A Section 508–conformant HTML version of this article is available at https://doi.org/10.1289/EHP1992.

Arsenic Exposure from Drinking Water and Urinary Metabolomics: Associations and Long-Term Reproducibility in Bangladesh Adults

Fen Wu,[1,2] Liang Chi,[3] Hongyu Ru,[3] Faruque Parvez,[4] Vesna Slavkovich,[4] Mahbub Eunus,[5] Alauddin Ahmed,[5] Tariqul Islam,[5] Muhammad Rakibuz-Zaman,[5] Rabiul Hasan,[5] Golam Sarwar,[5] Joseph H. Graziano,[4] Habibul Ahsan,[6] Kun Lu,[3] and Yu Chen[1,2]

[1]Department of Population Health, New York University School of Medicine, New York, New York, USA
[2]Department of Environmental Medicine, New York University School of Medicine, New York, New York, USA
[3]Department of Environmental Sciences and Engineering, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA
[4]Department of Environmental Health Sciences, Mailman School of Public Health, Columbia University, New York, New York, USA
[5]U-Chicago Research Bangladesh, Ltd., Dhaka, Bangladesh
[6]Department of Health Studies, Center for Cancer Epidemiology and Prevention, University of Chicago, Chicago, Illinois, USA

Background: Chronic exposure to inorganic arsenic from drinking water has been associated with a host of cancer and noncancer diseases. The application of metabolomics in epidemiologic studies may allow researchers to identify biomarkers associated with arsenic exposure and its health effects.

Objective: Our goal was to evaluate the long-term reproducibility of urinary metabolites and associations between reproducible metabolites and arsenic exposure.

Methods: We studied samples and data from 112 nonsmoking participants (58 men and 54 women) who were free of any major chronic diseases and who were enrolled in the Health Effects of Arsenic Longitudinal Study (HEALS), a large prospective cohort study in Bangladesh. Using a global gas chromatography–mass spectrometry platform, we measured metabolites in their urine samples, which were collected at baseline and again 2 y apart, and estimated intraclass correlation coefficients (ICCs). Linear regression was used to assess the association between arsenic exposure at baseline and metabolite levels in baseline urine samples.

Results: We identified 2,519 molecular features that were present in all 224 urine samples from the 112 participants, of which 301 had an ICC of $\geq$0.60. Of the 301 molecular features, water arsenic was significantly related to 31 molecular features and urinary arsenic was significantly related to 74 molecular features after adjusting for multiple comparisons. Six metabolites with a confirmed identity were identified from the 82 molecular features that were significantly associated with either water arsenic or urinary arsenic after adjustment for multiple comparisons.

Conclusions: Our study identified urinary metabolites with long-term reproducibility that were associated with arsenic exposure. The data established the feasibility of using metabolomics in future larger studies. https://doi.org/10.1289/EHP1992

Introduction

Inorganic arsenic (iAs) occurs naturally in groundwater in many parts of the world, affecting millions of people worldwide. Chronic exposure to iAs from drinking water has been associated with a host of human diseases, including cancer and cardiovascular disease (CVD) (Chen et al. 2009). Metabolomics, or metabolite profiling, refers to the systematic analysis of low molecular weight metabolites (the entire set of metabolites constitute the metabolome) in a biological sample that are intermediates or endpoints of metabolism (Wang et al. 2011). Downstream of genomic, transcriptomic, and proteomic perturbations, metabolites represent the most proximal reporters of alterations in the body in response to external stimuli (Lindon et al. 2003). Metabolomics has the potential to help identify the causes of environmentally mediated disease. Emerging evidence indicates metabolic perturbations associated with exposure to environmental contaminants, including welding fumes (Wang et al. 2012), cadmium (Ellis et al. 2012; Gao et al. 2014; Xu et al. 2016), tobacco smoking (Ellis et al. 2012; Hsu et al. 2013), phthalate (Xu et al. 2016), pesticides (Bonvallot et al. 2013), and arsenic (Li et al. 2017; Martin et al. 2015; Zhang et al. 2014). Epidemiologic studies of arsenic exposure from drinking water and metabolomics are limited. A recent study of 246 pregnant Chinese women identified 9 urinary metabolites that could be used to classify the women into different arsenic exposure levels (Li et al. 2017). Another study in diabetes cases and controls from México found associations between arsenic exposure from drinking water and relative amounts of 61 metabolites in urine (Martin et al. 2015). However, additional population-based studies are needed.

In epidemiologic studies, the usual level of a biomarker is of key interest and most likely to be associated with disease risk or exposure. However, a single measurement in time may not be representative of the usual level, thus reducing the power for epidemiologic studies to detect associations with disease (Rosner et al. 1992). Therefore, it is critical to evaluate the long-term reproducibility of new biomarkers before including them in large epidemiologic studies. Temporal reproducibility refers to the consistency of measurements of more than one sample from the same person at different times (Willett and Lenart 1998) and is expressed by the intraclass correlation coefficient (ICC) as the ratio of between-subject variation to total variation (sum of within- and between-subject variation). The closer the ICC is to 1, indicating little within-subject variation relative to the between-subject variation, the better a single measurement of a biomarker is at differentiating the relative ordering of the level among the individuals (Willett and Lenart 1998). Metabolites in serum and plasma were reported to have, on average, moderate reproducibility (ICC median value 0.4–0.5 covering several months to a year (Floegel et al. 2011; Sampson et al. 2013; Townsend et al. 2013). Urine is easy to collect with a large volume and it is largely free

Address correspondence to Y. Chen, Departments of Population Health and Environmental Medicine, New York University School of Medicine, 650 First Ave., Room 510, New York, NY 10016 USA. Telephone: (212) 263-4839. Email: yu.chen@nyumc.org

Supplemental Material is available online (https://doi.org/10.1289/EHP1992).

The authors declare they have no actual or potential competing financial interests.

Received 5 April 2017; Revised 24 November 2017; Accepted 27 November 2017; Published 12 January 2018.

Note to readers with disabilities: *EHP* strives to ensure that all journal content is accessible to all readers. However, some figures and Supplemental Material published in *EHP* articles may not conform to 508 standards due to the complexity of the information being presented. If you need assistance accessing journal content, please contact ehponline@niehs.nih.gov. Our staff will work with you to assess and meet your accessibility needs within 3 working days.

from interfering proteins or lipids, presenting opportunities for biomarker discovery in epidemiologic studies. However, the long-term reproducibility of urinary metabolites has yet to be evaluated.

