BMJ Open

Low level of serum aromatic amino acid associated with elevated risk of arsenic-induced skin lesions: data from a metabolomics study.

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2018-025336
Article Type:	Research
Date Submitted by the Author:	13-Jul-2018
Complete List of Authors:	wei, yaping; Wenzhou Medical University, Chashan University town, Ouhai district, Wenzhou, China., Jia, Chaonan; School of Public Health & Management, Wenzhou Medical University, Department of Preventive Medicine Lan, Yuan; School of Ophthalmology & Optometry, Wenzhou Medical University, Center on Clinical Research, the Eye Hospital of Wenzhou Medical University Hou, Xiangqing; School of Public Health & Management, Wenzhou Medical Univeristy, Department of Preventive Medicine Zuo, Jingjing; Wenzhou Medical University School of Optometry and Ophthalmology, Center on Clinical Research, the Eye Hospital of Wenzhou Medical University Li, Jushuang; School of Public Health & Management, Wenzhou Medical Univeristy, Department of Preventive Medicine Wang, Tao; School of Public Health & Management, Wenzhou Medical University, Department of Preventive Medicine Mao, Guangyun; School of Public Health & Management, Preventive Medicine
Keywords:	Metabolomics, Arsenic, Skin lesions, Amino acid, UPLC-MS/MS

Low level of serum aromatic amino acid associated with elevated risk of arsenic-induced skin lesions: data from a metabolomics study.

Yaping Wei, MS[a, b], Chaonan Jia, MS[a,b], Yuan Lan, MS[c], Xiangqing Hou, MS[a,b], Jingjing Zuo, MS[c], Jushuang Li[a,b], Tao Wang[a,b], Guangyun Mao, MD, PhD[a,b,d]✉

[a] Department of Preventive Medicine, School of Public Health and Management, Wenzhou Medical University, Wenzhou, China.

[b] Center on Evidence-Based Medicine & Clinical Epidemiological Research, School of Public Health, Wenzhou Medical University, Wenzhou, China.

[c] School of Ophthalmology & Optometry, Wenzhou Medical University, Wenzhou, China.

[d] Center on Clinical Research, the Affiliated Eye Hospital of Wenzhou Medical University, Wenzhou, China.

Yaping Wei and Chaonan Jia contributed equally to this paper.

✉Corresponding author: Guangyun Mao, Room 7A321, School of Public Health and Management，Wenzhou Medical University, Chashan University town, Ouhai district, Wenzhou, China. 325035

E-mail: maogy@wmu.edu.cn

Telephone: +86-577-86689848

Fax numbers: +86-577-86699122

ABSTRACT:

Objectives To investigate the association of specific serum amino acids (AA) with arsenic-induced skin lesions (AISL) risk and their ability to distinguish AISL from the counterparts.

Design Case-control study.

Setting 3 arsenic exposed villages in Wuyuan county of Hetao Plain, Inner Mongolia, China.

Participants Among 450 residents aged 18 to 79 years chronically exposed to arsenic via drinking water, 56 of them were diagnosed as AISL and selected as the case. Another 56 participants without AISL matched by gender and similar age (±1 year) from the same population were picked out as the control. The inclusion criteria were subjects having the metabolomic test. The exclusion criteria included unmatched participants and those without serum metabolites data.

Primary and secondary outcome measures The outcome measure was whether it suffered from AISL. Generalized linear models and receiver operating characteristic curve analysis were performed to investigate the relationship between AISL and AA metabolism.

Results Two aromatic amino acid (tryptophan, phenylalanine) level were both negatively associated with AISL (P<0.05). As compared to the 1st quartile, the adjusted risk of AISL in the 2nd, 3rd and 4th quartile decreased by 56%, 89% and 82% for tryptophan, and 37%, 84% and 83% for phenylalanine, respectively. The combination of higher level of the two above-mentioned AA revealed the lowest probability to develop AISL (OR=0.06; 95%CI: 0.02, 0.22; P<0.001)). Furthermore, both AA showed moderate ability to distinguish AISL from the control, with area-under-curve (95%CI) as 0.67 (0.57, 0.77) for tryptophan and 0.70 (0.60, 0.80) for phenylalanine, respectively (all P<0.05). The combined pattern with AUC (95%CI) was 0.72 (0.62, 0.81).

Conclusions Current study suggests that AA metabolism may be linked to AISL onset and may play an important role in AISL's early identification. Additional studies are needed to confirm our findings.

Keywords: Metabolomics; Arsenic; skin lesions; Amino acid ; UPLC-MS/MS.

Strengths and limitations of this study

- The main strength of this study might be that the findings was depended on a community-based long-term arsenic exposure cohort with well-designed quality assurance and quality control throughout the study.
- Multivariable logistic models were performed to examine the association between the contributing amino acids (AA) levels and arsenic-induced skin lesions (AISL) after adjusting for some potential confounding factors, and receiver operator characteristic analysis was applied to evaluate the value and feasibility of the AA to distinguish AISL from the counterparts.
- The association between amino acid metabolism and AISL was unclear and comprehensively assessed in this study.
- The findings were mainly based on case-control study, which only revealed the association between amino acid metabolism and the risk of AISL rather than confirming their causal relationship.
- The participants were mainly exposed to arsenic via drinking water, which would limit the findings extrapolated to the other arsenic exposure population via food and other ways.

INTRODUCTION

Chronic arsenic exposure via drinking water is widely believed as a global health concern, affecting a large amount of people worldwide. It may give rise to several human health effects and has been documented to associate with cardiovascular disease, diabetes, cancer and others[1 2]. With the industrial boom and dramatic rise of worldwide water pollution including arsenic contamination in the past, the prevalence and burden of arsenic-induced health damage will continue to increase. Skin has been confirmed as one of the most common and susceptible targets of arsenic-induced health lesions. Cutaneous skin lesions are typical signs of arsenicosis after persistent arsenic exposure for a long term, characterized by hyperkeratosis and hyperpigmentation. Considerable evidences of the prevalence of arsenical skin lesions have been observed in many countries[3-5].

As arsenic induced skin lesions (AISL) have been widely accepted as the major early manifestation of arsenic toxicity[6] and may be indicators of susceptibility to more serious arsenic-induced health hazards[7], it is particularly crucial to identify participants at risk as early as possible for preventing the onset or delaying the progression of the serious health problems effectively. Several possible mechanism may explain arsenic poisoning lesion, such as genetic differences[8], oxidative stress[9] and epigenetic dysregulation[10] and so on. Studies have shown that arsenic methylation in vivo is tightly associated with metabolic syndrome[11 12], which is believed to be closely related to many kinds of metabolites.

Amino acids (AA) are the "basic unit" that make up body's various proteins and necessary to maintain the health status. Some AA are important regulators of key metabolic pathways and of great importance in maximizing efficiency of food utilization, enhancing protein accretion and health improvement[13 14]. Abnormal metabolism of AA will disturbs whole body homeostasis, impairs growth and development, and even causes death[15]. So, serum AA level may be an important implication for the metabolic status and disease conditions. As a powerful tool in system biology research, metabolomic approach is beneficial on unbiased monitoring changes in endogenous

metabolism-related physiological processes, providing integrative information on the distinctive features across multiple functional levels, and offering a window to capture core attributes responsible for various phenotypes, which is particularly important in understanding the relevant pathophysiological changes in a disease state, identifying novel biomarkers for person at risk screening, diagnosis, treatment and prognosis of important human diseases[16-18].

Animal experiments and epidemiological study have reported obvious arsenic-related metabolomics perturbations [19 20]. All of these researches substantially suggest that the relationship between specific metabolites and arsenic-induced health lesions should be investigated. However, few works have been conducted to comprehensively examine the metabolic mechanism relevant to AISL, especially for the AA metabolism. Based on our previous non-targeted metabolomics data, the present study aims to quantitatively examine the association of several specific AA with the risk of AISL and their ability to predict AISL.

Methods

Patient and Public Involvement

No patients were involved.

Study Population

Participants enrolled in the present study came from our previous population-based chronic arsenic exposure cohort established in the Hetao plain of Inner Mongolia, China in 2010. Among 450 residents aged 18 to 79 years chronically exposed to arsenic via drinking water, 56 of them were diagnosed as AISL and selected as the case. Another 56 participants without AISL matched by gender and similar age (±1 year) from the same population were picked out as the control. The inclusion criteria were subjects having the metabolomic test. The exclusion criteria included unmatched participants and those without serum metabolites data. The AISL was diagnosed as the presence of arsenic induced keratosis, hyperpigmentation or depigmentation by a physician from Wenzhou

medical university strictly follow the criteria of arsenicosis[21]. Informed consent was obtained from all participates and this study was approved by ethics committee of Wenzhou Medical University, Wenzhou, China.

Data Collection and Assessment

Detailed data collection of blood and urine samples and assessment methods for clinical and sociodemographic variables had been published previously[22]. The epuration of various urinary arsenic species was conducted by means of a high-performance liquid chromatography coupled mass spectrometer system for separation and detection[23]. The species of arsenic in urine samples consisted of inorganic arsenic (iAs, [iAs$^{\square}$ plus iAs$^{\square}$]), monomethyl arsenate (MMA, [MMA$^{\square}$ plus MMA$^{\square}$]) and dimethyl arsenate (DMA, [DMA$^{\square}$ plus DMA$^{\square}$]). All arsenic species corrected by creatinine.

Distinct Metabolites Identification

Detection of serum metabolites were performed using ultra-performance liquid chromatography couple with tandem mass spectrometry (UPLC-MS/MS). The serum samples preparation and metabolites determination were described in our previous report [24]. The peak intensity of metabolites for the 56 pairs were acquired and then imported to MetaboAnalyst 4.0 (http://www.metaboanalyst.ca/) for statistical analysis. A partial least-squares discriminant analysis (PLS-DA), a supervised and well accepted pattern recognition approach, was used for the differentiation between the cases and controls. False discovery rate (FDR) adjustment p-value in univariate analysis were performed to reduce the potential impact induced by false positive on the results. The criteria used for distinct metabolites screening were defined as both variable importance plot (VIP) scores >1 in PLS-DA and the crude or FDR adjusted p-value all < 0.05. We identified a total of 114 extracted small molecular metabolites significantly contributed to the cases recognition. Among them, 4 metabolites were confirmed as amino acids and obviously down-regulated in the AISL group.

Statistical Analysis

The normality of continuous data was assessed using Shapiro-Wilk test. The comparison of them

between the case and control was applied with paired *t*-test if they met normal or similar normal distribution. Otherwise, Wilcoxon signed rank test would be used. Differences in the proportion of categorical variables between the two groups were evaluated by McNemar-Bowker tests. As the distinct metabolites might be high related with each other, spearman correlation analysis was used to investigate relationship among the various screened AA. To avoid or reduce the influences due to collinearity among the AA, a multiple stepwise regression analysis was applied to screen relevant AA independently associated with AISL. The inclusion and exclusion criteria were 0.05 and 0.10, respectively. Author firstly used the locally weighted regression model to estimate the relationship between serum AA level and the risk to develop AISL. Then, generalized linear models (GLMs) were performed to examine the association between the contributing AA levels and AISL after adjusting for some potential confounding factors. The individual impacts of AA metabolites on the risk of AISL were quantified separately by odds ratio (OR) and 95% confidence interval (CI) with GLMs in the following two ways: with AA as a categorical variable (quartiles) and as a continuous variable [scaled to interquartile range (IQR)]. The combined effects of relevant AA on AISL were also performed. In addition, receiver operator characteristic (ROC) analysis was applied to evaluate the value and feasibility of the AA as the potential sensitive and specific biomarker to predict AISL. Data management, analysis and figure drawing were finished using R version 3.4.4 (Copyright © 2018 The R Foundation for Statistical Computing). All tests were two-sides and P<=0.05 was set as significant level.

RESULTS

Table 1 summarizes the general characteristics of the study population. The comparison of demographical, clinical features and urinary arsenic species in the 56 pairs of subjects were presented in Table 1. The median (1st quartile, 3rd quartile) of age was 50.30 (44.70, 58.70) years for cases and 50.40 (44.60, 58.70) years for the controls with female proportion as 58.93% and there was no statistically significant difference in urinary arsenic levels between the two groups. More than half of

them had no history of smoking or alcohol consumption. As compared to the counterpart, the serum triglycerides level in AISL participants was significantly lower (P=0.041). While the other variables were similar between AISL participants and their control (P>0.05), which indicated that the participants in the two groups were quite comparable to some extent.

Table 1 Demographic characteristics of the study population[ξ].

Variables	AISL[ζ] (n=56)	Non-AISL[ζ] (n=56)	P
Clinical Characteristics			
Age (years)	50.30(44.70,58.70)	50.40(44.60,58.70)	0.425
Exposure year (years)	48.19±11.53	47.62±10.97	0.489
Body mass index (kg/m^2)	24.12±3.14	23.91±2.86	0.697
Fasting plasma glucose (mmol/L)	4.89(4.60,5.25)	5.12(4.53,5.40)	0.137
Folate (ng/mL)	4.00(3.20,5.10)	4.25(3.35,5.40)	0.392
Total homocysteine (μmol/L)	12.30(10.32,16.50)	12.67(11.21,14.69)	0.961
Blood urea nitrogen (mmol/L)	6.45(5.42,7.69)	6.84(5.36,8.80)	0.603
Total cholesterol (mmol/L)	4.58(4.10,5.69)	4.65(3.96,5.95)	0.904
Triglycerides (mmol/L)	1.41(0.90,1.74)	1.45(1.09,2.29)	0.041
High-density lipoprotein (mmol/L)	1.19±0.34	1.16±0.31	0.675
Low-density lipoprotein (mmol/L)	3.04±0.80	3.25±0.84	0.110
Women [# (%)]	33(58.93)	33(58.93)	1.000
Cigarette smoking [# (%)]	20(35.71)	22(39.29)	0.696
Alcohol consumption [# (%)]	17(30.91)	21(37.50)	0.464
Illiteracy [# (%)]	21(37.50)	15(25.00)	0.252
Urinary arsenic species[ζ]			
iAS%	0.12(0.10,0.17)	0.12(0.08,0.15)	0.148
MMA%	0.25(0.20,0.30)	0.26(0.21,0.32)	0.420
DMA%	0.62(0.57,0.71)	0.62(0.48,0.65)	0.096
tAs (μg/g creatinine)	140.93(104.41,208.53)	186.77(80.11,217.30)	0.445

[ξ] AISL: arsenic induced skin lesions; the variables met normal distribution was described with mean± standard deviation; otherwise, median (1st quartile, 3rd quartile) was used to describe their features. Number of cases (percentage) was used to describe the proportion of categorical variables between the two groups.

[ζ] iAS: inorganic arsenic (iAs$^{\square}$+iAs$^{\square}$); MMA: monomethyl arsenate (MMA$^{\square}$+MMA$^{\square}$); DMA: dimethyl arsenate (DMA$^{\square}$+DMA$^{\square}$); tAs: total arsenic (iAs$^{\square}$+iAs$^{\square}$+MMA+DMA); iAs%= iAS/tAs*100%; MMA%=MMA/tAs*100% and DMA%=DMA/tAs*100%.

Table 2 shows the Four AA with FDR adjusted p-value <0.05 and VIP>1 in the cases were observed significantly lower than those of their counterparts. Two of them were aromatic amino acids (AAA) identified as phenylalanine and tryptophan, one is branched-chain amino acids (BCAA) appraised as leucine and the last one is phenylalanyl phenylalanine. Spearman's rank correlation to investigate the relationships among these four AA before the following analyzes and observed that the AA were obviously correlated to each other with the correlation coefficient as 0.18 to 0.86 **(Table**

S1). Through a multivariable logistic stepwise regression model to obtain variables independently associated with the risk to develop AISL and to avoid the potential collinearity among all independent variables in the model. respectively. Finally, we found that tryptophan and phenylalanine were significantly related to the AISL risk.

Table 2 Distinct metabolites in population with arsenic induced skin lesions and their counterparts.

Serum amino acid metabolites	Retention time (min)	Mass-to-Charge Ratio	VIP value	p-value$^{\xi}$	Adjusted p-values$^{\zeta}$
Phenylalanine	3.402	166.087	1.508	<0.001	0.009
Tryptophan	3.886	203.082	1.046	0.003	0.014
Leucine	2.642	132.102	1.014	0.001	0.020
Phenylalanyl phenylalanine	5.048	313.155	1.833	0.004	0.033

$^{\xi}$ From Wilcoxon signed-rank test; $^{\zeta}$ Adjusted by false discovery rate (FDR).

Figure 1 presents the individual association between AA and AISL. Based on locally weighted regression curve, we observed obvious "dose-response" relationships between the levels of serum tryptophan and phenylalanine with AISL. The risk to get AISL was significantly decreased in participants with high serum tryptophan and phenylalanine levels. **Table 3** shows that As compared to the lowest quartile, participants with the 3rd and 4th quartiles of tryptophan were significantly linked to the decreased risk of AISL, with odd radio (OR) and 95% confidence interval (CI) as 0.11(0.02, 0.54) for the former and 0.18(0.05, 0.70) for the latter, after adjusting for the potential confounders such as age, gender, cigarette smoking status, alcohol consumption, fasting plasma glucose, triglycerides, low-density lipoprotein and iAS% (percentage of inorganic arsenic by total arsenic). A significant linear trend existed between the risk to develop AISL and serum tryptophan, decreased by 56% (OR: 0.44; 95% CI: 0.24, 0.84) with per IQR increment of serum tryptophan (P=0.013). Meanwhile, the similar association between serum phenylalanine level and AISL was also found in the present study. After adjusting for the above-mentioned potential confounding factors, the risk to get AISL for subjects in the 3rd and 4th quartiles significantly decreased by 84% (OR: 0.16; 95% CI: 0.04, 0.69) and 83% (OR: 0.17; 95% CI: 0.04, 0.68), respectively. There also existed a linear trend between the serum phenylalanine concentration and the risk of AISL, significantly

decreased by 45% (OR: 0.55; 95% CI: 0.32, 0.94) with per IQR increment of serum phenylalanine level (*P*=0.028) after adjusting for the impacts due to some potential confounding factors.

Table 3 Relationship of amino acids levels with the risk of arsenic-induced skin lesions[ξ]

Model	Tryptophan		Phenylalanine	
	Crude	Adjusted[ζ]	Crude	Adjusted[ζ]
Metabolite as continuous variable				
Per IQR	0.48 (0.27, 0.84)	0.44(0.24,0.84)	0.57 (0.36, 0.91)	0.55(0.32,0.94)
P	0.005	0.013	0.010	0.028
Metabolite as categorical variable				
Q1	1.00(referent)	1.00(referent)	1.00(referent)	1.00(referent)
Q2	0.50(0.16, 1.54)	0.44(0.14,1.43)	0.79(0.22, 2.82)	0.63(0.15,2.59)
Q3	0.12(0.03, 0.53)	0.11 (0.02,0.54)	0.18(0.05, 0.66)	0.16(0.04,0.69)
Q4	0.19(0.05 ,0.71)	0.18(0.05,0.70)	0.25(0.08, 0.79)	0.17(0.04,0.68)
P for trend	0.008	0.005	<0.001	<0.001

[ξ] Values are odds ratios (95% confidence intervals) for arsenic-induced skin lesions from conditional logistic regression. IQR: interquartile range; Q1: the first quartering; Q2: the second quartering; Q3: the third quartering; Q4: the fourth quartering.
[ζ] Adjusted for: age, gender, cigarette smoking status, alcohol consumption, fasting plasma glucose, triglycerides, low-density lipoprotein and percentage of inorganic arsenic by total arsenic (iAS%).

Table 4 shows the joint effect of tryptophan and phenylalanine levels on AISL We firstly classified the serum tryptophan and phenylalanine into two categories with the cut-off value of the mass spectrum peak area based on the ROC analysis as 1730858.446 for the former and 1649420.815 for the latter, respectively. The higher level of these two serum AA was separately defined as equal to or over the cut-off value, while the lower categories were considered as less than the associated cut-off values. As compared to the reference group (both serum tryptophan and phenylalanine were at the lower levels), after adjusting for the impacts induced by some potential confounding factors, participants with higher level of both tryptophan and phenylalanine had the lowest risk to get AISL, significantly decreased by 94% (OR=0.06; 95%CI: 0.02, 0.22; P<0.001). This suggested that AISL jointly and negatively associated with serum tryptophan and phenylalanine concentration. While no significant interaction between tryptophan and phenylalanine on the occurrence of skin damage could be observed (P=0.462).

Table 4 Joint association between tryptophan and phenylalanine levels with arsenic-induced skin lesions.

Tryptophan	Phenylalanine	N	Cases (%)	Crude		Adjusted[ζ]	
<cut-off value[ξ]	<cut-off value[ξ]			OR(95%CI)	P	OR(95%CI)	P
Yes	Yes	35	26(74.3)	1.00(1.00,1.00)	Ref.	1.00(1.00,1.00)	Ref.

Yes	No	20	12(60.0)	0.52(0.16,1.68)	0.273	0.45(0.13,1.55)	0.204
No	Yes	24	12(50.0)	0.35(0.12,1.04)	0.059	0.28(0.09,0.91)	0.034
No	No	33	6(18.2)	0.08(0.02,0.25)	<0.001	0.06(0.02,0.22)	<0.001
Interaction					0.320		0.462

[ξ] Cut-off value was determined by means of receiver operator characteristic (ROC) analysis.
[ζ] Adjusted for age, gender, cigarette smoking status, alcohol consumption, fasting plasma glucose, triglycerides, low-density lipoprotein and percentage of inorganic arsenic by total arsenic (iAS%).

Table 5 shows that based on the ROC analysis, both serum tryptophan and phenylalanine were suggested to be potential biomarkers in distinguishing AISL from a chronic arsenic exposure population (P=0.0020 and P=0.0017). The area under the curve (AUC) and its related 95% CI, sensitivity, specificity, positive predictive value and negative predictive value were 0.67 (0.57, 0.77), 69.64%, 62.50%, 65.00 and 67.31% for tryptophan, and 0.70 (0.60, 0.80), 69.64%, 69.64%, 69.64% and 69.64% for phenylalanine, respectively. For the combined prediction of tryptophan and phenylalanine on the AISL risk, the AUC (95% CI), sensitivity, specificity, positive predictive value and negative predictive value were 0.72 (0.62, 0.81), 76.79%, 58.93%, 65.15 and 71.74%, respectively. Our results suggested that these two AA could be either individually or jointly used as indicators of arsenic induced skin lesions identification.

Table 5 Combination of diagnostic indicators and ROC analysis results[ξ].

Indicators	AUC(95%CI)	Sensitivity, %	Specificity, %	Predict+, %	Predict-, %	P
Tryptophan	0.67(0.57,0.77)	69.64	62.50	65.00	67.31	0.0020
Phenylalanine	0.70(0.60,0.80)	69.64	69.64	69.64	69.64	0.0017
Combined[ζ]	0.72(0.62,0.81)	76.79	58.93	65.15	71.74	<0.001

[ξ] ROC: a receiver operator characteristic; AUC: area under the roc curve; CI: confidence interval; The sensitivities, specificity, positive predictive value and negative predictive value were calculated at their best cut-off points; Predict+: positive predictive value; Predict-: negative predictive value.

[ζ] Combined: tryptophan and phenylalanine. The combination is modeled according to the formula $\beta 1X1 + \beta 2X2$, with Xj denoting the standardized value for the j^{th} amino acid and βj denoting the regression coefficient from the logistic regression model.

DISCUSSION

In the present study, the association of serum tryptophan and phenylalanine, screened in our previous non-targeted metabolomics study using UPLC-MS/MS, with AISL risk and their ability to indicate AISL occurrence were quantitatively evaluated in individual and joint modes. Our results clearly showed that the risk of AISL significantly and negatively associated with serum tryptophan and phenylalanine levels in a chronic arsenic exposure population via drinking water. It was suggested that abnormal amino acids metabolism independently linked to AISL and some specific AA might play an important role in the early identification of AISL.

Participants enrolled in the current study were chronically exposed to arsenic via drinking water. The geometric mean (GM) and its related 95% CI of urinary iAs/creatinine and tAs/creatinine in this population were 17.49 (14.90, 20.53) μg/g and 147.20 (129.00, 167.97) μg/g, respectively. They were much higher than those in the 20-μg/L–exposed to arsenic via drinking water [GM (95% CI): 0.4 (0.3,0.5) μg/g for iAs and 9.1 (6.5,12.7) μg/g for tAs], while obviously lower than those in the 90-μg/L-exposed group [GM (95% CI): 39.4 (31.4, 49.6) μg/g for iAs and 248.7 (208.8, 296.3) μg/g][25]. The available reports show that arsenic induced health lesions can scarcely been cured though the medical technology has made great progress in the past[26]. It is crucial to identify those who are most likely to progress to overt arsenic damages including AISL among people at risk as early as possible. Due to the arsenic methylation[9], genetic differences[8], and other possible metabolic mechanisms[10], some residents exposed to arsenic will not develop to severe arsenic poisoning, especially for those at the early stages of arsenic exposure[4]. Metabolomics study, which mainly focus

on thoroughly assessing the variation of metabolites possibly linked to diseases occurrence and development, has been widely utilized to help us understanding the pathogenesis and others because of its much closer to the phenotypes as compared to other 'OMICs' study[27]. Moreover, mathematical modeling to assess the linkage between small molecular metabolites and arsenic toxicity has grown[28]. Developing a simple and interpretable modeling approach for the early detection of arsenic induced health lesions is of great theoretical value and realistic meaning[29], though it might be difficult due to population specific complexities and the impacts due to some potential unmeasured covariates such as diet and genetic determinants.

Previous studies reported that gene-gene and gene-environment interaction were involved in arsenicosis through toxicological mechanisms including genomic instability[30] and oxidative stress[31]. Skin hyperpigmentation and palmoplantar hyperkeratosis could be biomarkers for long-term (many years) internal dose. Identifying the differences in metabolites that are really associated with phenotypes through metabolites analysis may promote our understanding and identification of AISL. Animal study reveals that the disruption of amino acid metabolism upon arsenic exposure in rat which may be beneficial on understanding the arsenic toxicity[32]. A recent population-based metabolomics study also shows that the serum metabolites alteration significantly related to the risk of arsenic-induced health damages[24]. In the current study, three of the four amino acids are BCAA or AAA are observed to be relevant to AISL occurrence. Several studies across numerous ethnic backgrounds supports the usage of BCAA including leucine, isoleucine as well as valine and AAA profile such as phenylalanine, tryptophan and tyrosine as biomarkers in determining metabolic diseases[27 33]. Simultaneously, Zhou *et al* report that arsenic-induced transformed cells exhibit apparent alterations in metabolite profiles including down-regulated of leucine, tryptophan, and phenylalanine in skin lesions group[34]. Consistent with Zhou's findings, two serum AAA (tryptophan, phenylalanine) levels significantly associated with elevated risk to get AISL in our study.

Normal metabolism of AA are necessary for whole body homeostasis, growth and development, and body health[15]. Studies have reported that the changes in the availability of AAA lead to a profound effect on cell signaling, gene expression, brain, and neuroendocrine function[35]. Tryptophan, an amino acid metabolism related biomarker, is also a sensitive and specific indicator of oxidation. Tryptophan metabolism in mammals is a physiological means of preserving immune homeostasis and associated with oxidative stress, and inflammation[36 37]. In addition, phenylalanine can be transformed into specific neurotransmitters such as dopamine and adrenaline by the action of related enzymes. Wu and colleagues[38] report that arsenic exposure will cause neurotransmitter metabolism disturbance, which may explained the reduction of phenylalanine. Furthermore, as one of the peptide-bound phenylalanine, phenylalanyl phenylalanine has been reported to affect protein synthesis and secretion[39], potentially indicating the possible relation between endothelium dysfunction and phenylalanine metabolism disorder.

The relationship between amino acid metabolism and AISL was unclear. The notable alteration of tryptophan and phenylalanine in this study may well indicate the occurrence of metabolic disorders due to arsenic exposure, is beneficial on intensive understanding the effects of arsenic toxicity and is of great importance and realistic value in the early identification of the occurrence as well as delaying the progression of various arsenic-induced health lesions including AISL. The main strength of this study might be that the findings were depended on a community-based long-term arsenic exposure cohort with well-designed quality assurance and quality control throughout the study. However, the findings are mainly based on case-control study, which only revealed the association between amino acid metabolism and the risk of AISL rather than confirming their causal relationship. The participants were mainly exposed to arsenic via drinking water, which would limit the findings extrapolated to the other arsenic exposure population via food and other ways. Therefore, Additional elaborate population-based studies are needed to verify our discoveries.

Taken together, two AAA's (tryptophan, phenylalanine) reduction were closely linked to the higher risk of AISL. It also suggests that tryptophan and phenylalanine are useful for distinguishing AISL earlier or screening of the high-risk individuals from their counterparts in a long-term exposed to low-level arsenic population via drinking water.

Contributors: Guangyun Mao and Yaping Wei designed the study. Chaonan Jia participated in collecting data. Yuan Lan and Chaonan Jia audited the data. Yaping Wei, Xiangqing hou, Jushuang Li, Tao Wang conducted the literature search, Yaping Wei, Chaonan Jia conducted statistical analysis and interpreted the results. Yaping Wei and Chaonan Jia wrote the first draft of the manuscript. Guangyun Mao reviewed the final manuscript and did substantial contributions.

Funding: This work was supported by the Science Technology Department of Zhejiang Province grant (2013C33169), the initial Scientific Research Fund (KYQD170301) and the major project of the Eye Hospital of Wenzhou Medical University (YNZD201602).

Acknowledgments: We gratefully acknowledge all of the study participants of this project. We wish to thank all of the dedication and hard work of the field team at the department of preventive medicine, school of public health and management, Wenzhou medical university.

Conflicts of Interest: The authors declare no conflict of interest.

Data Sharing Statement: No additional unpublished data are available.

REFERENCES

1. Mandal BK, Suzuki KT. Arsenic round the world: a review. Talanta. 2002; 58:201-35.
2. Smith AH, Steinmaus CM. Health Effects of Arsenic and Chromium in Drinking Water: Recent Human Findings. Annual Review Of Public Health. Palo Alto: Annual Reviews, 2009:107-22.
3. Guha Mazumder DN, Haque R, Ghosh N, et al. Arsenic levels in drinking water and the prevalence of skin lesions in West Bengal, India. International journal of epidemiology. 1998; 27:871-7.
4. Guo XJ, Fujino Y, Kaneko S, et al. Arsenic contamination of groundwater and prevalence of arsenical dermatosis in the Hetao plain area, Inner Mongolia, China. Mol Cell Biochem. 2001; 222:137-40.
5. Lindberg AL, Rahman M, Persson LA, et al. The risk of arsenic induced skin lesions in Bangladeshi men and women is affected by arsenic metabolism and the age at first exposure. Toxicology and Applied Pharmacology. 2008; 230:9-16.
6. Tondel M, Rahman M, Magnuson A, et al. The relationship of arsenic levels in drinking water and the prevalence rate of skin lesions in Bangladesh. Environmental Health Perspectives. 1999; 107:727-29.
7. Hsu LI, Chen GS, Lee CH, et al. Use of Arsenic-Induced Palmoplantar Hyperkeratosis and Skin Cancers to Predict Risk of Subsequent Internal Malignancy. Am J Epidemiol. 2013; 177:202-12.
8. Luo LR, Li YY, Gao YH, et al. Association between arsenic metabolism gene polymorphisms and arsenic-induced skin lesions in individuals exposed to high-dose inorganic arsenic in northwest China. Sci Rep. 2018; 8:12.
9. Hughes MF, Beck BD, Chen Y, et al. Arsenic Exposure and Toxicology: A Historical Perspective. Toxicol Sci. 2011; 123:305-32.
10. Ren X, McHale CM, Skibola CF, et al. An Emerging Role for Epigenetic Dysregulation in Arsenic Toxicity and Carcinogenesis. Environmental Health Perspectives. 2011; 119:11-19.
11. Chen J-W, Wang S-L, Wang Y-H, et al. Arsenic methylation, GSTO1 polymorphisms, and metabolic syndrome in an arseniasis endemic area of southwestern Taiwan. Chemosphere. 2012; 88:432-38.
12. Wang S-L, Chang F-H, Liou S-H, et al. Inorganic arsenic exposure and its relation to metabolic syndrome in an industrial area of Taiwan. Environment International. 2007; 33:805-11.
13. Liu R, Li H, Fan WJ, et al. Leucine Supplementation Differently Modulates Branched-Chain Amino Acid Catabolism, Mitochondrial Function and Metabolic Profiles at the Different Stage of Insulin Resistance in Rats on High-Fat Diet. Nutrients. 2017; 9:20.
14. Platten M, Wick W, Van den Eynde BJ. Tryptophan Catabolism in Cancer: Beyond IDO and Tryptophan Depletion. Cancer Res. 2012; 72:5435-40.
15. Wu G. Amino acids: metabolism, functions, and nutrition. Amino Acids. 2009; 37:1-17.
16. Casadei L, Valerio M, Manetti C. Metabolomics: Challenges and Opportunities in Systems Biology Studies. Methods Mol Biol. 2018; 1702:327-36.
17. Johnson CH, Gonzalez FJ. Challenges and opportunities of metabolomics. J Cell Physiol. 2012; 227:2975-81.
18. Hui AY, McCarty WJ, Masuda K, et al. A systems biology approach to synovial joint lubrication in health, injury, and disease. Wiley Interdiscip Rev Syst Biol Med. 2012; 4:15-37.
19. Lu K, Abo RP, Schlieper KA, et al. Arsenic Exposure Perturbs the Gut Microbiome and Its Metabolic Profile in Mice: An Integrated Metagenomics and Metabolomics Analysis. Environmental Health Perspectives. 2014; 122:284-91.
20. Zhang J, Shen H, Xu W, et al. Urinary Metabolomics Revealed Arsenic Internal Dose-Related Metabolic Alterations: A Proof-of-Concept Study in a Chinese Male Cohort. Environmental Science & Technology. 2014; 48:12265-74.
21. Standard of diagnosis for endemic arsenism of China (WS/T211-2001),the Ministry of Health of the People's Republic of China. Beijing. 2001.
22. Guo X, Cui H, Zhang H, et al. Protective Effect of Folic Acid on Oxidative DNA Damage: A Randomized, Double-Blind, and Placebo Controlled Clinical Trial. Medicine (Baltimore). 2015; 94:e1872.
23. Chen X, Guo X, He P, et al. Interactive Influence of N6AMT1 and As3MT Genetic Variations on Arsenic Metabolism in the Population of Inner Mongolia, China. Toxicol Sci. 2017; 155:124-34.

