. Sci. Total Environ 722, 137907. 10.1016/j.scitotenv.2020.137907. [PubMed: 32208265]

Scalbert A, Brennan L, Manach C, Andres-Lacueva C, Dragsted LO, Draper J, Rappaport SM, van der Hooft JJ, Wishart DS, 2014. The food metabolome: a window over dietary exposure. Am. J. Clin. Nutr 99 (6), 1286–1308. 10.3945/ajcn.113.076133. [PubMed: 24760973]

Singh DP, Chong HH, Pitt KM, Cleary M, Dokoozlian NK, Downey MO, 2011. Guaiacol and 4-methylguaiacol accumulate in wines made from smoke-affected fruit because of hydrolysis of their conjugates. Aust. J. Grape Wine Res 17 (2), S13–S21. 10.1111/j.1755-0238.2011.00128.x.

Srivastava S, Chandra A, Wang LF, Seifert WE, 1998. Metabolism of the lipid peroxidation product, 4-hydroxy-trans-2-nonenal, in isolated perfused rat heart. Metabolism of the lipid peroxidation product, 4-hydroxy-trans-2-nonenal, in isolated perfused rat heart. doi: 10.1074/jbc.273.18.10893 READCUBE.

Srivastava S, Watowich SJ, Petrash JM, Srivastava SK, Bhatnagar A, 1999. Structural and kinetic determinants of aldehyde reduction by aldose reductase. Biochemistry-US 38 (1), 42–54. 10.1021/bi981794l.

Succop PA, Clark S, Chen M, Galke W, 2004. Imputation of data values that are less than a detection limit. J. Occup. Environ. Hyg 1 (7), 436–441. 10.1080/15459620490462797. [PubMed: 15238313]

Sun J, Fang R, Wang H, Xu DX, Yang J, Huang X, Cozzolino D, Fang M, Huang Y, 2022. A review of environmental metabolism disrupting chemicals and effect biomarkers associating disease risks: where exposomics meets metabolomics. Environ. Int 158, 106941. 10.1016/j.envint.2021.106941. [PubMed: 34689039]

Taylor R, Pergolizzi JV, Raffa RB, 2012. Acetaminophen (paracetamol): properties, clinical uses, and adverse effects. In: Acetaminophen: Properties, Clinical Uses and Adverse Effects. Nova Science Publishers Inc, pp. 1–24.

Testa B, 2007. 5.06 - Principles of drug metabolism 2: hydrolysis and conjugation reactions. Compreh. Med. Chem II (5), 133–166. 10.1016/B0-08-045044-X/00120-6.

Vladykovskaya E, Sithu SD, Haberzettl P, Wickramasinghe NS, Merchant ML, Hill BG, McCracken J, Agarwal A, Dougherty S, Gordon SA, et al. , 2012. Lipid peroxidation product 4-hydroxy-trans-2-nonenal causes endothelial activation by inducing endoplasmic reticulum stress. J. Biol. Chem 287 (14), 11398–11409. 10.1074/jbc.M111.320416. [PubMed: 22228760]

Wagner S, Scholz K, Donegan M, Burton L, Wingate J, Volkel W, 2006. Metabonomics and biomarker discovery: LC-MS metabolic profiling and constant neutral loss scanning combined with multivariate data analysis for mercapturic acid analysis. Anal. Chem 78 (4), 1296–1305. 10.1021/ac051705s. [PubMed: 16478125]

Wahlang B, Gao H, Rai SN, Keith RJ, McClain CJ, Srivastava S, Cave MC, Bhatnagar A, 2023. Associations between residential volatile organic compound exposures and liver injury markers: the role of biological sex and race. Environ. Res 221, 115228. 10.1016/j.envres.2023.115228. [PubMed: 36610539]

Walker AM, Cohen AJ, Loughlin JE, Rothman KJ, DeFonso LR, 1991. Mortality from cancer of the colon or rectum among workers exposed to ethyl acrylate and methyl methacrylate. Scand. J. Work Environ. Health 17 (1), 7–19. 10.5271/sjweh.1731.