We have established the Health Effects of Arsenic Longitudinal Study (HEALS), a well characterized cohort in Bangladesh with >20,000 participants recruited since year 2000. With repeated measures of urinary arsenic at baseline and every follow-up for more than 95% of the participants, we have the unique opportunity to evaluate the long-term reproducibility of urinary metabolites. We conducted a study of urinary metabolite profiling in 124 HEALS participants with iAs exposure at low-to-moderate levels (0.1 to <300 μg/L). We measured the metabolites in their urine samples collected 2 y apart using a global gas chromatography–mass spectrometry (GC-MS) platform and evaluated long-term reproducibility. We also examined the associations of both water arsenic and urinary arsenic at baseline with metabolites with sufficient reproducibility. In a subset of 84 participants for whom urinary arsenic metabolites had been measured, we also assessed the relationship between monomethylarsonic acid (MMA) percentage, an indicator of arsenic methylation capacity, and these reproducible metabolites.

Methods

Subject Selection

The parent study, the HEALS, is an ongoing prospective cohort study designed to investigate the health effects of arsenic exposure from drinking water in Araihazar, Bangladesh. Details of the HEALS have been described previously (Ahsan et al. 2006). Briefly, between October 2000 and May 2002, we recruited 11,746 married adults (original cohort) 18–75 y of age who were primarily drinking water from a local tube well, from a well-defined 25-km^2 geographical area. During 2006–2008, the cohort was expanded to include an additional 8,287 participants (expansion cohort) following the same methodologies. The overall participation rate was 97%. At baseline, water samples from all 10,971 tube wells in the study area were collected, and trained clinicians collected demographic and lifestyle data using a standardized questionnaire and collected spot urine samples from participants using structured protocols. The cohort is being actively followed up biennially with similar in-person visits (Ahsan et al. 2006). Spot urine samples were collected at follow-up visits. Informed consent was obtained from the study participants, and the study procedures were approved by the ethical committee of the Bangladesh Medical Research Council and the institutional review boards of Columbia University and the University of Chicago.

All participants in the present study used the tube wells as their exclusive source of drinking water for a long period of time before baseline (on average 7.8 y prior to baseline), and they did not switch wells during the follow-up. Also, analyses of time-series samples collected from 20 tube wells monitored for 3 y in the study area showed that the arsenic concentration in well water was relatively stable over time (Cheng et al. 2005). Therefore, baseline water arsenic is an indicator for long-term exposure in our study population. We aimed to include a homogeneous subpopulation with a wide range of iAs exposure in this study; therefore, we excluded smokers and those with any major chronic diseases such as cancer, CVD, and diabetes from overall HEALS participants. Then we randomly selected a total of 124 participants, consisting of 62 male and 62 female nonsmokers 25–45 y of age. We also frequency matched them by sex, age (±5 y), water arsenic levels (±10 μg/L), and cohort memberships (original vs. expansion cohort).

Arsenic Measurement

Details of the methods have been described (Chen et al. 2013). Briefly, total water arsenic concentration was analyzed by high-resolution inductively coupled plasma mass spectrometry with a detection limit of <0.2 μg/L. Total urinary arsenic concentration was measured by graphite furnace atomic absorption, using a Perkin-Elmer Analyst 600 graphite furnace system (Waltham, MA, USA) with a detection limit of 2 μg/L (Nixon et al. 1991). Urinary creatinine was analyzed using a method based on the Jaffe reaction (Slot 1965). In a subset of 84 HEALS participants, urinary arsenic metabolites were measured by high-performance liquid chromatography (HPLC) separation of arsenobetaine (AsB), arsenocholine (AsC), As^V, As^{III}, MMA, and dimethylarsinic acid (DMA), followed by detection by inductively coupled plasma mass spectrometry (Reuter et al. 2003). The percentage of MMA was calculated by dividing MMA by the sum of the metabolites as total arsenic after subtracting AsB and AsC (i.e., nontoxic organic As from dietary sources). Based on our data, urinary MMA% does not change much over time, with an ICC of 0.85 (Ahsan et al. 2007).

Metabolite Measurement Using GC-MS

Urinary metabolites were measured as described previously (Gao et al. 2017). Cold methanol (80 μL) was added to 20 μL urine. After vortexing at maximum speed for 1 min, the samples were incubated at 4°C for 20 min and then centrifuged for 10 min at 12,000 rpm. The supernatant was collected and dried in a SpeedVac (Savant SC110A; Thermo Electron), followed by derivatization using methoxyamine-HCL and BSTFA. The derivatized samples were analyzed using an Agilent Technologies 6890N Network GC System/5,973 Mass-Selective Detector (Agilent Technologies) with an Agilent J&W GC column [30 m length; 0.250 mm diameter (narrow bore); film thickness 0.25 μm] (Agilent Technologies) under the following conditions: initial oven temperature was set at 60°C for 2 min, ramped to 320°C by 8°C/min, and then held at 320°C for 10.5 min. Two microliters of sample solution was injected with helium as the carrier gas at a flow rate of 0.8 mL/min. The temperature of the injector, ion source, and MS Quadrupole were set at 275°C, 230°C, and 150°C, respectively. The mass spectrometer was operated in full scan mode from 50 to 600 m/z. The resultant data were processed with XCMS (https://xcmsonline.scripps.edu) for peak picking, alignment, and extraction of peak intensities. We used molecular features to refer fragment ions obtained by mass spectrometry (Alonso et al. 2015; Lu et al. 2014; Smith et al. 2006) that included both the ions that were assigned to specific metabolites and those with unknown identities. Normalization was performed by dividing the peak area of each molecular feature by the sum of peak areas of all molecular features. The molecular features with an ICC of ≥0.60 were selected for metabolite identification by comparing both the MS spectra and retention time with those in the National Institute of Standards and Technology (NIST) Standard Reference Database.

Statistical Analyses

We used PROC VARCOMP and PROC GLM in SAS (version 9.3; SAS Institute Inc.) to estimate the ICCs and their 95% confidence intervals, respectively, for the normalized peak intensity of each molecular feature detected in the two yearly urine samples. The molecular features with an ICC of ≥0.60 were selected for further statistical analyses. We used linear regression models to

Table 1. Distribution of selected variables by sex.

Variables	Men ($n=58$)	Women ($n=54$)	p-Value[a]
Age (y)	34.4 ± 5.7	34.9 ± 5.5	0.63
Body mass index (kg/m^2)	19.9 ± 2.6	21.1 ± 4.0	0.07
Education (y)	4.8 ± 4.0	2.7 ± 3.4	0.003
Systolic blood pressure (mmHg)	117.6 ± 13.1	115.9 ± 15.7	0.52
Diastolic blood pressure (mmHg)	74.6 ± 10.0	77.3 ± 11.1	0.18
Baseline total water arsenic (μg/L)	47.2 ± 51.7	44.8 ± 51.4	0.81
Baseline total urinary arsenic (μg/g creatinine)	194.3 ± 170.1	206.7 ± 178.2	0.71
Follow-up total urinary arsenic (μg/g creatinine)	171.3 ± 131.2	199.6 ± 162.0	0.31
Cohort [n (%)]			
Original	25 (43.1)	25 (46.3)	0.74
Expansion	33 (56.9)	29 (53.7)	

[a]p-Values were computed with the chi-square test or analysis of variance.