24. Jia C, Wei Y, Guan X, et al. Comprehensive Analysis of the Metabolomic Characteristics on the Health Lesions Induced by Chronic Arsenic Exposure: A Metabolomics Study. （under review）.
25. Sun G, Xu Y, Li X, et al. Urinary arsenic metabolites in children and adults exposed to arsenic in drinking water in Inner Mongolia, China. Environmental Health Perspectives. 2007; 115:648-52.
26. Kuo CC, Moon KA, Wang SL, et al. The Association of Arsenic Metabolism with Cancer, Cardiovascular Disease, and Diabetes: A Systematic Review of the Epidemiological Evidence. Environmental Health Perspectives. 2017; 125:15.
27. Wang TJ, Larson MG, Vasan RS, et al. Metabolite profiles and the risk of developing diabetes. Nature Medicine. 2011; 17:448-U83.
28. Lawley SD, Cinderella M, Hall MN, et al. Mathematical model insights into arsenic detoxification. Theor Biol Med Model. 2011; 8:31.
29. Juhasz AL, Herde P, Herde C, et al. Validation of the predictive capabilities of the Sbrc-G in vitro assay for estimating arsenic relative bioavailability in contaminated soils. Environ Sci Technol. 2014; 48:12962-9.
30. Kibriya MG, Jasmine F, Parvez F, et al. Association between genome-wide copy number variation and arsenic-induced skin lesions: a prospective study. Environ Health. 2017; 16:13.
31. Priori R, Scrivo R, Brandt J, et al. Metabolomics in rheumatic diseases: the potential of an emerging methodology for improved patient diagnosis, prognosis, and treatment efficacy. Autoimmun Rev. 2013; 12:1022-30.
32. Wang X, Mu X, Zhang J, et al. Serum metabolomics reveals that arsenic exposure disrupted lipid and amino acid metabolism in rats: a step forward in understanding chronic arsenic toxicity. Metallomics Integrated Biometal Science. 2015; 7:544.
33. Weng L, Quinlivan E, Gong Y, et al. Association of Branched and Aromatic Amino Acids Levels with Metabolic Syndrome and Impaired Fasting Glucose in Hypertensive Patients. Metabolic Syndrome And Related Disorders. 2015; 13:195-202.
34. Zhou YY, Wang YF, Su J, et al. Integration of microRNAome, proteomics and metabolomics to analyze arsenic-induced malignant cell transformation. Oncotarget. 2017; 8:90879-96.
35. Lee JC, Chen MJ, Chang CH, et al. Plasma amino acid levels in patients with colorectal cancers and liver cirrhosis with hepatocellular carcinoma. Hepato-Gastroenterol. 2003; 50:1269-73.
36. Munn DH, Mellor AL. Indoleamine 2,3 dioxygenase and metabolic control of immune responses. Trends In Immunology. 2013; 34:137-43.
37. Pawlak K, Domaniewski T, Mysliwiec M, et al. The kynurenines are associated with oxidative stress, inflammation and the prevalence of cardiovascular disease in patients with end-stage renal disease. Atherosclerosis. 2009; 204:309-14.
38. Wu L-L, Gong W, Shen S-P, et al. Multiple metal exposures and their correlation with monoamine neurotransmitter metabolism in Chinese electroplating workers. Chemosphere. 2017; 182:745-52.
39. Zhou MM, Wu YM, Liu HY, et al. Effects of phenylalanine and threonine oligopeptides on milk protein synthesis in cultured bovine mammary epithelial cells. J Anim Physiol Anim Nutr. 2015; 99:215-20.

Figure 1 Association between the peak intensity of tryptophan and phenylalanine and arsenic-induced skin lesions based on multivariable locally weighted regression models. a:tryptophan; b:phenylalanine.

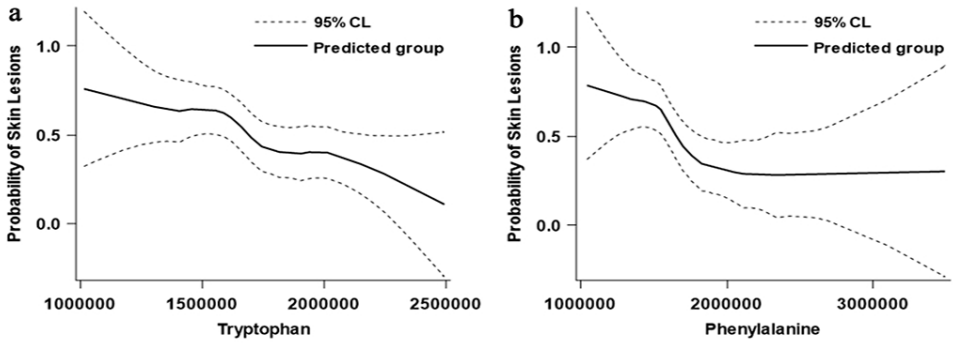

Figure 1 Association between the peak intensity of tryptophan and phenylalanine and arsenic-induced skin lesions based on multivariable locally weighted regression models. a: tryptophan; b: phenylalanine.

90x31mm (300 x 300 DPI)

Table S1

Correlation matrix among four screened amino acids [ξ].

Metabolites	Tryptophan	Leucine	Phenylalanine	Phenylalanyl phenylalanine
Tryptophan	-	0.20(0.038)	0.22(0.022)	0.18(0.061)
Leucine		-	0.86(<.0001)	0.57(<.0001)
Phenylalanine			-	0.60(<.0001)
Phenylalanyl phenylalanine				-

[ξ] Data were presented as the coefficient of correlation (p-value).

BMJ Open

Tryptophan, phenylalanine and arsenic-induced skin lesions in a chronic arsenic exposure Chinese population via drinking water: data from a metabolomics study

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2018-025336.R1
Article Type:	Original research
Date Submitted by the Author:	12-Dec-2018
Complete List of Authors:	wei, yaping; Wenzhou Medical University, Chashan University town, Ouhai district, Wenzhou, China., Jia, Chaonan; School of Public Health & Management, Wenzhou Medical University, Department of Preventive Medicine Lan, Yuan; School of Ophthalmology & Optometry, Wenzhou Medical University, Center on Clinical Research, the Eye Hospital of Wenzhou Medical University Hou, Xiangqing; School of Public Health & Management, Wenzhou Medical Univeristy, Department of Preventive Medicine Zuo, Jingjing; Wenzhou Medical University School of Optometry and Ophthalmology, Center on Clinical Research, the Eye Hospital of Wenzhou Medical University Li, Jushuang; School of Public Health & Management, Wenzhou Medical Univeristy, Department of Preventive Medicine Wang, Tao; School of Public Health & Management, Wenzhou Medical University, Department of Preventive Medicine Mao, Guangyun; School of Public Health & Management, Preventive Medicine
Primary Subject Heading:	Occupational and environmental medicine
Secondary Subject Heading:	Epidemiology, Public health, Occupational and environmental medicine
Keywords:	Metabolomics, Chronic arsenic exposure, Skin lesions, Amino acid, UPLC-MS/MS

Tryptophan, phenylalanine and arsenic-induced skin lesions in a chronic arsenic exposure Chinese population via drinking water: data from a metabolomics study

Yaping Wei[1,2], Chaonan Jia,[1,2], Yuan Lan[3], Xiangqing Hou[1,2], Jingjing Zuo[3], Jushuang Li[1,2], Tao Wang[1,2], Guangyun Mao[1,2,4]✉

[1] Department of Preventive Medicine, School of Public Health and Management, Wenzhou Medical University, Wenzhou, China.

[2] Center on Evidence-Based Medicine & Clinical Epidemiological Research, School of Public Health, Wenzhou Medical University, Wenzhou, China.

[3] School of Ophthalmology & Optometry, Wenzhou Medical University, Wenzhou, China.

[4] Center on Clinical Research, the Affiliated Eye Hospital of Wenzhou Medical University, Wenzhou, China.

Yaping Wei and Chaonan Jia contributed equally to this paper.

✉Corresponding author: Guangyun Mao, Room 7A321, School of Public Health and Management，Wenzhou Medical University, Chashan University town, Ouhai district, Wenzhou, China. 325035

E-mail: maogy@wmu.edu.cn

Telephone: +86-577-86689848

Fax numbers: +86-577-86699122

ABSTRACT:

Objectives To investigate the association of specific serum amino acids (AAs) with the odds of arsenic-induced skin lesions (AISL) and their ability to distinguish AISL from the counterparts.

Design Case-control study.

Setting Three arsenic exposed villages in Wuyuan county of Hetao Plain, Inner Mongolia, China.

Participants Among 450 residents aged 18 to 79 years chronically exposed to arsenic via drinking water, 56 of them were diagnosed as AISL and defined as the cases. Another 56 participants without AISL matched by gender and similar age (±1 year) from the same population were picked out as the controls. The inclusion criteria were subjects having the metabolomics determination. Unmatched participants and those without serum metabolites data were excluded.

Primary and secondary outcome measures The outcome was whether it suffered from AISL. Multivariable conditional logistic regression models and receiver operating characteristic curve (ROC) analysis were performed to investigate the relationship between specific AAs and AISL.

Results The level of two aromatic AAs (tryptophan, phenylalanine) were both negatively associated with AISL ($P<0.05$). As compared to the 1st quartile, the adjusted odds of AISL in the 2nd, 3rd and 4th quartile decreased by 69%, 90% and 84% for tryptophan, and 14%, 80% and 76% for phenylalanine, respectively. The combination of the two aforementioned higher level AAs revealed the lowest odds of AISL (OR=0.06; 95%CI: 0.02, 0.22; $P<0.001$)). Furthermore, both AAs showed moderate ability to distinguish AISL from the control, with area-under-curve [(AUC), 95%CI] as 0.67 (0.57, 0.77) for tryptophan and 0.70 (0.60, 0.80) for phenylalanine, respectively (all $P<0.05$). The combined pattern with AUC (95%CI) was 0.72 (0.62, 0.81), sensitivity of 76.79% and specificity of 58.93% ($P<0.001$).

Conclusions Specific AAs might be linked to AISL and play an important role in its early identification. Additional studies are needed to confirm our findings.

Keywords: Metabolomics; Chronic arsenic exposure; Skin lesions; Amino acid; UPLC-MS/MS.

Strengths and limitations of this study

- Our findings were depended on a community-based metabolomics study with paired-design, strictly quality assurance and quality control.
- Multivariable conditional logistic models were performed to examine the association between specific amino acid levels and AISL, and ROC analysis was applied to evaluate the value and feasibility of the AA to distinguish AISL from the counterparts.
- Although the AAs were determined by untargeted metabolomics approach, which can assess a large amount of metabolites precisely and efficiently, we only can obtain relative levels of AAs instead of their accurate quantitative concentration.
- The findings were based on a case-control study, which only revealed the association between the AAs and the odds of AISL rather than confirming their causal relationship.
- The participants were mainly chronically exposed to arsenic via drinking water, which would limit the findings extrapolated to another arsenic exposure population via food or other ways.

INTRODUCTION

Chronic arsenic exposure via drinking water is widely believed as a global health concern, affecting a large amount of people worldwide. It may give rise to several human health issues and has been documented to associate with cardiovascular disease, diabetes, cancer and others[1 2]. With the industrial boom and dramatic rise of worldwide water pollution including arsenic contamination in the past, the prevalence and burden of arsenic-induced health damage will continue to increase. Skin has been confirmed as one of the most common and susceptible target of arsenic-induced health lesions. Cutaneous skin lesions are typical signs of arsenicosis after persistent arsenic exposure for a long term which are characterized by hyperkeratosis and hyperpigmentation. Considerable evidences of the prevalence of arsenical skin lesions had been observed in many countries[3-5].

As arsenic-induced skin lesions (AISL) have been widely accepted as the major early manifestation of arsenic toxicity[6] and might be indicators of susceptibility to more serious arsenic-induced health hazards[7], it is particularly crucial to identify participants at risk as early as possible for preventing the onset or delaying the progression of the serious health problems effectively. Several possible mechanism such as genetic differences[8], oxidative stress[9] and epigenetic dysregulation[10] and others may explain arsenic poisoning. Previous studies also reported that arsenic methylation in vivo might be associated with metabolic syndrome[11 12].

Amino acids (AAs) are the "basic unit" that make up the body's various proteins and necessary to maintain the health. Some AAs are important regulators of some key metabolic pathways and have great importance in maximizing efficiency of food utilization, enhancing protein accretion and health improvement[13 14]. Abnormal metabolism of AAs will disturb the homeostasis of the body, impairs growth and development, and even causes death[15]. So, the levels of serum AAs may be an important implication for the metabolic status and disease condition. As a powerful tool in system biology research, metabolomics approach is beneficial on unbiased monitoring changes in endogenous metabolism-related physiological processes, providing integrative information on the distinctive

features across multiple functional levels, and offering a window to capture the core attributes responsible for various phenotypes, which are particularly important in understanding the relevant pathophysiological changes of a disease and its status, identifying novel biomarkers for risk screening, diagnosis, treatment and prognosis of important human diseases[16-18].

Animal experiments and epidemiological study have reported obvious arsenic-related metabolomics perturbations[19 20]. All of these researches substantially suggests that the relationship between specific metabolites and arsenic-induced health lesions should be investigated. However, few works have been conducted to comprehensively examine the metabolic mechanism relevant to AISL, especially for the AAs metabolism. The present study aims to quantitatively examine the association of several specific AAs with AISL and the ability to identify AISL.

Methods

Patient and Public Involvement

No patients were involved.

Study Population

This study was originally from a randomized, double-blind, and placebo controlled clinical trial (NCT02235948) in 2010, in which all subjects were randomly selected using permuted block randomization from a single rural area in a population chronically exposed to low-level arsenic drinking water, had similar life style and influences under similar environmental factors. Information on the inclusion and exclusion criteria of the participants could be found in our previous study[21]. Strictly following the criteria of arsenicosis[22]. AISL was diagnosed as the presence of arsenic-induced keratosis, hyperpigmentation or depigmentation by a physician from Wenzhou medical university at the beginning of the trial. Among 450 residents aged 18 to 79 years old enrolled in the above-mentioned trial, 56 of them were diagnosed as AISL and selected as the case. Another 56 participants without AISL matched by gender and similar age (±1 year) from the same population

were picked out as the control. The inclusion criteria were subjects having the metabolomic test. Unmatched participants and those without serum metabolites data were excluded. Informed consent was obtained from all participants and this study was approved by the ethics committee of Wenzhou Medical University, Wenzhou, China.

Data Collection and Assessment

Blood and urine samples were also collected at the time of participants' enrollment. Detailed data collection of blood and urine samples and assessment methods for clinical and sociodemographic variables had been published previously[21]. The epuration of various urinary arsenic species were conducted by means of a high-performance liquid chromatography coupled mass spectrometer system for separation and detection[23]. The species of arsenic in urine samples consisted of inorganic arsenic (iAs, [$iAs^{Ⅲ}$ plus $iAs^{Ⅴ}$]), monomethyl arsenate (MMA, [$MMA^{Ⅲ}$ plus $MMA^{Ⅴ}$]) and dimethyl arsenate (DMA, [$DMA^{Ⅲ}$ plus $DMA^{Ⅴ}$]). All arsenic species were corrected by creatinine.

UPLC-MS/MS Metabonomic Profiling

Serum samples（200 μL in microcentrifuge tubes) were thawed to room temperature (25°C) and 600 μL mixture (90% acetonitrile - 10%water) were added to each sample. The samples were vigorously mixed for 20 seconds and centrifuged for 5 min at 12000 rpm (20°C). The top 400 μL of each supernatant were then transferred and dried down in a vacuum concentrator centrifuge. The dried samples were re-suspended in 130 μL of water (including 15% acetonitrile), mixed vigorously for 20 seconds and repeated the centrifugation method described above. Two μL of the supernatant were collected as samples to be determined. Serum metabolic profile acquisition was performed by using ACQUITY UPLC®/Xevo® G2 QTof/MS^E (Waters Corp., Milford, MA, USA). Chromatographic separation was performed at 50°C using a WATERS HSS T3 column (2.1×100 mm, 1.7 μm) with a flow rate of 0.4mL/min. The mobile phase was a mixture of (A) H_2O with 0.1% formic acid and (B) methanol with 0.1% formic acid. Elution was in linear gradient with the

programmed gradient at 0 min with 100% A and 0% B, 1.00min with 100%A and 0% B, 8 min with 0%A and 100% B, 13.00 min with 0% A and 100% B. The mass spectrometer was operated under both positive-ion (ESI^+) mode and negative-ion (ESI^-) mode electrospray ionization. The scan range was from 50 to 1200 m/z. Data was collected in both ESI^+ and ESI^- modes. Capillary voltage was set at 3000 V and 2500 V, respectively. The desolvation flow rate was 800 L/h at 350°C. Argon was used as a collision gas, and the collision energy was adjusted from 10 eV to 40 eV for each analysis. Quantum clustering (QC) samples were prepared by pooling aliquots of each sample and used to reflect the reliability of further metabolomics analysis. After peak deconvolution, alignment, integration and normalization, the data including retention time (RT), mass to charge ratio(m/z), and peak intensity were extracted from raw chromatograms using Progenesis QI 2.0 (Waters Corp., Milford, MA, USA). The MS/MS mode was performed to obtain metabolites levels processed with MarkerLynx Applications Manager Version 4.1 (Waters Corp., Milford, MA, USA).

Distinct Metabolites Identification

The peak intensity of metabolites for the 56 pairs were acquired and then imported to MetaboAnalyst 4.0 (http://www.metaboanalyst.ca/) for statistical analysis. A partial least-squares discriminant analysis (PLS-DA), which is a supervised and well accepted pattern recognition approach, was used for the differentiation between the cases and controls. False discovery rate (FDR) adjusted p-value in univariate analysis were performed to reduce the potential impact induced by false positive of the results. The criteria used in the selection of metabolites include variable importance in the project (VIP) scores >1 in PLS-DA and the crude or FDR adjusted p-value all < 0.05. We identified a total of 70 extracted small molecular metabolites that were linked to the recognition of AISL. Among them, 4 metabolites were confirmed as AAs and obviously down-regulated in the AISL group.

Statistical Analysis

The normality of continuous data was assessed using both QQ-plots and Shapiro-Wilk test. The

comparison between the cases and controls was performed with the paired *t*-test if they met normal or similar normal distribution. Otherwise, Wilcoxon signed rank test would be used. Differences in the proportion of categorical variables between the two groups were evaluated by McNemar-Bowker tests. We firstly used locally weighted regression (Loess) models to estimate the "real" relationship between serum AAs levels and the odds of AISL. Then, multivariable conditional logistic regression models were performed to examine the association between the contributing AAs levels and AISL after adjusting for some potential confounding factors. The individual impacts of AA metabolites on the risk of AISL were quantified separately by odds ratio (OR) and 95% confidence interval (CI) in the following two ways: with AA as a categorical variable (quartiles) and as a continuous variable [scaled to interquartile range (IQR)]. As the distinct metabolites might be high related to each other, collinearity should be well considered. To screen appropriate covariates in the logistic models, some potential risk factors of AISL such as the duration of arsenic exposure, serum folate, cigarette smoking, alcohol consumption, blood urea nitrogen and others were added in the model as covariate, respectively. As too many covariates in a multiple regression model will lead to overfitting to some extent[24], we finally select no more than 5 variables as confounding factors to decrease the potential overfitting when assessing the association between AAs and AISL. Meanwhile, we used the variance inflation factor (VIF) based on the VIF package of R software to detect potential collinearity among the AAs. When the VIF is greater than 1.5, it would be considered as collinearity existed in the model and the associated variable would be removed from the model. The combined effects of relevant AAs on AISL were also performed. In addition, receiver operator characteristic (ROC) analysis was applied to evaluate the value and feasibility of the AAs as the potential sensitive and specific biomarker to recognize AISL. Data management, analysis and figure drawing were finished using R version 3.4.4 (Copyright © 2018 The R Foundation for Statistical Computing). All tests were two-sides and P<=0.05 was set as significant level.

RESULTS

Table 1 summarizes the general characteristics of the study population. The comparison of demographical, clinical features and urinary arsenic species in the 56 pairs of subjects were presented in Table 1. The median (1st quartile, 3rd quartile) age of AISL population was 50.30 (44.70, 58.70) and 50.40 (44.60, 58.70) years for the controls. Both groups have the same proportion of female population (58.93%), and there were no obvious statistical difference in the urinary arsenic levels between the two groups. More than half of them had no history of smoking or alcohol consumption. When compared to the controls, the serum triglycerides level in AISL participants was significantly lower (P=0.041). While the other variables were similar between AISL participants and the control (P>0.05). This indicates that the participants in two groups are comparable to some extent.

Table 1 Demographic characteristics of the study population[ξ].

Variables	AISL[ζ] (n=56)	Non-AISL[ζ] (n=56)	P
Clinical Characteristics			
Age (years)	50.30(44.70,58.70)	50.40(44.60,58.70)	0.425
Exposure year (years)	48.19±11.53	47.62±10.97	0.489
Body mass index (kg/m^2)	24.12±3.14	23.91±2.86	0.697
Fasting plasma glucose (mmol/L)	4.89(4.60,5.25)	5.12(4.53,5.40)	0.137
Folate (ng/mL)	4.00(3.20,5.10)	4.25(3.35,5.40)	0.392
Total homocysteine (μmol/L)	12.30(10.32,16.50)	12.67(11.21,14.69)	0.961
Blood urea nitrogen (mmol/L)	6.45(5.42,7.69)	6.84(5.36,8.80)	0.603
Total cholesterol (mmol/L)	4.58(4.10,5.69)	4.65(3.96,5.95)	0.904
Triglycerides (mmol/L)	1.41(0.90,1.74)	1.45(1.09,2.29)	0.041
High-density lipoprotein (mmol/L)	1.19±0.34	1.16±0.31	0.675
Low-density lipoprotein (mmol/L)	3.04±0.80	3.25±0.84	0.110
Women [# (%)]	33(58.93)	33(58.93)	1.000
Cigarette smoking [# (%)]	20(35.71)	22(39.29)	0.696
Alcohol consumption [# (%)]	17(30.91)	21(37.50)	0.464
Illiteracy [# (%)]	21(37.50)	15(25.00)	0.252
Urinary arsenic species[ζ]			
iAS%	0.12(0.10,0.17)	0.12(0.08,0.15)	0.148
MMA%	0.25(0.20,0.30)	0.26(0.21,0.32)	0.420
DMA%	0.62(0.57,0.71)	0.62(0.48,0.65)	0.096
tAs (μg/g creatinine)	140.93(104.41,208.53)	186.77(80.11,217.30)	0.445

[ξ] AISL: arsenic-induced skin lesions; the variables met normal distribution was described with mean± standard deviation; otherwise, median (1st quartile, 3rd quartile) was used to describe their features. Number of cases (percentage) was used to describe the proportion of categorical variables between the two groups.

[ζ] iAS: inorganic arsenic ($iAs^{Ⅲ}$+$iAs^{Ⅴ}$); MMA: monomethyl arsenate ($MMA^{Ⅲ}$+$MMA^{Ⅴ}$); DMA: dimethyl arsenate ($DMA^{Ⅲ}$+$DMA^{Ⅴ}$); tAs: total arsenic ($iAs^{Ⅲ}$+$iAs^{Ⅴ}$+MMA+DMA); iAs%= iAS/tAs*100%; MMA%=MMA/tAs*100% and DMA%=DMA/tAs*100%.

Table 2 shows that the four AAs, which FDR adjusted p-value <0.05 and VIP>1, in the cases are observed significantly lower than those of the controls. Two of them are aromatic amino acids (AAA) identified as phenylalanine and tryptophan, one is branched-chain amino acids (BCAA) appraised as leucine and the last one is phenylalanyl phenylalanine. The individual association of AAs with AISL were presented in figure1, which clearly reveals obvious "dose-response" relationships between them.

Table 2 Distinct metabolites in population with arsenic-induced skin lesions and their counterparts.

Serum amino acid metabolites	Retention time (min)	Mass-to-Charge Ratio	VIP value	p-value$^{\xi}$	Adjusted p-values$^{\zeta}$
Phenylalanine	3.402	166.087	1.508	<0.001	0.009
Tryptophan	3.886	203.082	1.046	0.003	0.014
Leucine	2.642	132.102	1.014	0.001	0.020
Phenylalanyl phenylalanine	5.048	313.155	1.833	0.004	0.033

VIP: variable importance in the project; $^{\xi}$ Wilcoxon signed-rank test; $^{\zeta}$ Adjusted by false discovery rate (FDR).

The relationships between the levels of serum AAs and the odds of AISL were presented in **table S1**, **table S2**, **table S3** and **table S4** with adding different covariates in the model, respectively. Finally, 4 variables including body mass index, serum folate, serum triglycerides and urinary total arsenic were selected as the confounding factors to avoid potential overfitting of the models. **Table 3** clearly shows that participants in the 3rd and 4th quartiles of the 4 specific AAs were all significantly linked to the decreased odds of AISL after adjusting for the potential confounders as compared to their lowest quartiles, respectively. Significant linear trends existed between AISL and the 4 serum AAs. Meanwhile, same linear negative association between AISL and per IQR rise of the 4 serum AAs were observed when these AAs were considered as continuous variables in the present study.

Table 3 Relationship of amino acids levels with the odds of arsenic-induced skin lesions[ξ].

Models	Tryptophan		Phenylalanine		Leucine		Phenylalanyl phenylalanine	
	Crude	Adjusted[ζ]	Crude	Adjusted[ζ]	Crude	Adjusted[ζ]	Crude	Adjusted[ζ]
Amino acids as continuous variable								
Per IQR	0.48 (0.27,0.84)	0.44 (0.24,0.80)	0.57 (0.36,0.91)	0.54 (0.32,0.89)	0.45 (0.25,0.82)	0.45 (0.24,0.83)	0.62 (0.36,1.04)	0.62 (0.36,1.08)
P	0.011	0.007	0.019	0.016	0.010	0.011	0.070	0.090
Amino acids as categorical variable								
Q_1	1.00 (referent)	1.00 (referent)	1.00 (referent)	1.00 (referent)	1.00 (referent)	1.00 (referent)	1.00 (referent)	1.00 (referent)
Q_2	0.50 (0.16,1.54)	0.41 (0.12,1.41)	0.79 (0.22,2.82)	0.86 (0.21,3.51)	0.39 (0.13,1.14)	0.41 (0.12,1.34)	0.16 (0.03,0.75)	0.16 (0.03,0.79)
Q_3	0.12 (0.03,0.53)	0.10 (0.02,0.50)	0.18 (0.05,0.66)	0.20 (0.05,0.79)	0.17 (0.05,0.64)	0.14 (0.04,0.58)	0.08 (0.02,0.39)	0.08 (0.01,0.40)
Q_4	0.19 (0.05,0.71)	0.16 (0.04,0.67)	0.25 (0.08,0.79)	0.24 (0.07,0.84)	0.18 (0.05,0.61)	0.15 (0.04,0.59)	0.12 (0.02,0.57)	0.13 (0.03,70)
P trend	0.007	0.007	<0.001	<0.001	0.002	0.002	0.005	0.005

[ξ]Values are odds ratio (95% confidence intervals) for arsenic-induced skin lesions from conditional logistic regression. IQR: interquartile range; Q_1: the 1st quartile; Q_2: the 2nd quartile; Q_3: the 3rd quartile; Q_4: the 4th quartile.
[ζ] Adjusted for: body mass index, serum folate, serum triglycerides and urinary total arsenic.

As these 4 specific AAs are significantly or marginal significantly associated with the odds of AISL, so it is needed to examine the joint impacts among them on AISL. However, the results of potential collinearity examination revealed that among these 4 specific AAs, both tryptophan and phenylalaine had the smallest VIF value (VIF=1.04) and no obvious collinearity existed (**Table S5**). Hence, we mainly focus on tryptophan and phenylalaine when assessing the joint impacts of AAs on AISL and only presented the results associated with these two AAs in the current study. To avoid the impacts due to insufficient power because of unreasonable grouping on the results, we classified both tryptophan and phenylalanine into two categories, according to the cut-off values of their mass spectrum peak area based on the ROC analysis, respectively. The higher levels of these two serum AA were defined as equal to or over the cut-off values, while the lower categories were considered as less than the associated values. **Table 4** shows the joint impacts of tryptophan and phenylalanine levels on AISL after considering the collinearity of variables in the model. The proportions of AISL were 74.3%, 60.0%, 50.0% and 18.2% for participants with lower levels of both tryptophan and phenylalanine (category A), with higher tryptophan and lower phenylalanine (category B), with

lower tryptophan and higher phenylalanine (category C), and higher levels of both tryptophan and phenylalanine (category D), respectively. Obvious decrease trend of the probability of AISL was observed among these 4 categories. As compared to the category A, adjusted OR (95% CI) for participants in the category B, C and D were 0.47(0.14, 1.60), 0.36(0.11, 1.15) and 0.06(0.02, 0.22). Subjects with higher levels of both tryptophan and phenylalanine had the lowest odds of AISL, significantly decreased by 94% (OR=0.06; 95%CI: 0.02, 0.22; P<0.001), after adjusting for the impacts induced by some potential confounding factors. This suggested that tryptophan and phenylalanine were jointly associated with the presence of AISL. While no significant interaction between the two AAs on the occurrence of AISL could be observed (P=0.270), which indicated that each AA was independently associated with AISL though their joint impact was significant.

Table 4 Joint association between tryptophan and phenylalanine levels with arsenic-induced skin lesions.

Tryptophan	Phenylalanine	N	Cases (%)	Crude		Adjusted$^{\zeta}$	
<cut-off value$^{\xi}$	<cut-off value$^{\xi}$			OR (95%CI)	P	OR (95%CI)	P
Yes	Yes	35	26(74.3)	1.00(1.00,1.00)	Ref.	1.00(1.00,1.00)	Ref.
No	Yes	20	12(60.0)	0.52(0.16,1.68)	0.273	0.47(0.14,1.60)	0.227
Yes	No	24	12(50.0)	0.35(0.12,1.04)	0.059	0.36(0.11,1.15)	0.085
No	No	33	6(18.2)	0.08(0.02,0.25)	<0.001	0.06(0.02,0.22)	<0.001
Interaction					0.320		0.270

$^{\xi}$ Cut-off value was determined by means of receiver operator characteristic analysis.
$^{\zeta}$ Adjusted for body mass index, serum folate, triglycerides, total arsenic.

Table 5 shows that, based on the ROC analysis, both serum tryptophan and phenylalanine might be potential biomarkers in distinguishing AISL from a chronic arsenic exposure population (P=0.0020 and P=0.0017). The area under the curve (AUC) and its related 95% CI, sensitivity, specificity, positive predictive value and negative predictive value were 0.67 (0.57, 0.77), 69.64%, 62.50%, 65.00 and 67.31% for tryptophan, and 0.70 (0.60, 0.80), 69.64%, 69.64%, 69.64% and 69.64% for phenylalanine, respectively. The AUC (95% CI), sensitivity, specificity, positive predictive value and negative predictive value of the combination of them were 0.72 (0.62, 0.81), 76.79%, 58.93%, 65.15 and 71.74%, respectively. Our results suggested that these two AAs could be either individually or jointly used as indicators of AISL identification.

Table 5 Combination of diagnostic indicators and ROC analysis results[ξ].

Indicators	AUC (95%CI)	Sensitivity, %	Specificity, %	Predict+, %	Predict-, %	P
Tryptophan	0.67(0.57,0.77)	69.64	62.50	65.00	67.31	0.002
Phenylalanine	0.70(0.60,0.80)	69.64	69.64	69.64	69.64	0.002
Combined[ζ]	0.72(0.62,0.81)	76.79	58.93	65.15	71.74	<0.001

[ξ] ROC: a receiver operator characteristic; AUC: area under the roc curve; CI: confidence interval; The sensitivities, specificity, positive predictive value and negative predictive value were calculated at their best cut-off points; Predict+: positive predictive value; Predict-: negative predictive value.

ζ Combined: tryptophan and phenylalanine. The combination is modeled according to the formula $\beta 1X_1 + \beta 2X_2$, with Xj denoting the standardized value for the j^{th} amino acid and βj denoting the regression coefficient from the logistic regression model.

DISCUSSION

In the present study, the association of serum tryptophan and phenylalanine, screened in our previous non-targeted metabolomics study using UPLC-MS/MS, with AISL and their ability to indicate AISL occurrences were quantitatively evaluated in individual and joint modes. Our results clearly showed that AISL are significantly and negatively associated with serum tryptophan and phenylalanine levels in a chronic arsenic exposure population via drinking water. Participants with higher level of both AAs would had lowest odds of AISL. These two AAs might also be able to serve as the indicators of AISL.

The probability of the initiation and development of AISL would be affected by a large number of factors including age, gender, life styles, arsenic exposure, metabolism and others. These factors would be important confounding factors and will largely affect our results. To adjust for the impacts due to these cofactors, we firstly selected all participants using permuted block randomization from a single rural area in which population were chronically exposed to arsenic in a same way, had similar life style and environmental factors. Secondly, the cases and controls were matched by gender and age (±1 year). All of these may be the reason why so many potential confounders including arsenic exposure do not differ significantly between the cases and controls (table 1).

Participants enrolled in the current study were chronically exposed to arsenic via drinking water. The geometric mean (GM) and its related 95% CI of urinary iAs/creatinine and tAs/creatinine in this population were 17.49 (14.90, 20.53) μg/g and 147.20 (129.00, 167.97) μg/g, respectively. They were much higher than those in the 20 μg/L exposed to arsenic via drinking water [GM (95% CI):

0.4 (0.3,0.5) μg/g for iAs and 9.1 (6.5,12.7) μg/g for tAs], while obviously lower than those in the 90 μg/L exposed group [GM (95% CI): 39.4 (31.4, 49.6) μg/g for iAs and 248.7 (208.8, 296.3) μg/g][25]. An available report has shown hat AISL cannot be completely cured even though the medical technology has already made great progress[26]. So, it is crucial to identify those who are most likely to progress to overt arsenic damages including AISL among people at risk as early as possible. Metabolomics study, which mainly focus on thoroughly assessing the variation of metabolites possibly linked to diseases occurrence and development, has been widely utilized to help us understand pathogenesis of diseases because of its relevance to the phenotypes as compared to other 'OMICs' study[27]. Moreover, mathematical modeling to assess the linkage between small molecular metabolites and arsenic toxicity has grown[28]. Developing a simple and interpretable modeling approach for the early detection of arsenic induced health lesions is of great theoretical value and realistic meaning[29], though it might be difficult due to population specific complexities and the impacts due to some potential unmeasured covariates such as diet and genetic determinants.

Previous studies reported that gene-gene and gene-environment interaction were involved in arsenicosis through toxicological mechanisms including genomic instability[30] and oxidative stress[31]. Skin hyperpigmentation and palmoplantar hyperkeratosis could be biomarkers for long-term arsenic exposure identifying the differences in metabolites that are really associated with phenotypes through metabolites analysis may promote our understanding and identification of AISL. Animal study reveals that the disruption of amino acids metabolism upon arsenic exposure in rat which may be beneficial on understanding arsenic toxicity[32]. In our previous population-based metabolomics study, we found that serum metabolites alteration were significantly related to the risk of arsenic-induced health damages. In the current study, we detected that BCAA or AAA were also signinificantly relevant to AISL occurrence. Several studies across numerous ethnic backgrounds supports the usage of BCAA including leucine, isoleucine as well as valine and AAA profile such as phenylalanine, tryptophan and tyrosine as biomarkers in determining metabolic diseases[27 33]. Simultaneously, Zhou

et al reports that arsenic-induced transformed cells exhibit apparent alterations in metabolite profiles including down-regulated of leucine, tryptophan, and phenylalanine in skin lesions group[34]. Consistent with Zhou's findings, two serum AAA (tryptophan, phenylalanine) levels were also significantly associated with AISL in our study.