Wells PG, Mackenzie PI, Chowdhury JR, Guillemette C, Gregory PA, Ishii Y, Hansen AJ, Kessler FK, Kim PM, Chowdhury NR, et al. , 2004. Glucuronidation and the UDP-glucuronosyltransferases in health and disease. Drug Metab. Dispos 32 (3), 281–290. 10.1124/dmd.32.3.281. [PubMed: 14977861]

Xie Z, Chen JY, Gao H, Keith RJ, Bhatnagar A, Lorkiewicz P, Srivastava S, 2023. Global profiling of urinary mercapturic acids using integrated library-guided analysis. Environ. Sci. Technol 57 (29), 10563–10573. 10.1021/acs.est.2c09554. [PubMed: 37432892]

Xie Z, Sutaria SR, Chen JY, Gao H, Conklin DJ, Keith RJ, Srivastava S, Lorkiewicz P, Bhatnagar A, 2024. Evaluation of urinary limonene metabolites as biomarkers of exposure to greenness. Environ. Res 245, 117991. 10.1016/j.envres.2023.117991. [PubMed: 38141921]

Yan Z, Caldwell GW, Jones WJ, Masucci JA, 2003. Cone voltage induced in-source dissociation of glucuronides in electrospray and implications in biological analyses. Rapid Commun. Mass Spectrom 17 (13), 1433–1442. 10.1002/rcm.1071. [PubMed: 12820208]

Yeager R, Browning MHEM, Breyer E, Ossola A, Larson LR, Riggs DW, Rigolon A, Chandler C, Fleischer D, Keith R, et al. , 2023. Greenness and equity: complex connections between intra-neighborhood contexts and residential tree planting implementation. Environ. Int 176, 107955. 10.1016/j.envint.2023.107955. [PubMed: 37196566]

Yu J, Han JC, Hua LM, Gao YJ, 2013. In vitro characterization of glucuronidation of vanillin: identification of human UDP-glucuronosyltransferases and species differences. Phytother. Res 27 (9), 1392–1397. 10.1002/ptr.4885. [PubMed: 23184728]

Yu WH, Zhang Z, Wang H, Ge ZH, Pinnavala TJ, 2007. Peroxide oxidation of 4-tert-butyltoluene to 4-tert-butylbenzaldehyde over titanium(IV)-functionalized mesostructured silica. Micropor. Mesopor. Mater 104 (1–3), 151–158. 10.1016/j.micromeso.2007.01.022.

Zhou W, Yang S, Wang PG, 2017. Matrix effects and application of matrix effect factor. Taylor & Francis: Bioanalysis 9, 1839–1844.