estimate the associations of continuous measure of water arsenic, urinary arsenic, and urinary MMA% with each molecular feature adjusting for sex, age, and cohort memberships. Assumptions of linear regression such as normal distribution of residuals, homocedasticity, and colinearity were checked and none was violated. The results with additional adjustment for body mass index (BMI) were similar and are therefore not shown. The threshold for the significance of the association was adjusted for multiple testing by controlling the false-discovery rate (FDR) (Benjamini and Hochberg 1995). A Venn diagram was used to illustrate the overlap of the metabolites that had a significant association with water arsenic, urinary arsenic, or urinary MMA%. For the reproducible metabolites that were nominally significantly associated with water arsenic or urinary arsenic, we included a heatmap to present multivariable Pearson correlations between these metabolites and arsenic measures with adjustment for sex, age, and cohort memberships, using the heatmap.2 function from the gplots package in R (version 3.4.0; R Core Team). In addition, we computed least squares means of urinary levels of the metabolite L-threonine that was significantly associated with both water arsenic and urinary arsenic after adjustment for multiple comparisons by quartiles of baseline water arsenic and urinary arsenic levels adjusting for sex, age, and cohort memberships.

Results

Characteristics of the Selected Subjects

A total of 12 participants were excluded from the analysis because metabolites were undetectable in both or one of their two yearly urine samples and the ICC could not be calculated. The present study consisted of 58 (51.8%) men and 54 (48.2%) women who were thin with a mean BMI of 20.5 kg/m^2, low-educated, and exposed to a mean level of <50 μg/L water arsenic at baseline (Table 1). Men and women did not differ appreciably regarding the matching factors of age, cohort memberships, and baseline total water arsenic as well as other variables such as BMI, systolic blood pressure, diastolic blood pressure, baseline total urinary arsenic and total urinary arsenic at the first follow-up 2 y later; however, men had significantly more years of formal education than women ($p<0.01$). Participants in the original cohort had significantly lower systolic and diastolic blood pressure compared with participants in the expansion cohort ($p<0.01$), but they did not differ by other variables such as age, BMI, baseline total water arsenic, baseline total urinary arsenic, and total urinary arsenic at the first follow-up (Table 2).

ICCs of the Molecular Features and Their Associations with Water Arsenic, Urinary Arsenic, and Urinary MMA%

We identified 2,519 molecular features that were present in all 224 urine samples from the 112 participants. The ICCs of these molecular features are shown in Excel Table S1. Of these molecular features, 688 (27%) had an ICC of ≥0.50 and 301 (12%) of ≥0.60. We then explored the associations of the 301 molecular features with an ICC of ≥0.60 with water arsenic, urinary arsenic, and urinary MMA% at baseline. Water arsenic was nominally significantly related to 89 features and 31 (34.8%) had an FDR $p<0.05$; 126 features were nominally significantly associated with urinary arsenic and 74 (58.7%) had an FDR $p<0.05$ (Figure 1A). A total of 142 features were significantly associated with either water arsenic or urinary arsenic at the nominal level; most of these features were correlated with water arsenic and urinary arsenic similarly (Figure 1B, see also Excel Table S2); the ICCs of these molecular features were similar by high and low levels of exposure (see Excel Table S3). Analyses based on log-transformation of the metabolites generated similar results (see Excel Table S4). Of the 142 features, 82 remained significant after adjustment for multiple comparisons (Figure 1A). The ICCs of these 82 features ranged from 0.60 to 0.83, with 26 features (31.7%) having an ICC of >0.70. Of the 82 features, 23 were related to both water arsenic and urinary arsenic in a consistent direction (Figure 1A). In addition, a total of 33 molecular features were nominally significantly associated with urinary MMA%, though none of the associations remained significant after adjustment for multiple comparisons (Figure 1A, see also Excel Table S5), probably because of the small sample size ($n=84$). Most of these features ($n=29$, 87.9%) were also significantly associated with either water arsenic ($n=21$, 63.6%) or urinary arsenic ($n=27$, 81.8%), and 19 (57.6%) were significantly related to both water arsenic and urinary arsenic at the nominal level (Figure 1A, see also Excel Table S5).

Identities of the Reproducible Molecular Features

We also identified the metabolite identities of the 142 molecular features that had an ICC of ≥0.60 and were significantly associated with either water arsenic or urinary arsenic at the nominal level by searching the NIST Standard Reference Database. A

Table 2. Distribution of selected variables by cohort.

Variables	Original cohort ($n=50$)	Expansion cohort ($n=62$)	p-Value[a]
Age (y)	34.3 ± 5.8	35.0 ± 5.4	0.50
Body mass index (kg/m^2)	19.8 ± 2.5	21.0 ± 3.8	0.07
Education (y)	3.8 ± 4.0	3.9 ± 3.7	0.92
Systolic blood pressure (mmHg)	112.1 ± 14.5	120.5 ± 13.3	0.002
Diastolic blood pressure (mmHg)	73.6 ± 9.4	77.8 ± 11.2	0.04
Baseline total water arsenic (μg/L)	55.6 ± 59.4	38.4 ± 42.7	0.08
Baseline total urinary arsenic (μg/g creatinine)	222.9 ± 191.5	182.0 ± 156.4	0.22
Follow-up total urinary arsenic (μg/g creatinine)	197.6 ± 162.4	174.8 ± 133.4	0.42

[a]p-Values were computed with the chi-square test or analysis of variance.