Normal metabolism of amino acids are necessary for whole body homeostasis, growth and development, and health status[15]. Studies have reported that the changes in the availability of AAA will affect cell signaling, gene expression, brain, and neuroendocrine function[35]. Tryptophan, an amino acid metabolism related biomarker, is also a sensitive and specific indicator of oxidation. Tryptophan metabolism in mammals is a physiological means of preserving immune homeostasis associated with oxidative stress and inflammation[36 37]. In addition, phenylalanine can be transformed into specific neurotransmitters such as dopamine and adrenaline by the action of related enzymes. Wu and colleagues[38] reported that arsenic exposure would lead to neurotransmitter metabolism disturbance, which might explain the reduction of phenylalanine. Furthermore, as one of the peptide-bound phenylalanine, phenylalanyl phenylalanine has been reported to affect protein synthesis and secretion[39], potentially indicating the possible relation between endothelium dysfunction and phenylalanine metabolism disorder. The relationship between amino acid metabolism and AISL was still unclear. The notable alteration of tryptophan and phenylalanine in this study may well indicate the occurrence of metabolic disorders due to arsenic exposure. It is also beneficial to understand the effects of arsenic toxicity and of great importance in early identification of occurrences as well as delaying the progression of various arsenic-induced health lesions including AISL.

The current study included 56 AISL cases matched 56 non-AISL controls and the sample size might be potentially insufficient. To estimate the impact due to this potential insufficient sample size on our conclusion, the PROC POWER procedure for paired design study in SAS 9.4 (SAS Institute Inc.) was applied to assess the power of the 4 AAs when assessing their associations with AISL in this study. The results showed that the lowest power associated with all of these four AAs was 0.911 based on 56 pairs of participants (Figure 1S). It

suggested that with type I error as 0.05, total sample size as 56 pairs and two-sided test, the powers associated with these 4 amino acids were all great than 0.8. So, we believed that the sample size for the present study, 56 pairs, would well balance the power of tests. Furthermore, previous metabolomics studies usually have sample size no more than 40 cases in each group[40 41].

The main strength of this study is that the findings were depended on a community-based, long-term arsenic exposure cohort with well-designed quality assurance and quality control throughout the study, and the AAs were detected with non-targeted metabolomics approach through the discovery and validation phases. However, there are also several limitations to this study. Firstly, although it can assess a large amount of metabolites precisely and efficiently, untargeted metabolomics approach only provide relative levels of AAs instead of their accurate quantitative concentration. Secondly, these findings are mainly based on a case-control study, which only reveals the association between amino acid metabolism and the odds of AISL rather than confirming their causal relationship. Finally, the participants were mainly exposed to arsenic via drinking water, which would limit the findings extrapolated to the other arsenic exposure population via food and other ways. Therefore, additional elaborate population-based studies are needed to verify our discoveries.

In conclusion, specific amino acids might be linked to AISL and amino acids metabolism may play an important role in AISL early identification. Additional studies are needed to confirm our findings

Contributors: Guangyun Mao and Yaping Wei designed the study. Chaonan Jia participated in collecting data. Yuan Lan and Chaonan Jia audited the data. Yaping Wei, Xiangqing hou, Jushuang Li, Tao Wang conducted the literature search, Yaping Wei, Chaonan Jia conducted statistical analysis and interpreted the results. Yaping Wei and Chaonan Jia wrote the first draft of the manuscript. Jingjing Zuo helped with copyediting. Guangyun Mao reviewed the final manuscript and did substantial contributions.

Funding: This work was supported by the Science Technology Department of Zhejiang Province grant (2013C33169), the initial Scientific Research Fund (KYQD170301) and the major project of the Eye Hospital of Wenzhou Medical University (YNZD201602).

Acknowledgments: We gratefully acknowledge all of the study participants of this project. We wish to thank all of the dedication and hard work of the field team at the department of preventive medicine, school of public health and management, Wenzhou medical university. Our thanks must also go to Dr. Kendrick Hii, a native-English speaker, for his valuable suggestions and edits to the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

Data Sharing Statement: No additional unpublished data are available.

REFERENCES

1. Mandal BK, Suzuki KT. Arsenic round the world: a review. Talanta. 2002; 58:201-35.
2. Smith AH, Steinmaus CM. Health Effects of Arsenic and Chromium in Drinking Water: Recent Human Findings. Annual Review Of Public Health. Palo Alto: Annual Reviews 2009:107-22.
3. Guha Mazumder DN, Haque R, Ghosh N, et al. Arsenic levels in drinking water and the prevalence of skin lesions in West Bengal, India. International journal of epidemiology. 1998; 27:871-7.
4. Guo XJ, Fujino Y, Kaneko S, et al. Arsenic contamination of groundwater and prevalence of arsenical dermatosis in the Hetao plain area, Inner Mongolia, China. Mol Cell Biochem. 2001; 222:137-40.
5. Lindberg A-L, Rahman M, Persson L-A, et al. The risk of arsenic induced skin lesions in Bangladeshi men and women is affected by arsenic metabolism and the age at first exposure. Toxicology and Applied Pharmacology. 2008; 230:9-16.
6. Tondel M, Rahman M, Magnuson A, et al. The relationship of arsenic levels in drinking water and the prevalence rate of skin lesions in Bangladesh. Environmental Health Perspectives. 1999; 107:727-29.
7. Hsu LI, Chen GS, Lee CH, et al. Use of Arsenic-Induced Palmoplantar Hyperkeratosis and Skin Cancers to Predict Risk of Subsequent Internal Malignancy. Am J Epidemiol. 2013; 177:202-12.
8. Luo LR, Li YY, Gao YH, et al. Association between arsenic metabolism gene polymorphisms and arsenic-induced skin lesions in individuals exposed to high-dose inorganic arsenic in northwest China. Sci Rep. 2018; 8:12.
9. Hughes MF, Beck BD, Chen Y, et al. Arsenic Exposure and Toxicology: A Historical Perspective. Toxicol Sci. 2011; 123:305-32.
10. Ren X, McHale CM, Skibola CF, et al. An Emerging Role for Epigenetic Dysregulation in Arsenic Toxicity and Carcinogenesis. Environmental Health Perspectives. 2011; 119:11-19.
11. Chen J-W, Wang S-L, Wang Y-H, et al. Arsenic methylation, GSTO1 polymorphisms, and metabolic syndrome in an arseniasis endemic area of southwestern Taiwan. Chemosphere. 2012; 88:432-38.
12. Wang S-L, Chang F-H, Liou S-H, et al. Inorganic arsenic exposure and its relation to metabolic syndrome in an industrial area of Taiwan. Environment International. 2007; 33:805-11.
13. Liu R, Li H, Fan WJ, et al. Leucine Supplementation Differently Modulates Branched-Chain Amino Acid Catabolism, Mitochondrial Function and Metabolic Profiles at the Different Stage of Insulin Resistance in Rats on High-Fat Diet. Nutrients. 2017; 9:20.
14. Platten M, Wick W, Van den Eynde BJ. Tryptophan Catabolism in Cancer: Beyond IDO and Tryptophan Depletion. Cancer Res. 2012; 72:5435-40.
15. Wu G. Amino acids: metabolism, functions, and nutrition. Amino Acids. 2009; 37:1-17.
16. Casadei L, Valerio M, Manetti C. Metabolomics: Challenges and Opportunities in Systems Biology Studies. Methods Mol Biol. 2018; 1702:327-36.
17. Johnson CH, Gonzalez FJ. Challenges and opportunities of metabolomics. J Cell Physiol. 2012; 227:2975-81.
18. Hui AY, McCarty WJ, Masuda K, et al. A systems biology approach to synovial joint lubrication in health, injury, and disease. Wiley Interdiscip Rev Syst Biol Med. 2012; 4:15-37.
19. Lu K, Abo RP, Schlieper KA, et al. Arsenic Exposure Perturbs the Gut Microbiome and Its Metabolic Profile in Mice: An Integrated Metagenomics and Metabolomics Analysis. Environmental Health Perspectives. 2014; 122:284-91.
20. Zhang J, Shen H, Xu W, et al. Urinary metabolomics revealed arsenic internal dose-related metabolic alterations: a proof-of-concept study in a Chinese male cohort. Environ Sci Technol. 2014; 48:12265-74.
21. Guo X, Cui H, Zhang H, et al. Protective Effect of Folic Acid on Oxidative DNA Damage: A

Randomized, Double-Blind, and Placebo Controlled Clinical Trial. Medicine (Baltimore). 2015; 94:e1872.
22. Standard of diagnosis for endemic arsenism of China (WS/T211-2001). the Ministry of Health of the People's Republic of China. Beijing 2001.
23. Chen X, Guo X, He P, et al. Interactive Influence of N6AMT1 and As3MT Genetic Variations on Arsenic Metabolism in the Population of Inner Mongolia, China. Toxicol Sci. 2017; 155:124-34.
24. Peduzzi P, Concato J, Kemper E, et al. A simulation study of the number of events per variable in logistic regression analysis. J Clin Epidemiol. 1996; 49:1373-9.
25. Sun G, Xu Y, Li X, et al. Urinary arsenic metabolites in children and adults exposed to arsenic in drinking water in Inner Mongolia, China. Environmental Health Perspectives. 2007; 115:648-52.
26. Kuo CC, Moon KA, Wang SL, et al. The Association of Arsenic Metabolism with Cancer, Cardiovascular Disease, and Diabetes: A Systematic Review of the Epidemiological Evidence. Environmental Health Perspectives. 2017; 125:15.
27. Wang TJ, Larson MG, Vasan RS, et al. Metabolite profiles and the risk of developing diabetes. Nature Medicine. 2011; 17:448-U83.
28. Lawley SD, Cinderella M, Hall MN, et al. Mathematical model insights into arsenic detoxification. Theor Biol Med Model. 2011; 8.
29. Juhasz AL, Herde P, Herde C, et al. Validation of the predictive capabilities of the Sbrc-G in vitro assay for estimating arsenic relative bioavailability in contaminated soils. Environ Sci Technol. 2014; 48:12962-9.
30. Kibriya MG, Jasmine F, Parvez F, et al. Association between genome-wide copy number variation and arsenic-induced skin lesions: a prospective study. Environ Health. 2017; 16:13.
31. Priori R, Scrivo R, Brandt J, et al. Metabolomics in rheumatic diseases: the potential of an emerging methodology for improved patient diagnosis, prognosis, and treatment efficacy. Autoimmun Rev. 2013; 12:1022-30.
32. Wang X, Mu X, Zhang J, et al. Serum metabolomics reveals that arsenic exposure disrupted lipid and amino acid metabolism in rats: a step forward in understanding chronic arsenic toxicity. Metallomics Integrated Biometal Science. 2015; 7:544.
33. Weng L, Quinlivan E, Gong Y, et al. Association of Branched and Aromatic Amino Acids Levels with Metabolic Syndrome and Impaired Fasting Glucose in Hypertensive Patients. Metabolic Syndrome And Related Disorders. 2015; 13:195-202.
34. Zhou YY, Wang YF, Su J, et al. Integration of microRNAome, proteomics and metabolomics to analyze arsenic-induced malignant cell transformation. Oncotarget. 2017; 8:90879-96.
35. Lee JC, Chen MJ, Chang CH, et al. Plasma amino acid levels in patients with colorectal cancers and liver cirrhosis with hepatocellular carcinoma. Hepato-Gastroenterol. 2003; 50:1269-73.
36. Munn DH, Mellor AL. Indoleamine 2,3 dioxygenase and metabolic control of immune responses. Trends In Immunology. 2013; 34:137-43.
37. Pawlak K, Domaniewski T, Mysliwiec M, et al. The kynurenines are associated with oxidative stress, inflammation and the prevalence of cardiovascular disease in patients with end-stage renal disease. Atherosclerosis. 2009; 204:309-14.
38. Wu L-L, Gong W, Shen S-P, et al. Multiple metal exposures and their correlation with monoamine neurotransmitter metabolism in Chinese electroplating workers. Chemosphere. 2017; 182:745-52.
39. Zhou MM, Wu YM, Liu HY, et al. Effects of phenylalanine and threonine oligopeptides on milk protein synthesis in cultured bovine mammary epithelial cells. J Anim Physiol Anim Nutr. 2015; 99:215-20.
40. Chen L, Cheng CY, Choi H, et al. Plasma Metabonomic Profiling of Diabetic Retinopathy. Diabetes. 2016; 65:1099-108.
41. Piszcz J, Lemancewicz D, Dudzik D, et al. Differences and similarities between LC-MS derived serum fingerprints of patients with B-cell malignancies. Electrophoresis. 2013; 34:2857-64.

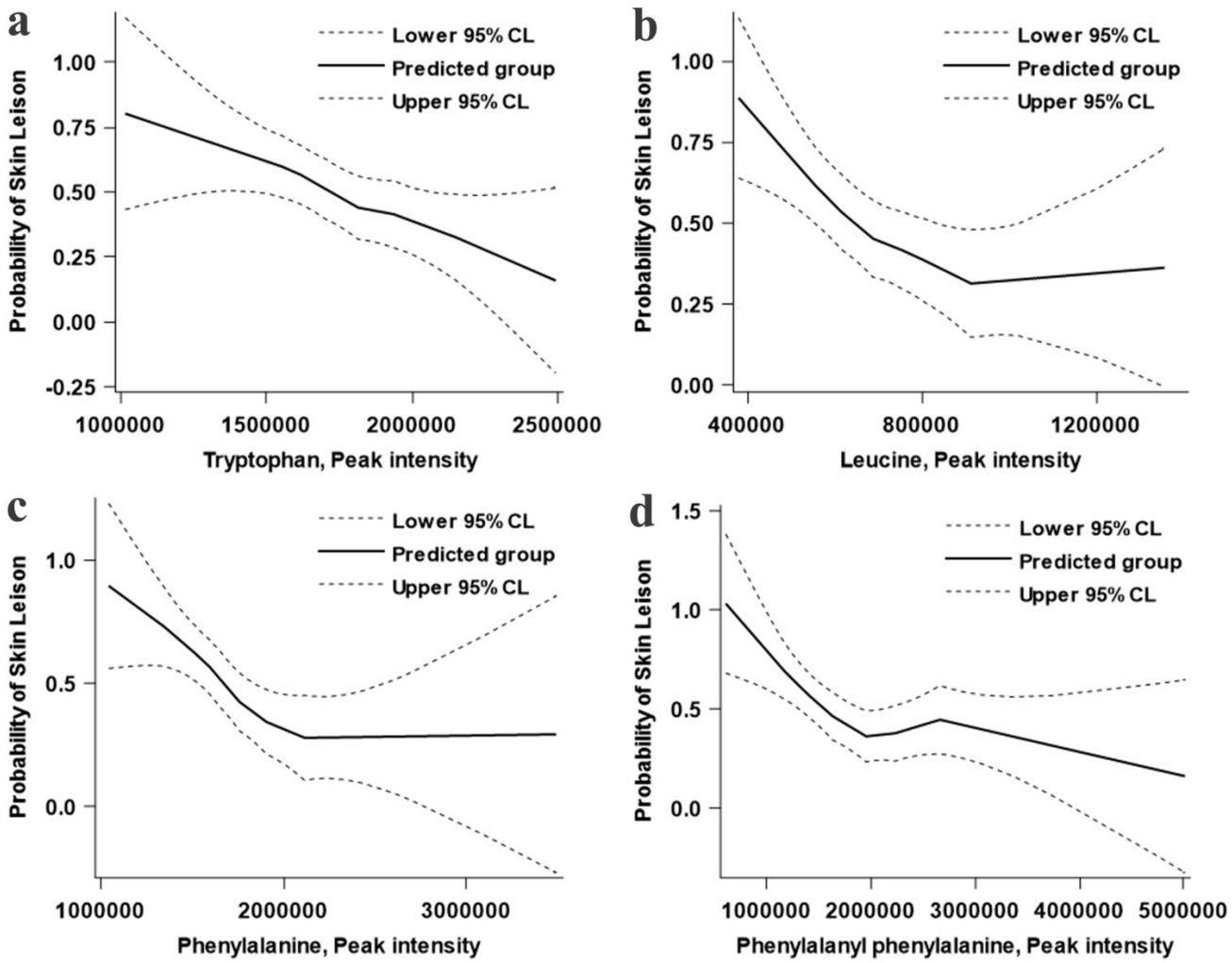

Figure 1. Association between the peak intensity of tryptophan and phenylalanine and arsenic-induced skin lesions based on multivariable locally weighted regression models. a: tryptophan; b: phenylalanine;c: leucine;d:phenylalanyl phenylalanine

SAS 系统

The POWER Procedure
Paired t Test for Mean Difference

Fixed Scenario Elements	
Distribution	Normal
Method	Exact
Mean 1	1832402
Mean 2	1655560
Standard Deviation 1	306551.7
Standard Deviation 2	289730.1
Correlation	0.4
Number of Pairs	56
Number of Sides	2
Null Difference	0
Alpha	0.05

Computed Power
Power
0.978

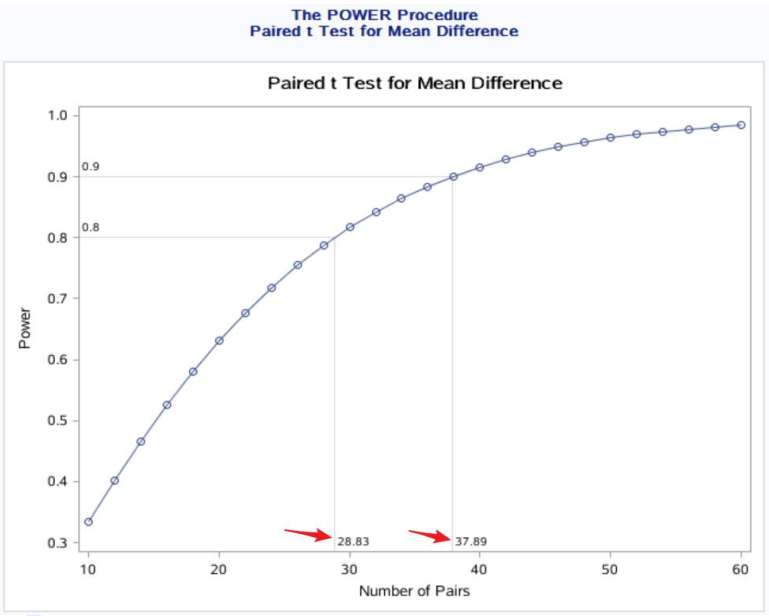

Figure 1S_A. Sample size and power estimation__Tryptophan

SAS 系统

The POWER Procedure
Paired t Test for Mean Difference

Fixed Scenario Elements	
Distribution	Normal
Method	Exact
Mean 1	768959
Mean 2	656243.1
Standard Deviation 1	206347.3
Standard Deviation 2	216983.4
Correlation	0.4
Number of Pairs	56
Number of Sides	2
Null Difference	0
Alpha	0.05

Computed Power
Power
0.946

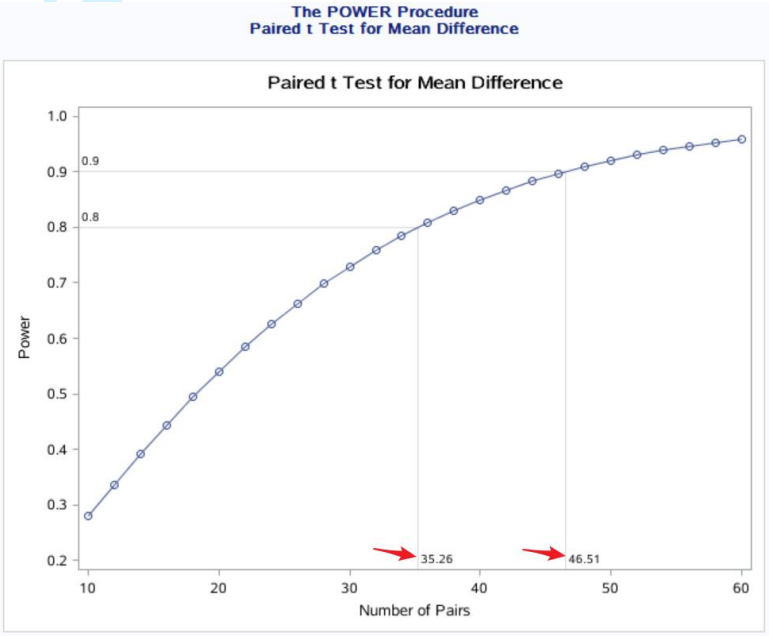

Figure 1S_B. Sample size and power estimation__Leucinev

SAS 系统

The POWER Procedure
Paired t Test for Mean Difference

Fixed Scenario Elements	
Distribution	Normal
Method	Exact
Mean 1	1865229
Mean 2	1637107
Standard Deviation 1	420123.2
Standard Deviation 2	356546.1
Correlation	0.4
Number of Pairs	56
Number of Sides	2
Null Difference	0
Alpha	0.05

Computed Power
Power
0.975

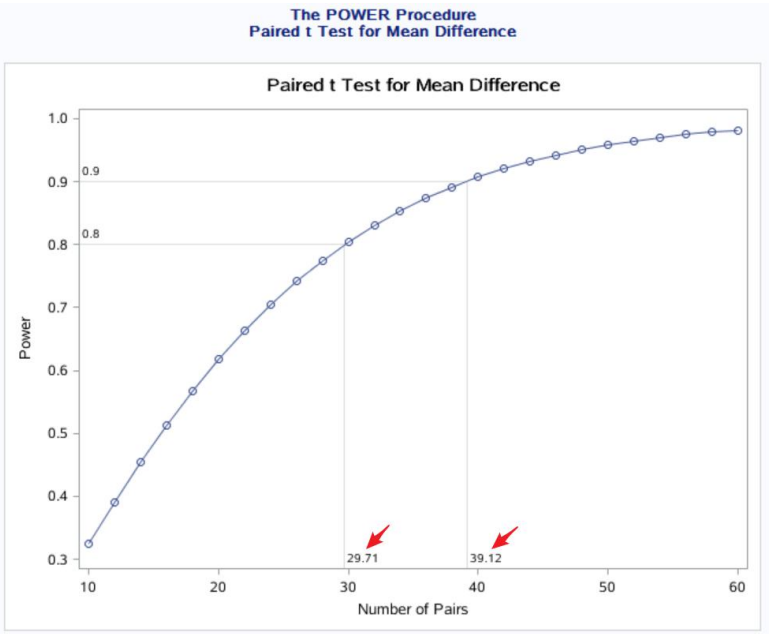

Figure 1S_C. Sample size and power estimation__Phenylalanine

SAS 系统

The POWER Procedure
Paired t Test for Mean Difference

Fixed Scenario Elements	
Distribution	Normal
Method	Exact
Mean 1	2056703
Mean 2	1664489
Standard Deviation 1	880518
Standard Deviation 2	682913.7
Correlation	0.4
Number of Pairs	56
Number of Sides	2
Null Difference	0
Alpha	0.05

Computed Power
Power
0.911

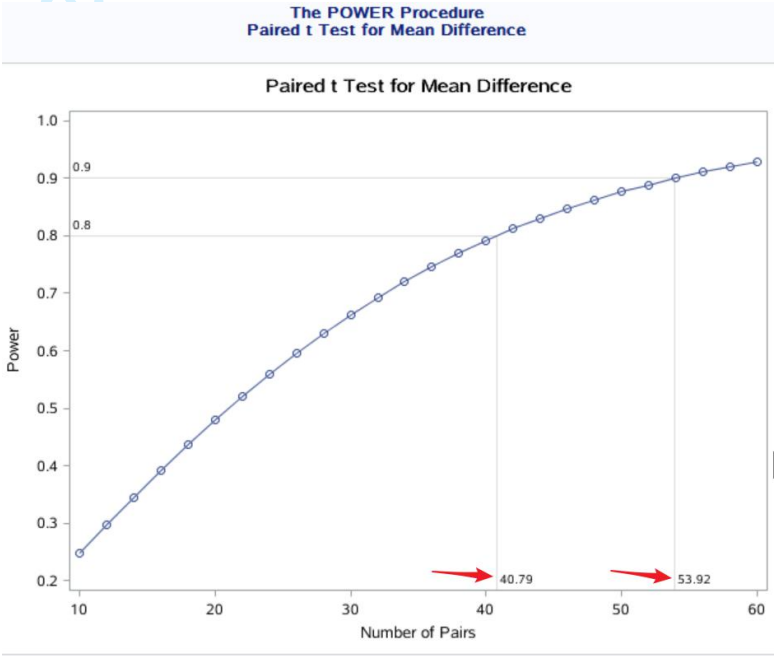

Figure 1S_D. Sample size and power estimation__Phenylalanyl phenylalanine

Table S1 Relationship of the levels of serum amino acids with arsenic-induced skin lesions[ξ].

Metabolite	OR (95% CI) in basic model	+Exposure year	+Body mass index	+Fasting plasma glucose
Tryptophan	0.48(0.27,0.84)*	0.48(0.26,0.85)*	0.47(0.27,0.84)#	0.49(0.28,0.87)*
Phenylalanine	0.57(0.36,0.91)*	0.57(0.36,0.92)*	0.57(0.35,0.91)*	0.57(0.35,0.93)*
Leucine	0.45(0.25,0.82)#	0.45(0.25,0.83)#	0.45(0.25,0.83)#	0.42(0.22,0.81)#
Phenylalanyl phenylalanine	0.62(0.36,1.04)	0.60(0.35,1.02)	0.60(0.35,1.02)	0.65(0.38,1.12)

[ξ] OR: odds ratio; CI: confidence interval. The table are presented as OR (95% CI);
* P-value<0.05; [#]P-value<0.01.

Table S2 Relationship of the levels of serum amino acids with arsenic-induced skin lesions[ξ].

Metabolite	Odds Ratios (95% CI) in Basic Model	+Serum folate	+Blood urea nitrogen	+smoking	+drinking
Tryptophan	0.48(0.27,0.84)*	0.47(0.27,0.84)*	0.47(0.27,0.84)*	0.47(0.26,0.84)*	0.49(0.26,0.87)*
Phenylalanine	0.57(0.36,0.91)*	0.59(0.37,0.94)*	0.57(0.36,0.91)*	0.56(0.35,0.90)*	0.55(0.34,0.89)*
Leucine	0.45(0.25,0.82)#	0.46(0.25,0.84)*	0.45(0.25,0.83)#	0.45(0.25,0.83)#	0.42(0.23,0.78)#
Phenylalanyl phenylalanine	0.62(0.36,1.04)	0.63(0.37,1.07)	0.61(0.36,1.04)	0.62(0.36,1.05)	0.63(0.37,1.07)

[ξ] OR: odds ratio; CI: confidence interval. The table are presented as OR (95% CI);
* P-value<0.05; [#]P-value<0.01.

Table S3 Relationship of the levels of serum amino acids with arsenic-induced skin lesions[ξ].

Amino acids	OR (95% CI) in basic model	+TC	+TG	+HDL	+LDL
Tryptophan	0.48(0.27,0.84)*	0.48(0.27,0.85)*	0.46(0.26,0.82)#	0.48(0.27,0.86)*	0.48(0.27,0.84)*
Phenylalanine	0.57(0.36,0.91)*	0.57(0.35,0.91)*	0.57(0.36,0.91)*	0.55(0.34,0.89)*	0.57(0.36,0.92)*
Leucine	0.45(0.25,0.82)#	0.44(0.24,0.81)#	0.45(0.25,0.84)*	0.45(0.25,0.83)*	0.44(0.24,0.83)*
Phenylalanyl phenylalanine	0.62(0.36,1.04)	0.62(0.37,1.05)	0.63(0.37,1.08)	0.61(0.35,1.05)	0.65(0.38,1.11)

[ξ] OR: odds ratio; CI: confidence interval. Values are odds ratio (95% confidence intervals); TC: Total cholesterol;TG:Triglycerides; LDL:Low-density lipoprotein ; HDL:High-density lipoprotein; OR: odds ratio; CI: confidence interval.
* P-value<0.05; [#]P-value<0.01.

Table S4 Relationship of the levels of serum amino acids with arsenic-induced skin lesions[ζ].

Amino acids	OR (95% CI) in basic model	+iAs%	+MMA%	+DMA%	+tAS
Tryptophan	0.48(0.27,0.84)*	0.47(0.27,0.84)*	0.49(0.28,0.85)*	0.48(0.28,0.84)*	0.47(0.27,0.83)#
Phenylalanine	0.57(0.36,0.91)*	0.59(0.36,0.95)*	0.57(0.35,0.91)*	0.59(0.36,0.95)*	0.52(0.32,0.87)*
Leucine	0.45(0.25,0.82)#	0.46(0.25,0.84)*	0.46(0.25,0.86)*	0.47(0.25,0.88)*	0.43(0.24,0.80)#
Phenylalanyl phenylalanine	0.62(0.36,1.04)	0.62(0.36,1.05)	0.64(0.37,1.09)	0.64(0.38,1.10)	0.61(0.36,1.03)

[ζ] OR: odds ratio; CI: confidence interval. iAS: inorganic arsenic (iAs^{III}+iAs^{V}); MMA: monomethyl arsenate (MMA^{III}+MMA^{V}); DMA: dimethyl arsenate (DMA^{III}+DMA^{V}); tAs: total arsenic (iAs^{III}+iAs^{V}+MMA+DMA); iAs%= iAS/tAs*100%; MMA%=MMA/tAs*100% and DMA%=DMA/tAs*100%;
* P-value<0.05; [#]P-value<0.01.

Table S5 Variance inflation factor of amino acids in different models[ξ].

Amino acids	Model 1	Model 2	Model 3	Model 4
Tryptophan	1.04	1.04	1.04	1.04
Phenylalanine	4.56		1.50	1.04
Leucine	4.21	1.39		
Phenylalanyl Phenylalanine	1.49	1.38	1.48	

[ξ]Model 1: Tryptophan, Phenylalanine, Leucine and Phenylalanyl Phenylalanine;
Model 2: Tryptophan, Leucine and Phenylalanyl Phenylalanine;
Model 3: Tryptophan, Phenylalanine and Phenylalanyl Phenylalanine;
Model 4: Tryptophan and Phenylalanine.

BMJ Open

Tryptophan, phenylalanine and arsenic-induced skin lesions in a chronic arsenic exposure Chinese population via drinking water: data from a metabolomics study

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2018-025336.R2
Article Type:	Original research
Date Submitted by the Author:	28-Feb-2019
Complete List of Authors:	wei, yaping; Wenzhou Medical University, Chashan University town, Ouhai district, Wenzhou, China., Jia, Chaonan; School of Public Health & Management, Wenzhou Medical University, Department of Preventive Medicine Lan, Yuan; School of Ophthalmology & Optometry, Wenzhou Medical University, Center on Clinical Research, the Eye Hospital of Wenzhou Medical University Hou, Xiangqing; School of Public Health & Management, Wenzhou Medical Univeristy, Department of Preventive Medicine Zuo, Jingjing; Wenzhou Medical University School of Optometry and Ophthalmology, Center on Clinical Research, the Eye Hospital of Wenzhou Medical University Li, Jushuang; School of Public Health & Management, Wenzhou Medical Univeristy, Department of Preventive Medicine Wang, Tao; School of Public Health & Management, Wenzhou Medical University, Department of Preventive Medicine Mao, Guangyun; School of Public Health & Management, Preventive Medicine
Primary Subject Heading:	Occupational and environmental medicine
Secondary Subject Heading:	Epidemiology, Public health, Occupational and environmental medicine
Keywords:	Metabolomics, Chronic arsenic exposure, Skin lesions, Amino acid, UPLC-MS/MS

Tryptophan, phenylalanine and arsenic-induced skin lesions in a chronic arsenic exposure Chinese population via drinking water: data from a metabolomics study

Yaping Wei[1,2], Chaonan Jia,[1,2], Yuan Lan[3], Xiangqing Hou[1,2], Jingjing Zuo[3], Jushuang Li[1,2], Tao Wang[1,2], Guangyun Mao[1,2,4]✉

[1] Department of Preventive Medicine, School of Public Health and Management, Wenzhou Medical University, Wenzhou, China.

[2] Center on Evidence-Based Medicine & Clinical Epidemiological Research, School of Public Health, Wenzhou Medical University, Wenzhou, China.

[3] School of Ophthalmology & Optometry, Wenzhou Medical University, Wenzhou, China.

[4] Center on Clinical Research, the Affiliated Eye Hospital of Wenzhou Medical University, Wenzhou, China.

Yaping Wei and Chaonan Jia contributed equally to this paper.

✉Corresponding author: Guangyun Mao, Room 7A321, School of Public Health and Management， Wenzhou Medical University, Chashan University town, Ouhai district, Wenzhou, China. 325035

E-mail: maogy@wmu.edu.cn

Telephone: +86-577-86689848

Fax numbers: +86-577-86699122

ABSTRACT:

Objectives To investigate the association of specific serum amino acids (AAs) with the odds of arsenic-induced skin lesions (AISL) and their ability to distinguish AISL from the counterparts.

Design Case-control study.

Setting Three arsenic exposed villages in Wuyuan county of Hetao Plain, Inner Mongolia, China.

Participants Among 450 residents aged 18 to 79 years chronically exposed to arsenic via drinking water, 56 of them were diagnosed as AISL and defined as the cases. Another 56 participants without AISL matched by gender and similar age (±1 year) from the same population were picked out as the controls. The inclusion criteria were subjects having the metabolomics determination. Unmatched participants and those without serum metabolites data were excluded.

Primary and secondary outcome measures The outcome was whether it suffered from AISL. Multivariable conditional logistic regression models and receiver operating characteristic curve (ROC) analysis were performed to investigate the relationship between specific AAs and AISL.

Results The levels of tryptophan and phenylalanine were both negatively associated with AISL (P<0.05). As compared to the 1st quartile, the adjusted odds of AISL in the 2nd, 3rd and 4th quartile decreased by 69%, 90% and 84% for tryptophan, and 14%, 80% and 76% for phenylalanine, respectively. The combination of the two aforementioned higher-level AAs revealed the lowest odds of AISL (OR=0.06; 95%CI: 0.02, 0.22; P<0.001)). Furthermore, both AAs showed moderate ability to distinguish AISL from the control, with area-under-curve [(AUC), 95%CI] as 0.67 (0.57, 0.77) for tryptophan and 0.70 (0.60, 0.80) for phenylalanine, respectively (all P<0.05). The combined pattern

with AUC (95%CI) was 0.72 (0.62, 0.81), sensitivity of 76.79% and specificity of 58.93% (P<0.001).

Conclusions Specific AAs might be linked to AISL and play an important role in its early identification. Additional studies are needed to confirm our findings.

Keywords: Metabolomics; Chronic arsenic exposure; Skin lesions; Amino acid; UPLC-MS/MS.

Strengths and limitations of this study

- Our findings were depended on a community-based metabolomics study with paired-design, strictly quality assurance and quality control.
- Multivariable conditional logistic models were performed to examine the association between specific amino acid levels and AISL, and ROC analysis was applied to evaluate the value and feasibility of the AA to distinguish AISL from the counterparts.
- Although the AAs were determined by untargeted metabolomics approach, which can assess a large number of metabolites precisely and efficiently, we only can obtain relative levels of AAs instead of their accurate quantitative concentration.
- The findings were based on a case-control study, which only revealed the association between the AAs and the odds of AISL rather than confirming their causal relationship.
- The participants were mainly chronically exposed to arsenic via drinking water, which would limit the findings extrapolated to another arsenic exposure population via food or other ways.

INTRODUCTION

Chronic arsenic exposure via drinking water is widely believed as a global health concern, affecting a large amount of people worldwide. It may give rise to several human health issues and has been documented to associate with cardiovascular disease, diabetes, cancer and others[1 2]. With the industrial boom and dramatic rise of worldwide water pollution including arsenic contamination in the past, the prevalence and burden of arsenic-induced health damage will continue to increase. Skin has been confirmed as one of the most common and susceptible target of arsenic-induced health lesions. Cutaneous skin lesions are typical signs of arsenicosis after persistent arsenic exposure for a long term which are characterized by hyperkeratosis and hyperpigmentation. Considerable evidences of the prevalence of arsenical skin lesions had been observed in many countries[3-5].