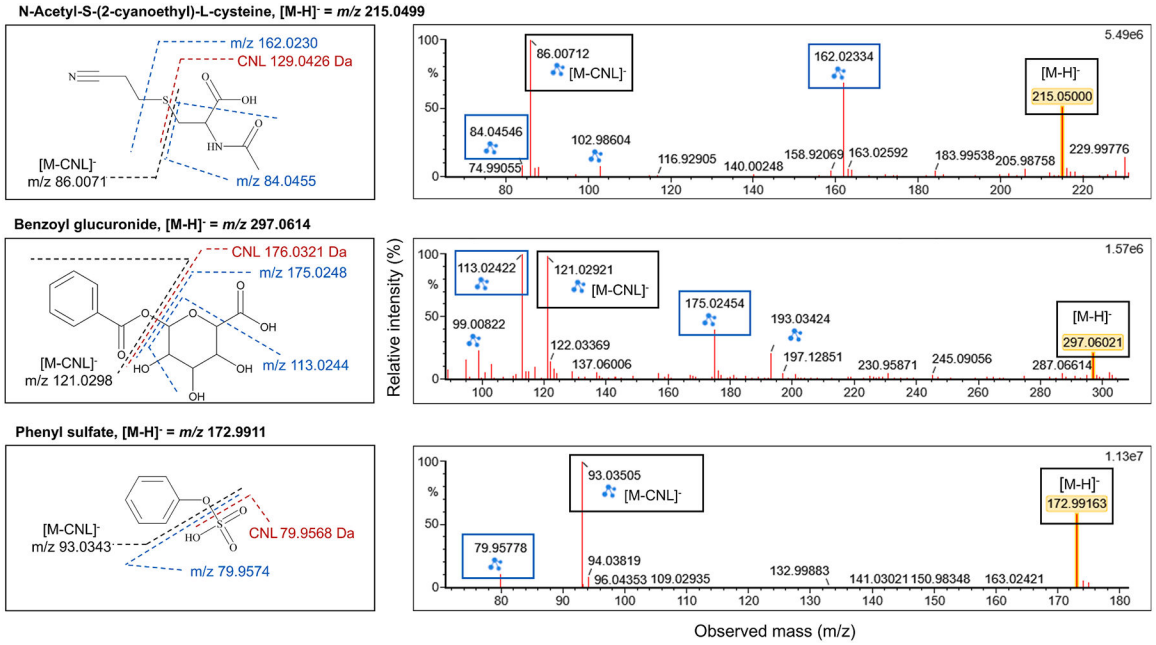

Fig. 1.

Characteristic fragmentations and mass spectra of a representative mercapturic acid (MA – N-acetyl-S-(2-cyanoethyl)-L-cysteine), glucuronic acid (GA – benzoyl glucuronide), and sulfate (SA – phenyl sulfate) metabolite. Complete list of common neutral losses and ion fragments used for metabolite annotation are provided in Table 1.

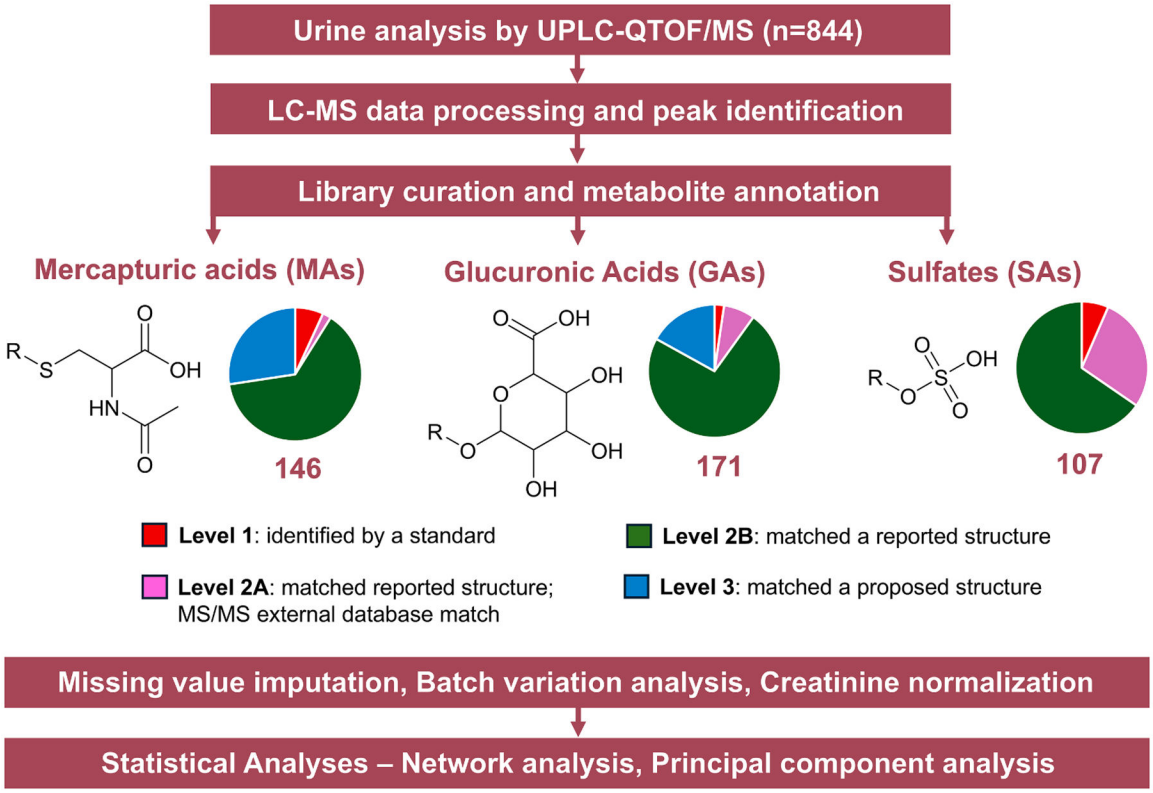

Fig. 2.