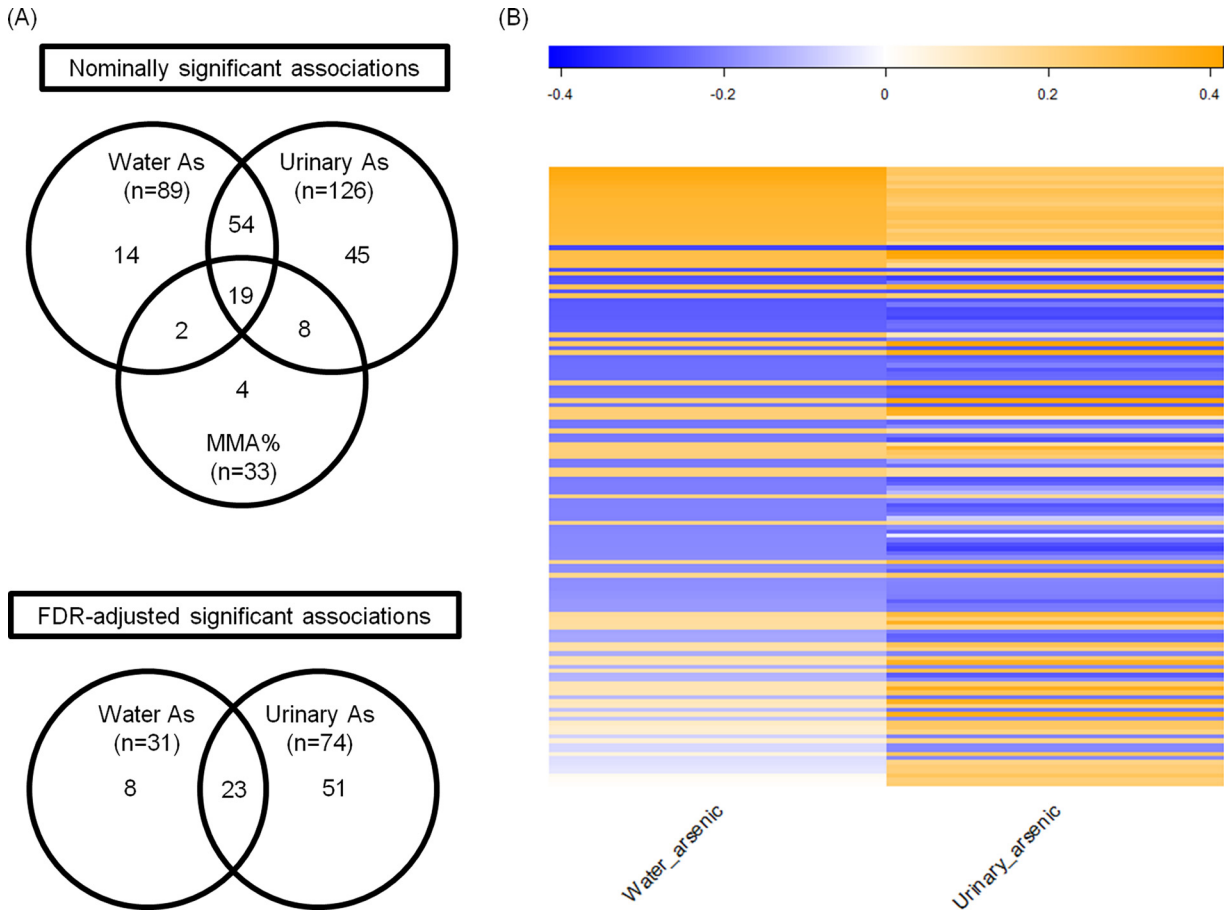

Figure 1. Associations of reproducible molecular features with baseline total water arsenic and baseline total urinary arsenic. (*A*) A Venn diagram shows the overlap of the metabolites that had a significant association with water arsenic, urinary arsenic, and urinary MMA%. (*B*) Heatmap of multivariable Pearson correlations of baseline total water arsenic and urinary arsenic with the reproducible metabolites that were nominally significantly associated with water arsenic or urinary arsenic. The coefficients were adjusted for sex, age, and cohort memberships. Note: MMA%, percent monomethylarsonic acid.

total of 16 metabolites had a confirmed identity (Table 3), namely, aminoethanol, β-amino isobutyric acid, citric acid, 1,2-dithiane-4,5-diol, ethanedioic acid, glycine, 3-hydroxyisovaleric acid, indole-3-acetic acid, L-threonine, phosphoric acid, pyroglutamic acid, (R*,S*)-3,4-dihydroxybutanoic acid, serine, succinic acid, uracil, and uric acid. Furthermore, 6 of the 16 metabolites (1,2-dithiane-4,5-diol, L-threonine, phosphoric acid, pyroglutamic acid, (R*,S*)-3,4-dihydroxybutanoic acid, and succinic acid) were significantly associated with either water arsenic or urinary arsenic after adjustment for multiple comparisons. The metabolite

Table 3. Nominally significant associations of reproducible molecular features with baseline total water arsenic and baseline total urinary arsenic.

		Water arsenic			Urinary arsenic		
Metabolite	ICC (95% CI)	β[a] (95% CI)	Raw *p*-value	FDR *p*-value	β[a] (95% CI)	Raw *p*-value	FDR *p*-value
Aminoethanol	0.60 (0.46, 0.70)	0.48 (0.06, 0.89)	0.025	0.11	0.50 (0.09, 0.91)	0.018	0.07
β-Amino isobutyric acid	0.68 (0.57, 0.77)	−0.02 (−0.05, 0.01)	0.211	0.39	−0.03 (−0.06, −0.00)	0.043	0.11
Citric acid	0.62 (0.49, 0.72)	−0.03 (−0.05, 0.00)	0.051	0.17	−0.03 (−0.05, −0.00)	0.048	0.12
1,2-Dithiane-4,5-diol	0.70 (0.59, 0.78)	0.35 (0.09, 0.62)	0.010	0.07	0.58 (0.34, 0.83)	<0.001	0.001
Ethanedioic acid	0.67 (0.56, 0.76)	−0.00 (−0.01, 0.00)	0.200	0.38	−0.00 (−0.01, −0.00)	0.049	0.12
3-Hydroxyisovaleric acid	0.63 (0.51, 0.73)	−0.02 (−0.04, −0.00)	0.022	0.11	−0.02 (−0.04, 0.00)	0.052	0.12
Glycine	0.60 (0.47, 0.71)	−0.09 (−0.19, 0.01)	0.072	0.20	−0.11 (−0.20, −0.01)	0.031	0.09
Indole-3-acetic acid	0.61 (0.48, 0.71)	−0.04 (−0.09, 0.00)	0.060	0.17	−0.05 (−0.09, −0.00)	0.035	0.10
L-Threonine	0.60 (0.46, 0.70)	−0.13 (−0.22, −0.04)	0.006	0.05	−0.15 (−0.24, −0.06)	0.001	0.02
Phosphoric acid	0.67 (0.55, 0.76)	1.05 (0.13, 1.96)	0.025	0.11	1.25 (0.35, 2.15)	0.007	0.04
Pyroglutamic acid	0.64 (0.51, 0.73)	−0.47 (−0.93, −0.01)	0.047	0.16	−0.68 (−1.13, −0.23)	0.003	0.03
(R*,S*)-3,4-Dihydroxybutanoic acid	0.61 (0.48, 0.72)	−0.09 (−0.16, −0.02)	0.016	0.10	−0.09 (−0.17, −0.02)	0.010	0.04
Serine	0.63 (0.51, 0.73)	−0.18 (−0.37, 0.02)	0.070	0.20	−0.21 (−0.40, −0.01)	0.035	0.10
Succinic acid	0.63 (0.50, 0.73)	−0.02 (−0.04, −0.01)	0.003	0.04	−0.02 (−0.04, −0.01)	<0.001	0.02
Uracil	0.63 (0.50, 0.73)	−0.05 (−0.11, 0.01)	0.127	0.28	−0.07 (−0.13, −0.01)	0.031	0.09
Uric acid	0.64 (0.51, 0.73)	2.56 (0.39, 4.74)	0.022	0.11	1.27 (−0.93, 3.47)	0.255	0.35

[a]Coefficient from linear regression model indicates difference in peak intensity of urinary metabolites in relation to per 1-SD increase in water arsenic (51.3 μg/L) and per 1-SD increase in urinary arsenic (173.4 μg/g creatinine), adjusting for sex, age, and cohort memberships.