As arsenic-induced skin lesions (AISL) have been widely accepted as the major early manifestation of arsenic toxicity[6] and might be indicators of susceptibility to more serious arsenic-induced health hazards[7], it is particularly crucial to identify participants at risk as early as possible for preventing the onset or delaying the progression of the serious health problems effectively. Several possible mechanisms such as genetic differences[8], oxidative stress[9] and epigenetic dysregulation[10] and others may explain arsenic poisoning. Previous studies also reported that arsenic methylation in vivo might be associated with metabolic syndrome[11 12].

Amino acids (AAs) are the "basic unit" that make up the body's various proteins and necessary to maintain the health. Some AAs are important regulators of some key metabolic pathways and have great importance in maximizing efficiency of food utilization, enhancing protein accretion and health improvement[13 14]. Abnormal metabolism of AAs will disturb the homeostasis of the body, impairs growth and development, and even causes death[15]. So, the levels of serum AAs may be an important implication for the metabolic status and disease condition. As a powerful tool in system biology research, metabolomics approach is beneficial on unbiased monitoring changes in endogenous metabolism-related physiological processes, providing integrative information on the distinctive

features across multiple functional levels, and offering a window to capture the core attributes responsible for various phenotypes, which are particularly important in understanding the relevant pathophysiological changes of a disease and its status, identifying novel biomarkers for risk screening, diagnosis, treatment and prognosis of important human diseases[16-18].

Animal experiments and epidemiological study have reported obvious arsenic-related metabolomics perturbations[19 20]. All of these researches substantially suggests that the relationship between specific metabolites and arsenic-induced health lesions should be investigated. However, few works have been conducted to comprehensively examine the metabolic mechanism relevant to AISL, especially for the AAs metabolism. The present study aims to quantitatively examine the association of several specific AAs with AISL and the ability to identify AISL.

METHODS

Patient and Public Involvement

No patients were involved.

Study Population

This study was originally from a randomized, double-blind, and placebo controlled clinical trial (NCT02235948) in 2010, in which all subjects were randomly selected using permuted block randomization from a single rural area in a population chronically exposed to low-level arsenic drinking water, had similar life style and influences under similar environmental factors. Information on the inclusion and exclusion criteria of the participants could be found in our previous study[21]. Strictly following the criteria of arsenicosis[22]. AISL was diagnosed as the presence of arsenic- induced keratosis, hyperpigmentation or depigmentation by a physician from Wenzhou medical university at the beginning of the trial. This was a matched case-control study (1:1 matching). Among 450 residents aged 18 to 79 years old enrolled in the above-mentioned trial, 56 of them were diagnosed as AISL and selected as the case. Another 56 participants without AISL matched by gender and similar age (±1

year) from the same population were picked out as the control. The inclusion criteria were subjects having the metabolomic test. Unmatched participants and those without serum metabolites data were excluded. Informed consent was obtained from all participants and this study was approved by the ethics committee of Wenzhou Medical University, Wenzhou, China.

Data Collection and Assessment

The information on age, gender, exposure year, body mass index, smoking, alcohol consumption, education level, etc. was collected with a standardized questionnaire. Blood and urine samples were also collected at the time of participants' enrollment. Detailed data collection of blood and urine samples and assessment methods for clinical variables including plasma fasting glucose (FPG), serum urea nitrogen, serum folate, total homocysteine, total cholesterol (TC), triglycerides (TG), high-density lipoprotein (HDL), Low-density lipoprotein (LDL) and others had been published previously[21]. The epuration of various urinary arsenic species were conducted by means of a high-performance liquid chromatography coupled mass spectrometer system for separation and detection[23]. The species of arsenic in urine samples consisted of inorganic arsenic (iAs, [$iAs^{Ⅲ}$ plus $iAs^{Ⅴ}$]), monomethyl arsenate (MMA, [$MMA^{Ⅲ}$ plus $MMA^{Ⅴ}$]) and dimethyl arsenate (DMA, [$DMA^{Ⅲ}$ plus $DMA^{Ⅴ}$]). All arsenic species were corrected by creatinine. The total arsenic (tAs) was the sum of iAs, MMA and DMA. The percentages of arsenic species were defined as: iAs%=iAs/tAs*100%, MMA% =MMA/tAs*100% and DMA%=DMA/tAs*100%, respectively.

UPLC-MS/MS Metabonomic Profiling

Serum samples（200 μL in microcentrifuge tubes) were thawed to room temperature (25°C) and 600 μL mixture (90% acetonitrile - 10%water) were added to each sample. The samples were vigorously mixed for 20 seconds and centrifuged for 5 min at 12000 rpm (20°C). The top 400 μL of each supernatant were then transferred and dried down in a vacuum concentrator centrifuge. The dried samples were re-suspended in 130 μL of water (including 15% acetonitrile), mixed vigorously for 20 seconds and repeated the centrifugation method described above. Two μL of the supernatant were

collected as samples to be determined. Serum metabolic profile acquisition was performed by using ACQUITY UPLC®/Xevo® G2 QTof/MS^E (Waters Corp., Milford, MA, USA). Chromatographic separation was performed at 50°C using a WATERS HSS T3 column (2.1×100 mm, 1.7 μm) with a flow rate of 0.4mL/min. The mobile phase was a mixture of (A) H_2O with 0.1% formic acid and (B) methanol with 0.1% formic acid. Elution was in linear gradient with the programmed gradient at 0 min with 100% A and 0% B, 1.00min with 100%A and 0% B, 8 min with 0%A and 100% B, 13.00 min with 0% A and 100% B. The mass spectrometer was operated under both positive-ion (ESI^+) mode and negative-ion (ESI^-) mode electrospray ionization. The scan range was from 50 to 1200 m/z. Data was collected in both ESI^+ and ESI^- modes. Capillary voltage was set at 3000 V and 2500 V, respectively. The desolvation flow rate was 800 L/h at 350°C. Argon was used as a collision gas, and the collision energy was adjusted from 10 eV to 40 eV for each analysis. Quantum clustering (QC) samples were prepared by pooling aliquots of each sample and used to reflect the reliability of further metabolomics analysis. After peak deconvolution, alignment, integration and normalization, the data including retention time (RT), mass to charge ratio(m/z), and peak intensity were extracted from raw chromatograms using Progenesis QI 2.0 (Waters Corp., Milford, MA, USA). The MS/MS mode was performed to obtain metabolites levels processed with MarkerLynx Applications Manager Version 4.1 (Waters Corp., Milford, MA, USA).

Distinct Metabolites Identification

The peak intensity of metabolites for the 56 pairs were acquired and then imported to MetaboAnalyst 4.0 (http://www.metaboanalyst.ca/) for statistical analysis. A partial least-squares discriminant analysis (PLS-DA), which is a supervised and well accepted pattern recognition approach, was used for the differentiation between the cases and controls. False discovery rate (FDR) adjusted p-value in univariate analysis were performed to reduce the potential impact induced by false positive of the results. The criteria used in the selection of metabolites include variable importance in the project (VIP) scores >1 in PLS-DA and the crude or FDR adjusted p-value all < 0.05 in Wilcoxon signed rank

test. We identified a total of 70 extracted small molecular metabolites that were linked to the recognition of AISL. Among them, four amino acid metabolites (Phenylalanine, Tryptophan, Leucine, Phenylalanylphenylalanine) were identified.

Statistical Analysis

The normality of continuous data was assessed using both QQ-plots and Shapiro-Wilk test. The comparison between the cases and controls was performed with the paired *t*-test if they met normal or similar normal distribution. Otherwise, Wilcoxon signed rank test would be used. Differences in the proportion of categorical variables between the two groups were evaluated by McNemar-Bowker tests. We firstly used locally weighted scatterplot smoothing (LOESS) models to estimate the "real" relationship between serum AAs levels and the odds of AISL. Then, multivariable conditional logistic regression models were performed to examine the association between the contributing AAs levels and AISL after adjusting for some potential confounding factors. The individual impacts of AA metabolites on the risk of AISL were quantified separately by odds ratio (OR) and 95% confidence interval (CI) in the following two ways: with AA as a categorical variable (quartiles) and as a continuous variable [scaled to interquartile range (IQR)]. Variables with p-value less than 0.2 in the comparison between two groups were selected as potential confounders because the sample size of the current study was not too large. Then variance inflation factor (VIF) was used to examine the potential collinarity among them. As too many covariates in a multiple regression model will lead to overfitting to some extent[24], we finally select no more than ~~5~~ 4 variables as confounding factors to decrease the potential overfitting when assessing the association between AAs and AISL. Furthermore, as the distinct metabolites might be high related to each other, collinearity should be well considered. So, we used the variance inflation factor (VIF) based on the VIF package of R software to detect potential collinearity among the AAs. When the VIF is greater than 1.5, it was considered as collinearity existed in the model and the associated variable would be removed. The combined effect of relevant AAs on AISL were also performed using a multivariable logistic regression model. In addition, receiver

operator characteristic (ROC) analysis was applied to evaluate the value and feasibility of the AAs as the potential sensitive and specific biomarker to recognize AISL. Data management, analysis and figure drawing were finished using R version 3.4.4 (Copyright © 2018 The R Foundation for Statistical Computing). All tests were two-sides and P<=0.05 was set as significant level.

RESULTS

Table 1 summarizes the general characteristics of the study population. The comparison of demographical, clinical features and urinary arsenic species in the 56 pairs of subjects were presented in Table 1. The median (1st quartile, 3rd quartile) age of AISL population was 50.30 (44.70, 58.70) for the cases and 50.40 (44.60, 58.70) years for the controls. Both groups have the same proportion of female population (58.93%), and there was no obvious statistical difference in the urinary arsenic levels between the two groups. More than half of them had no history of smoking or alcohol consumption. When compared to the controls, the serum triglycerides level in AISL participants was significantly lower (P=0.041). While the other variables were similar between AISL participants and the control (P>0.05). This indicates that the participants in two groups are comparable to some extent.

Table 1 Demographic characteristics of the study population[ξ].

Variables	AISL (n=56)	Non-AISL (n=56)	P
Clinical Characteristics			
Age (years)	50.30(44.70,58.70)	50.40(44.60,58.70)	0.425
Exposure year (years)	48.19±11.53	47.62±10.97	0.489
Body mass index (kg/m^2)	24.12±3.14	23.91±2.86	0.697
Fasting plasma glucose (mmol/L)	4.89(4.60,5.25)	5.12(4.53,5.40)	0.137
Folate (ng/mL)	4.00(3.20,5.10)	4.25(3.35,5.40)	0.392
Total homocysteine (μmol/L)	12.30(10.32,16.50)	12.67(11.21,14.69)	0.961
Blood urea nitrogen (mmol/L)	6.45(5.42,7.69)	6.84(5.36,8.80)	0.603
Total cholesterol (mmol/L)	4.58(4.10,5.69)	4.65(3.96,5.95)	0.904
Triglycerides (mmol/L)	1.41(0.90,1.74)	1.45(1.09,2.29)	0.041
High-density lipoprotein (mmol/L)	1.19±0.34	1.16±0.31	0.675
Low-density lipoprotein (mmol/L)	3.04±0.80	3.25±0.84	0.110
Women [# (%)]	33(58.93)	33(58.93)	1.000
Cigarette smoking [# (%)]	20(35.71)	22(39.29)	0.696
Alcohol consumption [# (%)]	17(30.91)	21(37.50)	0.464
Illiteracy [# (%)]	21(37.50)	15(25.00)	0.252
Urinary arsenic species[ζ]			
iAS%	12.26(8.13,14.68)	12.31(10.04,16.54)	0.148
MMA%	24.68(20.11,29.68)	25.85(20.90,31.66)	0.420
DMA%	61.84(56.62,71.01)	61.84(47.85,64.99)	0.096
tAs (μg/g creatinine)	140.93(104.41,208.53)	186.77(80.11,217.30)	0.445

[ξ] AISL: arsenic-induced skin lesions; the variables met normal distribution was described with mean± standard deviation; otherwise, median (1st quartile, 3rd quartile) was used to describe their features. Number of cases (percentage) was used to describe the proportion of categorical variables between the two groups.

*[ζ] iAS: inorganic arsenic (iAs^{III}+iAs^{V}); MMA: monomethyl arsenate (MMA^{III}+MMA^{V}); DMA: dimethyl arsenate (DMA^{III}+DMA^{V}); tAs: total arsenic (iAs^{III}+iAs^{V}+MMA+DMA); iAs%= iAS/tAs*100%; MMA%=MMA/tAs*100% and DMA%=DMA/tAs*100%.*

Table 2 shows that the four AAs, which FDR adjusted p-value <0.05 and VIP>1, in the cases are observed significantly lower than those of the controls. Two of them are aromatic amino acids (AAA) identified as phenylalanine and tryptophan, one of them belongs to aromatic amino acids branched-chain amino acids (BCAA) appraised as leucine and the last one is phenylalanylphenylalanine. The individual association of AAs with AISL were presented in figure1, which clearly reveals obvious "dose-response" relationships between them.

Table 2 Distinct metabolites in population with arsenic-induced skin lesions and their counterparts.

Serum amino acid metabolites	Retention time (min)	Mass-to-Charge Ratio	VIP value	p-value$^{\xi}$	Adjusted p-values$^{\zeta}$
Phenylalanine	3.402	166.087	1.508	<0.001	0.009
Tryptophan	3.886	203.082	1.046	0.003	0.014
Leucine	2.642	132.102	1.014	0.001	0.020
Phenylalanylphenylalanine	5.048	313.155	1.833	0.004	0.033

VIP: variable importance in the project; ξ Wilcoxon signed-rank test; ζ Adjusted by false discovery rate (FDR).

Table 3 clearly shows that participants in the 3rd and 4th quartiles of the four specific AAs were all significantly linked to the decreased odds of AISL after adjusting for FPG, LDL, TG and DMA%, as compared to their lowest quartiles, respectively. The category boundaries of the quartiles were showed in the Table S1. Significant linear trends existed between AISL and those four serum AAs. Meanwhile, same linear negative association between AISL and per IQR rise of the four serum AAs were observed when these AAs were considered as continuous variables in the present study.

Table 3 Relationship of amino acids levels with the odds of arsenic-induced skin lesions[ξ]

Amino acids	N	Cases (%)	Crude		Adjusted[ζ]	
			OR (95%CI)	P	OR (95%CI)	P
Tryptophan						
Per IQR	112	56(50)	0.48(0.27,0.84)	0.011	0.48(0.27,0.86)	0.013
Quartiles						
Q_1	28	20(71.40)	1.00(1.00,1.00)	Ref.	1.00(1.00,1.00)	Ref.
Q_2	28	15(53.60)	0.50(0.16,1.54)	0.225	0.56(0.16,1.98)	0.370
Q_3	28	10(35.70)	0.12(0.03,0.53)	0.005	0.12(0.02,0.60)	0.010
Q_4	28	11(39.30)	0.19(0.05,0.71)	0.014	0.21(0.05,0.84)	0.028
P for trend				0.008		0.012
Phenylalanine						
Per IQR	112	56(50)	0.57(0.36,0.91)	0.019	0.56(0.33,0.94)	0.028
Quartiles						
Q_1	28	20(71.40)	1.00(1.00,1.00)	Ref.	1.00(1.00,1.00)	Ref.
Q_2	28	19(67.90)	0.79(0.22,2.82)	0.712	0.70(0.18,2.76)	0.609
Q_3	28	8(28.60)	0.18(0.05,0.66)	0.010	0.20(0.05,0.77)	0.019
Q_4	28	9(32.10)	0.25(0.08,0.79)	0.018	0.20(0.05,0.75)	0.017
P for trend				<0.001		0.001
Leucine						
Per IQR	112	56(50)	0.45(0.25,0.82)	0.019	0.43(0.21,0.86)	0.016
Quartiles						
Q_1	28	21(75.00)	1.00(1.00,1.00)	Ref.	1.00(1.00,1.00)	Ref.
Q_2	28	14(50.00)	0.33(0.11,1.03)	0.057	0.31(0.10,1.01)	0.052
Q_3	28	11(39.30)	0.22(0.07,0.68)	0.009	0.22(0.07,0.73)	0.014
Q_4	28	10(35.70)	0.19(0.06,0.59)	0.004	0.19(0.06,0.65)	0.008
P for trend				0.003		0.007
Phenylalanylphenylalanine						
Per IQR	112	56(50)	0.62(0.36,1.04)	0.070	0.71(0.41,1.24)	0.227
Quartiles						
Q_1	27	21(77.80)	1.00(1.00,1.00)	Ref.	1.00(1.00,1.00)	Ref.
Q_2	29	14(48.30)	0.16(0.03,0.75)	0.021	0.14(0.03,0.73)	0.019
Q_3	28	9(32.10)	0.08(0.02,0.39)	0.002	0.09(0.02,0.52)	0.007
Q_4	28	12(42.90)	0.12(0.02,0.57)	0.008	0.11(0.02,0.66)	0.016
P for trend				0.006		0.023

[ξ]Values are odds ratio (95% confidence intervals) for arsenic-induced skin lesions from conditional logistic regression. IQR: interquartile range; Q_1: the 1st quartile; Q_2: the 2nd quartile; Q_3: the 3rd quartile; Q_4: the 4th quartile.

[ζ] Adjusted for plasma glucose, low-density lipoprotein, triglyceride and urinary and dimethyl arsenate.

As these 4 specific AAs are significantly or marginal significantly associated with the odds of AISL, so it is needed to examine the joint impacts among them on AISL. However, the results of potential collinearity examination revealed that among these 4 specific AAs, both tryptophan and phenylalaine had the smallest VIF value (VIF=1.04) and no obvious collinearity existed (**Table S2**). Hence, we mainly focus on tryptophan and phenylalaine when assessing the joint impacts of AAs on AISL and only presented the results associated with these two AAs in the current study. To avoid the impacts due to insufficient power because of unreasonable grouping on the results, we classified both tryptophan and phenylalanine into two categories, according to the cut-off values of their mass

spectrum peak area based on the ROC analysis, respectively. The higher levels of these two serum AA were defined as equal to or over the cut-off values, while the lower categories were considered as less than the associated values.

Table 4 shows the joint impacts of tryptophan and phenylalanine levels on AISL after considering the collinearity of variables in the model. The proportions of AISL were 74.3%, 60.0%, 50.0% and 18.2% for participants with lower levels of both tryptophan and phenylalanine (category A), with higher tryptophan and lower phenylalanine (category B), with lower tryptophan and higher phenylalanine (category C), and higher levels of both tryptophan and phenylalanine (category D), respectively. Obvious decrease trend of the probability of AISL was observed among these 4 categories. As compared to the category A, adjusted OR (95% CI) for participants in the category B, C and D were 0.49(0.15, 1.63), 0.32(0.10, 1.02) and 0.08(0.02, 0.25). Subjects with higher levels of both tryptophan and phenylalanine had the lowest odds of AISL, significantly decreased by 92% (OR=0.02; 95%CI: 0.02, 0.25; P<0.001), after adjusting for the impacts induced by some potential confounding factors. This suggested that tryptophan and phenylalanine were jointly associated with the presence of AISL. While no significant interaction between the two AAs on the occurrence of AISL could be observed (P=0.419), which indicated that each AA was independently associated with AISL though their joint impact was significant.

Table 4 Joint association between tryptophan and phenylalanine levels with arsenic-induced skin lesions.

Tryptophan <cut-off value[ξ]	Phenylalanine <cut-off value[ξ]	N	Cases (%)	Crude		Adjusted[ζ]	
				OR (95%CI)	P	OR (95%CI)	P
Yes	Yes	35	26(74.3)	1.00(1.00,1.00)	Ref.	1.00(1.00,1.00)	Ref.
No	Yes	20	12(60.0)	0.52(0.16,1.68)	0.273	0.49(0.15,1.63)	0.244
Yes	No	24	12(50.0)	0.35(0.12,1.04)	0.059	0.32(0.10,1.02)	0.053
No	No	33	6(18.2)	0.08(0.02,0.25)	<0.001	0.08(0.02,0.25)	<0.001
Interaction					0.320	0.49(0.09,2.78)	0.419

[ξ] Cut-off value was determined by means of receiver operator characteristic analysis.
[ζ] Adjusted for plasma glucose, low-density lipoprotein, triglyceride and urinary and dimethyl arsenate.

Table 5 shows that, based on the ROC analysis, both serum tryptophan and phenylalanine might be potential biomarkers in distinguishing AISL from a chronic arsenic exposure population (P=0.0020 and P=0.0017). The area under the curve (AUC) and its related 95% CI, sensitivity, specificity, positive predictive value and negative predictive value were 0.67 (0.57, 0.77), 69.64%, 62.50%, 65.00 and 67.31% for tryptophan, and 0.70 (0.60, 0.80), 69.64%, 69.64%, 69.64% and 69.64% for phenylalanine, respectively. The AUC (95% CI), sensitivity, specificity, positive predictive value and negative predictive value of the combination of them were 0.72 (0.62, 0.81), 76.79%, 58.93%, 65.15 and 71.74%, respectively. Our results suggested that these two AAs could be either individually or jointly used as indicators of AISL identification.

Table 5 Combination of diagnostic indicators and ROC analysis results[ξ].

Indicators	AUC (95%CI)	Sensitivity, %	Specificity, %	Predict$^+$, %	Predict$^-$, %	P
Tryptophan	0.67(0.57,0.77)	69.64	62.50	65.00	67.31	0.002
Phenylalanine	0.70(0.60,0.80)	69.64	69.64	69.64	69.64	0.002
Combined[ζ]	0.72(0.62,0.81)	76.79	58.93	65.15	71.74	<0.001

[ξ] ROC: a receiver operator characteristic; AUC: area under the roc curve; CI: confidence interval; The sensitivities, specificity, positive predictive value and negative predictive value were calculated at their best cut-off points; Predict+: positive predictive value; Predict-: negative predictive value.
[ζ] Combined: tryptophan and phenylalanine. The combination is modeled according to the formula $\beta_1 X_1 + \beta_2 X_2$, with X_j denoting the standardized value for the j^{th} amino acid and βj denoting the regression coefficient from the logistic regression model.

DISCUSSION

In the present study, the association of serum tryptophan and phenylalanine, screened in our previous non-targeted metabolomics study using UPLC-MS/MS, with AISL and their ability to indicate AISL occurrences were quantitatively evaluated in individual and joint modes. Our results clearly showed that AISL are significantly and negatively associated with serum tryptophan and phenylalanine levels in a chronic arsenic exposure population via drinking water. Participants with

higher level of both AAs would have lowest odds of AISL. These two AAs might also be able to serve as the indicators of AISL.

The probability of the initiation and development of AISL would be affected by a large number of factors including age, gender, life styles, arsenic exposure, metabolism and others. These factors would be important confounding factors and will largely affect our results. To adjust for the impacts due to these cofactors, we firstly selected all participants using permuted block randomization from a single rural area in which population were chronically exposed to arsenic in a same way, had similar life style and environmental factors. Secondly, the cases and controls were matched by gender and age (±1 year). All of these may be the reason why so many potential confounders including arsenic exposure do not differ significantly between the cases and controls (table 1).

Participants enrolled in the current study were chronically exposed to arsenic via drinking water. The geometric mean (GM) and its related 95% CI of urinary iAs/creatinine and tAs/creatinine in this population were 17.49 (14.90, 20.53) μg/g and 147.20 (129.00, 167.97) μg/g, respectively. They were much higher than those in the 20 μg/L exposed to arsenic via drinking water [GM (95% CI): 0.4 (0.3,0.5) μg/g for iAs and 9.1 (6.5,12.7) μg/g for tAs], while obviously lower than those in the 90 μg/L exposed group [GM (95% CI): 39.4 (31.4, 49.6) μg/g for iAs and 248.7 (208.8, 296.3) μg/g][25]. An available report has shown that AISL cannot be completely cured even though the medical technology has already made great progress[26]. So, it is crucial to identify those who are most likely to progress to overt arsenic damages including AISL among people at risk as early as possible. Metabolomics study, which mainly focus on thoroughly assessing the variation of metabolites possibly linked to diseases occurrence and development, has been widely utilized to help us understand pathogenesis of diseases because of its relevance to the phenotypes as compared to other 'OMICs' study[27]. Moreover, mathematical modeling to assess the linkage between small molecular metabolites and arsenic toxicity has grown[28]. Developing a simple and interpretable modeling approach for the early detection of arsenic induced health lesions is of great theoretical value and realistic meaning[29], though it might be

difficult due to population specific complexities and the impacts due to some potential unmeasured covariates such as diet and genetic determinants.

Previous studies reported that gene-gene and gene-environment interaction were involved in arsenicosis through toxicological mechanisms including genomic instability[30] and oxidative stress[31]. Skin hyperpigmentation and palmoplantar hyperkeratosis could be biomarkers for long-term arsenic exposure identifying the differences in metabolites that are really associated with phenotypes through metabolites analysis may promote our understanding and identification of AISL. Animal study reveals that the disruption of amino acids metabolism upon arsenic exposure in rat which may be beneficial on understanding arsenic toxicity[32]. In our previous population-based metabolomics study, we found that serum metabolites alteration was significantly related to the risk of arsenic-induced health damages. In the current study, we detected that BCAA or AAA were also significantly relevant to AISL occurrence. Several studies across numerous ethnic backgrounds supports the usage of BCAA including leucine, isoleucine as well as valine and AAA profile such as phenylalanine, tryptophan and tyrosine as biomarkers in determining metabolic diseases[27 33]. Simultaneously, Zhou *et al* reports that arsenic-induced transformed cells exhibit apparent alterations in metabolite profiles including down-regulated of leucine, tryptophan, and phenylalanine in skin lesions group[34]. Consistent with Zhou's findings, two serum AAA (tryptophan, phenylalanine) levels were also significantly associated with AISL in our study.

Normal metabolism of amino acids are necessary for whole body homeostasis, growth and development, and health status[15]. Studies have reported that the changes in the availability of AAA will affect cell signaling, gene expression, brain, and neuroendocrine function[35]. Tryptophan, an amino acid metabolism related biomarker, is also a sensitive and specific indicator of oxidation. Tryptophan metabolism in mammals is a physiological means of preserving immune homeostasis associated with oxidative stress and inflammation[36 37]. In addition, phenylalanine can be transformed into specific neurotransmitters such as dopamine and adrenaline by the action of related enzymes. Wu and

colleagues[38] reported that arsenic exposure would lead to neurotransmitter metabolism disturbance, which might explain the reduction of phenylalanine. Furthermore, as one of the peptide-bound phenylalanine, phenylalanylphenylalanine has been reported to affect protein synthesis and secretion[39], potentially indicating the possible relation between endothelium dysfunction and phenylalanine metabolism disorder. The relationship between amino acid metabolism and AISL was still unclear. The notable alteration of tryptophan and phenylalanine in this study may well indicate the occurrence of metabolic disorders due to arsenic exposure. It is also beneficial to understand the effects of arsenic toxicity and of great importance in early identification of occurrences as well as delaying the progression of various arsenic-induced health lesions including AISL.

The current study included 56 AISL cases matched 56 non-AISL controls and the sample size might be potentially insufficient. To estimate the impact due to this potential insufficient sample size on our conclusion, the PROC POWER procedure for paired design study in SAS 9.4 (SAS Institute Inc.) was applied to assess the power of the 4 AAs when assessing their associations with AISL in this study. The results showed that the lowest power associated with all of these four AAs was 0.911 based on 56 pairs of participants. It suggested that with type I error as 0.05, total sample size as 56 pairs and two-sided test, the powers associated with these 4 amino acids were all great than 0.8. So, we believed that the sample size for the present study, 56 pairs, would well balance the power of tests. Furthermore, previous metabolomics studies usually have sample size no more than 40 cases in each group[40 41].

The main strength of this study is that the findings were depended on a community-based, long-term arsenic exposure cohort with well-designed quality assurance and quality control throughout the study, and the AAs were detected with non-targeted metabolomics approach through the discovery and validation phases. However, there are also several limitations to this study. Firstly, although untargeted metabolomics approach can assess a large amount number of metabolites precisely and efficiently, it only provides relative levels of AAs instead of their accurate quantitative concentration. Secondly, these findings are mainly based on a case-control study, which only reveals the association between

amino acid metabolism and the odds of AISL rather than confirming their causal relationship. Furthermore, as it is suggested that the ratio of approximately 10 to 15 observations per predictor in a logistic regression model will produce reasonably stable estimations[24], we selected only 4 covariates in the models due to the small sample size and large number of predictors to avoid potential overfitting and obtain a more stable estimation. Finally, the participants were mainly exposed to arsenic via drinking water, which would limit the findings extrapolated to the other arsenic exposure population via food and other ways. Therefore, additional elaborate population-based studies are needed to verify our discoveries.

In conclusion, specific amino acids might be linked to AISL and amino acids metabolism may play an important role in AISL early identification. Additional studies may be needed to confirm our findings.

Contributors: Guangyun Mao and Yaping Wei designed the study. Chaonan Jia participated in collecting data. Yuan Lan and Chaonan Jia audited the data. Yaping Wei, Xiangqing hou, Jushuang Li, Tao Wang conducted the literature search, Yaping Wei, Chaonan Jia conducted statistical analysis and interpreted the results. Yaping Wei and Chaonan Jia wrote the first draft of the manuscript. Jingjing Zuo helped with copyediting. Guangyun Mao reviewed the final manuscript and did substantial contributions.

Funding: This work was supported by the basic public welfare research project of Zhejiang province (LGF19H260011), the initial scientific research fund (KYQD170301) and the major project of the eye hospital of Wenzhou medical university (YNZD201602).

Acknowledgments: We gratefully acknowledge all of the study participants of this project. We wish to thank all of the dedication and hard work of the field team at the department of preventive medicine, school of public health and management, Wenzhou medical university. Our thanks must also go to Dr. Kendrick Hii, a native-English speaker, for his valuable suggestions and edits to the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

Data Sharing Statement: No additional unpublished data are available.

REFERENCES

1. Mandal BK, Suzuki KT. Arsenic round the world: a review. Talanta. 2002; 58:201-35.
2. Smith AH, Steinmaus CM. Health Effects of Arsenic and Chromium in Drinking Water: Recent Human Findings. Annual Review Of Public Health. Palo Alto: Annual Reviews, 2009:107-22.
3. Guha Mazumder DN, Haque R, Ghosh N, et al. Arsenic levels in drinking water and the prevalence of skin lesions in West Bengal, India. International journal of epidemiology. 1998; 27:871-7.
4. Guo XJ, Fujino Y, Kaneko S, et al. Arsenic contamination of groundwater and prevalence of arsenical dermatosis in the Hetao plain area, Inner Mongolia, China. Mol Cell Biochem. 2001; 222:137-40.
5. Lindberg A-L, Rahman M, Persson L-A, et al. The risk of arsenic induced skin lesions in Bangladeshi men and women is affected by arsenic metabolism and the age at first exposure. Toxicology and Applied Pharmacology. 2008; 230:9-16.
6. Tondel M, Rahman M, Magnuson A, et al. The relationship of arsenic levels in drinking water and the prevalence rate of skin lesions in Bangladesh. Environmental Health Perspectives. 1999; 107:727-29.
7. Hsu LI, Chen GS, Lee CH, et al. Use of Arsenic-Induced Palmoplantar Hyperkeratosis and Skin Cancers to Predict Risk of Subsequent Internal Malignancy. Am J Epidemiol. 2013; 177:202-12.
8. Luo LR, Li YY, Gao YH, et al. Association between arsenic metabolism gene polymorphisms and arsenic-induced skin lesions in individuals exposed to high-dose inorganic arsenic in northwest China. Sci Rep. 2018; 8:12.
9. Hughes MF, Beck BD, Chen Y, et al. Arsenic Exposure and Toxicology: A Historical Perspective. Toxicol Sci. 2011; 123:305-32.
10. Ren X, McHale CM, Skibola CF, et al. An Emerging Role for Epigenetic Dysregulation in Arsenic Toxicity and Carcinogenesis. Environmental Health Perspectives. 2011; 119:11-19.
11. Chen J-W, Wang S-L, Wang Y-H, et al. Arsenic methylation, GSTO1 polymorphisms, and metabolic syndrome in an arseniasis endemic area of southwestern Taiwan. Chemosphere. 2012; 88:432-38.
12. Wang S-L, Chang F-H, Liou S-H, et al. Inorganic arsenic exposure and its relation to metabolic syndrome in an industrial area of Taiwan. Environment International. 2007; 33:805-11.
13. Liu R, Li H, Fan WJ, et al. Leucine Supplementation Differently Modulates Branched-Chain Amino Acid Catabolism, Mitochondrial Function and Metabolic Profiles at the Different Stage of Insulin Resistance in Rats on High-Fat Diet. Nutrients. 2017; 9:20.
14. Platten M, Wick W, Van den Eynde BJ. Tryptophan Catabolism in Cancer: Beyond IDO and Tryptophan Depletion. Cancer Res. 2012; 72:5435-40.
15. Wu G. Amino acids: metabolism, functions, and nutrition. Amino Acids. 2009; 37:1-17.
16. Casadei L, Valerio M, Manetti C. Metabolomics: Challenges and Opportunities in Systems Biology Studies. Methods Mol Biol. 2018; 1702:327-36.
17. Johnson CH, Gonzalez FJ. Challenges and opportunities of metabolomics. J Cell Physiol. 2012; 227:2975-81.
18. Hui AY, McCarty WJ, Masuda K, et al. A systems biology approach to synovial joint lubrication in health, injury, and disease. Wiley Interdiscip Rev Syst Biol Med. 2012; 4:15-37.
19. Lu K, Abo RP, Schlieper KA, et al. Arsenic Exposure Perturbs the Gut Microbiome and Its Metabolic Profile in Mice: An Integrated Metagenomics and Metabolomics Analysis. Environmental Health Perspectives. 2014; 122:284-91.
20. Zhang J, Shen H, Xu W, et al. Urinary Metabolomics Revealed Arsenic Internal Dose-Related Metabolic Alterations: A Proof-of-Concept Study in a Chinese Male Cohort. Environmental Science & Technology. 2014; 48:12265-74.
21. Guo X, Cui H, Zhang H, et al. Protective Effect of Folic Acid on Oxidative DNA Damage: A Randomized, Double-Blind, and Placebo Controlled Clinical Trial. Medicine (Baltimore). 2015; 94:e1872.
22. Standard of diagnosis for endemic arsenism of China (WS/T211-2001). the Ministry of Health of the People's Republic of China. Beijing 2001.
23. Chen X, Guo X, He P, et al. Interactive Influence of N6AMT1 and As3MT Genetic Variations on Arsenic Metabolism in the Population of Inner Mongolia, China. Toxicol Sci. 2017; 155:124-34.
24. Peduzzi P, Concato J, Kemper E, et al. A simulation study of the number of events per variable in logistic

regression analysis. J Clin Epidemiol. 1996; 49:1373-9.