General workflow showing mass spectrometric analysis of urine samples, annotation of phase II metabolites (MAs, GAs and SAs) and post-analytical data and statistical analyses.

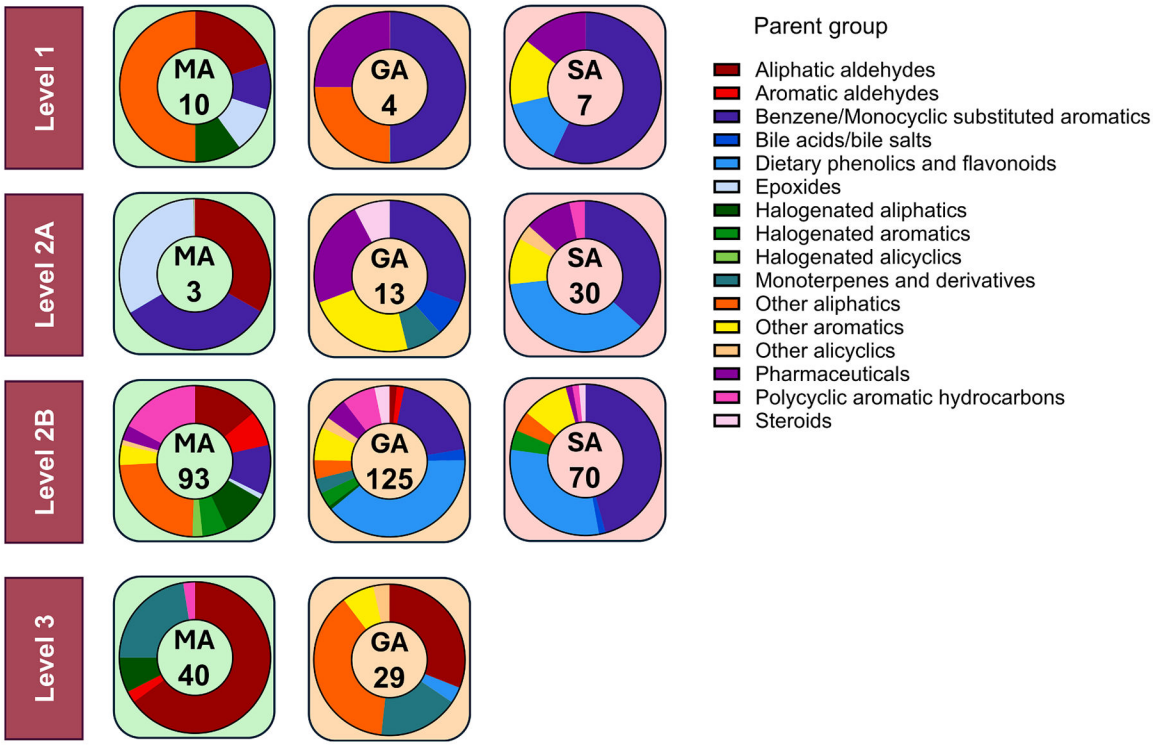

Fig. 3.

Metabolites detected in all participant samples (n = 844) categorized by annotation confidence levels and parent groups.

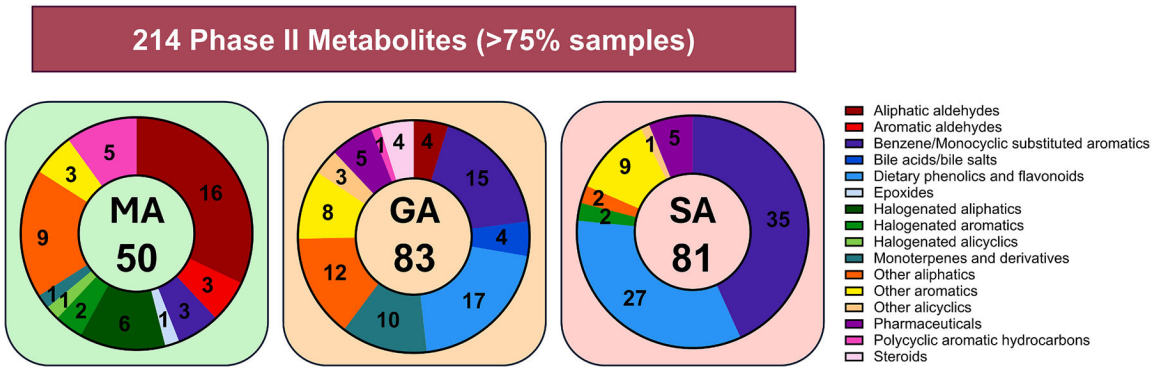

Fig. 4.