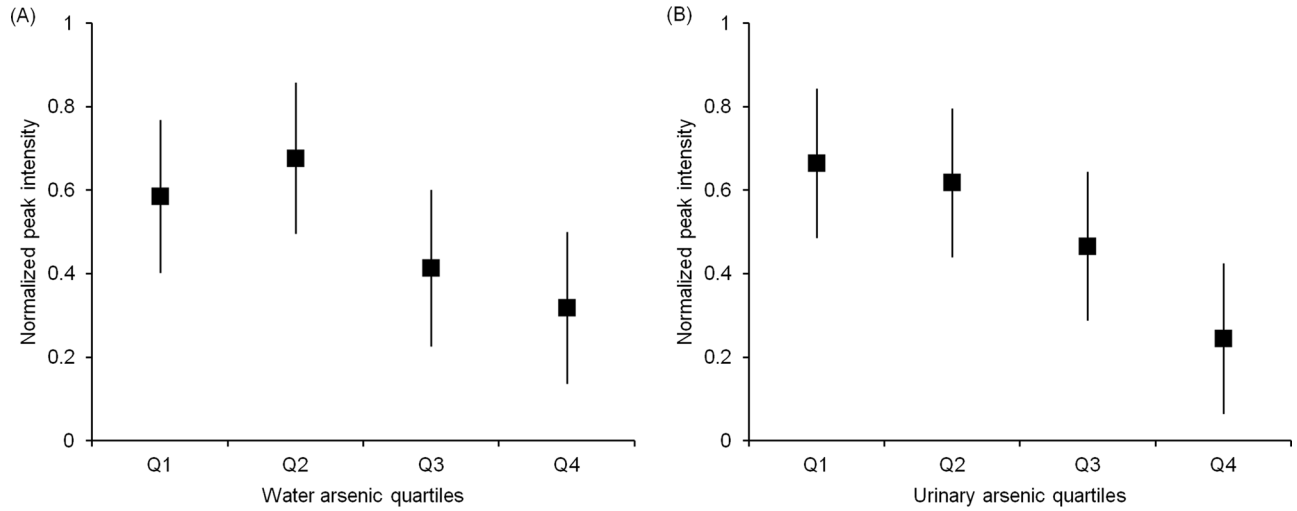

Figure 2. Adjusted means of normalized peak intensity of L-threonine by quartiles of baseline (*A*) total water arsenic levels and (*B*) total urinary arsenic levels. Means were adjusted for sex, age, and cohort memberships.

L-threonine was significantly associated with both water arsenic and urinary arsenic after adjustment for multiple comparisons (Table 3). The adjusted means of urinary levels of L-threonine according to quartiles of baseline water arsenic and urinary arsenic levels are shown in Figure 2. Overall, L-threonine levels were inversely related to water arsenic and urinary arsenic in a similar pattern and were significantly lower in the highest quartile compared with the lowest quartile ($p < 0.05$ for water arsenic and $p < 0.01$ for urinary arsenic).

Discussion

To investigate the suitability of urinary metabolite profiling for use in searching for biomarkers of arsenic-related health effects, we conducted a study to evaluate the long-term reproducibility of the metabolites using repeated urine samples collected 2 y apart. Our data showed that there are substantial known and unknown metabolites with sufficient reproducibility and strong associations with iAs exposure, presenting future opportunities of biomarker discovery for epidemiologic studies.

We found 301 molecular features (27% of the detected features) with excellent reproducibility over time (ICCs $\geq$0.60). These ICCs compare favorably with the reproducibility over a several-year period of serum cholesterol (ICC = 0.65) (Shekelle et al. 1981), blood pressure (ICC = 0.60–0.64) (Rosner et al. 1977), blood glucose (ICC = 0.52) (Gordon and Shurtleff 1973), pulse (ICC = 0.49) (Gordon and Shurtleff 1973), and plasma estradiol in postmenopausal women (ICC = 0.68) (Hankinson et al. 1995), all of which are exposures considered to be reasonably well-measured and reliable predictors of disease in epidemiologic studies. More than 10% of the 301 molecular features were also associated with urinary MMA%, a biomarker specific for susceptibility to iAs exposure that has been related to cancers (Chen et al. 2003; Huang et al. 2008; Steinmaus et al. 2010; Yu et al. 2000). Furthermore, 82 of the 301 molecular features were significantly associated with either water arsenic or urinary arsenic after controlling for the influence of sex, age, cohort memberships and multiple comparisons. Although a limited number of metabolites were identified from these reproducible molecular features, possibly because of the intrinsic limitation of GC-MS, our results suggest that within the context of a prospective epidemiologic study, a single urine measurement of certain molecular features may adequately represent their longer-term usual (i.e., at least 2 y) levels, which may serve as intermediate biomarkers linking arsenic exposure and chronic diseases.

Consistent with our finding of reproducible molecular features, a previous study investigated the source of variability of 539 metabolites measured by LC-MS and GC-MS in urine samples and found that a large proportion (81%) of the metabolites had an ICC exceeding 0.50. However, the evaluation of reproducibility was based on 17 male subjects over 2 to 10 d. In our study, reproducibility of molecular features were estimated in urine samples collected 2 y apart and we found 27% molecular features having an ICC over 0.50. The difference in time intervals between sample collections may partially explain the much higher portion of reproducible metabolites in the previous study as compared with that of our study. For many metabolites, the correlation between samples would be expected to be higher when time intervals between sample collections are short. As a result, measures over a few days may not capture the true temporal variability around the "usual" long-term level of these metabolites. Therefore, our study adds to the evidence that certain urinary metabolites may be more relatively stable than others over a longer period of time and have the potential to serve as long-term biomarkers associated with exposures and/or diseases.

We identified several urinary metabolites that were dose-dependently associated with either water arsenic or urinary arsenic, some of which may be of biological significance. For instance, three amino acids (glycine, L-threonine, and serine) that are involved in one-carbon metabolism—the central pathway that facilitates arsenic methylation and elimination—were inversely related to water or urinary arsenic levels. Glycine and serine participate in the metabolism of methionine as a methyl-group acceptor and as a substrate for cystathionine synthesis, respectively (Benevenga and Harper 1970; Stead et al. 2000). Both animal and human studies (Benevenga and Harper 1970; Fukada et al. 2006; Girard-Globa et al. 1972; Stead et al. 2000; Verhoef et al. 2004) have shown that serine can lower homocysteine—a risk factor for CVD. Glycine is synthesized endogenously from serine, threonine, choline, or glyoxylate in the liver and kidney (Wang et al. 2013). Glycine exerts anti-inflammatory and antioxidative effects (McCarty and DiNicolantonio 2014; Senthilkumar et al. 2004) and has been shown to reduce plasma insulin, fat mass, and blood pressure in rodents (Alvarado-Vásquez et al. 2003; El Hafidi et al. 2004). Lower glycine concentrations have been associated with several traditional cardiovascular risk

factors, including obesity (Oberbach et al. 2011; Tastesen et al. 2014; Zhao et al. 2016), hypertension (El Hafidi et al. 2006; El Hafidi et al. 2004), and diabetes mellitus (De Luca et al. 2001; Palmer et al. 2015; Wang-Sattler et al. 2012). Previous studies also demonstrated that greater dietary intake of threonine was associated with lower blood pressure in a cohort study of patients with CVD (Tuttle et al. 2012) and circulating serine levels were inversely associated with BMI in nonsmoking healthy women (Zhao et al. 2016) and youth with obesity and type 2 diabetes (Mihalik et al. 2012).