25. Sun G, Xu Y, Li X, et al. Urinary arsenic metabolites in children and adults exposed to arsenic in drinking water in Inner Mongolia, China. Environmental Health Perspectives. 2007; 115:648-52.
26. Kuo CC, Moon KA, Wang SL, et al. The Association of Arsenic Metabolism with Cancer, Cardiovascular Disease, and Diabetes: A Systematic Review of the Epidemiological Evidence. Environmental Health Perspectives. 2017; 125:15.
27. Wang TJ, Larson MG, Vasan RS, et al. Metabolite profiles and the risk of developing diabetes. Nature Medicine. 2011; 17:448-U83.
28. Lawley SD, Cinderella M, Hall MN, et al. Mathematical model insights into arsenic detoxification. Theor Biol Med Model. 2011; 8.
29. Juhasz AL, Herde P, Herde C, et al. Validation of the predictive capabilities of the Sbrc-G in vitro assay for estimating arsenic relative bioavailability in contaminated soils. Environ Sci Technol. 2014; 48:12962-9.
30. Kibriya MG, Jasmine F, Parvez F, et al. Association between genome-wide copy number variation and arsenic-induced skin lesions: a prospective study. Environ Health. 2017; 16:13.
31. Priori R, Scrivo R, Brandt J, et al. Metabolomics in rheumatic diseases: the potential of an emerging methodology for improved patient diagnosis, prognosis, and treatment efficacy. Autoimmun Rev. 2013; 12:1022-30.
32. Wang X, Mu X, Zhang J, et al. Serum metabolomics reveals that arsenic exposure disrupted lipid and amino acid metabolism in rats: a step forward in understanding chronic arsenic toxicity. Metallomics Integrated Biometal Science. 2015; 7:544.
33. Weng L, Quinlivan E, Gong Y, et al. Association of Branched and Aromatic Amino Acids Levels with Metabolic Syndrome and Impaired Fasting Glucose in Hypertensive Patients. Metabolic Syndrome And Related Disorders. 2015; 13:195-202.
34. Zhou YY, Wang YF, Su J, et al. Integration of microRNAome, proteomics and metabolomics to analyze arsenic-induced malignant cell transformation. Oncotarget. 2017; 8:90879-96.
35. Lee JC, Chen MJ, Chang CH, et al. Plasma amino acid levels in patients with colorectal cancers and liver cirrhosis with hepatocellular carcinoma. Hepato-Gastroenterol. 2003; 50:1269-73.
36. Munn DH, Mellor AL. Indoleamine 2,3 dioxygenase and metabolic control of immune responses. Trends In Immunology. 2013; 34:137-43.
37. Pawlak K, Domaniewski T, Mysliwiec M, et al. The kynurenines are associated with oxidative stress, inflammation and the prevalence of cardiovascular disease in patients with end-stage renal disease. Atherosclerosis. 2009; 204:309-14.
38. Wu L-L, Gong W, Shen S-P, et al. Multiple metal exposures and their correlation with monoamine neurotransmitter metabolism in Chinese electroplating workers. Chemosphere. 2017; 182:745-52.
39. Zhou MM, Wu YM, Liu HY, et al. Effects of phenylalanine and threonine oligopeptides on milk protein synthesis in cultured bovine mammary epithelial cells. J Anim Physiol Anim Nutr. 2015; 99:215-20.
40. Chen LY, Cheng CY, Choi HW, et al. Plasma Metabonomic Profiling of Diabetic Retinopathy. Diabetes. 2016; 65:1099-108.
41. Piszcz J, Lemancewicz D, Dudzik D, et al. Differences and similarities between LC-MS derived serum fingerprints of patients with B-cell malignancies. Electrophoresis. 2013; 34:2857-64.

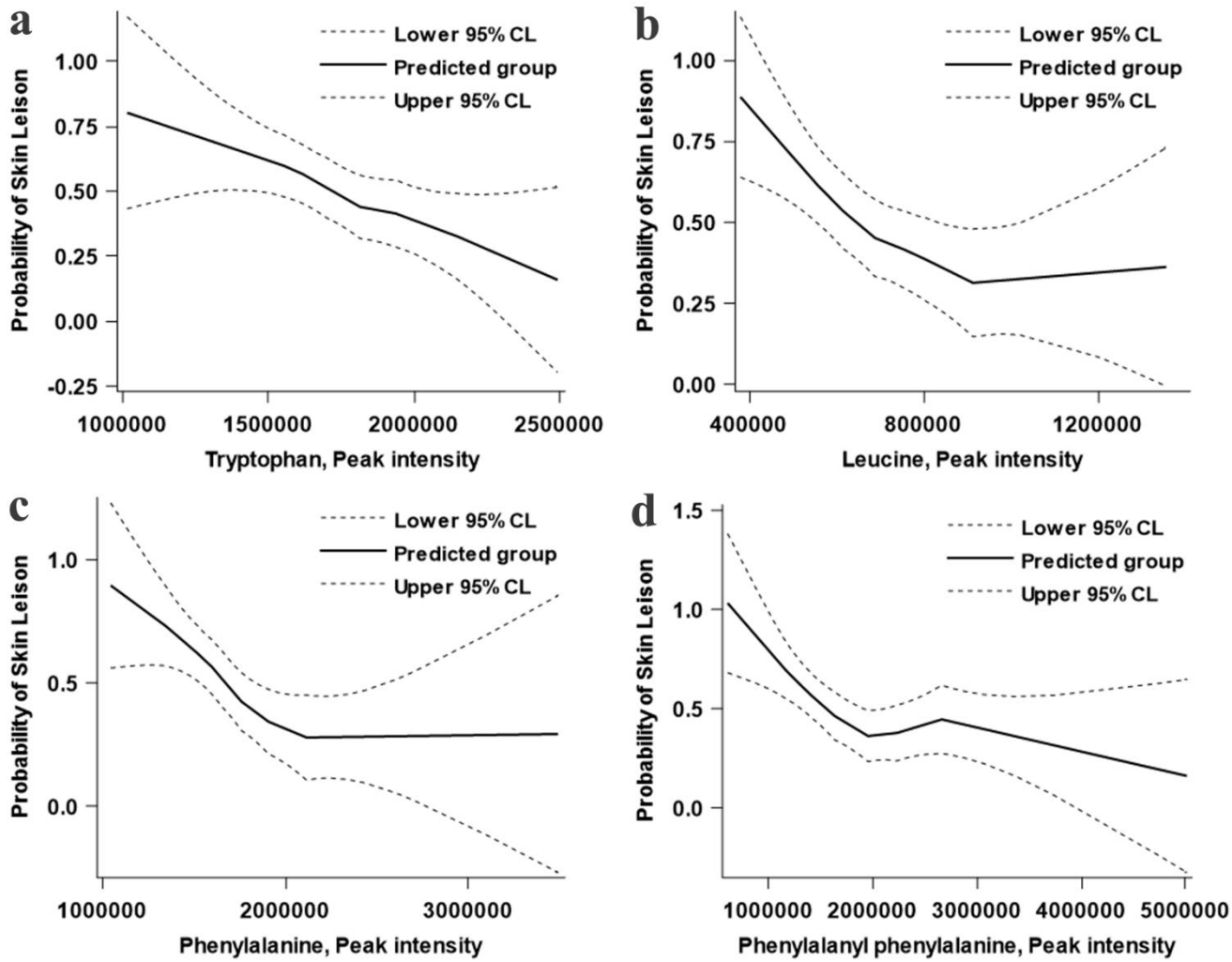

Figure 1. Association between the peak intensity of tryptophan and phenylalanine and arsenic-induced skin lesions based on multivariable locally weighted regression models. a: Tryptophan; b: Phenylalanine;c: Leucine;d: Phenylalanylphenylalanine

Table S1 Amino acids metabolites in serum.

Amino acids	Quartile	Quartile Range	Number
Tryptophan	1	1015821.54-1555040.97	28
	2	1555370.88-1701256.07	28
	3	1704298.84-1937437.56	28
	4	1945755.37-2492319.05	28
Phenylalanine	1	1044964.46-1516708.76	28
	2	1519993.60-1649420.82	28
	3	1659818.85-1908125.28	28
	4	1929946.22-3489918.62	28
Leucine	1	379957.04-556314.46	28
	2	558744.08-688759.83	28
	3	689470.83-806033.16	28
	4	821492.90-1356991.53	28
Phenylalanylphenylalanine	1	608200.15-1291853.09	27
	2	1303597.59-1632678.66	29
	3	1638204.36-2234283.97	28
	4	2235250.29-5011775.70	28

Table S2 Variance inflation factor of amino acids in different models[ξ].

Amino acids	Model 1	Model 2	Model 3	Model 4
Tryptophan	1.04	1.04	1.04	1.04
Phenylalanine	4.56		1.50	1.04
Leucine	4.21	1.39		
Phenylalanylphenylalanine	1.49	1.38	1.48	

[ξ]*Model 1: Tryptophan, Phenylalanine, Leucine and Phenylalanyl Phenylalanine;*
Model 2: Tryptophan, Leucine and Phenylalanyl Phenylalanine;
Model 3: Tryptophan, Phenylalanine and Phenylalanylphenylalanine;
Model 4: Tryptophan and Phenylalanine.

Table1 The STROBE checklist in this study.

	Item No	Recommendation
Title and abstract	1	(a) Indicate the study's design with a commonly used term in the title or the abstract (b) Provide in the abstract an informative and balanced summary of what was done and what was found **Please see detail in the "ABSTRACT" section in the manuscript.**
Introduction		
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported **Please see detail in the second, third, fourth paragraphs of the "INTRODUCTION" section in the manuscript.**
Objectives	3	State specific objectives, including any prespecified hypotheses **Please see detail in the fourth paragraph of the "INTRODUCTION" section in the manuscript.**
Methods		
Study design	4	Present key elements of study design early in the paper **Please see detail in the "Study Population "of the "METHODS" section in the manuscript.**
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection **Please see detail in the "Study Population "of the "METHODS" section in the manuscript.**
Participants	6	(a) Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls (*b*) For matched studies, give matching criteria and the number of controls per case **Please see detail in the "Study Population "of the "METHODS" section in the manuscript.**
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable **Please see detail in the "Study Population ", "Data Collection and Assessment" and "Distinct Metabolites Identification" of the "METHODS" section in the manuscript.**
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group **Please see detail in the "Study Population ", "Data Collection and**

		Assessment" and "Distinct Metabolites Identification" of the "METHODS" section in the manuscript.
Bias	9	Describe any efforts to address potential sources of bias **Please see detail in the "Study Population "of the "METHODS" section in the manuscript.**
Study size	10	Explain how the study size was arrived at **Please see detail in the sixth paragraph of the of the "DISCUSSION" section in the manuscript.**
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why **Please see detail in the "Statistical Analysis" of the "METHODS" section in the manuscript.**
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding (*b*) Describe any methods used to examine subgroups and interactions (*c*) Explain how missing data were addressed (*d*) If applicable, explain how matching of cases and controls was addressed (*e*) Describe any sensitivity analyses **No missing values were observed in our database. Others please see detail in the "Statistical Analysis" of the "METHODS" section in the manuscript.**
Results		
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed; (b) Give reasons for non-participation at each stage; (c) Consider use of a flow diagram **Please see detail in the "Study Population "of the "METHODS" section in the manuscript.**
Descriptive data	**14***	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders (b) Indicate number of participants with missing data for each variable of interest **See Table 1 in the manuscript.**
Outcome data	15*	Report numbers in each exposure category, or summary measures of exposure **Please see Table 3 in the manuscript.**
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were

		included **Please see Table 3 in the manuscript.**
		(*b*) Report category boundaries when continuous variables were categorized **Please see Table S1 in the manuscript.**
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses **Please see Table 4 in the manuscript.**
Discussion		
Key results	18	Summarise key results with reference to study objectives **Please see detail in the first paragraph of the of the "DISCUSSION" section in the manuscript.**
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias **Please see detail in the seventh paragraph of the of the "DISCUSSION" section in the manuscript.**
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence **Please see detail in the fifth paragraph of the of the "DISCUSSION" section in the manuscript.**
Generalisability	21	Discuss the generalisability (external validity) of the study results **Please see detail in the seventh paragraph of the "DISCUSSION" section in the manuscript.**
Other information		
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based **Please see detail in the "Funding" section in the manuscript.**

*Give information separately for cases and controls.

BMJ Open

Tryptophan, phenylalanine and arsenic-induced skin lesions in a chronic arsenic exposure Chinese population via drinking water: data from a metabolomics study

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2018-025336.R3
Article Type:	Original research
Date Submitted by the Author:	24-Apr-2019
Complete List of Authors:	wei, yaping; Wenzhou Medical University, Chashan University town, Ouhai district, Wenzhou, China., Jia, Chaonan; School of Public Health & Management, Wenzhou Medical University, Department of Preventive Medicine Lan, Yuan; School of Ophthalmology & Optometry, Wenzhou Medical University, Center on Clinical Research, the Eye Hospital of Wenzhou Medical University Hou, Xiangqing; School of Public Health & Management, Wenzhou Medical Univeristy, Department of Preventive Medicine Zuo, Jingjing; Wenzhou Medical University School of Optometry and Ophthalmology, Center on Clinical Research, the Eye Hospital of Wenzhou Medical University Li, Jushuang; School of Public Health & Management, Wenzhou Medical Univeristy, Department of Preventive Medicine Wang, Tao; School of Public Health & Management, Wenzhou Medical University, Department of Preventive Medicine Mao, Guangyun; School of Public Health & Management, Preventive Medicine
Primary Subject Heading:	Occupational and environmental medicine
Secondary Subject Heading:	Epidemiology, Public health, Occupational and environmental medicine
Keywords:	Metabolomics, Chronic arsenic exposure, Skin lesions, Amino acid, UPLC-MS/MS

Tryptophan, phenylalanine and arsenic-induced skin lesions in a chronic arsenic exposure Chinese population via drinking water: data from a metabolomics study

Yaping Wei[1,2], Chaonan Jia,[1,2], Yuan Lan[3], Xiangqing Hou[1,2], Jingjing Zuo[3], Jushuang Li[1,2], Tao Wang[1,2], Guangyun Mao[1,2,4]✉

[1] Department of Preventive Medicine, School of Public Health and Management, Wenzhou Medical University, Wenzhou, China.

[2] Center on Evidence-Based Medicine & Clinical Epidemiological Research, School of Public Health, Wenzhou Medical University, Wenzhou, China.

[3] School of Ophthalmology & Optometry, Wenzhou Medical University, Wenzhou, China.

[4] Center on Clinical Research, the Affiliated Eye Hospital of Wenzhou Medical University, Wenzhou, China.

Yaping Wei and Chaonan Jia contributed equally to this paper.

✉Corresponding author: Guangyun Mao, Room 7A321, School of Public Health and Management, Wenzhou Medical University, Chashan University town, Ouhai district, Wenzhou, China. 325035

E-mail: maogy@wmu.edu.cn

Telephone: +86-577-86689848

Fax numbers: +86-577-86699122

ABSTRACT:

Objectives To investigate the association of specific serum amino acids (AAs) with the odds of arsenic-induced skin lesions (AISL) and their ability to distinguish AISL from the counterparts.

Design Case-control study.

Setting Three arsenic exposed villages in Wuyuan county of Hetao Plain, Inner Mongolia, China.

Participants Among 450 residents aged 18 to 79 years chronically exposed to arsenic via drinking water, 56 of them were diagnosed as AISL and defined as the cases. Another 56 participants without AISL matched by gender and similar age (±1 year) from the same population were picked out as the controls. The inclusion criteria were subjects having the metabolomics determination. Unmatched participants and those without serum metabolites data were excluded.

Primary and secondary outcome measures The outcome was whether it suffered from AISL. Multivariable conditional logistic regression models and receiver operating characteristic curve (ROC) analysis were performed to investigate the relationship between specific AAs and AISL.

Results The levels of tryptophan and phenylalanine were both negatively associated with AISL ($P<0.05$). As compared to the 1st quartile, the adjusted odds of AISL in the 2nd, 3rd and 4th quartile decreased by 69%, 90% and 84% for tryptophan, and 14%, 80% and 76% for phenylalanine, respectively. The combination of the two aforementioned higher-level AAs revealed the lowest odds of AISL (OR=0.06; 95%CI: 0.02, 0.22; $P<0.001$)). Furthermore, both AAs showed moderate ability to distinguish AISL from the control, with area-under-curve [(AUC), 95%CI] as 0.67 (0.57, 0.77) for tryptophan and 0.70 (0.60, 0.80) for phenylalanine, respectively (all $P<0.05$). The combined pattern with AUC (95%CI) was 0.72 (0.62, 0.81), sensitivity of 76.79% and specificity of 58.93% ($P<0.001$).

Conclusions Specific AAs might be linked to AISL and play an important role in its early identification. Additional studies are needed to confirm our findings.

Keywords: Metabolomics; Chronic arsenic exposure; Skin lesions; Amino acid; UPLC-MS/MS.

Strengths and limitations of this study

- Our findings were depended on a community-based metabolomics study with paired-design, strictly quality assurance and quality control.
- Multivariable conditional logistic models were performed to examine the association between specific levels of AA and AISL, and ROC analysis was applied to evaluate the value and feasibility of the AA to distinguish AISL from the counterparts.
- Although the AAs were determined by untargeted metabolomics approach, which can assess a large number of metabolites precisely and efficiently, we only can obtain relative levels of AAs instead of their accurate quantitative concentration.
- The findings were based on a case-control study, which only revealed the association between the AAs and the odds of AISL rather than confirming their causal relationship.
- The participants were mainly chronically exposed to arsenic via drinking water, which would limit the findings extrapolated to another arsenic exposure population via food or other ways.

INTRODUCTION

Chronic arsenic exposure via drinking water is widely believed as a global health concern, affecting a large amount of people worldwide. It may give rise to several human health issues and has been documented to associate with cardiovascular disease, diabetes, cancer and others[1 2]. With the industrial boom and dramatic rise of worldwide water pollution including arsenic contamination in the past, the prevalence and burden of arsenic-induced health damage will continue to increase. Skin has been confirmed as one of the most common and susceptible target of arsenic-induced health lesions. Cutaneous skin lesions are typical signs of arsenicosis after persistent arsenic exposure for a long term which are characterized by hyperkeratosis and hyperpigmentation. Considerable evidences of the prevalence of arsenical skin lesions had been observed in many countries[3-5].

As arsenic-induced skin lesions (AISL) have been widely accepted as the major early manifestation of arsenic toxicity[6] and might be indicators of susceptibility to more serious arsenic-induced health hazards[7], it is particularly crucial to identify participants at risk as early as possible for preventing the onset or delaying the progression of the serious health problems effectively. Several possible mechanisms such as genetic differences[8], oxidative stress[9] and epigenetic dysregulation[10] and others may explain arsenic poisoning. Previous studies also reported that arsenic methylation in vivo might be associated with metabolic syndrome[11 12].

Amino acids (AAs) are the "basic unit" that make up the body's various proteins and necessary to maintain the health. Some AAs are important regulators of some key metabolic pathways and have great importance in maximizing efficiency of food utilization, enhancing protein accretion and health improvement[13 14]. Abnormal metabolism of AAs will disturb the homeostasis of the body, impairs growth and development, and even causes death[15]. So, the levels of serum AAs may be an important implication for the metabolic status and disease condition. As a powerful tool in system biology research, metabolomics approach is beneficial on unbiased monitoring changes in endogenous metabolism-related physiological processes, providing integrative information on the distinctive

features across multiple functional levels, and offering a window to capture the core attributes responsible for various phenotypes, which are particularly important in understanding the relevant pathophysiological changes of a disease and its status, identifying novel biomarkers for risk screening, diagnosis, treatment and prognosis of important human diseases[16-18].

Animal experiments and epidemiological study have reported obvious arsenic-related metabolomics perturbations[19 20]. All of these researches substantially suggests that the relationship between specific metabolites and arsenic-induced health lesions should be investigated. However, few works have been conducted to comprehensively examine the metabolic mechanism relevant to AISL, especially for the AAs metabolism. The present study aims to quantitatively examine the association of several specific AAs with AISL and the ability to identify AISL.

METHODS

Patient and Public Involvement

No patients were involved.

Study Population

This study was originally from a randomized, double-blind, and placebo controlled clinical trial (NCT02235948) in 2010, in which all subjects were randomly selected using permuted block randomization from a single rural area in a population chronically exposed to low-level arsenic drinking water, had similar life style and influences under similar environmental factors. Information on the inclusion and exclusion criteria of the participants could be found in our previous study[21]. Strictly following the criteria of arsenicosis[22]. AISL was diagnosed as the presence of arsenic-induced keratosis, hyperpigmentation or depigmentation by a physician from Wenzhou medical university at the beginning of the trial. This was a matched case-control study (1:1 matching). Among 450 residents aged 18 to 79 years old enrolled in the above-mentioned trial, 56 of them were diagnosed as AISL and selected as the case. Another 56 participants without AISL matched by

gender and similar age (±1 year) from the same population were picked out as the control. The inclusion criteria were subjects having the metabolomic test. Unmatched participants and those without serum metabolites data were excluded. Informed consent was obtained from all participants and this study was approved by the ethics committee of Wenzhou Medical University, Wenzhou, China.

Data Collection and Assessment

The information on age, gender, exposure year, body mass index, smoking, alcohol consumption, education level, etc. was collected with a standardized questionnaire. Blood and urine samples were also collected at the time of participants' enrollment. Detailed data collection of blood and urine samples and assessment methods for clinical variables including plasma fasting glucose (FPG), serum urea nitrogen, serum folate, total homocysteine, total cholesterol (TC), triglycerides (TG), high-density lipoprotein (HDL), low-density lipoprotein (LDL) and others had been published previously[21]. The epuration of various urinary arsenic species were conducted by means of a high-performance liquid chromatography coupled mass spectrometer system for separation and detection[23]. The species of arsenic in urine samples consisted of inorganic arsenic (iAs, [$iAs^{Ⅲ}$ plus $iAs^{Ⅴ}$]), monomethyl arsenate (MMA, [$MMA^{Ⅲ}$ plus $MMA^{Ⅴ}$]) and dimethyl arsenate (DMA, [$DMA^{Ⅲ}$ plus $DMA^{Ⅴ}$]). All arsenic species were corrected by creatinine. The total arsenic (tAs) was the sum of iAs, MMA and DMA. The percentages of arsenic species were defined as: iAs%=iAs/tAs*100%, MMA% =MMA/tAs*100% and DMA%=DMA/tAs*100%, respectively.

UPLC-MS/MS Metabonomic Profiling

Serum samples（200 μL in microcentrifuge tubes) were thawed to room temperature (25°C) and 600 μL mixture (90% acetonitrile - 10%water) were added to each sample. The samples were vigorously mixed for 20 seconds and centrifuged for 5 min at 12000 rpm (20°C). The top 400 μL of each supernatant were then transferred and dried down in a vacuum concentrator centrifuge. The dried samples were re-suspended in 130 μL of water (including 15% acetonitrile), mixed vigorously

for 20 seconds and repeated the centrifugation method described above. Two μL of the supernatant were collected as samples to be determined. Serum metabolic profile acquisition was performed by using ACQUITY UPLC®/Xevo® G2 QTof/MS^E (Waters Corp., Milford, MA, USA). Chromatographic separation was performed at 50°C using a WATERS HSS T3 column (2.1×100 mm, 1.7 μm) with a flow rate of 0.4mL/min. The mobile phase was a mixture of (A) H_2O with 0.1% formic acid and (B) methanol with 0.1% formic acid. Elution was in linear gradient with the programmed gradient at 0 min with 100% A and 0% B, 1.00min with 100%A and 0% B, 8 min with 0%A and 100% B, 13.00 min with 0% A and 100% B. The mass spectrometer was operated under both positive-ion (ESI^+) mode and negative-ion (ESI^-) mode electrospray ionization. The scan range was from 50 to 1200 m/z. Data was collected in both ESI^+ and ESI^- modes. Capillary voltage was set at 3000 V and 2500 V, respectively. The desolvation flow rate was 800 L/h at 350°C. Argon was used as a collision gas, and the collision energy was adjusted from 10 eV to 40 eV for each analysis. Quantum clustering (QC) samples were prepared by pooling aliquots of each sample and used to reflect the reliability of further metabolomics analysis. After peak deconvolution, alignment, integration and normalization, the data including retention time (RT), mass to charge ratio(m/z), and peak intensity were extracted from raw chromatograms using Progenesis QI 2.0 (Waters Corp., Milford, MA, USA). The MS/MS mode was performed to obtain metabolites levels processed with MarkerLynx Applications Manager Version 4.1 (Waters Corp., Milford, MA, USA).

Distinct Metabolites Identification

The peak intensity of metabolites for the 56 pairs were acquired and then imported to MetaboAnalyst 4.0 (http://www.metaboanalyst.ca/) for statistical analysis. A partial least-squares discriminant analysis (PLS-DA), which is a supervised and well accepted pattern recognition approach, was used for the differentiation between the cases and controls. False discovery rate (FDR) adjusted p-value in univariate analysis were performed to reduce the potential impact induced by false positive of the results. The criteria used in the selection of metabolites include variable

importance in the project (VIP) scores >1 in PLS-DA and the crude or FDR adjusted p-value all < 0.05 in Wilcoxon signed rank test. We identified a total of 70 extracted small molecular metabolites that were linked to the recognition of AISL. The Human Metabolome Database (http://www.hmdb.ca) was used to identify the name of metabolites. Among them, there were four amino acid metabolites (Phenylalanine, Tryptophan, Leucine, Phenylalanylphenylalanine).

Statistical Analysis

The normality of continuous data was assessed using both QQ-plots and Shapiro-Wilk test. The comparison between the cases and controls was performed with the paired *t*-test if they met normal or similar normal distribution. Otherwise, Wilcoxon signed rank test would be used. Differences in the proportion of categorical variables between the two groups were evaluated by McNemar-Bowker tests. We firstly used locally weighted scatterplot smoothing (LOESS) models to estimate the "real" relationship between serum AAs levels and the probability of AISL. Then, multivariable conditional logistic regression models were performed to examine the association between the contributing AAs levels and AISL after adjusting for some potential confounding factors. The individual impacts of AA metabolites on the risk of AISL were quantified separately by odds ratio (OR) and 95% confidence interval (CI) in the following two ways: with AA as a categorical variable (quartiles) and as a continuous variable [scaled to interquartile range (IQR)]. Variables with p-value less than 0.2 in the comparison between two groups were selected as potential confounders, which had been widely performed in many studies especially when the sample size of the current study was not too large. Then variance inflation factor (VIF) was used to examine the potential collinarity among them. As too many covariates in a multiple regression model will lead to overfitting to some extent[24], we finally select no more than 4 variables as confounding factors to decrease the potential overfitting when assessing the association between AAs and AISL. Furthermore, as the distinct metabolites might be high related to each other, collinearity should be well considered. So, we used the VIF based on the VIF package of R software to detect potential collinearity among the AAs. When the

VIF is greater than 1.5, it was considered as collinearity existed in the model and the associated variable would be removed. The combined effect of relevant AAs on AISL were also performed using a multivariable logistic regression model. In addition, receiver operator characteristic (ROC) analysis was applied to evaluate the value and feasibility of the AAs as the potential sensitive and specific biomarker to recognize AISL. Data management, analysis and figure drawing were finished using R version 3.4.4 (Copyright © 2018 The R Foundation for Statistical Computing). All tests were two-sides and P<=0.05 was set as significant level.

RESULTS

Table 1 summarizes the general characteristics of the study population. The comparison of demographical, clinical features and urinary arsenic species in the 56 pairs of subjects were presented in Table 1. The median (1st quartile, 3rd quartile) age of AISL population was 50.30 (44.70, 58.70) for the cases and 50.40 (44.60, 58.70) years for the controls. Both groups have the same proportion of female population (58.93%), and there was no obvious statistical difference in the urinary arsenic levels between the two groups. More than half of them had no history of smoking or alcohol consumption. When compared to the controls, the serum triglycerides level in AISL participants was significantly lower (P=0.041). While the other variables were similar between AISL participants and the control (P>0.05). This indicates that the participants in two groups are comparable to some extent.

Table 1 Demographic characteristics of the study population[ξ].

Variables	AISL (n=56)	Non-AISL (n=56)	P
Clinical Characteristics			
Age (years)	50.30(44.70,58.70)	50.40(44.60,58.70)	0.425
Exposure year (years)	48.19±11.53	47.62±10.97	0.489
Body mass index (kg/m^2)	24.12±3.14	23.91±2.86	0.697
Fasting plasma glucose (mmol/L)	4.89(4.60,5.25)	5.12(4.53,5.40)	0.137
Folate (ng/mL)	4.00(3.20,5.10)	4.25(3.35,5.40)	0.392
Total homocysteine (μmol/L)	12.30(10.32,16.50)	12.67(11.21,14.69)	0.961
Blood urea nitrogen (mmol/L)	6.45(5.42,7.69)	6.84(5.36,8.80)	0.603
Total cholesterol (mmol/L)	4.58(4.10,5.69)	4.65(3.96,5.95)	0.904
Triglycerides (mmol/L)	1.41(0.90,1.74)	1.45(1.09,2.29)	0.041
High-density lipoprotein (mmol/L)	1.19±0.34	1.16±0.31	0.675
Low-density lipoprotein (mmol/L)	3.04±0.80	3.25±0.84	0.110
Women [# (%)]	33(58.93)	33(58.93)	1.000
Cigarette smoking [# (%)]	20(35.71)	22(39.29)	0.696
Alcohol consumption [# (%)]	17(30.91)	21(37.50)	0.464
Illiteracy [# (%)]	21(37.50)	15(25.00)	0.252
Urinary arsenic species[ζ]			
iAS%	12.26(8.13,14.68)	12.31(10.04,16.54)	0.148
MMA%	24.68(20.11,29.68)	25.85(20.90,31.66)	0.420
DMA%	61.84(56.62,71.01)	61.84(47.85,64.99)	0.096
tAs (μg/g creatinine)	140.93(104.41,208.53)	186.77(80.11,217.30)	0.445

[ξ] AISL: arsenic-induced skin lesions; the variables met normal distribution was described with mean± standard deviation; otherwise, median (1st quartile, 3rd quartile) was used to describe their features. Number of cases (percentage) was used to describe the proportion of categorical variables between the two groups.

*[ζ] iAS: inorganic arsenic ($iAs^{III}+iAs^{V}$); MMA: monomethyl arsenate ($MMA^{III}+MMA^{V}$); DMA: dimethyl arsenate ($DMA^{III}+DMA^{V}$); tAs: total arsenic ($iAs^{III}+iAs^{V}$+MMA+DMA); iAs%= iAS/tAs*100%; MMA%=MMA/tAs*100% and DMA%=DMA/tAs*100%.*

Table 2 shows that the four AAs, which FDR adjusted p-value <0.05 and VIP>1, in the cases are observed significantly lower than those of the controls. Two of them are aromatic amino acids (AAA) identified as phenylalanine and tryptophan, one of them belongs to aromatic amino acids branched-chain amino acids (BCAA) appraised as leucine and the last one is phenylalanylphenylalanine. The individual association of AAs with AISL were presented in figure1, which clearly reveals obvious "dose-response" relationships between them.

Table 2 Distinct metabolites in population with arsenic-induced skin lesions and their counterparts.

Serum amino acid metabolites	Retention time (min)	Mass-to-Charge Ratio	VIP value	p-value$^{\xi}$	Adjusted p-values$^{\zeta}$
Phenylalanine	3.402	166.087	1.508	<0.001	0.009
Tryptophan	3.886	203.082	1.046	0.003	0.014
Leucine	2.642	132.102	1.014	0.001	0.020
Phenylalanylphenylalanine	5.048	313.155	1.833	0.004	0.033

VIP: variable importance in the project; ξ Wilcoxon signed-rank test; ζ Adjusted by false discovery rate (FDR).

Table 3 clearly shows that participants in the 3rd and 4th quartiles of the four specific AAs were all significantly linked to the decreased odds of AISL after adjusting for FPG, LDL, TG and DMA%, as compared to their lowest quartiles, respectively. The category boundaries of the quartiles were showed in the Table S1. Significant linear trends existed between AISL and those four serum AAs. Meanwhile, same linear negative association between AISL and per IQR rise of the four serum AAs were observed when these AAs were considered as continuous variables in the present study.

Table 3 Relationship of amino acids levels with the odds of arsenic-induced skin lesions[ξ]

Amino acids	N	Cases (%)	Crude		Adjusted[ζ]	
			OR (95%CI)	P	OR (95%CI)	P
Tryptophan						
Per IQR	112	56(50)	0.48(0.27,0.84)	0.011	0.48(0.27,0.86)	0.013
Quartiles						
Q_1	28	20(71.40)	1.00(1.00,1.00)	Ref.	1.00(1.00,1.00)	Ref.
Q_2	28	15(53.60)	0.50(0.16,1.54)	0.225	0.56(0.16,1.98)	0.370
Q_3	28	10(35.70)	0.12(0.03,0.53)	0.005	0.12(0.02,0.60)	0.010
Q_4	28	11(39.30)	0.19(0.05,0.71)	0.014	0.21(0.05,0.84)	0.028
P for trend				0.008		0.012
Phenylalanine						
Per IQR	112	56(50)	0.57(0.36,0.91)	0.019	0.56(0.33,0.94)	0.028
Quartiles						
Q_1	28	20(71.40)	1.00(1.00,1.00)	Ref.	1.00(1.00,1.00)	Ref.
Q_2	28	19(67.90)	0.79(0.22,2.82)	0.712	0.70(0.18,2.76)	0.609
Q_3	28	8(28.60)	0.18(0.05,0.66)	0.010	0.20(0.05,0.77)	0.019
Q_4	28	9(32.10)	0.25(0.08,0.79)	0.018	0.20(0.05,0.75)	0.017
P for trend				<0.001		0.001
Leucine						
Per IQR	112	56(50)	0.45(0.25,0.82)	0.019	0.43(0.21,0.86)	0.016
Quartiles						
Q_1	28	21(75.00)	1.00(1.00,1.00)	Ref.	1.00(1.00,1.00)	Ref.
Q_2	28	14(50.00)	0.33(0.11,1.03)	0.057	0.31(0.10,1.01)	0.052
Q_3	28	11(39.30)	0.22(0.07,0.68)	0.009	0.22(0.07,0.73)	0.014
Q_4	28	10(35.70)	0.19(0.06,0.59)	0.004	0.19(0.06,0.65)	0.008
P for trend				0.003		0.007
Phenylalanylphenylalanine						
Per IQR	112	56(50)	0.62(0.36,1.04)	0.070	0.71(0.41,1.24)	0.227
Quartiles						
Q_1	27	21(77.80)	1.00(1.00,1.00)	Ref.	1.00(1.00,1.00)	Ref.
Q_2	29	14(48.30)	0.16(0.03,0.75)	0.021	0.14(0.03,0.73)	0.019
Q_3	28	9(32.10)	0.08(0.02,0.39)	0.002	0.09(0.02,0.52)	0.007
Q_4	28	12(42.90)	0.12(0.02,0.57)	0.008	0.11(0.02,0.66)	0.016
P for trend				0.006		0.023

[ξ]Values are odds ratio (95% confidence intervals) for arsenic-induced skin lesions from conditional logistic regression. IQR: interquartile range; Q_1: the 1st quartile; Q_2: the 2nd quartile; Q_3: the 3rd quartile; Q_4: the 4th quartile.

[ζ] Adjusted for plasma glucose, low-density lipoprotein, triglyceride and urinary dimethyl arsenate.