Metabolites uncovered in >75 % of participant samples categorized by parent groups.

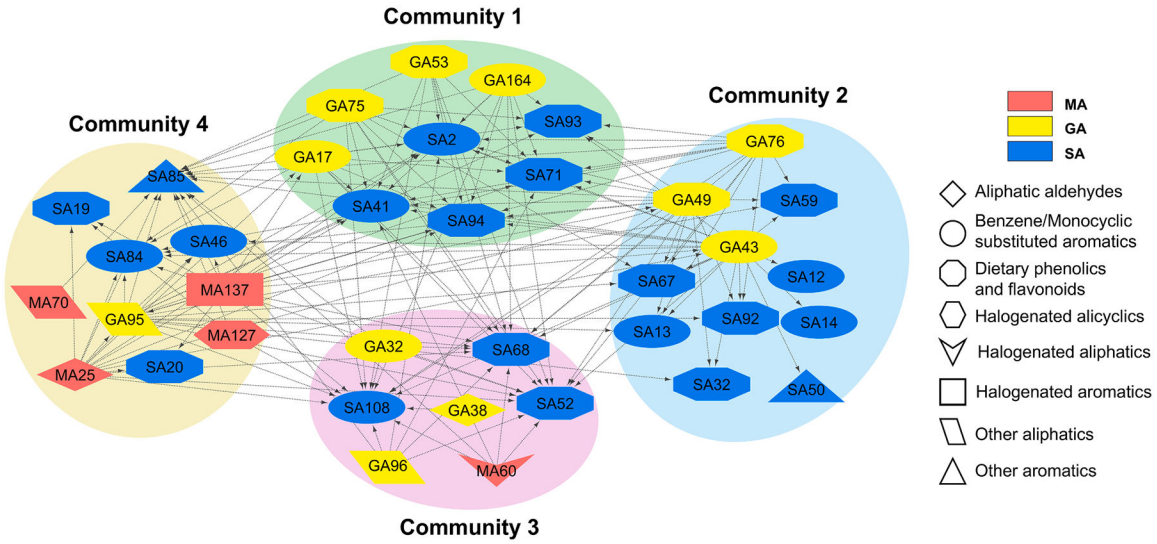

Fig. 5.

Network analysis map showing 37 positively correlated metabolites in four communities. Nodes represent individual metabolites. All metabolites have a positive correlation that met the correlation threshold of 0.6 and were statistically significant ($p < 0.05$).