Two epidemiologic studies have investigated the impact of arsenic exposure from drinking water on metabolite profiles. A recent study of 246 pregnant Chinese women identified nine urinary metabolites that could be used to classify the women into a low (the first tertile of urinary total creatinine-adjusted arsenic) or high (the third tertile of urinary arsenic) exposure category using UPLC/Q-TOF MS (ultra performance liquid chromatography coupled to quadrupole with time-of-flight mass spectrometry) (Li et al. 2017). The identified metabolites were potentially related to endocrine disruption and oxidative stress. Another study of 86 Mexican individuals with exposure to low-to-moderate arsenic levels in drinking water (0.1 to 285 μg/L) reported 61 altered metabolites in urine associated with urinary total unadjusted arsenic using GC- and LC-TOF-MS; these metabolites were associated with amino acid metabolism, carbohydrate/energy metabolism, and vitamin (riboflavin) metabolism (Martin et al. 2015). However, there was no overlap of the identified metabolites between these studies and our study. It should be noted that we evaluated the exposure–metabolite associations only for the metabolites that were relatively stable over time (ICC $\geq$0.60). Taken together, these data point to metabolic disruption by arsenic exposure, though specific metabolites shared across all studies are difficult to identify because of differences in exposure background, population characteristics, and metabolomics platforms. Larger studies are needed to characterize the interplay between arsenic exposure, metabolite profile, and disease outcomes.

Strengths of this study include a sufficient number of subjects for a metabolomics study to assess reproducibility, a wide range of arsenic exposure levels, and the long interval between collections of repeated urine samples for evaluation of long-term reproducibility. This study is among the first to evaluate the long-term reproducibility of metabolomics data in human urine samples from a prospective cohort study. We also acknowledge several limitations. First, GC-MS data contain complexity such as that a single metabolite can produce multiple fragments. It is suggested that a simple strategy is to combine these fragments with the same retention time into a single metabolite. We did not use this strategy to identify metabolites from all the 2,519 molecular features because our aim was to evaluate the existence of reproducible metabolomics data using urine samples that are readily available from our large parent cohort study. Second, we identified only a small number of metabolites from the reproducible molecular features that prevented us from further pathway analysis. We therefore acknowledge that a complementary platform such as LC-MS should be employed in future studies for comprehensive understanding of metabolic alterations in response to arsenic exposure.

In summary, our study identified urinary metabolites with long-term reproducibility that were associated with arsenic exposure using a global GC-MS metabolomics platform. The data established the feasibility of using metabolomics platform in future larger studies to assess alterations in urinary metabolites in relation to arsenic exposure and their associations with the risk of CVD or cancer.

Acknowledgments

This study was supported by the National Institutes of Health (NIH)/National Institute of Environmental Health Sciences (NIEHS) grants P42ES010349, P30ES000260, P30ES009089, and R01ES024950.

References

Ahsan H, Chen Y, Kibriya MG, Slavkovich V, Parvez F, Jasmine F. 2007. Arsenic metabolism, genetic susceptibility, and risk of premalignant skin lesions in Bangladesh. Cancer Epidemiol Biomarkers Prev 16(6):1270–1278, PMID: 17548696, https://doi.org/10.1158/1055-9965.EPI-06-0676.

Ahsan H, Chen Y, Parvez F, Argos M, Hussain AI, Momotaj H, et al. 2006. Health Effects of Arsenic Longitudinal Study (HEALS): description of a multidisciplinary epidemiologic investigation. J Expo Sci Environ Epidemiol 16(2):191–205, PMID: 16160703, https://doi.org/10.1038/sj.jea.7500449.

Alonso A, Marsal S, Julià A. 2015. Analytical methods in untargeted metabolomics: state of the art in 2015. Front Bioeng Biotechnol 3:23, PMID: 25798438, https://doi.org/10.3389/fbioe.2015.00023.

Alvarado-Vásquez N, Zamudio P, Cerón E, Vanda B, Zenteno E, Carvajal-Sandoval G. 2003. Effect of glycine in streptozotocin-induced diabetic rats. Comp Biochem Physiol C Toxicol Pharmacol 134(4):521–527, PMID: 12727302, https://doi.org/10.1016/S1532-0456(03)00046-2.

Benevenga NJ, Harper AE. 1970. Effect of glycine and serine on methionine metabolism in rats fed diets high in methionine. J Nutr 100(10):1205–1214, PMID: 5471055.

Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Series B Methodol 57(1):289–300, https://doi.org/10.2307/2346101.

Bonvallot N, Tremblay-Franco M, Chevrier C, Canlet C, Warembourg C, Cravedi JP, et al. 2013. Metabolomics tools for describing complex pesticide exposure in pregnant women in Brittany (France). PLoS One 8(5):e64433, PMID: 23704985, https://doi.org/10.1371/journal.pone.0064433.

Chen Y, Parvez F, Gamble M, Islam T, Ahmed A, Argos M, et al. 2009. Arsenic exposure at low-to-moderate levels and skin lesions, arsenic metabolism, neurological functions, and biomarkers for respiratory and cardiovascular diseases: review of recent findings from the Health Effects of Arsenic Longitudinal Study (HEALS) in Bangladesh. Toxicol Appl Pharmacol 239(2):184–192, PMID: 19371619, https://doi.org/10.1016/j.taap.2009.01.010.

Chen YC, Su HJ, Guo YL, Hsueh YM, Smith TJ, Ryan LM, et al. 2003. Arsenic methylation and bladder cancer risk in Taiwan. Cancer Causes Control 14(4):303–310, PMID: 12846360.

Chen Y, Wu F, Graziano JH, Parvez F, Liu M, Paul RR, et al. 2013. Arsenic exposure from drinking water, arsenic methylation capacity, and carotid intima-media thickness in Bangladesh. Am J Epidemiol 178(3):372–381, PMID: 23788675, https://doi.org/10.1093/aje/kwt001.

Cheng Z, van Geen A, Seddique AA, Ahmed KM. 2005. Limited temporal variability of arsenic concentrations in 20 wells monitored for 3 years in Araihazar, Bangladesh. Environ Sci Technol 39(13):4759–4766, PMID: 16053073, https://doi.org/10.1021/es048065f.