As these 4 specific AAs are significantly or marginal significantly associated with the odds of AISL, so it is needed to examine the joint impacts among them on AISL. However, the results of potential collinearity examination revealed that among these 4 specific AAs, both tryptophan and phenylalaine had the smallest VIF value (VIF=1.04) and no obvious collinearity existed (**Table S2**). Hence, we mainly focus on tryptophan and phenylalaine when assessing the joint impacts of AAs on AISL and only presented the results associated with these two AAs in the current study. To avoid the impacts due to insufficient power because of unreasonable grouping on the results, we classified both tryptophan and phenylalanine into two categories, according to the cut-off values of their mass

spectrum peak area based on the ROC analysis, respectively. The higher levels of these two serum AA were defined as equal to or over the cut-off values, while the lower categories were considered as less than the associated values.

Table 4 shows the joint impacts of tryptophan and phenylalanine levels on AISL after considering the collinearity of variables in the model. The proportions of AISL were 74.3%, 60.0%, 50.0% and 18.2% for participants with lower levels of both tryptophan and phenylalanine (category A), with higher tryptophan and lower phenylalanine (category B), with lower tryptophan and higher phenylalanine (category C), and higher levels of both tryptophan and phenylalanine (category D), respectively. Obvious decrease trend of the probability of AISL was observed among these 4 categories. As compared to the category A, adjusted OR (95% CI) for participants in the category B, C and D were 0.49(0.15, 1.63), 0.32(0.10, 1.02) and 0.08(0.02, 0.25). Subjects with higher levels of both tryptophan and phenylalanine had the lowest odds of AISL, significantly decreased by 92% (OR=0.02; 95%CI: 0.02, 0.25; P<0.001), after adjusting for the impacts induced by some potential confounding factors. This suggested that tryptophan and phenylalanine were jointly associated with the presence of AISL. While no significant interaction between the two AAs on the occurrence of AISL could be observed (P=0.419), which indicated that each AA was independently associated with AISL though their joint impact was significant.

Table 4 Joint association between tryptophan and phenylalanine levels with arsenic-induced skin lesions.

Tryptophan <cut-off value[ξ]	Phenylalanine <cut-off value[ξ]	N	Cases (%)	Crude		Adjusted[ζ]	
				OR (95%CI)	P	OR (95%CI)	P
Yes	Yes	35	26(74.3)	1.00(1.00,1.00)	Ref.	1.00(1.00,1.00)	Ref.
No	Yes	20	12(60.0)	0.52(0.16,1.68)	0.273	0.49(0.15,1.63)	0.244
Yes	No	24	12(50.0)	0.35(0.12,1.04)	0.059	0.32(0.10,1.02)	0.053
No	No	33	6(18.2)	0.08(0.02,0.25)	<0.001	0.08(0.02,0.25)	<0.001
Interaction					0.320	0.49(0.09,2.78)	0.419

[ξ] Cut-off value was determined by means of receiver operator characteristic analysis.
[ζ] Adjusted for plasma glucose, low-density lipoprotein, triglyceride and urinary dimethyl arsenate.

Table 5 shows that, based on the ROC analysis, both serum tryptophan and phenylalanine might be potential biomarkers in distinguishing AISL from a chronic arsenic exposure population (P=0.0020 and P=0.0017). The area under the curve (AUC) and its related 95% CI, sensitivity, specificity, positive predictive value and negative predictive value were 0.67 (0.57, 0.77), 69.64%, 62.50%, 65.00 and 67.31% for tryptophan, and 0.70 (0.60, 0.80), 69.64%, 69.64%, 69.64% and 69.64% for phenylalanine, respectively. The AUC (95% CI), sensitivity, specificity, positive predictive value and negative predictive value of the combination of them were 0.72 (0.62, 0.81), 76.79%, 58.93%, 65.15 and 71.74%, respectively. Our results suggested that these two AAs could be either individually or jointly used as indicators of AISL identification.

Table 5 Combination of diagnostic indicators and ROC analysis results[ξ].

Indicators	AUC (95%CI)	Sensitivity, %	Specificity, %	Predict$^+$, %	Predict$^-$, %	P
Tryptophan	0.67(0.57,0.77)	69.64	62.50	65.00	67.31	0.002
Phenylalanine	0.70(0.60,0.80)	69.64	69.64	69.64	69.64	0.002
Combined[ζ]	0.72(0.62,0.81)	76.79	58.93	65.15	71.74	<0.001

[ξ] ROC: a receiver operator characteristic; AUC: area under the roc curve; CI: confidence interval; The sensitivities, specificity, positive predictive value and negative predictive value were calculated at their best cut-off points; Predict+: positive predictive value; Predict-: negative predictive value.
[ζ] Combined: tryptophan and phenylalanine. The combination is modeled according to the formula $\beta_1 X_1 + \beta_2 X_2$, with X_j denoting the standardized value for the j^{th} amino acid and βj denoting the regression coefficient from the logistic regression model.

DISCUSSION

In the present study, the association of serum tryptophan and phenylalanine, screened in our previous non-targeted metabolomics study using UPLC-MS/MS, with AISL and their ability to indicate AISL occurrences were quantitatively evaluated in individual and joint modes. Our results clearly showed that AISL are significantly and negatively associated with serum tryptophan and phenylalanine levels in a chronic arsenic exposure population via drinking water. Participants with

higher level of both AAs would have lowest odds of AISL. These two AAs might also be able to serve as the indicators of AISL.

The probability of the initiation and development of AISL would be affected by a large number of factors including age, gender, life styles, arsenic exposure, metabolism and others. These factors would be important confounding factors and will largely affect our results. To adjust for the impacts due to these cofactors, we firstly selected all participants using permuted block randomization from a single rural area in which population were chronically exposed to arsenic in a same way, had similar life style and environmental factors. Secondly, the cases and controls were matched by gender and age (±1 year). All of these may be the reason why so many potential confounders including arsenic exposure do not differ significantly between the cases and controls (table 1).

Participants enrolled in the current study were chronically exposed to arsenic via drinking water. The geometric mean (GM) and its related 95% CI of urinary iAs/creatinine and tAs/creatinine in this population were 17.49 (14.90, 20.53) μg/g and 147.20 (129.00, 167.97) μg/g, respectively. They were much higher than those in the 20 μg/L exposed to arsenic via drinking water [GM (95% CI): 0.4 (0.3,0.5) μg/g for iAs and 9.1 (6.5,12.7) μg/g for tAs], while obviously lower than those in the 90 μg/L exposed group [GM (95% CI): 39.4 (31.4, 49.6) μg/g for iAs and 248.7 (208.8, 296.3) μg/g][25]. An available report has shown that AISL cannot be completely cured even though the medical technology has already made great progress[26]. So, it is crucial to identify those who are most likely to progress to overt arsenic damages including AISL among people at risk as early as possible. Metabolomics study, which mainly focus on thoroughly assessing the variation of metabolites possibly linked to diseases occurrence and development, has been widely utilized to help us understand pathogenesis of diseases because of its relevance to the phenotypes as compared to other 'OMICs' study[27]. Moreover, mathematical modeling to assess the linkage between small molecular metabolites and arsenic toxicity has grown[28]. Developing a simple and interpretable modeling approach for the early detection of arsenic induced health lesions is of great theoretical value and

realistic meaning[29], though it might be difficult due to population specific complexities and the impacts due to some potential unmeasured covariates such as diet and genetic determinants.

Previous studies reported that gene-gene and gene-environment interaction were involved in arsenicosis through toxicological mechanisms including genomic instability[30] and oxidative stress[31]. Skin hyperpigmentation and palmoplantar hyperkeratosis could be biomarkers for long-term arsenic exposure identifying the differences in metabolites that are really associated with phenotypes through metabolites analysis may promote our understanding and identification of AISL. Animal study reveals that the disruption of amino acids metabolism upon arsenic exposure in rat which may be beneficial on understanding arsenic toxicity[32]. In our previous population-based metabolomics study, we found that serum metabolites alteration was significantly related to the risk of arsenic-induced health damages. In the current study, we detected that BCAA or AAA were also significantly relevant to AISL occurrence. Several studies across numerous ethnic backgrounds supports the usage of BCAA including leucine, isoleucine as well as valine and AAA profile such as phenylalanine, tryptophan and tyrosine as biomarkers in determining metabolic diseases[27 33]. Simultaneously, Zhou *et al* reports that arsenic-induced transformed cells exhibit apparent alterations in metabolite profiles including down-regulated of leucine, tryptophan, and phenylalanine in skin lesions group[34]. Consistent with Zhou's findings, two serum AAA (tryptophan, phenylalanine) levels were also significantly associated with AISL in our study.

Normal metabolism of amino acids are necessary for whole body homeostasis, growth and development, and health status[15]. Studies have reported that the changes in the availability of AAA will affect cell signaling, gene expression, brain, and neuroendocrine function[35]. Tryptophan, an amino acid metabolism related biomarker, is also a sensitive and specific indicator of oxidation. Tryptophan metabolism in mammals is a physiological means of preserving immune homeostasis associated with oxidative stress and inflammation[36 37]. In addition, phenylalanine can be transformed into specific neurotransmitters such as dopamine and adrenaline by the action of related enzymes.

Wu and colleagues[38] reported that arsenic exposure would lead to neurotransmitter metabolism disturbance, which might explain the reduction of phenylalanine. Furthermore, as one of the peptide-bound phenylalanine, phenylalanylphenylalanine has been reported to affect protein synthesis and secretion[39], potentially indicating the possible relation between endothelium dysfunction and phenylalanine metabolism disorder. The relationship between amino acid metabolism and AISL was still unclear. The notable alteration of tryptophan and phenylalanine in this study may well indicate the occurrence of metabolic disorders due to arsenic exposure. It is also beneficial to understand the effects of arsenic toxicity and of great importance in early identification of occurrences as well as delaying the progression of various arsenic-induced health lesions including AISL.

The current study included 56 AISL cases matched 56 non-AISL controls and the sample size might be potentially insufficient. To estimate the impact due to this potential insufficient sample size on our conclusion, the PROC POWER procedure for paired design study in SAS 9.4 (SAS Institute Inc.) was applied to assess the power of the 4 AAs when assessing their associations with AISL in this study. The results showed that the lowest power associated with all of these four AAs was 0.911 based on 56 pairs of participants (Figure 1S). It suggested that with type I error as 0.05, total sample size as 56 pairs and two-sided test, the powers associated with these 4 amino acids were all great than 0.8. So, we believed that the sample size for the present study, 56 pairs, would well balance the power of tests. Furthermore, previous metabolomics studies usually have sample size no more than 40 cases in each group[40 41].

The main strength of this study is that the findings were depended on a community-based, long-term arsenic exposure cohort with well-designed quality assurance and quality control throughout the study, and the AAs were detected with non-targeted metabolomics approach through the discovery and validation phases. However, there are also several limitations to this study. Firstly, although untargeted metabolomics approach can assess a large amount number of metabolites

precisely and efficiently, it only provides relative levels of AAs instead of their accurate quantitative concentration. Secondly, these findings are mainly based on a case-control study, which only reveals the association between amino acid metabolism and the odds of AISL rather than confirming their causal relationship. Furthermore, as it is suggested that the ratio of approximately 10 to 15 observations per predictor in a logistic regression model will produce reasonably stable estimations[24], we selected only 4 covariates in the models due to the small sample size and these results need to be confirmed in new studies. Finally, the participants were mainly exposed to arsenic via drinking water, which would limit the findings extrapolated to the other arsenic exposure population via food and other ways. Therefore, additional elaborate population-based studies are needed to verify our discoveries.

In conclusion, specific AAs might be linked to AISL and amino acids metabolism may play an important role in AISL early identification. Additional studies may be needed to confirm our findings.

Contributors: Guangyun Mao and Yaping Wei designed the study. Chaonan Jia participated in collecting data. Yuan Lan and Chaonan Jia audited the data. Yaping Wei, Xiangqing hou, Jushuang Li, Tao Wang conducted the literature search, Yaping Wei, Chaonan Jia conducted statistical analysis and interpreted the results. Yaping Wei and Chaonan Jia wrote the first draft of the manuscript. Jingjing Zuo helped with copyediting. Guangyun Mao reviewed the final manuscript and did substantial contributions.

Funding: This work was supported by the basic public welfare research project of Zhejiang province (LGF19H260011), the initial scientific research fund (KYQD170301) and the major project of the eye hospital of Wenzhou medical university (YNZD201602).

Acknowledgments: We gratefully acknowledge all of the study participants of this project. We wish to thank all of the dedication and hard work of the field team at the department of preventive medicine, school of public health and management, Wenzhou medical university. Our thanks must also go to Dr. Kendrick Hii, a native-English speaker, for his valuable suggestions and edits to the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

Data Sharing Statement: No additional unpublished data are available.

REFERENCES

1. Mandal BK, Suzuki KT. Arsenic round the world: a review. Talanta. 2002; 58:201-35.
2. Smith AH, Steinmaus CM. Health Effects of Arsenic and Chromium in Drinking Water: Recent Human Findings. Annual Review Of Public Health. Palo Alto: Annual Reviews, 2009:107-22.
3. Guha Mazumder DN, Haque R, Ghosh N, et al. Arsenic levels in drinking water and the prevalence of skin lesions in West Bengal, India. International journal of epidemiology. 1998; 27:871-7.
4. Guo XJ, Fujino Y, Kaneko S, et al. Arsenic contamination of groundwater and prevalence of arsenical dermatosis in the Hetao plain area, Inner Mongolia, China. Mol Cell Biochem. 2001; 222:137-40.
5. Lindberg A-L, Rahman M, Persson L-A, et al. The risk of arsenic induced skin lesions in Bangladeshi men and women is affected by arsenic metabolism and the age at first exposure. Toxicology and Applied Pharmacology. 2008; 230:9-16.
6. Tondel M, Rahman M, Magnuson A, et al. The relationship of arsenic levels in drinking water and the prevalence rate of skin lesions in Bangladesh. Environmental Health Perspectives. 1999; 107:727-29.
7. Hsu LI, Chen GS, Lee CH, et al. Use of Arsenic-Induced Palmoplantar Hyperkeratosis and Skin Cancers to Predict Risk of Subsequent Internal Malignancy. Am J Epidemiol. 2013; 177:202-12.
8. Luo LR, Li YY, Gao YH, et al. Association between arsenic metabolism gene polymorphisms and arsenic-induced skin lesions in individuals exposed to high-dose inorganic arsenic in northwest China. Sci Rep. 2018; 8:12.
9. Hughes MF, Beck BD, Chen Y, et al. Arsenic Exposure and Toxicology: A Historical Perspective. Toxicol Sci. 2011; 123:305-32.
10. Ren X, McHale CM, Skibola CF, et al. An Emerging Role for Epigenetic Dysregulation in Arsenic Toxicity and Carcinogenesis. Environmental Health Perspectives. 2011; 119:11-19.
11. Chen J-W, Wang S-L, Wang Y-H, et al. Arsenic methylation, GSTO1 polymorphisms, and metabolic syndrome in an arseniasis endemic area of southwestern Taiwan. Chemosphere. 2012; 88:432-38.
12. Wang S-L, Chang F-H, Liou S-H, et al. Inorganic arsenic exposure and its relation to metabolic syndrome in an industrial area of Taiwan. Environment International. 2007; 33:805-11.
13. Liu R, Li H, Fan WJ, et al. Leucine Supplementation Differently Modulates Branched-Chain Amino Acid Catabolism, Mitochondrial Function and Metabolic Profiles at the Different Stage of Insulin Resistance in Rats on High-Fat Diet. Nutrients. 2017; 9:20.
14. Platten M, Wick W, Van den Eynde BJ. Tryptophan Catabolism in Cancer: Beyond IDO and Tryptophan Depletion. Cancer Res. 2012; 72:5435-40.
15. Wu G. Amino acids: metabolism, functions, and nutrition. Amino Acids. 2009; 37:1-17.
16. Casadei L, Valerio M, Manetti C. Metabolomics: Challenges and Opportunities in Systems Biology Studies. Methods Mol Biol. 2018; 1702:327-36.
17. Johnson CH, Gonzalez FJ. Challenges and opportunities of metabolomics. J Cell Physiol. 2012; 227:2975-81.
18. Hui AY, McCarty WJ, Masuda K, et al. A systems biology approach to synovial joint lubrication in health, injury, and disease. Wiley Interdiscip Rev Syst Biol Med. 2012; 4:15-37.
19. Lu K, Abo RP, Schlieper KA, et al. Arsenic Exposure Perturbs the Gut Microbiome and Its Metabolic Profile in Mice: An Integrated Metagenomics and Metabolomics Analysis. Environmental Health Perspectives. 2014; 122:284-91.
20. Zhang J, Shen H, Xu W, et al. Urinary Metabolomics Revealed Arsenic Internal Dose-Related Metabolic Alterations: A Proof-of-Concept Study in a Chinese Male Cohort. Environmental Science & Technology. 2014; 48:12265-74.
21. Guo X, Cui H, Zhang H, et al. Protective Effect of Folic Acid on Oxidative DNA Damage: A Randomized, Double-Blind, and Placebo Controlled Clinical Trial. Medicine (Baltimore). 2015; 94:e1872.
22. Standard of diagnosis for endemic arsenism of China (WS/T211-2001). the Ministry of Health of the People's Republic of China. Beijing 2001.
23. Chen X, Guo X, He P, et al. Interactive Influence of N6AMT1 and As3MT Genetic Variations on Arsenic Metabolism in the Population of Inner Mongolia, China. Toxicol Sci. 2017; 155:124-34.
24. Peduzzi P, Concato J, Kemper E, et al. A simulation study of the number of events per variable in logistic

regression analysis. J Clin Epidemiol. 1996; 49:1373-9.

25. Sun G, Xu Y, Li X, et al. Urinary arsenic metabolites in children and adults exposed to arsenic in drinking water in Inner Mongolia, China. Environmental Health Perspectives. 2007; 115:648-52.
26. Kuo CC, Moon KA, Wang SL, et al. The Association of Arsenic Metabolism with Cancer, Cardiovascular Disease, and Diabetes: A Systematic Review of the Epidemiological Evidence. Environmental Health Perspectives. 2017; 125:15.
27. Wang TJ, Larson MG, Vasan RS, et al. Metabolite profiles and the risk of developing diabetes. Nature Medicine. 2011; 17:448-U83.
28. Lawley SD, Cinderella M, Hall MN, et al. Mathematical model insights into arsenic detoxification. Theor Biol Med Model. 2011; 8.
29. Juhasz AL, Herde P, Herde C, et al. Validation of the predictive capabilities of the Sbrc-G in vitro assay for estimating arsenic relative bioavailability in contaminated soils. Environ Sci Technol. 2014; 48:12962-9.
30. Kibriya MG, Jasmine F, Parvez F, et al. Association between genome-wide copy number variation and arsenic-induced skin lesions: a prospective study. Environ Health. 2017; 16:13.
31. Priori R, Scrivo R, Brandt J, et al. Metabolomics in rheumatic diseases: the potential of an emerging methodology for improved patient diagnosis, prognosis, and treatment efficacy. Autoimmun Rev. 2013; 12:1022-30.
32. Wang X, Mu X, Zhang J, et al. Serum metabolomics reveals that arsenic exposure disrupted lipid and amino acid metabolism in rats: a step forward in understanding chronic arsenic toxicity. Metallomics Integrated Biometal Science. 2015; 7:544.
33. Weng L, Quinlivan E, Gong Y, et al. Association of Branched and Aromatic Amino Acids Levels with Metabolic Syndrome and Impaired Fasting Glucose in Hypertensive Patients. Metabolic Syndrome And Related Disorders. 2015; 13:195-202.
34. Zhou YY, Wang YF, Su J, et al. Integration of microRNAome, proteomics and metabolomics to analyze arsenic-induced malignant cell transformation. Oncotarget. 2017; 8:90879-96.
35. Lee JC, Chen MJ, Chang CH, et al. Plasma amino acid levels in patients with colorectal cancers and liver cirrhosis with hepatocellular carcinoma. Hepato-Gastroenterol. 2003; 50:1269-73.
36. Munn DH, Mellor AL. Indoleamine 2,3 dioxygenase and metabolic control of immune responses. Trends In Immunology. 2013; 34:137-43.
37. Pawlak K, Domaniewski T, Mysliwiec M, et al. The kynurenines are associated with oxidative stress, inflammation and the prevalence of cardiovascular disease in patients with end-stage renal disease. Atherosclerosis. 2009; 204:309-14.
38. Wu L-L, Gong W, Shen S-P, et al. Multiple metal exposures and their correlation with monoamine neurotransmitter metabolism in Chinese electroplating workers. Chemosphere. 2017; 182:745-52.
39. Zhou MM, Wu YM, Liu HY, et al. Effects of phenylalanine and threonine oligopeptides on milk protein synthesis in cultured bovine mammary epithelial cells. J Anim Physiol Anim Nutr. 2015; 99:215-20.
40. Chen LY, Cheng CY, Choi HW, et al. Plasma Metabonomic Profiling of Diabetic Retinopathy. Diabetes. 2016; 65:1099-108.
41. Piszcz J, Lemancewicz D, Dudzik D, et al. Differences and similarities between LC-MS derived serum fingerprints of patients with B-cell malignancies. Electrophoresis. 2013; 34:2857-64.

Figure 1. Association between the peak intensity of tryptophan and phenylalanine and arsenic-induced skin lesions based on multivariable locally weighted regression models. a: Tryptophan; b: Phenylalanine;c: Leucine;d: Phenylalanylphenylalanine

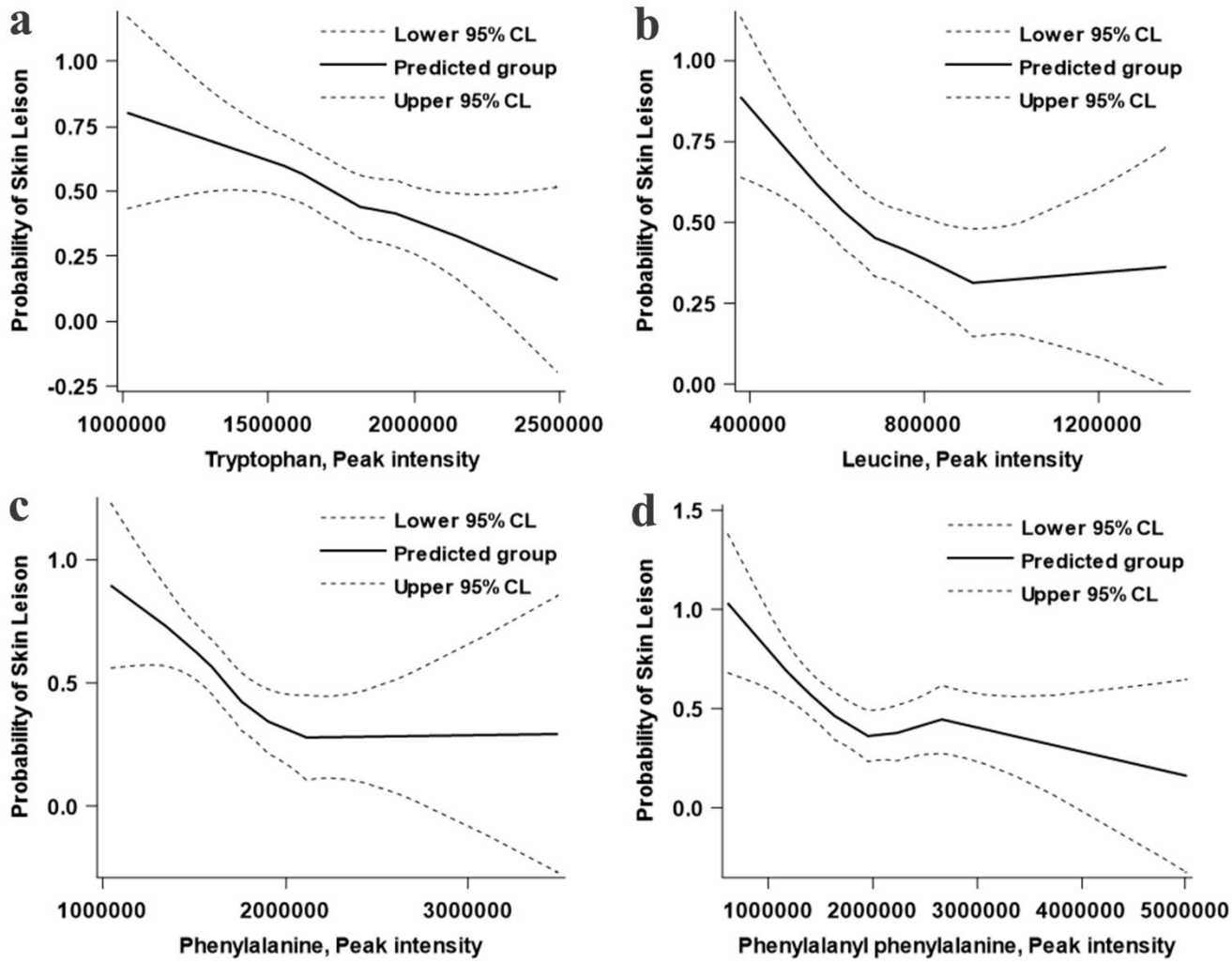

Figure 1. Association between the peak intensity of tryptophan and phenylalanine and arsenic-induced skin lesions based on multivariable locally weighted regression models. a: Tryptophan; b: Phenylalanine;c: Leucine;d: Phenylalanylphenylalanine

Table S1 Amino acids metabolites in serum.

Amino acids	Quartile	Quartile Range	Number
Tryptophan	1	1015821.54-1555040.97	28
	2	1555370.88-1701256.07	28
	3	1704298.84-1937437.56	28
	4	1945755.37-2492319.05	28
Phenylalanine	1	1044964.46-1516708.76	28
	2	1519993.60-1649420.82	28
	3	1659818.85-1908125.28	28
	4	1929946.22-3489918.62	28
Leucine	1	379957.04-556314.46	28
	2	558744.08-688759.83	28
	3	689470.83-806033.16	28
	4	821492.90-1356991.53	28
Phenylalanylphenylalanine	1	608200.15-1291853.09	27
	2	1303597.59-1632678.66	29
	3	1638204.36-2234283.97	28
	4	2235250.29-5011775.70	28

Table S2 Variance inflation factor of amino acids in different models[ξ].

Amino acids	Model 1	Model 2	Model 3	Model 4
Tryptophan	1.04	1.04	1.04	1.04
Phenylalanine	4.56		1.50	1.04
Leucine	4.21	1.39		
Phenylalanylphenylalanine	1.49	1.38	1.48	

[ξ]*Model 1: Tryptophan, Phenylalanine, Leucine and Phenylalanyl Phenylalanine;*
Model 2: Tryptophan, Leucine and Phenylalanyl Phenylalanine;
Model 3: Tryptophan, Phenylalanine and Phenylalanylphenylalanine;
Model 4: Tryptophan and Phenylalanine.

SAS 系统

The POWER Procedure
Paired t Test for Mean Difference

Fixed Scenario Elements	
Distribution	Normal
Method	Exact
Mean 1	768959
Mean 2	656243.1
Standard Deviation 1	206347.3
Standard Deviation 2	216983.4
Correlation	0.4
Number of Pairs	56
Number of Sides	2
Null Difference	0
Alpha	0.05

Computed Power
Power
0.946

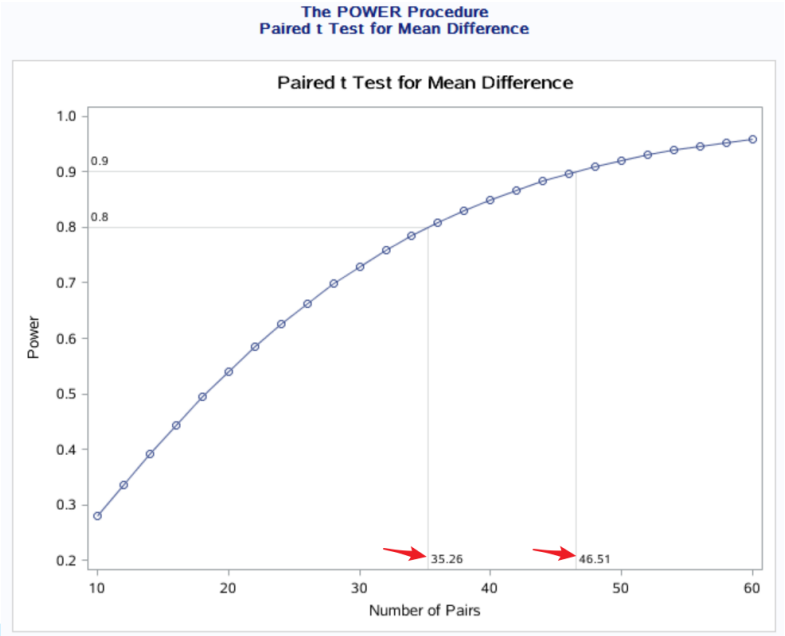

Figure 1S_A. Sample size and power estimation__Leucine

SAS 系统

The POWER Procedure
Paired t Test for Mean Difference

Fixed Scenario Elements	
Distribution	Normal
Method	Exact
Mean 1	1865229
Mean 2	1637107
Standard Deviation 1	420123.2
Standard Deviation 2	356546.1
Correlation	0.4
Number of Pairs	56
Number of Sides	2
Null Difference	0
Alpha	0.05

Computed Power
Power
0.975

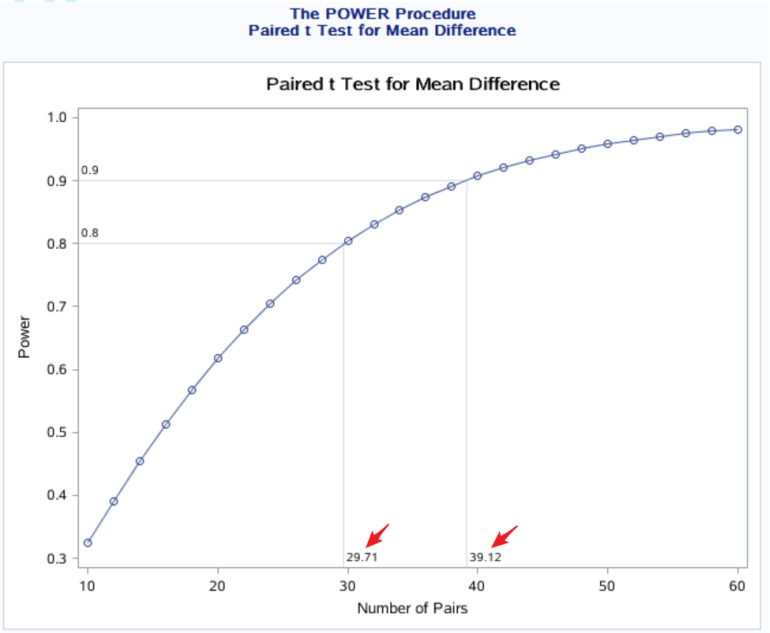

Figure 1S_B. Sample size and power estimation__Phenylalanine

SAS 系统

The POWER Procedure
Paired t Test for Mean Difference

Fixed Scenario Elements	
Distribution	Normal
Method	Exact
Mean 1	2056703
Mean 2	1664489
Standard Deviation 1	880518
Standard Deviation 2	682913.7
Correlation	0.4
Number of Pairs	56
Number of Sides	2
Null Difference	0
Alpha	0.05

Computed Power
Power
0.911

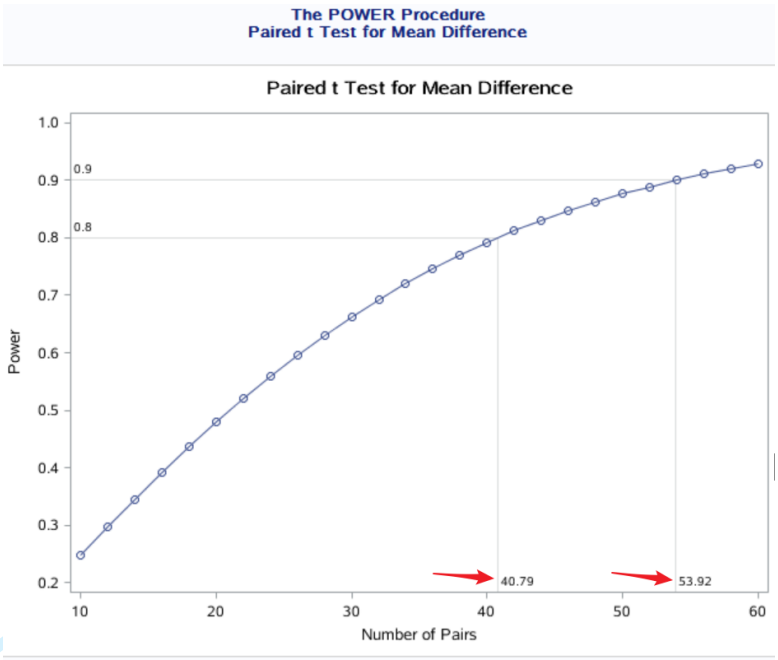

Figure 1S_C. Sample size and power estimation__Phenylalanylphenylalanine

SAS 系统

The POWER Procedure
Paired t Test for Mean Difference

Fixed Scenario Elements	
Distribution	Normal
Method	Exact
Mean 1	1832402
Mean 2	1655560
Standard Deviation 1	306551.7
Standard Deviation 2	289730.1
Correlation	0.4
Number of Pairs	56
Number of Sides	2
Null Difference	0
Alpha	0.05

Computed Power
Power
0.978

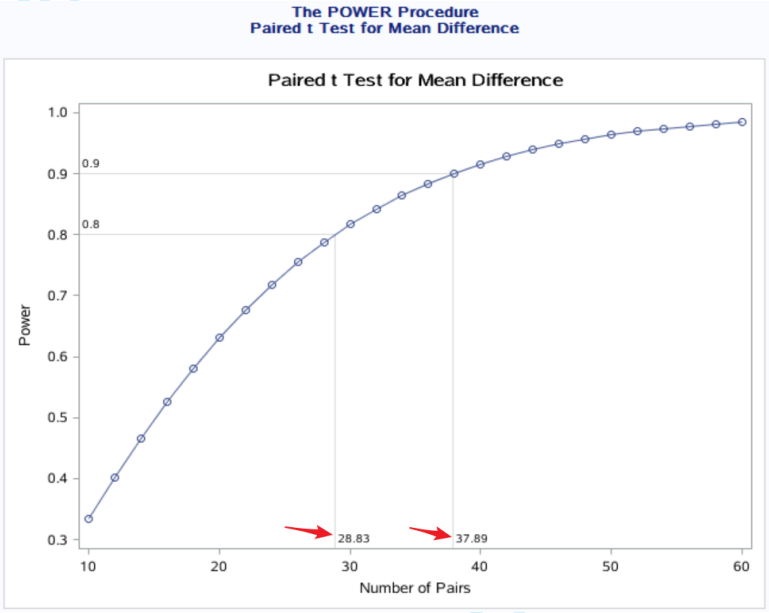

Figure 1S_D. Sample size and power estimation__Tryptophan

Table1 The STROBE checklist in this study.