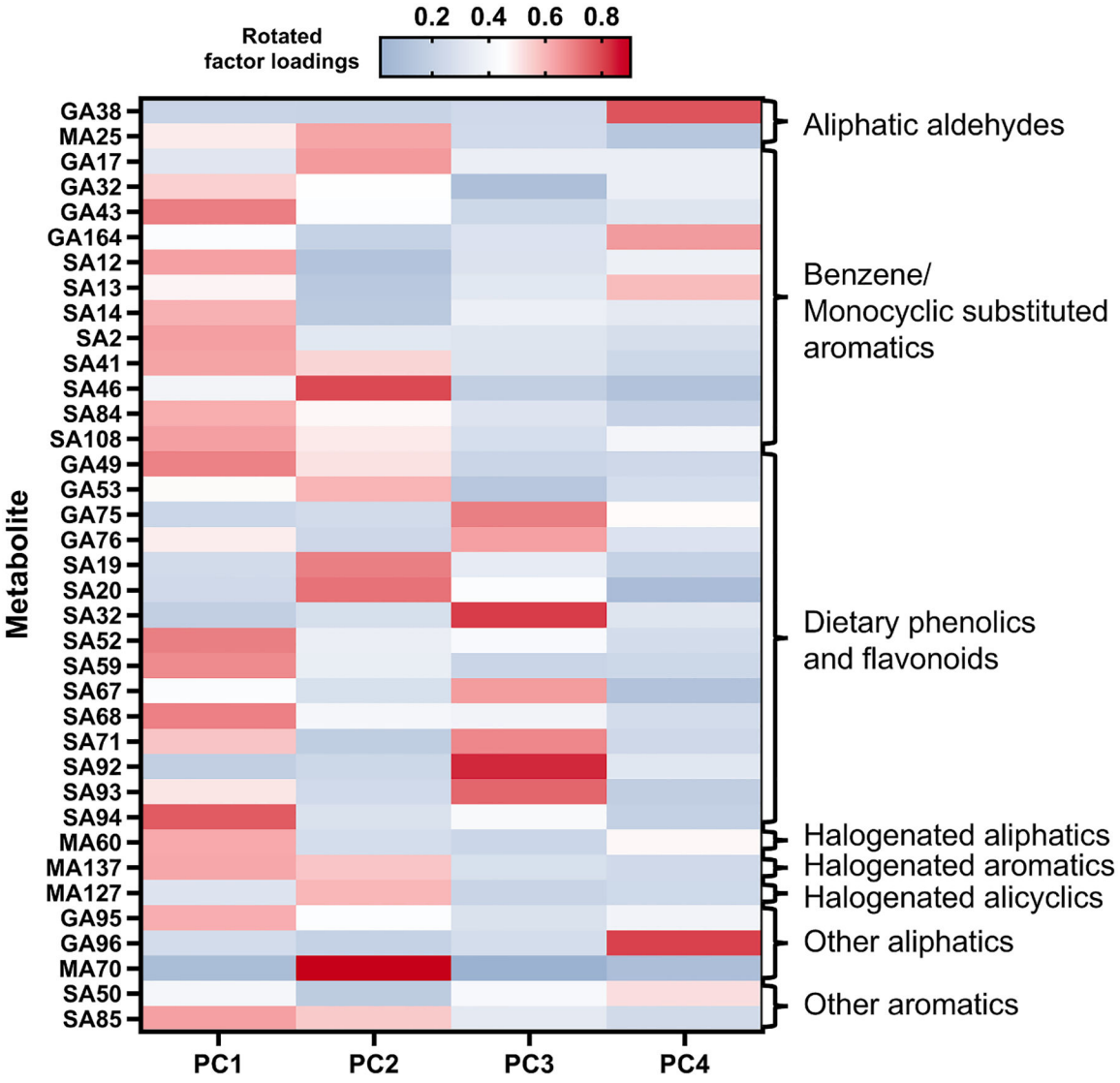

Fig. 6.

Principal component analysis (PCA) showing the relationship between four PCs and levels of 37 metabolites from the network analysis.

Table 1

Reported and observed fragmentation signatures – common neutral loss (CNL) and common ion fragment (CIF), of mercapturic acid (MA), glucuronic acid (GA), and sulfate (SA) used in this study for metabolite annotation.

Metabolite group	CNL (formula, Da) Reported	CIF (formula, *m/z*)	
		Reported	Observed in this study
Mercapturic acid (MA) (Xie et al., 2023; Wagner et al., 2006)	$C_5H_7NO_3$ (129.0426)	$C_5H_8NO_3S^-$ (162.0230) $C_5H_6NO_3^-$ (128.0353) $C_4H_6NO^-$ (84.0455) $C_3H_6S^-$ (74.0196)	
Glucuronic acid (GA) (Levsen et al., 2005; Huber et al., 2022; Yan et al., 2003; Evich et al., 2022)	$C_6H_{10}O_7$ (194.0427) $C_6H_8O_6$ (176.0321)	$C_6H_7O_6^-$ (175.0248) $C_5H_5O_3^-$ (113.0244) $C_4H_3O_3^-$ (99.0088) $C_5H_3O_2^-$ (95.0139) $C_4H_5O_2^-$ (85.0295) $C_2H_3O_3^-$ (75.0088)	$C_6H_9O_7^-$ (193.0354) $C_4H_5O_4^-$ (117.0193) $C_4H_7O_3^-$ (103.0401) $C_3H_3O_3^-$ (87.0088)
Sulfate (SA) (Levsen et al., 2005; Lafaye et al., 2004)	H_2SO_4 (97.9674) SO_3 (79.9568)	HSO_4^- (96.9601) SO_4^- (95.9523) HSO_3 (80.9652) SO_3^- (79.9574)	

Table 2

Description and data requirements for annotation confidence levels, and number of annotated metabolites for each level.