De Luca G, Calpona PR, Caponetti A, Macaione V, Di Benedetto A, Cucinotta D, et al. 2001. Preliminary report: amino acid profile in platelets of diabetic patients. Metab Clin Exp 50(7):739–741, PMID: 11436175, https://doi.org/10.1053/meta.2001.24193.

El Hafidi M, Pérez I, Baños G. 2006. Is glycine effective against elevated blood pressure? Curr Opin Clin Nutr Metab Care 9(1):26–31, PMID: 16444815, https://doi.org/10.1097/01.mco.0000196143.72985.9a.

El Hafidi M, Pérez I, Zamora J, Soto V, Carvajal-Sandoval G, Baños G. 2004. Glycine intake decreases plasma free fatty acids, adipose cell size, and blood pressure in sucrose-fed rats. Am J Physiol Regul Integr Comp Physiol 287(6):R1387–R1393, PMID: 15331379, https://doi.org/10.1152/ajpregu.00159.2004.

Ellis JK, Athersuch TJ, Thomas LDK, Teichert F, Pérez-Trujillo M, Svendsen C, et al. 2012. Metabolic profiling detects early effects of environmental and lifestyle exposure to cadmium in a human population. BMC Med 10(1):61, PMID: 22713677, https://doi.org/10.1186/1741-7015-10-61.

Floegel A, Drogan D, Wang-Sattler R, Prehn C, Illig T, Adamski J, et al. 2011. Reliability of serum metabolite concentrations over a 4-month period using a targeted metabolomic approach. PloS One 6(6):e21103, PMID: 21698256, https://doi.org/10.1371/journal.pone.0021103.

Fukada S, Shimada Y, Morita T, Sugiyama K. 2006. Suppression of methionine-induced hyperhomocysteinemia by glycine and serine in rats. Biosci Biotechnol Biochem 70(10):2403–2409, PMID: 17031061, https://doi.org/10.1271/bbb.60130.

Gao B, Bian X, Mahbub R, Lu K. 2017. Sex-specific effects of organophosphate diazinon on the gut microbiome and its metabolic functions. Environ Health Perspect 125(2):198–206, PMID: 27203275, https://doi.org/10.1289/EHP202.

Gao Y, Lu Y, Huang S, Gao L, Liang X, Wu Y, et al. 2014. Identifying early urinary metabolic changes with long-term environmental exposure to cadmium by mass-spectrometry-based metabolomics. Environ Sci Technol 48(11):6409–6418, PMID: 24834460, https://doi.org/10.1021/es500750w.

Girard-Globa A, Robin P, Forestier M. 1972. Long-term adaptation of weanling rats to high dietary levels of methionine and serine. J Nutr 102(2):209–217, PMID: 4400227.

Gordon T, Shurtleff D. 1973. The Framingham Study: An Epidemiologic Investigation of Cardiovascular Disease. Section 29: Means at Each Examination and Inter-Examination Variation of Specified Characteristics: Framingham Study Exam 1 to Exam 10: DHEW Pub No. (NIH) 74-478.

Hankinson SE, Manson JE, Spiegelman D, Willett WC, Longcope C, Speizer FE. 1995. Reproducibility of plasma hormone levels in postmenopausal women over a 2-3-year period. Cancer Epidemiol Biomarkers Prev 4(6):649–654, PMID: 8547832.

Hsu PC, Zhou B, Zhao Y, Ressom HW, Cheema AK, Pickworth W, et al. 2013. Feasibility of identifying the tobacco-related global metabolome in blood by UPLC-QTOF-MS. J Proteome Res 12(2):679–691, PMID: 23240883, https://doi.org/10.1021/pr3007705.

Huang YK, Huang YL, Hsueh YM, Yang MH, Wu MM, Chen SY, et al. 2008. Arsenic exposure, urinary arsenic speciation, and the incidence of urothelial carcinoma: a twelve-year follow-up study. Cancer Causes Control 19(8):829–839, PMID: 18351295, https://doi.org/10.1007/s10552-008-9146-5.

Li H, Wang M, Liang Q, Jin S, Sun X, Jiang Y, et al. 2017. Urinary metabolomics revealed arsenic exposure related to metabolic alterations in general Chinese pregnant women. J Chromatogr A 1479:145–152, PMID: 27988079, https://doi.org/10.1016/j.chroma.2016.12.007.

Lindon JC, Holmes E, Nicholson JK. 2003. So what's the deal with metabonomics? Anal Chem 75(17):384A–391A, PMID: 14632032.

Lu K, Abo RP, Schlieper KA, Graffam ME, Levine S, Wishnok JS, et al. 2014. Arsenic exposure perturbs the gut microbiome and its metabolic profile in mice: an integrated metagenomics and metabolomics analysis. EnvironHealth Perspect 122(3):284–291, PMID: 24413286, https://doi.org/10.1289/ehp.1307429.

Martin E, González-Horta C, Rager J, Bailey KA, Sánchez-Ramírez B, Ballinas-Casarrubias L, et al. 2015. Metabolomic characteristics of arsenic-associated diabetes in a prospective cohort in Chihuahua, Mexico. Toxicol Sci 144(2):338–346, PMID: 25577196, https://doi.org/10.1093/toxsci/kfu318.

McCarty MF, DiNicolantonio JJ. 2014. The cardiometabolic benefits of glycine: is glycine an 'antidote' to dietary fructose? Open Heart 1(1):e000103, PMID: 25332814, https://doi.org/10.1136/openhrt-2014-000103.

Mihalik SJ, Michaliszyn SF, de las Heras J, Bacha F, Lee S, Chace DH, et al. 2012. Metabolomic profiling of fatty acid and amino acid metabolism in youth with obesity and type 2 diabetes: evidence for enhanced mitochondrial oxidation. Diabetes Care 35(3):605–611, PMID: 22266733, https://doi.org/10.2337/DC11-1577.

Nixon DE, Mussmann GV, Eckdahl SJ, Moyer TP. 1991. Total arsenic in urine: palladium-persulfate vs nickel as a matrix modifier for graphite furnace atomic absorption spectrophotometry. Clin Chem 37(9):1575–1579, PMID: 1893592.

Oberbach A, Blüher M, Wirth H, Till H, Kovacs P, Kullnick Y, et al. 2011. Combined proteomic and metabolomic profiling of serum reveals association of the complement system with obesity and identifies novel markers of body fat mass changes. J Proteome Res 10(10):4769–4788, PMID: 21823675, https://doi.org/10.1021/pr2005555.

Palmer ND, Stevens RD, Antinozzi PA, Anderson A, Bergman RN, Wagenknecht LE, et al. 2015. Metabolomic profile associated with insulin resistance and conversion to diabetes in the Insulin Resistance Atherosclerosis Study. The Journal of clinical endocrinology and metabolism 100(3):E463–E468, PMID: 25423564, https://doi.org/10.1210/jc.2014-2357.