	Item No	Recommendation
Title and abstract	1	(a) Indicate the study's design with a commonly used term in the title or the abstract (b) Provide in the abstract an informative and balanced summary of what was done and what was found **Please see detail in the "ABSTRACT" section in the manuscript.**
Introduction		
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported **Please see detail in the second, third, fourth paragraphs of the "INTRODUCTION" section in the manuscript.**
Objectives	3	State specific objectives, including any prespecified hypotheses **Please see detail in the fourth paragraph of the "INTRODUCTION" section in the manuscript.**
Methods		
Study design	4	Present key elements of study design early in the paper **Please see detail in the "Study Population "of the "METHODS" section in the manuscript.**
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection **Please see detail in the "Study Population "of the "METHODS" section in the manuscript.**
Participants	6	(a) Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls (*b*) For matched studies, give matching criteria and the number of controls per case **Please see detail in the "Study Population "of the "METHODS" section in the manuscript.**
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable **Please see detail in the "Study Population ", "Data Collection and Assessment" and "Distinct Metabolites Identification" of the "METHODS" section in the manuscript.**
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group **Please see detail in the "Study Population ", "Data Collection and**

		Assessment" and "Distinct Metabolites Identification" of the "METHODS" section in the manuscript.
Bias	9	Describe any efforts to address potential sources of bias **Please see detail in the "Study Population "of the "METHODS" section in the manuscript.**
Study size	10	Explain how the study size was arrived at **Please see detail in the sixth paragraph of the of the "DISCUSSION" section in the manuscript.**
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why **Please see detail in the "Statistical Analysis" of the "METHODS" section in the manuscript.**
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding (*b*) Describe any methods used to examine subgroups and interactions (*c*) Explain how missing data were addressed (*d*) If applicable, explain how matching of cases and controls was addressed (*e*) Describe any sensitivity analyses **No missing values were observed in our database. Others please see detail in the "Statistical Analysis" of the "METHODS" section in the manuscript.**
Results		
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed; (b) Give reasons for non-participation at each stage; (c) Consider use of a flow diagram **Please see detail in the "Study Population "of the "METHODS" section in the manuscript.**
Descriptive data	**14***	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders (b) Indicate number of participants with missing data for each variable of interest **See Table 1 in the manuscript.**
Outcome data	15*	Report numbers in each exposure category, or summary measures of exposure **Please see Table 3 in the manuscript.**
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were

		included **Please see Table 3 in the manuscript.**
		(*b*) Report category boundaries when continuous variables were categorized **Please see Table S1 in the manuscript.**
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses **Please see Table 4 in the manuscript.**
Discussion		
Key results	18	Summarise key results with reference to study objectives **Please see detail in the first paragraph of the of the "DISCUSSION" section in the manuscript.**
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias **Please see detail in the seventh paragraph of the of the "DISCUSSION" section in the manuscript.**
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence **Please see detail in the fifth paragraph of the of the "DISCUSSION" section in the manuscript.**
Generalisability	21	Discuss the generalisability (external validity) of the study results **Please see detail in the seventh paragraph of the "DISCUSSION" section in the manuscript.**
Other information		
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based **Please see detail in the "Funding" section in the manuscript.**

*Give information separately for cases and controls.

BMJ Open

Tryptophan, phenylalanine and arsenic-induced skin lesions in a chronic arsenic exposure Chinese population via the drinking water: a case-control study.

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2018-025336.R4
Article Type:	Original research
Date Submitted by the Author:	07-Aug-2019
Complete List of Authors:	wei, yaping; Wenzhou Medical University, Chashan University town, Ouhai district, Wenzhou, China., Jia, Chaonan; School of Public Health & Management, Wenzhou Medical University, Department of Preventive Medicine Lan, Yuan; School of Ophthalmology & Optometry, Wenzhou Medical University, Center on Clinical Research, the Eye Hospital of Wenzhou Medical University Hou, Xiangqing; School of Public Health & Management, Wenzhou Medical Univeristy, Department of Preventive Medicine Zuo, Jingjing; Wenzhou Medical University School of Optometry and Ophthalmology, Center on Clinical Research, the Eye Hospital of Wenzhou Medical University Li, Jushuang; School of Public Health & Management, Wenzhou Medical Univeristy, Department of Preventive Medicine Wang, Tao; School of Public Health & Management, Wenzhou Medical University, Department of Preventive Medicine Mao, Guangyun; School of Public Health & Management, Preventive Medicine
Primary Subject Heading:	Occupational and environmental medicine
Secondary Subject Heading:	Epidemiology, Public health, Occupational and environmental medicine
Keywords:	Metabolomics, Chronic arsenic exposure, Skin lesions, Amino acid

Tryptophan, phenylalanine and arsenic-induced skin lesions in a chronic arsenic exposure Chinese population via the drinking water: a case-control study.

Yaping Wei[1,2], Chaonan Jia,[1,2], Yuan Lan[3], Xiangqing Hou[1,2], Jingjing Zuo[3], Jushuang Li[1,2], Tao Wang[1,2], Guangyun Mao[1,2,4]✉

[1] Department of Preventive Medicine, School of Public Health and Management, Wenzhou Medical University, Wenzhou, China.

[2] Center on Evidence-Based Medicine & Clinical Epidemiological Research, School of Public Health, Wenzhou Medical University, Wenzhou, China.

[3] School of Ophthalmology & Optometry, Wenzhou Medical University, Wenzhou, China.

[4] Center on Clinical Research, the Affiliated Eye Hospital of Wenzhou Medical University, Wenzhou, China.

Yaping Wei and Chaonan Jia contributed equally to this paper.

✉Corresponding author: Guangyun Mao, Room 7A321, School of Public Health and Management, Wenzhou Medical University, Chashan University Town, Ouhai District, Wenzhou, China. 325035

E-mail: maogy@wmu.edu.cn

Telephone: +86-577-86689848

Fax numbers: +86-577-86699122

ABSTRACT

Objectives To investigate the association of specific serum amino acids (AAs) with the odds of arsenic-induced skin lesions (AISL) and ability to distinguish AISL from their counterparts.

Design Case-control study.

Setting Three arsenic exposed villages in Wuyuan county of Hetao Plain, Inner Mongolia, China.

Participants Among 450 residents aged 18 to 79 years chronically exposed to arsenic via drinking water, 56 of them were diagnosed as AISL and defined as the cases. Another 56 participants without AISL matched by gender and similar age (±1 year) from the same population were picked out as the controls.

Main outcome measures and methods The intensities of the AAs were determined by ultra-high-performance liquid chromatography-quadrupole time-of-flight mass spectrometry (UPLC-QTOF-MS)-based metabolomics approach. Potential confounding variables were obtained via a standardized questionnaire and clinical examination. Multivariable conditional logistic regression models and receiver operating characteristic curve (ROC) analysis were performed to investigate the relationship between specific AAs and AISL.

Results The levels of tryptophan and phenylalanine were both negatively associated with AISL ($P<0.05$). As compared to the 1st quartile, the adjusted odds of AISL in the 2nd, 3rd, and 4th quartile decreased by 44%, 88%, and 79% for tryptophan, and 30%, 80%, and 80% for phenylalanine, respectively. The combination of the two aforementioned higher-level AAs revealed the lowest odds of AISL (OR=0.08; 95% CI 0.02 to 0.25; $P<0.001$). Furthermore, both AAs showed moderate ability to distinguish AISL from the control, with area-under-curve [(AUC), 95% CI] as 0.67 (0.57, 0.77) for tryptophan and 0.70 (0.60, 0.80) for phenylalanine, respectively (all $P<0.05$). The combined pattern with AUC (95% CI) was 0.72 (0.62, 0.81), sensitivity of 76.79% and specificity of 58.93% ($P<0.001$).

Conclusions Specific AAs may be linked to AISL and play important roles in its early identification. Additional studies are needed to confirm our findings.

Keywords: Chronic arsenic exposure; Skin lesions; Metabolomics; Amino acid.

Strengths and limitations of this study

- Our findings were based on a community-based metabolomics study with paired-design, strictly quality assurance and quality control.
- Multivariable conditional logistic models were performed to examine the association between specific levels of AAs and AISL, and ROC analysis was applied to evaluate the feasibility of the AAs to distinguish AISL from their counterparts.
- Although the AAs were determined by untargeted metabolomics approach, which can assess a large number of metabolites precisely and efficiently, only relative levels of AAs could be obtained instead of their accurate quantitative concentration.
- As based on a case control-study, the findings only revealed the association between the AAs and the odds of AISL rather than confirming their causal relationship.
- The participants were mainly chronically exposed to arsenic via drinking water, which may limit the findings extrapolated to another arsenic exposure population via food or other ways.

INTRODUCTION

Chronic arsenic exposure via drinking water is widely believed as a global health concern, affecting a large number of people worldwide. It may give rise to several human health issues and has been documented to associate with cardiovascular disease, diabetes, cancer and others[1 2]. With the industrial boom and dramatic rise of worldwide water pollution including arsenic contamination in the past, the prevalence and burden of arsenic-induced health damage will continue to increase. The skin has been confirmed as one of the most common and susceptible targets of arsenic-induced health lesions. Cutaneous skin lesions are typical signs of arsenicosis after persistent arsenic exposure for the long term which are characterized by hyperkeratosis and hyperpigmentation. Considerable evidence of the prevalence of arsenical skin lesions had been observed in many countries[3-5].

As arsenic-induced skin lesions (AISL) have been widely accepted as the major early manifestation of arsenic toxicity[6] and might be indicators of susceptibility to more serious arsenic-induced health hazards[7], it is particularly crucial to identify participants at risk as early as possible for preventing the onset or delaying the progression of the serious health problems effectively. Several possible mechanisms such as genetic differences[8], oxidative stress[9], and epigenetic dysregulation[10] and others may explain arsenic poisoning. Previous studies also reported that arsenic methylation in vivo might be associated with metabolic syndrome[11 12].

Amino acids (AAs) are the "basic unit" that make up the body's various proteins and necessary to maintain the health. Some AAs are important regulators of some key metabolic pathways and have great importance in maximizing the efficiency of food utilization, enhancing protein accretion and health improvement[13 14]. Abnormal metabolism of AAs will disturb the homeostasis of the body, impairs growth and development, and even causes death[15]. So, the levels of serum AAs may be an important implication for the metabolic status and disease condition. As a powerful tool in system biology research, metabolomics approach is beneficial on unbiased monitoring changes in

endogenous metabolism-related physiological processes, providing integrative information on the distinctive features across multiple functional levels, offering a window to capture the core attributes responsible for various phenotypes, which are particularly important in understanding the relevant pathophysiological changes of a disease and its status, and identifying novel biomarkers for risk screening, diagnosis, treatment and prognosis of important human diseases[16-18].

Animal experiments and epidemiological studies have reported obvious arsenic-related metabolomics perturbations[19 20]. All of these researches substantially suggest that the relationship between specific metabolites and arsenic-induced health lesions should be investigated. However, few works have been conducted to comprehensively examine the metabolic mechanism relevant to AISL, especially for amino acid's metabolism. The present study aims to quantitatively examine the association of several specific AAs with AISL and the ability to identify AISL.

METHODS

Study Population

This study was originally from a randomized, double-blind, and placebo-controlled clinical trial (NCT02235948) in 2010, in which all subjects were randomly selected using permuted block randomization from a single rural area in a population chronically exposed to low-level arsenic drinking water, had similar lifestyle and influences under similar environmental factors. Information on the inclusion and exclusion criteria of the participants could be found in our previous study[21]. Strictly following the criteria of arsenicosis[22]. AISL was diagnosed as the presence of arsenic-induced keratosis, hyperpigmentation or depigmentation by a physician from Wenzhou medical university at the beginning of the trial. This was a matched case-control study (1:1 matching). Among 450 residents aged 18 to 79 years old enrolled in the trial mentioned above, 56 of them were diagnosed as AISL and selected as the case. Another 56 participants without AISL matched by gender and similar age (±1 year) from the same population were picked out as the control. The

inclusion criteria were subjects having the metabolomic test. Unmatched participants and those without serum metabolites data were excluded.

Data Collection and Assessment

The information on age, gender, exposure year, body mass index, smoking, alcohol consumption, education level, etc. was collected with a standardized questionnaire. Blood and urine samples were also collected at the time of participants' enrollment. Detailed data collection of blood and urine samples and assessment methods for clinical variables including plasma fasting glucose (FPG), serum urea nitrogen, serum folate, total homocysteine, total cholesterol (TC), triglycerides (TG), high-density lipoprotein (HDL), low-density lipoprotein (LDL) and others had been published previously[21]. The epuration of various urinary arsenic species was conducted utilizing high-performance liquid chromatography coupled mass spectrometer system for separation and detection[23]. The species of arsenic in urine samples consisted of inorganic arsenic (iAs, [$iAs^{Ⅲ}$ plus $iAs^{Ⅴ}$]), monomethyl arsenate (MMA, [$MMA^{Ⅲ}$ plus $MMA^{Ⅴ}$]) and dimethyl arsenate (DMA, [$DMA^{Ⅲ}$ plus $DMA^{Ⅴ}$]). All arsenic species were corrected by creatinine. The total arsenic (tAs) was the sum of iAs, MMA and DMA. The percentages of arsenic species were defined as: iAs%=iAs/tAs*100%, MMA% =MMA/tAs*100% and DMA%=DMA/tAs*100%, respectively.

UPLC-QTOF-MS Metabonomic Profiling

Serum samples were thawed to a temperature of 4°C and 600 μL mixture (90% acetonitrile - 10%water) were added to each sample (200 μL in microcentrifuge tubes). The samples were vigorously mixed for 20 seconds and centrifuged for 5 min at 12000 rpm (20°C). The top 400 μL of each supernatant was then transferred and dried down in a vacuum concentrator centrifuge. The dried samples were re-suspended in 130 μL of water (including 15% acetonitrile), mixed vigorously for 20 seconds and repeat the centrifugation method described above. Two μL of the supernatant was collected as samples to be determined. Serum metabolic profile acquisition was performed by using ACQUITY UPLC®/Xevo® G2 QTof/MS^E (Waters Corp., Milford, MA, USA). Chromatographic

separation was performed at 50℃ using a WATERS HSS T3 column (2.1×100 mm, 1.7 μm) with a flow rate of 0.4mL/min. The mobile phase was a mixture of (A) H_2O with 0.1% formic acid and (B) methanol with 0.1% formic acid. Elution was in linear gradient with the programmed gradient at 0 min with 100% A and 0% B, 1.00min with 100%A and 0% B, 8 min with 0%A and 100% B, 13.00 min with 0% A and 100% B. The mass spectrometer was operated under both positive-ion (ESI^+) mode and negative-ion (ESI^-) mode electrospray ionization. The scan range was from 50 to 1200 m/z. Data were collected in both ESI^+ and ESI^- modes. The capillary voltage was set at 3000 V and 2500 V, respectively. The desolvation flow rate was 800 L/h at 350°C. Argon was used as collision gas, and the collision energy was adjusted from 10 eV to 40 eV for each analysis. Quantum clustering (QC) samples were prepared by pooling aliquots of each sample and used to reflect the reliability of further metabolomics analysis. After peak deconvolution, alignment, integration and normalization, the data including retention time (RT), mass to charge ratio(m/z), and peak intensity were extracted from raw chromatograms using Progenesis QI 2.0 (Waters Corp., Milford, MA, USA). The MS/MS mode was performed to obtain metabolites levels processed with MarkerLynx Applications Manager Version 4.1 (Waters Corp., Milford, MA, USA).

Distinct Metabolites Identification

The peak intensity of metabolites for the 56 pairs was acquired and then imported to MetaboAnalyst 4.0 (http://www.metaboanalyst.ca/) for statistical analysis. A partial least-squares discriminant analysis (PLS-DA), which is a supervised and well-accepted pattern recognition approach, was used for the differentiation between the cases and controls. False discovery rate (FDR) adjusted p-value in univariate analysis was performed to reduce the potential impact induced by the false positive of the results. The criteria used in the selection of metabolites include variable importance in the project (VIP) scores >1 in PLS-DA and the crude or FDR adjusted p-value all < 0.05 in Wilcoxon signed-rank test. We identified a total of 70 extracted small molecular metabolites that were linked to the recognition of AISL. The Human Metabolome Database (http://www.hmdb.ca)

was used to identify the name of metabolites. Among them, there were four amino acid metabolites (phenylalanine, tryptophan, leucine, phenylalanylphenylalanine).

Statistical Analysis

The normality of continuous data was assessed using both QQ-plots and Shapiro-Wilk test. The comparison between the cases and controls was performed with the paired *t*-test if they met normal or similar normal distribution. Otherwise, the Wilcoxon signed-rank test would be used. Differences in the proportion of categorical variables between the two groups were evaluated by McNemar-Bowker tests. We firstly used locally weighted scatterplot smoothing (LOESS) models to estimate the "real" relationship between serum AAs levels and the probability of AISL. Then, multivariable conditional logistic regression models were performed to examine the association between the contributing AAs levels and AISL after adjusting for some potentially confounding factors. The individual impacts of AA metabolites on the risk of AISL were quantified separately by odds ratio (OR) and 95% confidence interval (CI) in the following two ways: with AAs as a categorical variable (quartiles) and as a continuous variable [scaled to an interquartile range (IQR)]. Variables with a p-value less than 0.2 in the comparison between two groups were selected as potential confounders, which had been widely performed in many studies especially when the sample size of the current study was not too large. The variance inflation factor (VIF) was used to examine the potential collinearity among them. As too many covariates in a multiple regression model will lead to overfitting to some extent[24], we finally select no more than 4 variables as confounding factors to decrease the potential overfitting when assessing the association between AAs and AISL. Furthermore, as the distinct metabolites might be highly related to each other, collinearity should be well considered. So, we used the VIF based on the VIF package of R software to detect potential collinearity among the AAs. When the VIF is greater than 1.5, it was considered as collinearity existed in the model and the associated variable would be removed. The combined effect of relevant AAs on AISL was also performed using a multivariable logistic regression model. Besides, receiver

operator characteristic (ROC) analysis was applied to evaluate the value and feasibility of the AAs as the potentially sensitive and specific biomarker to recognize AISL. Data management, analysis, and figure drawing were finished using R version 3.4.4 (Copyright © 2018 The R Foundation for Statistical Computing). All tests were two-sided and P<=0.05 was set at a significant level.

Patient and Public Involvement

The present study was designed as an observational study, and as such patients and the public were not involved in the planning, recruitment and conduct of this study. All participants were informed about the purpose of this study and signed informed consent at the beginning of the study. The results of this study have not yet been disseminated to the relevant patient population.

RESULTS

Table 1 summarizes the general characteristics of the study population. The comparison of demographical, clinical features and urinary arsenic species in the 56 pairs of subjects were presented in Table 1. The median (1st quartile, 3rd quartile) age of AISL population was 50.30 (44.70, 58.70) for the cases and 50.40 (44.60, 58.70) years for the controls. Both groups have the same proportion of the female population (58.93%), and there was no statistical difference in the urinary arsenic levels between the two groups. More than half of them had no history of smoking or alcohol consumption. When compared to the controls, the serum triglycerides level in AISL participants was significantly lower (P=0.041). While the other variables were similar between AISL participants and the control (P>0.05). This indicates that the participants in the two groups are comparable to some extent.

Table 1 Demographic characteristics of the study population[ξ].

Variables	AISL (n=56)	Non-AISL (n=56)	P
Clinical Characteristics			
Age (years)	50.30(44.70,58.70)	50.40(44.60,58.70)	0.425
Exposure year (years)	48.19±11.53	47.62±10.97	0.489
Body mass index (kg/m^2)	24.12±3.14	23.91±2.86	0.697
Fasting plasma glucose (mmol/L)	4.89(4.60,5.25)	5.12(4.53,5.40)	0.137
Folate (ng/mL)	4.00(3.20,5.10)	4.25(3.35,5.40)	0.392
Total homocysteine (μmol/L)	12.30(10.32,16.50)	12.67(11.21,14.69)	0.961
Blood urea nitrogen (mmol/L)	6.45(5.42,7.69)	6.84(5.36,8.80)	0.603
Total cholesterol (mmol/L)	4.58(4.10,5.69)	4.65(3.96,5.95)	0.904
Triglycerides (mmol/L)	1.41(0.90,1.74)	1.45(1.09,2.29)	0.041
High-density lipoprotein (mmol/L)	1.19±0.34	1.16±0.31	0.675
Low-density lipoprotein (mmol/L)	3.04±0.80	3.25±0.84	0.110
Women [# (%)]	33(58.93)	33(58.93)	1.000
Cigarette smoking [# (%)]	20(35.71)	22(39.29)	0.696
Alcohol consumption [# (%)]	17(30.91)	21(37.50)	0.464
Illiteracy [# (%)]	21(37.50)	15(25.00)	0.252
Urinary arsenic species[ζ]			
iAS%	12.26(8.13,14.68)	12.31(10.04,16.54)	0.148
MMA%	24.68(20.11,29.68)	25.85(20.90,31.66)	0.420
DMA%	61.84(56.62,71.01)	61.84(47.85,64.99)	0.096
tAs (μg/g creatinine)	140.93(104.41,208.53)	186.77(80.11,217.30)	0.445

[ξ] AISL: arsenic-induced skin lesions; the variables met normal distribution was described with mean± standard deviation; otherwise, median (1st quartile, 3rd quartile) was used to describe their features. Number of cases (percentage) was used to describe the proportion of categorical variables between the two groups.
*[ζ] iAS: inorganic arsenic (iAs^{III}+iAs^{V}); MMA: monomethyl arsenate (MMA^{III}+MMA^{V}); DMA: dimethyl arsenate (DMA^{III}+DMA^{V}); tAs: total arsenic (iAs^{III}+iAs^{V}+MMA+DMA); iAs%= iAS/tAs*100%; MMA%=MMA/tAs*100% and DMA%=DMA/tAs*100%.*

Table 2 shows that the four AAs, which FDR adjusted p-value <0.05 and VIP>1, in the cases are observed significantly lower than those of the controls. Two of them are aromatic amino acids (AAAs) identified as phenylalanine and tryptophan, one of them belongs to aromatic amino acids branched-chain amino acids (BCAAs) appraised as leucine and the last one is phenylalanylphenylalanine. The individual association of AAs with AISL were presented in figure1, which reveal obvious "dose-response" relationships between them.

Table 2 Distinct metabolites in population with arsenic-induced skin lesions and their counterparts.

Serum amino acid metabolites	Retention time (min)	Mass-to-Charge Ratio	VIP value	p-values[ξ]	Adjusted p-values[ζ]
Phenylalanine	3.402	166.087	1.508	<0.001	0.009
Tryptophan	3.886	203.082	1.046	0.003	0.014
Leucine	2.642	132.102	1.014	0.001	0.020
Phenylalanylphenylalanine	5.048	313.155	1.833	0.004	0.033

VIP: variable importance in the project; ξ Wilcoxon signed-rank test; ζ Adjusted by false discovery rate (FDR).

Table 3 shows that participants in the 3rd and 4th quartile of the four specific AAs were all significantly linked to the decreased odds of AISL after adjusting for FPG, LDL, TG and DMA%, as

compared to their lowest quartiles, respectively. The category boundaries of the quartiles were shown in Table S1. Significant linear trends existed between AISL and those four serum AAs. Meanwhile, the same linear negative association between AISL and per IQR rise of the four serum AAs was observed when these AAs were considered as continuous variables in the present study.

Table 3 Relationship of amino acid levels with the odds of arsenic-induced skin lesions[ξ]

Amino acids	N	Cases (%)	Crude		Adjusted[ζ]	
			OR (95% CI)	P	OR (95% CI)	P
Tryptophan						
Per IQR	112	56(50)	0.48(0.27,0.84)	0.011	0.48(0.27,0.86)	0.013
Quartiles						
Q_1	28	20(71.40)	1.00(1.00,1.00)	Ref.	1.00(1.00,1.00)	Ref.
Q_2	28	15(53.60)	0.50(0.16,1.54)	0.225	0.56(0.16,1.98)	0.370
Q_3	28	10(35.70)	0.12(0.03,0.53)	0.005	0.12(0.02,0.60)	0.010
Q_4	28	11(39.30)	0.19(0.05,0.71)	0.014	0.21(0.05,0.84)	0.028
P for trend				0.008		0.012
Phenylalanine						
Per IQR	112	56(50)	0.57(0.36,0.91)	0.019	0.56(0.33,0.94)	0.028
Quartiles						
Q_1	28	20(71.40)	1.00(1.00,1.00)	Ref.	1.00(1.00,1.00)	Ref.
Q_2	28	19(67.90)	0.79(0.22,2.82)	0.712	0.70(0.18,2.76)	0.609
Q_3	28	8(28.60)	0.18(0.05,0.66)	0.010	0.20(0.05,0.77)	0.019
Q_4	28	9(32.10)	0.25(0.08,0.79)	0.018	0.20(0.05,0.75)	0.017
P for trend				<0.001		0.001
Leucine						
Per IQR	112	56(50)	0.45(0.25,0.82)	0.019	0.43(0.21,0.86)	0.016
Quartiles						
Q_1	28	21(75.00)	1.00(1.00,1.00)	Ref.	1.00(1.00,1.00)	Ref.
Q_2	28	14(50.00)	0.33(0.11,1.03)	0.057	0.31(0.10,1.01)	0.052
Q_3	28	11(39.30)	0.22(0.07,0.68)	0.009	0.22(0.07,0.73)	0.014
Q_4	28	10(35.70)	0.19(0.06,0.59)	0.004	0.19(0.06,0.65)	0.008
P for trend				0.003		0.007
Phenylalanylphenylalanine						
Per IQR	112	56(50)	0.62(0.36,1.04)	0.070	0.71(0.41,1.24)	0.227
Quartiles						
Q_1	27	21(77.80)	1.00(1.00,1.00)	Ref.	1.00(1.00,1.00)	Ref.
Q_2	29	14(48.30)	0.16(0.03,0.75)	0.021	0.14(0.03,0.73)	0.019
Q_3	28	9(32.10)	0.08(0.02,0.39)	0.002	0.09(0.02,0.52)	0.007
Q_4	28	12(42.90)	0.12(0.02,0.57)	0.008	0.11(0.02,0.66)	0.016
P for trend				0.006		0.023

[ξ]Values are odds ratio (95% confidence intervals) for arsenic-induced skin lesions from conditional logistic regression. IQR: interquartile range; Q_1: the 1st quartile; Q_2: the 2nd quartile; Q_3: the 3rd quartile; Q_4: the 4th quartile.

[ζ] Adjusted for plasma glucose, low-density lipoprotein, triglyceride, and urinary dimethyl arsenate.

As these four specific AAs are significantly or marginal significantly associated with the odds of AISL, so it is needed to examine the joint impacts among them on AISL. However, the results of potential collinearity examination revealed that among these four specific AAs, both tryptophan and phenylalanine had the smallest VIF value (VIF=1.04) and no obvious collinearity existed (**Table S2**).

Hence, we mainly focus on tryptophan and phenylalanine when assessing the joint impacts of AAs on AISL and only presented the results associated with these two AAs in the current study. To avoid the impacts due to insufficient power because of unreasonable grouping on the results, we classified both tryptophan and phenylalanine into two categories, according to the cut-off values of their mass spectrum peak area based on the ROC analysis, respectively. The higher levels of these two serum AAs were defined as equal to or over the cut-off values, while the lower categories were considered as less than the associated values.

Table 4 shows the joint impacts of tryptophan and phenylalanine levels on AISL after considering the collinearity of variables in the model. The proportions of AISL were 74.3%, 60.0%, 50.0% and 18.2% for participants with lower levels of both tryptophan and phenylalanine (category A), with higher tryptophan and lower phenylalanine (category B), with lower tryptophan and higher phenylalanine (category C), and higher levels of both tryptophan and phenylalanine (category D), respectively. An obvious decrease trend of the probability of AISL was observed among these four categories. As compared to the category A, adjusted OR (95% CI) for participants in the category B, C and D were 0.49(0.15, 1.63), 0.32(0.10, 1.02) and 0.08(0.02, 0.25). Subjects with higher levels of both tryptophan and phenylalanine had the lowest odds of AISL, which significantly decreased by 92% (OR=0.08; 95% CI 0.02 to 0.25; P<0.001), after adjusting for the impacts induced by some potentially confounding factors. This suggested that tryptophan and phenylalanine were jointly associated with the presence of AISL. While no significant interaction between the two AAs on the occurrence of AISL could be observed (P=0.419).

Table 4 Joint association between tryptophan and phenylalanine levels with arsenic-induced skin lesions.

Tryptophan <cut-off value[ξ]	Phenylalanine <cut-off value[ξ]	N	Cases (%)	Crude		Adjusted[ζ]	
				OR (95% CI)	P	OR (95% CI)	P
Yes	Yes	35	26(74.3)	1.00(1.00,1.00)	Ref.	1.00(1.00,1.00)	Ref.
No	Yes	20	12(60.0)	0.52(0.16,1.68)	0.273	0.49(0.15,1.63)	0.244
Yes	No	24	12(50.0)	0.35(0.12,1.04)	0.059	0.32(0.10,1.02)	0.053
No	No	33	6(18.2)	0.08(0.02,0.25)	<0.001	0.08(0.02,0.25)	<0.001
Interaction					0.320	0.49(0.09,2.78)	0.419

[ξ] Cut-off value was determined using receiver operator characteristic analysis.
[ζ] Adjusted for plasma glucose, low-density lipoprotein, triglyceride, and urinary dimethyl arsenate.

Table 5 shows that, based on the ROC analysis, both serum tryptophan and phenylalanine might be potential biomarkers in distinguishing AISL from a chronic arsenic exposure population (P=0.0020 and P=0.0017). The area under the curve (AUC) and its related 95% CI, sensitivity, specificity, positive predictive value and negative predictive value were 0.67 (0.57, 0.77), 69.64%, 62.50%, 65.00 and 67.31% for tryptophan, and 0.70 (0.60, 0.80), 69.64%, 69.64%, 69.64% and 69.64% for phenylalanine, respectively. The AUC (95% CI), sensitivity, specificity, positive predictive value and negative predictive value of the combination of them were 0.72 (0.62, 0.81), 76.79%, 58.93%, 65.15 and 71.74%, respectively. Our results suggest that these two AAs could be either individually or jointly used as indicators of AISL identification.

Table 5 Combination of diagnostic indicators and ROC analysis results[ξ].

Indicators	AUC (95% CI)	Sensitivity, %	Specificity, %	Predict$^+$, %	Predict$^-$, %	P
Tryptophan	0.67(0.57,0.77)	69.64	62.50	65.00	67.31	0.002
Phenylalanine	0.70(0.60,0.80)	69.64	69.64	69.64	69.64	0.002
Combined[ζ]	0.72(0.62,0.81)	76.79	58.93	65.15	71.74	<0.001

[ξ] ROC: a receiver operator characteristic; AUC: area under the roc curve; CI: confidence interval; The sensitivities, specificity, positive predictive value, and negative predictive value were calculated at their best cut-off points; Predict+: positive predictive value; Predict-: negative predictive value.
[ζ] Combined: tryptophan and phenylalanine. The combination is modeled according to the formula $\beta_1 X_1 + \beta_2 X_2$, with X_j denoting the standardized value for the j^{th} amino acid and βj denoting the regression coefficient from the logistic regression model.

DISCUSSION

In the present study, the association of serum tryptophan and phenylalanine, screened in our previous non-targeted metabolomics study using UPLC-MS/MS, with AISL and their ability to indicate AISL occurrences were quantitatively evaluated in individual and joint modes. Our results clearly showed that AISL is significantly and negatively associated with serum tryptophan and

phenylalanine levels in a chronic arsenic exposure population via drinking water. Participants with a higher level of both AAs would have the lowest odds of AISL. These two AAs might also be able to serve as the indicators of AISL.

The probability of the initiation and development of AISL would be affected by a large number of factors including age, gender, lifestyles, arsenic exposure, metabolism, and others. These factors would be important confounding factors and will largely affect our results. To adjust for the impacts due to these co-factors, we firstly selected all participants using permuted block randomization from a single rural area in which the population were chronically exposed to arsenic in the same way, had a similar lifestyle and environmental factors. Secondly, the cases and controls were matched by gender and age (±1 year). All of these may be the reason why so many potential confounders including arsenic exposure do not differ significantly between the cases and controls.

Participants enrolled in the current study were chronically exposed to arsenic via drinking water. The geometric mean (GM) and its related 95% CI of urinary iAs/creatinine and tAs/creatinine in this population were 17.49 (14.90, 20.53) μg/g and 147.20 (129.00, 167.97) μg/g, respectively. They were much higher than those in the 20 μg/L exposed to arsenic via drinking water [GM (95% CI): 0.4 (0.3,0.5) μg/g for iAs and 9.1 (6.5,12.7) μg/g for tAs], while obviously lower than those in the 90 μg/L exposed group [GM (95% CI): 39.4 (31.4, 49.6) μg/g for iAs and 248.7 (208.8, 296.3) μg/g for tAs][25]. An available report has shown that AISL cannot be completely cured even though medical technology has already made great progress[26]. So, it is crucial to identify those who are most likely to progress to overt arsenic damages including AISL among people at risk as early as possible. Metabolomics study, which mainly focuses on thoroughly assessing the variation of metabolites possibly linked to diseases occurrence and development, has been widely utilized to help us understand the pathogenesis of diseases because of its relevance to the phenotypes as compared to other 'OMICs' study[27]. Moreover, mathematical modeling to assess the linkage between small molecular metabolites and arsenic toxicity has grown[28]. Developing a simple and interpretable

modeling approach for the early detection of arsenic-induced health lesions is of great theoretical value and realistic meaning[29], though it might be difficult due to population-specific complexities and the impacts due to some potential unmeasured covariates such as diet and genetic determinants.

Previous studies reported that gene-gene and gene-environment interaction were involved in arsenicosis through toxicological mechanisms including genomic instability[30] and oxidative stress[31]. Skin hyperpigmentation and palmoplantar hyperkeratosis could be biomarkers for long-term arsenic exposure identifying the differences in metabolites that are associated with phenotypes through metabolites analysis may promote our understanding and identification of AISL. Animal study reveals that the disruption of amino acids metabolism upon arsenic exposure in the rat which may be beneficial in understanding arsenic toxicity[32]. In our previous population-based metabolomics study, we found that serum metabolites alteration was significantly related to the risk of arsenic-induced health damages. In the current study, we detected that some BCAAs or AAAs were also significantly relevant to AISL occurrence. Several studies across numerous ethnic backgrounds support the usage of BCAAs including leucine, isoleucine as well as valine and AAAs profile such as phenylalanine, tryptophan, and tyrosine as biomarkers in determining metabolic diseases[27 33]. Simultaneously, Zhou *et al* reported that arsenic-induced transformed cells exhibit apparent alterations in metabolite profiles including down-regulated of leucine, tryptophan, and phenylalanine in skin lesions group[34]. Consistent with Zhou's findings, two serum AAAs (tryptophan, phenylalanine) levels were also significantly associated with AISL in our study.

Normal metabolism of amino acids is necessary for whole-body homeostasis, growth and development, and health status[15]. Studies have reported that the changes in the availability of AAAs will affect cell signaling, gene expression, brain, and neuroendocrine function[35]. Tryptophan, an amino acid metabolism-related biomarker, is also a sensitive and specific indicator of oxidation. Tryptophan metabolism in mammals is a physiological means of preserving immune homeostasis associated with oxidative stress and inflammation[36 37]. Besides, phenylalanine can be transformed

into specific neurotransmitters such as dopamine and adrenaline by the action of related enzymes. Wu and colleagues[38] reported that arsenic exposure would lead to neurotransmitter metabolism disturbance, which might explain the reduction of phenylalanine. Furthermore, as one of the peptide-bound phenylalanine, phenylalanylphenylalanine has been reported to affect protein synthesis and secretion[39], potentially indicating the possible relation between endothelium dysfunction and phenylalanine metabolism disorder. The relationship between amino acid metabolism and AISL was still unclear. The notable alteration of tryptophan and phenylalanine in this study may well indicate the occurrence of metabolic disorders due to arsenic exposure. It is also beneficial to understand the effects of arsenic toxicity and of great importance in early identification of occurrences as well as delaying the progression of various arsenic-induced health lesions including AISL.