Annotation confidence	Description	Data requirement	MA	GA	SA
Level 1	Metabolite with confident 2D structure. Matches a reference standard or full structure elucidation.	MS/MS and retention time match.	10	4	7
Level 2A	Metabolite with probable structure. Matches an entry in the published databases and literature.	MS/MS match with *in silico* fragmentation based on the reported structure (.mol file) and external database elucidation (ChemBank, DrugBank, HMDB, MassBank, NIST Spectra, PubMed, and Urine Metabolome Database).	3	13	30
Level 2B	Metabolite with probable structure. Matches an entry in the published databases and literature.	MS/MS match with *in silico* fragmentation based on the reported structure (.mol file)	93	125	70
Level 3	Metabolite with proposed structure. Matches proposed structure in curated library.	One or more library entries with matching *m/z* and elemental composition based on the formula.	40	29	n/a
Level 4 *	Potential metabolite with undetermined structure.Match one of following selection criteria (refer to Table 1): • at least one CNL and one CIF • at least two CIFs	No library match.	n/a	n/a	n/a
		Total annotated	**146**	**171**	**107**

n/a: not applicable.

* Level 4 annotations are not discussed in this manuscript.

Table 3

Number of phase II metabolites sharing a common parent compound.

#	Parent	MA	GA	SA
1	Acetaminophen	2	2	5
2	Benzo[a]pyrene	1	2	1
3	Naphthalene	9	3	1
4	Trihydroxybenzene	2	3	4
5	4-Hydroxy-2-heptenal	4	4	
6	4-Hydroxy-2-nonenal	3	2	
7	4-Hydroxy-2-octenal	5	3	
8	Cinnamaldehyde	3	2	
9	Hydroquinone	1	1	
10	Limonene	3	3	
11	Phenanthrene	6	3	
12	Styrene	1	1	
13	3-Methoxy-4-hydroxyphenylglycol		1	6
14	Benzoic acid		1	3
15	Caffeic acid		3	8
16	Catechol		1	1
17	Cresol		1	1
18	Dopamine		1	4
19	Ferulic acid		2	3
20	Glycochenodeoxycholic acid		1	1
21	Guaiacol		1	1
22	Homovanillic acid		1	5
23	Indole		1	2
24	Phenol		1	1
	Total	**40**	**44**	**47**

Table 4

Level 1 MA, GA, and SA metabolites identified by matching retention times and mass spectra of authentic standards.

Short name	Full name	Formula	Observed [M-H]$^-$, *m/z*	CAS #	Metabolite parent
MAs					
MA38	N-Acetyl-S-(3,4-dihydroxybutyl)-L-cysteine	C9H17NO5S	250.0753	144889-50-9	1,3-Butadiene
MA15	N-Acetyl-S-(3-hydroxypropyl)-L-cysteine	C8H15NO4S	220.0647	23127-40-4	Acrolein
MA36	N-Acetyl-S-(2-carboxyethyl)-L-cysteine	C8H13NO5S	234.0439	51868-61-2	Acrolein
MA29	N-Acetyl-S-(2-carbamoylethyl)-L-cysteine	C8H14N2O4S	233.0608	81690-92-8	Acrylamide
MA40	N-Acetyl-S-(2-carbamoyl-2-hydroxyethyl)-L-cysteine	C8H14N2O5S	249.0552	137698-08-9	Acrylamide
MA37	N-Acetyl-S-(2-cyanoethyl)-L-cysteine	C8H12N2O3S	215.0511	74514-75-3	Acrylonitrile
MA35	N-Acetyl-S-(n-propyl)-L-cysteine	C8H15NO3S	204.0688	14402-54-1	Halogenated propane
MA32	N-Acetyl-S-(N-methylcarbamoyl)-L-cysteine	C7H12N2O4S	219.0448	103974-29-4	N,N-dimethylformamide
MA11	N-Acetyl-S-(2-hydroxypropyl)-L-cysteine	C8H15NO4S	220.0645	923-43-3	Propylene oxide
MA34	N-Acetyl-S-benzyl-L-cysteine	C12H15NO3S	252.0696	19542-77-9	Toluene
GAs					
GA108	Acetaminophen glucuronide	C14H17NO8	326.0879	16110-10-4	Acetaminophen
GA113	Benzoyl Glucuronide	C13H14O8	297.0611	19237-53-7	Benzoic acid
GA152	Ethyl β-D-glucuronide	C7H12O7	221.0664	17685-04-0	Ethanol
GA155	Phenyl-β-D-glucuronide	C12H14O7	269.0665	17685-05-1	Phenol
SAs					
SA86	1-[3-Methoxy-4-(sulfooxy)phenyl]-1,2-ethanediol_1.61	C9H12O7S	263.0208	3415-67-6	3-Methoxy-4-hydroxyphenylglycol
SA38	Acetaminophen sulfate_2.4	C8H9NO5S	230.0127	10066-90-7	Acetaminophen
SA2	Catechol sulfate	C6H6O5S	188.9861	4918-96-1	Catechol
SA106	p-Cresol sulfate	C7H8O4S	187.0069	3233-58-7	p-Cresol
SA16	Indoxyl sulfate_3.53	C8H7NO4S	212.0023	487-94-5	Indole
SA67	Isovanillic acid-3-sulfate_3.28	C8H8O7S	246.9909	78061-96-8	Isovanillic acid
SA37	Phenyl sulfate	C6H6O4S	172.9913	937-34-8	Phenol