Reuter W, Davidowski L, Neubauer K, Di Bussolo J. 2003. Speciation of five arsenic compounds in urine by HPLC/ICP-MS. PerkinElmer Application Note 2003. https://www.perkinelmer.com/PDFs/Downloads/app_speciationfivearseniccompounds.pdf [accessed 1 August 2003].

Rosner B, Hennekens CH, Kass EH, Miall WE. 1977. Age-specific correlation analysis of longitudinal blood pressure data. Am J Epidemiol 106(4):306–313, PMID: 910798, https://doi.org/10.1093/oxfordjournals.aje.a112466.

Rosner B, Spiegelman D, Willett WC. 1992. Correction of logistic regression relative risk estimates and confidence intervals for random within-person measurement error. Am J Epidemiol 136(11):1400–1413, PMID: 1488967, https://doi.org/10.1093/oxfordjournals.aje.a116453.

Sampson JN, Boca SM, Shu XO, Stolzenberg-Solomon RZ, Matthews CE, Hsing AW, et al. 2013. Metabolomics in epidemiology: sources of variability in metabolite measurements and implications. Cancer Epidemiol Biomarkers Prev 22(4):631–640, PMID: 23396963, https://doi.org/10.1158/1055-9965.EPI-12-1109.

Senthilkumar R, Sengottuvelan M, Nalini N. 2004. Protective effect of glycine supplementation on the levels of lipid peroxidation and antioxidant enzymes in the erythrocyte of rats with alcohol-induced liver injury. Cell Biochem Funct 22(2):123–128, PMID: 15027101, https://doi.org/10.1002/cbf.1062.

Shekelle RB, Shryock AM, Paul O, Lepper M, Stamler J, Liu S, et al. 1981. Diet, serum cholesterol, and death from coronary heart disease. The Western Electric study. N Engl J Med 30(4):65–70, PMID: 7442730, https://doi.org/10.1056/NEJM198101083040201.

Slot C. 1965. Plasma creatinine determination. A new and specific Jaffe reaction method. Scand J Clin Lab Invest 17(4):381–387, PMID: 5838275, https://doi.org/10.3109/00365516509077065.

Smith CA, Want EJ, O'Maille G, Abagyan R, Siuzdak G. 2006. XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. Anal Chem 78(3):779–787, PMID: 16448051, https://doi.org/10.1021/ac051437y.

Stead LM, Brosnan ME, Brosnan JT. 2000. Characterization of homocysteine metabolism in the rat liver. Biochem J 350(pt 3):685–692, PMID: 10970780, https://doi.org/10.1042/0264-6021:3500685.

Steinmaus C, Yuan Y, Kalman D, Rey OA, Skibola CF, Dauphine D, et al. 2010. Individual differences in arsenic metabolism and lung cancer in a case-control study in Cordoba, Argentina. Toxicol Appl Pharmacol 247(2):138–145, PMID: 20600216, https://doi.org/10.1016/j.taap.2010.06.006.

Tastesen HS, Keenan AH, Madsen L, Kristiansen K, Liaset B. 2014. Scallop protein with endogenous high taurine and glycine content prevents high-fat, high-sucrose-induced obesity and improves plasma lipid profile in male C57BL/6J mice. Amino acids 46(7):1659–1671, PMID: 24658997, https://doi.org/10.1007/s00726-014-1715-1.

Townsend MK, Clish CB, Kraft P, Wu C, Souza AL, Deik AA, et al. 2013. Reproducibility of metabolomic profiles among men and women in 2 large cohort studies. Clin Chem 59(11):1657–1667, PMID: 23897902, https://doi.org/10.1373/clinchem.2012.199133.

Tuttle KR, Milton JE, Packard DP, Shuler LA, Short RA. 2012. Dietary amino acids and blood pressure: a cohort study of patients with cardiovascular disease. Am J Kidney Dis 59(6):803–809, PMID: 22381643, https://doi.org/10.1053/j.ajkd.2011.12.026.

Verhoef P, Steenge GR, Boelsma E, van Vliet T, Olthof MR, Katan MB. 2004. Dietary serine and cystine attenuate the homocysteine-raising effect of dietary methionine: a randomized crossover trial in humans. Am J Clin Nutr 80(3):674–679, PMID: 15321808.

Wang KC, Kuo CH, Tian TF, Tsai MH, Chiung YM, Hsiech CM, et al. 2012. Metabolomic characterization of laborers exposed to welding fumes. Chem Res Toxicol 25(3):676–686, PMID: 22292500, https://doi.org/10.1021/tx200465e.

Wang TJ, Larson MG, Vasan RS, Cheng S, Rhee EP, McCabe E, et al. 2011. Metabolite profiles and the risk of developing diabetes. Nat Med 17(4):448–453, PMID: 21423183, https://doi.org/10.1038/nm.2307.

Wang W, Wu Z, Dai Z, Yang Y, Wang J, Wu G. 2013. Glycine metabolism in animals and humans: implications for nutrition and health. Amino Acids 45(3):463–477, PMID: 23615880, https://doi.org/10.1007/s00726-013-1493-1.

Wang-Sattler R, Yu Z, Herder C, Messias AC, Floegel A, He Y, et al. 2012. Novel biomarkers for pre-diabetes identified by metabolomics. Mol Syst Biol 8:615, PMID: 23010998, https://doi.org/10.1038/msb.2012.43.

Willett W, Lenart E. 1998. Reproducibility and validity of food-frequency questionnaires. In: *Nutritional Epidemiology*. Willett W, ed. 3rd Edition. New York:Oxford University Press, 101–147.

Xu Y, Wang J, Liang X, Gao Y, Chen W, Huang Q, et al. 2016. Urine metabolomics of women from small villages exposed to high environmental cadmium levels. Environ Toxicol Chem 35(5):1268–1275, PMID: 26450519, https://doi.org/10.1002/etc.3274.

Yu RC, Hsu KH, Chen CJ, Froines JR. 2000. Arsenic methylation capacity and skin cancer. Cancer Epidemiol Biomarkers Prev 9(11):1259–1262, PMID: 11097236.

Zhang J, Shen H, Xu W, Xia Y, Barr DB, Mu X, et al. 2014. Urinary metabolomics revealed arsenic internal dose-related metabolic alterations: a proof-of-concept study in a Chinese male cohort. Environ Sci Technol 48(20):12265–12274, PMID: 25233106, https://doi.org/10.1021/es503659w.

Zhao H, Shen J, Djukovic D, Daniel-MacDougall C, Gu H, Wu X, et al. 2016. Metabolomics-identified metabolites associated with body mass index and prospective weight gain among Mexican American women. Obes Sci Pract 2(3):309–317, PMID: 27708848, https://doi.org/10.1002/osp4.63.