Table 5

Association between demographic characteristics and metabolite principal components (n = 787).

	Total N = 787	PC1		PC2		PC3		PC4	
		Effect (95 % CI)	p-value	Effect (95 % CI)	p-value	Effect (95 % CI)	p-value	Effect (95 % CI)	p-value
Age-Mean (SD)	49(13)	0.12 (0.05, 0.19)	0.0009	0.13 (0.07, 0.19)	<0.0001	0.1 (0.03, 0.17)	0.007	−0.02 (−0.09, 0.05)	0.56
Sex									
Men	317 (40)	Ref		Ref		Ref		Ref	
Women	470 (60)	0.04 (−0.10, 0.17)	0.60	0.06 (−0.06, 0.17)	0.32	−0.09 (−0.23, 0.06)	0.23	0.11 (−0.04, 0.25)	0.15
Race									
NHW	644 (77)	Ref		Ref		Ref		Ref	
NHB	137 (16)	−0.38 (−0.58, −0.20)	<0.0001	−0.06 (−0.22, 0.10)	0.48	−0.04 (−0.24, 0.16)	0.68	−0.35 (−0.54, −0.15)	0.0007
Other	55 (7)	−0.18 (−0.45, 0.08)	0.17	−0.15 (−0.38, 0.07)	0.17	0.03 (−0.25, 0.31)	0.83	−0.07 (−0.35, 0.20)	0.61
Education									
High School, GED, or less	249 (30)	Ref		Ref		Ref		Ref	
Some college or 2-year degree	315 (38)	0.30 (0.13, 0.47)	0.005	−0.09 (−0.22, 0.05)	0.22	0.02 (−0.15, 0.20)	0.82	0.03 (−0.15, 0.20)	0.75
4-year College degree or more	267 (32)	0.58 (0.39, 0.77)	<0.0001	0.04 (−0.12, 0.20)	0.59	0.16 (−0.04, 0.36)	0.11	0.20 (0.00, 0.40)	0.046
Income									
<$50,000	414 (52)	Ref		Ref		Ref		Ref	
$50,000 or more	377 (48)	0.13 (−0.02, 0.28)	0.081	−0.08 (−0.21, 0.05)	0.21	0.13 (−0.03, 0.28)	0.11	−0.08 (−0.24, 0.08)	0.31
Tobacco Use									
No	502 (60)	Ref		Ref		Ref		Ref	
Yes	332 (40)	−0.01 (−0.16, 0.13)	0.87	1.24 (1.12, 1.36)	<0.0001	−0.20 (−0.35, −0.05)	0.01	−0.01 (−0.16, 0.15)	0.94

Abbreviations: NHW – non-Hispanic White; NHB – non-Hispanic Black.