The current study included 56 AISL cases matched 56 non-AISL controls and the sample size might be potentially insufficient. To estimate the impact due to this potential insufficient sample size on our conclusion, the PROC POWER procedure for paired design study in SAS 9.4 (SAS Institute Inc.) was applied to assess the power of the four AAs when assessing their associations with AISL in this study. The results showed that the lowest power associated with all of these four AAs was 0.911 based on 56 pairs of participants (Figure 1S). It suggested that with type I error as 0.05, the total sample size as 56 pairs and two-sided test, the powers associated with these four AAs were all great than 0.8. So, we believed that the sample size for the present study, 56 pairs, would well balance the power of tests. Furthermore, previous metabolomics studies usually have a sample size of no more than 40 cases in each group[40 41].

The main strength of this study is that the findings were based on a community-based, long-term arsenic exposure cohort with well-designed quality assurance and quality control throughout the study, and the AAs were detected with non-targeted metabolomics approach through the discovery and validation phases. However, there are also several limitations to this study. Firstly, although the

untargeted metabolomics approach can assess a large amount number of metabolites precisely and efficiently, it only provided relative levels of AAs instead of their accurate quantitative concentration. Secondly, these findings were mainly based on a case-control study, which only reveals the association between amino acid metabolism and the odds of AISL rather than confirming their causal relationship. Furthermore, as it is suggested that the ratio of approximately 10 to 15 observations per predictor in a logistic regression model will produce reasonably stable estimations[24], we selected only 4 covariates in the models due to the small sample size and these results need to be confirmed in new studies. Finally, the participants were mainly exposed to arsenic via drinking water, which would limit the findings extrapolated to the other arsenic exposure population via food and other ways. Therefore, additional elaborate population-based studies are needed to verify our discoveries.

In conclusion, specific AAs might be linked to AISL and amino acid metabolism may play an important role in AISL early identification. Additional studies may be needed to confirm our findings.

Contributors: Guangyun Mao and Yaping Wei designed the study. Chaonan Jia participated in collecting data. Yuan Lan and Chaonan Jia audited the data. Yaping Wei, Xiangqing hou, Jushuang Li, Tao Wang conducted the literature search, Yaping Wei, Chaonan Jia conducted statistical analysis and interpreted the results. Yaping Wei and Chaonan Jia wrote the first draft of the manuscript. Jingjing Zuo helped with copyediting. Guangyun Mao reviewed the final manuscript and did substantial contributions.

Funding: This work was supported by the basic public welfare research project of Zhejiang province (LGF19H260011), the initial scientific research fund (KYQD170301) and the major project of the eye hospital of Wenzhou medical university (YNZD201602).

Acknowledgments: We gratefully acknowledge all of the study participants of this project. We wish to thank all of the dedication and hard work of the field team at the department of preventive medicine, school of public health and management, Wenzhou medical university. Our thanks must also go to Dr. Kendrick Hii, a native-English speaker, for his valuable suggestions and edits to the manuscript.

Competing Interest: None declared.

Ethics approval: Ethical approval was received from the ethics committee of Wenzhou Medical University, Wenzhou, China.

Provenance and peer review: Not commissioned; externally peer reviewed.

Data Sharing Statement: No additional unpublished data are available.

Patient consent for publication: Not required.

REFERENCES

1. Mandal BK, Suzuki KT. Arsenic round the world: a review. Talanta. 2002; 58:201-35.
2. Smith AH, Steinmaus CM. Health Effects of Arsenic and Chromium in Drinking Water: Recent Human Findings. Annual Review Of Public Health. Palo Alto: Annual Reviews, 2009:107-22.
3. Guha Mazumder DN, Haque R, Ghosh N, et al. Arsenic levels in drinking water and the prevalence of skin lesions in West Bengal, India. International journal of epidemiology. 1998; 27:871-7.
4. Guo XJ, Fujino Y, Kaneko S, et al. Arsenic contamination of groundwater and prevalence of arsenical dermatosis in the Hetao plain area, Inner Mongolia, China. Mol Cell Biochem. 2001; 222:137-40.
5. Lindberg A-L, Rahman M, Persson L-A, et al. The risk of arsenic induced skin lesions in Bangladeshi men and women is affected by arsenic metabolism and the age at first exposure. Toxicology and Applied Pharmacology. 2008; 230:9-16.
6. Tondel M, Rahman M, Magnuson A, et al. The relationship of arsenic levels in drinking water and the prevalence rate of skin lesions in Bangladesh. Environmental Health Perspectives. 1999; 107:727-29.
7. Hsu LI, Chen GS, Lee CH, et al. Use of Arsenic-Induced Palmoplantar Hyperkeratosis and Skin Cancers to Predict Risk of Subsequent Internal Malignancy. Am J Epidemiol. 2013; 177:202-12.
8. Luo LR, Li YY, Gao YH, et al. Association between arsenic metabolism gene polymorphisms and arsenic-induced skin lesions in individuals exposed to high-dose inorganic arsenic in northwest China. Sci Rep. 2018; 8:12.
9. Hughes MF, Beck BD, Chen Y, et al. Arsenic Exposure and Toxicology: A Historical Perspective. Toxicol Sci. 2011; 123:305-32.
10. Ren X, McHale CM, Skibola CF, et al. An Emerging Role for Epigenetic Dysregulation in Arsenic Toxicity and Carcinogenesis. Environmental Health Perspectives. 2011; 119:11-19.
11. Chen J-W, Wang S-L, Wang Y-H, et al. Arsenic methylation, GSTO1 polymorphisms, and metabolic syndrome in an arseniasis endemic area of southwestern Taiwan. Chemosphere. 2012; 88:432-38.
12. Wang S-L, Chang F-H, Liou S-H, et al. Inorganic arsenic exposure and its relation to metabolic syndrome in an industrial area of Taiwan. Environment International. 2007; 33:805-11.
13. Liu R, Li H, Fan WJ, et al. Leucine Supplementation Differently Modulates Branched-Chain Amino Acid Catabolism, Mitochondrial Function and Metabolic Profiles at the Different Stage of Insulin Resistance in Rats on High-Fat Diet. Nutrients. 2017; 9:20.
14. Platten M, Wick W, Van den Eynde BJ. Tryptophan Catabolism in Cancer: Beyond IDO and Tryptophan Depletion. Cancer Res. 2012; 72:5435-40.
15. Wu G. Amino acids: metabolism, functions, and nutrition. Amino Acids. 2009; 37:1-17.
16. Casadei L, Valerio M, Manetti C. Metabolomics: Challenges and Opportunities in Systems Biology Studies. Methods Mol Biol. 2018; 1702:327-36.
17. Johnson CH, Gonzalez FJ. Challenges and opportunities of metabolomics. J Cell Physiol. 2012; 227:2975-81.
18. Hui AY, McCarty WJ, Masuda K, et al. A systems biology approach to synovial joint lubrication in health, injury, and disease. Wiley Interdiscip Rev Syst Biol Med. 2012; 4:15-37.
19. Lu K, Abo RP, Schlieper KA, et al. Arsenic Exposure Perturbs the Gut Microbiome and Its Metabolic Profile in Mice: An Integrated Metagenomics and Metabolomics Analysis. Environmental Health Perspectives. 2014; 122:284-91.
20. Zhang J, Shen H, Xu W, et al. Urinary Metabolomics Revealed Arsenic Internal Dose-Related Metabolic Alterations: A Proof-of-Concept Study in a Chinese Male Cohort. Environmental Science & Technology. 2014; 48:12265-74.
21. Guo X, Cui H, Zhang H, et al. Protective Effect of Folic Acid on Oxidative DNA Damage: A Randomized, Double-Blind, and Placebo Controlled Clinical Trial. Medicine (Baltimore). 2015; 94:e1872.
22. Standard of diagnosis for endemic arsenism of China (WS/T211-2001). the Ministry of Health of the People's Republic of China. Beijing 2001.
23. Chen X, Guo X, He P, et al. Interactive Influence of N6AMT1 and As3MT Genetic Variations on Arsenic Metabolism in the Population of Inner Mongolia, China. Toxicol Sci. 2017; 155:124-34.
24. Peduzzi P, Concato J, Kemper E, et al. A simulation study of the number of events per variable in logistic

regression analysis. J Clin Epidemiol. 1996; 49:1373-9.

25. Sun G, Xu Y, Li X, et al. Urinary arsenic metabolites in children and adults exposed to arsenic in drinking water in Inner Mongolia, China. Environmental Health Perspectives. 2007; 115:648-52.
26. Kuo CC, Moon KA, Wang SL, et al. The Association of Arsenic Metabolism with Cancer, Cardiovascular Disease, and Diabetes: A Systematic Review of the Epidemiological Evidence. Environmental Health Perspectives. 2017; 125:15.
27. Wang TJ, Larson MG, Vasan RS, et al. Metabolite profiles and the risk of developing diabetes. Nature Medicine. 2011; 17:448-U83.
28. Lawley SD, Cinderella M, Hall MN, et al. Mathematical model insights into arsenic detoxification. Theor Biol Med Model. 2011; 8.
29. Juhasz AL, Herde P, Herde C, et al. Validation of the predictive capabilities of the Sbrc-G in vitro assay for estimating arsenic relative bioavailability in contaminated soils. Environ Sci Technol. 2014; 48:12962-9.
30. Kibriya MG, Jasmine F, Parvez F, et al. Association between genome-wide copy number variation and arsenic-induced skin lesions: a prospective study. Environ Health. 2017; 16:13.
31. Priori R, Scrivo R, Brandt J, et al. Metabolomics in rheumatic diseases: the potential of an emerging methodology for improved patient diagnosis, prognosis, and treatment efficacy. Autoimmun Rev. 2013; 12:1022-30.
32. Wang X, Mu X, Zhang J, et al. Serum metabolomics reveals that arsenic exposure disrupted lipid and amino acid metabolism in rats: a step forward in understanding chronic arsenic toxicity. Metallomics Integrated Biometal Science. 2015; 7:544.
33. Weng L, Quinlivan E, Gong Y, et al. Association of Branched and Aromatic Amino Acids Levels with Metabolic Syndrome and Impaired Fasting Glucose in Hypertensive Patients. Metabolic Syndrome And Related Disorders. 2015; 13:195-202.
34. Zhou YY, Wang YF, Su J, et al. Integration of microRNAome, proteomics and metabolomics to analyze arsenic-induced malignant cell transformation. Oncotarget. 2017; 8:90879-96.
35. Lee JC, Chen MJ, Chang CH, et al. Plasma amino acid levels in patients with colorectal cancers and liver cirrhosis with hepatocellular carcinoma. Hepato-Gastroenterol. 2003; 50:1269-73.
36. Munn DH, Mellor AL. Indoleamine 2,3 dioxygenase and metabolic control of immune responses. Trends In Immunology. 2013; 34:137-43.
37. Pawlak K, Domaniewski T, Mysliwiec M, et al. The kynurenines are associated with oxidative stress, inflammation and the prevalence of cardiovascular disease in patients with end-stage renal disease. Atherosclerosis. 2009; 204:309-14.
38. Wu L-L, Gong W, Shen S-P, et al. Multiple metal exposures and their correlation with monoamine neurotransmitter metabolism in Chinese electroplating workers. Chemosphere. 2017; 182:745-52.
39. Zhou MM, Wu YM, Liu HY, et al. Effects of phenylalanine and threonine oligopeptides on milk protein synthesis in cultured bovine mammary epithelial cells. J Anim Physiol Anim Nutr. 2015; 99:215-20.
40. Chen LY, Cheng CY, Choi HW, et al. Plasma Metabonomic Profiling of Diabetic Retinopathy. Diabetes. 2016; 65:1099-108.
41. Piszcz J, Lemancewicz D, Dudzik D, et al. Differences and similarities between LC-MS derived serum fingerprints of patients with B-cell malignancies. Electrophoresis. 2013; 34:2857-64.

Figure 1. Association between the peak intensity of tryptophan and phenylalanine and arsenic-induced skin lesions based on multivariable locally weighted regression models. a: Tryptophan; b: Phenylalanine;c: Leucine;d: Phenylalanylphenylalanine

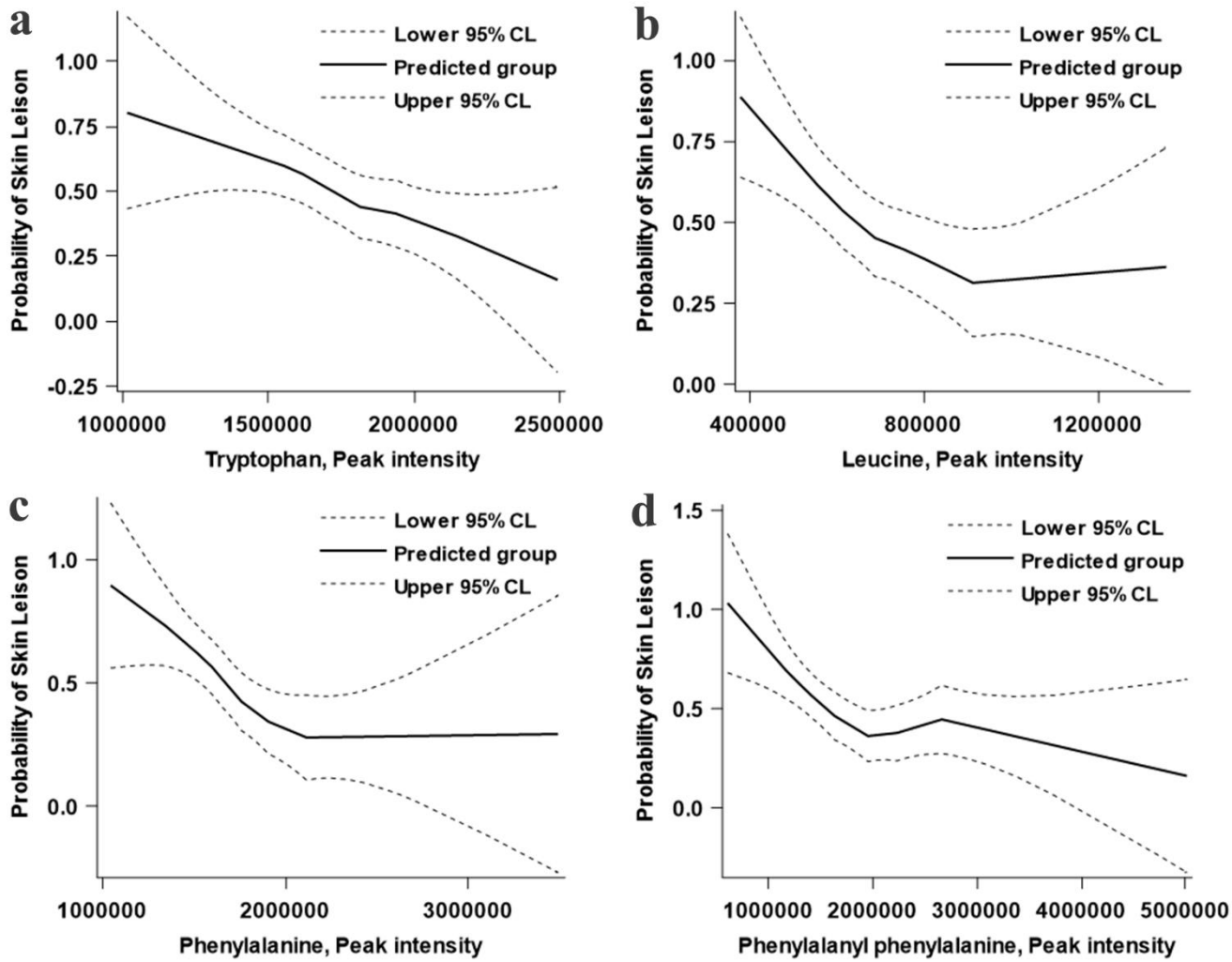

Figure 1. Association between the peak intensity of tryptophan and phenylalanine and arsenic-induced skin lesions based on multivariable locally weighted regression models. a: Tryptophan; b: Phenylalanine;c: Leucine;d: Phenylalanylphenylalanine

Table S1 Amino acids metabolites in serum.

Amino acids	Quartile	Quartile Range	Number
Tryptophan	1	1015821.54-1555040.97	28
	2	1555370.88-1701256.07	28
	3	1704298.84-1937437.56	28
	4	1945755.37-2492319.05	28
Phenylalanine	1	1044964.46-1516708.76	28
	2	1519993.60-1649420.82	28
	3	1659818.85-1908125.28	28
	4	1929946.22-3489918.62	28
Leucine	1	379957.04-556314.46	28
	2	558744.08-688759.83	28
	3	689470.83-806033.16	28
	4	821492.90-1356991.53	28
Phenylalanylphenylalanine	1	608200.15-1291853.09	27
	2	1303597.59-1632678.66	29
	3	1638204.36-2234283.97	28
	4	2235250.29-5011775.70	28

Table S2 Variance inflation factor of amino acids in different models[ξ].

Amino acids	Model 1	Model 2	Model 3	Model 4
Tryptophan	1.04	1.04	1.04	1.04
Phenylalanine	4.56		1.50	1.04
Leucine	4.21	1.39		
Phenylalanylphenylalanine	1.49	1.38	1.48	

[ξ]*Model 1: Tryptophan, phenylalanine, leucine and phenylalanylphenylalanine;*
Model 2: Tryptophan, leucine and phenylalanylphenylalanine;
Model 3: Tryptophan, phenylalanine and phenylalanylphenylalanine;
Model 4: Tryptophan and phenylalanine.

SAS 系统

The POWER Procedure
Paired t Test for Mean Difference

Fixed Scenario Elements	
Distribution	Normal
Method	Exact
Mean 1	768959
Mean 2	656243.1
Standard Deviation 1	206347.3
Standard Deviation 2	216983.4
Correlation	0.4
Number of Pairs	56
Number of Sides	2
Null Difference	0
Alpha	0.05

Computed Power
Power
0.946

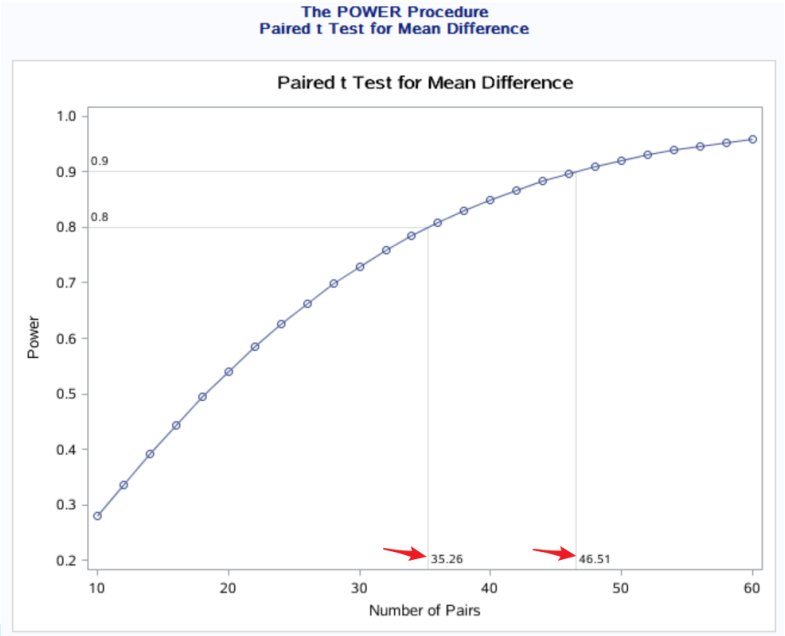

Figure 1S_A. Sample size and power estimation__Leucine

SAS 系统

The POWER Procedure
Paired t Test for Mean Difference

Fixed Scenario Elements	
Distribution	Normal
Method	Exact
Mean 1	1865229
Mean 2	1637107
Standard Deviation 1	420123.2
Standard Deviation 2	356546.1
Correlation	0.4
Number of Pairs	56
Number of Sides	2
Null Difference	0
Alpha	0.05

Computed Power
Power
0.975

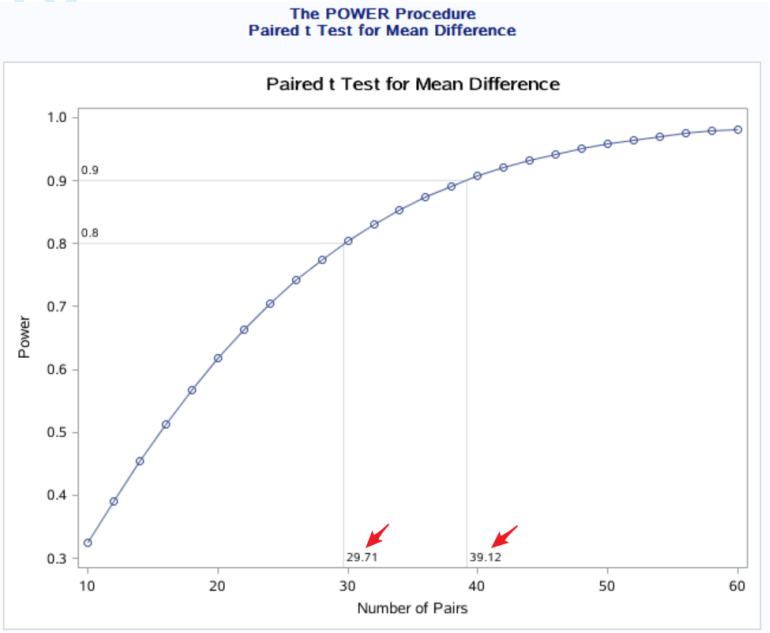

Figure 1S_B. Sample size and power estimation__Phenylalanine

SAS 系统

The POWER Procedure
Paired t Test for Mean Difference

Fixed Scenario Elements	
Distribution	Normal
Method	Exact
Mean 1	2056703
Mean 2	1664489
Standard Deviation 1	880518
Standard Deviation 2	682913.7
Correlation	0.4
Number of Pairs	56
Number of Sides	2
Null Difference	0
Alpha	0.05

Computed Power
Power
0.911

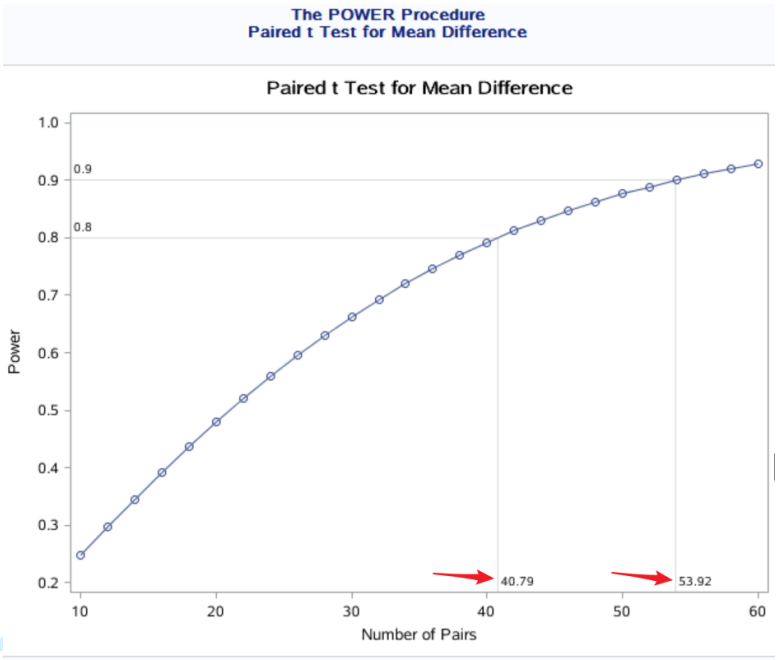

Figure 1S_C. Sample size and power estimation__Phenylalanylphenylalanine

SAS 系统

The POWER Procedure
Paired t Test for Mean Difference

Fixed Scenario Elements	
Distribution	Normal
Method	Exact
Mean 1	1832402
Mean 2	1655560
Standard Deviation 1	306551.7
Standard Deviation 2	289730.1
Correlation	0.4
Number of Pairs	56
Number of Sides	2
Null Difference	0
Alpha	0.05

Computed Power
Power
0.978

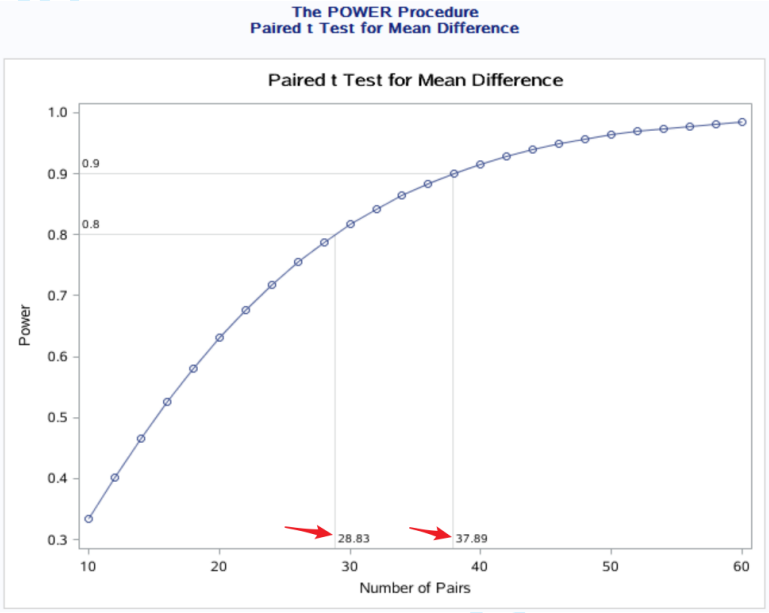

Figure 1S_D. Sample size and power estimation__Tryptophan

Table1 The STROBE checklist in this study.

	Item No	Recommendation
Title and abstract	1	(a) Indicate the study's design with a commonly used term in the title or the abstract (b) Provide in the abstract an informative and balanced summary of what was done and what was found **Please see detail in the "ABSTRACT" section in the manuscript.**
Introduction		
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported **Please see detail in the second, third, fourth paragraphs of the "INTRODUCTION" section in the manuscript.**
Objectives	3	State specific objectives, including any prespecified hypotheses **Please see detail in the fourth paragraph of the "INTRODUCTION" section in the manuscript.**
Methods		
Study design	4	Present key elements of study design early in the paper **Please see detail in the "Study Population "of the "METHODS" section in the manuscript.**
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection **Please see detail in the "Study Population "of the "METHODS" section in the manuscript.**
Participants	6	(a) Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls (*b*) For matched studies, give matching criteria and the number of controls per case **Please see detail in the "Study Population "of the "METHODS" section in the manuscript.**
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable **Please see detail in the "Study Population ", "Data Collection and Assessment" and "Distinct Metabolites Identification" of the "METHODS" section in the manuscript.**
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group **Please see detail in the "Study Population ", "Data Collection and**

		Assessment" and "Distinct Metabolites Identification" of the "METHODS" section in the manuscript.
Bias	9	Describe any efforts to address potential sources of bias **Please see detail in the "Study Population "of the "METHODS" section in the manuscript.**
Study size	10	Explain how the study size was arrived at **Please see detail in the sixth paragraph of the of the "DISCUSSION" section in the manuscript.**
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why **Please see detail in the "Statistical Analysis" of the "METHODS" section in the manuscript.**
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding (*b*) Describe any methods used to examine subgroups and interactions (*c*) Explain how missing data were addressed (*d*) If applicable, explain how matching of cases and controls was addressed (*e*) Describe any sensitivity analyses **No missing values were observed in our database. Others please see detail in the "Statistical Analysis" of the "METHODS" section in the manuscript.**
Results		
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed; (b) Give reasons for non-participation at each stage; (c) Consider use of a flow diagram **Please see detail in the "Study Population "of the "METHODS" section in the manuscript.**
Descriptive data	**14***	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders (b) Indicate number of participants with missing data for each variable of interest **See Table 1 in the manuscript.**
Outcome data	15*	Report numbers in each exposure category, or summary measures of exposure **Please see Table 3 in the manuscript.**
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were

		included **Please see Table 3 in the manuscript.**
		(*b*) Report category boundaries when continuous variables were categorized **Please see Table S1 in the manuscript.**
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses **Please see Table 4 in the manuscript.**
Discussion		
Key results	18	Summarise key results with reference to study objectives **Please see detail in the first paragraph of the of the "DISCUSSION" section in the manuscript.**
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias **Please see detail in the seventh paragraph of the of the "DISCUSSION" section in the manuscript.**
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence **Please see detail in the fifth paragraph of the of the "DISCUSSION" section in the manuscript.**
Generalisability	21	Discuss the generalisability (external validity) of the study results **Please see detail in the seventh paragraph of the "DISCUSSION" section in the manuscript.**
Other information		
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based **Please see detail in the "Funding" section in the manuscript.**

*Give information separately for cases and controls.

BMJ Open

Tryptophan and phenylalanine are associated with arsenic–induced skin lesions in a Chinese population chronically exposed to arsenic via drinking water: a case–control study

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2018-025336.R5
Article Type:	Original research
Date Submitted by the Author:	03-Sep-2019
Complete List of Authors:	wei, yaping; Wenzhou Medical University, Chashan University town, Ouhai district, Wenzhou, China., Jia, Chaonan; School of Public Health & Management, Wenzhou Medical University, Department of Preventive Medicine Lan, Yuan; School of Ophthalmology & Optometry, Wenzhou Medical University, Center on Clinical Research, the Eye Hospital of Wenzhou Medical University Hou, Xiangqing; School of Public Health & Management, Wenzhou Medical Univeristy, Department of Preventive Medicine Zuo, Jingjing; Wenzhou Medical University School of Optometry and Ophthalmology, Center on Clinical Research, the Eye Hospital of Wenzhou Medical University Li, Jushuang; School of Public Health & Management, Wenzhou Medical Univeristy, Department of Preventive Medicine Wang, Tao; School of Public Health & Management, Wenzhou Medical University, Department of Preventive Medicine Mao, Guangyun; School of Public Health & Management, Preventive Medicine
Primary Subject Heading:	Occupational and environmental medicine
Secondary Subject Heading:	Epidemiology, Public health, Occupational and environmental medicine
Keywords:	Metabolomics, Chronic arsenic exposure, Skin lesions, Amino acid

Tryptophan and phenylalanine are associated with arsenic–induced skin lesions in a Chinese population chronically exposed to arsenic via drinking water: a case–control study

Yaping Wei[1,2], Chaonan Jia,[1,2], Yuan Lan[3], Xiangqing Hou[1,2], Jingjing Zuo[3], Jushuang Li[1,2], Tao Wang[1,2], Guangyun Mao[1,2,4]✉

[1] Department of Preventive Medicine, School of Public Health and Management, Wenzhou Medical University, Wenzhou, China.

[2] Center on Evidence–Based Medicine & Clinical Epidemiological Research, School of Public Health, Wenzhou Medical University, Wenzhou, China.

[3] School of Ophthalmology & Optometry, Wenzhou Medical University, Wenzhou, China.

[4] Center on Clinical Research, the Affiliated Eye Hospital of Wenzhou Medical University, Wenzhou, China.

Yaping Wei and Chaonan Jia contributed equally to this paper.

✉Corresponding author: Guangyun Mao, Room 7A321, School of Public Health and Management, Wenzhou Medical University, Chashan University Town, Ouhai District, Wenzhou, China. 325035

E-mail: maogy@wmu.edu.cn

Telephone: +86 - 577 - 86689848

Fax numbers: +86 - 577 - 86699122

ABSTRACT

Objectives We investigated the association of specific serum amino acids (AAs) with the odds of arsenic–induced skin lesions (AISL) and their ability to distinguish AISL patients from people chronically exposed to arsenic.

Design Case–control study.

Setting Three arsenic–exposed villages in Wuyuan County, Hetao Plain, Inner Mongolia, China were evaluated.

Participants Among the 450 residents aged 18–79 years, who were chronically exposed to arsenic via drinking water, 56 were diagnosed as having AISL (defined as cases). Another 56 participants without AISL, matched by gender and age (±1 year) from the same population, were examined as controls.

Main outcome measures and methods AA levels were determined by ultra–high–performance liquid chromatography–quadrupole time–of–flight mass spectrometry–based metabolomics analysis. Potential confounding variables were identified via a standardized questionnaire and clinical examination. Multivariable conditional logistic regression model and receiver operating characteristic curve analyses were performed to investigate the relationship between specific AAs and AISL.

Results Tryptophan and phenylalanine levels were negatively associated with AISL ($P < 0.05$). Compared with that in the first quartile, the adjusted odds ratio of AISL in the second, third, and fourth quartiles were decreased by 44%, 88%, and 79% for tryptophan and 30%, 80%, and 80% for phenylalanine, respectively. The combination of these two higher–level AAs showed the lowest odds ratio for AISL (OR = 0.08; 95% CI 0.02–0.25; $P < 0.001$). Furthermore, both AAs showed a moderate ability to distinguish patients with AISL from the control, with the area under the curve [(AUC), 95% CI] as 0.67 (0.57–0.77) for tryptophan and 0.70 (0.60–0.80) for phenylalanine ($P <$

0.05). The combined pattern with AUC (95% CI) was 0.72 (0.62–0.81), showing a sensitivity of 76.79% and specificity of 58.93% ($P < 0.001$).

Conclusions Specific AAs may be linked to AISL and play important roles in early AISL identification.

Keywords: Chronic arsenic exposure; Skin lesions; Metabolomics; Amino acid.

Strengths and limitations of this study

- Our findings were based on a community–based metabolomics study with paired design, and strict quality assurance and quality control.
- Multivariable conditional logistic models were used to examine the association between specific levels of AAs and AISL, and the ROC analysis was applied to evaluate the feasibility of AAs to distinguish patients with AISL from their counterparts.
- Although AAs were determined by untargeted metabolomics approach, which can assess a large number of metabolites precisely and efficiently, only relative levels of AAs could be obtained instead of their accurate quantitative concentration.
- Based on a case control–study, the findings only revealed the association between AAs and the odds of AISL rather than confirming their causal relationship.
- The participants were mainly chronically exposed to arsenic via drinking water, which may limit the findings extrapolated to another arsenic exposure population via food or other routes.

INTRODUCTION

Chronic arsenic exposure via drinking water is widely considered a global health concern affecting several people worldwide. This exposure may cause various human health issues such as cardiovascular diseases, diabetes, and cancer[1 2]. With the industrial boom and considerable increase in global water pollution including arsenic contamination in the past, the prevalence and burden of arsenic–induced health damage will continue to increase. The skin has been confirmed as one of the most common and susceptible targets of arsenic–induced health effects. Cutaneous skin lesions are typical signs of arsenicosis after persistent and long–term arsenic exposure. These lesions are characterized by hyperkeratosis and hyperpigmentation. Considerable evidence of the prevalence of arsenic–induced skin lesions (AISL) has been reported in several countries[3–5].

As AISL are widely accepted as the major early manifestation of arsenic toxicity[6] and may be an indicator of susceptibility to more serious arsenic–induced health hazards[7], it is critical to identify people who are at risk as early as possible to prevent the onset or delay the progression of serious health problems. Several possible mechanisms such as genetic differences[8], oxidative stress[9], and epigenetic dysregulation[10] may explain arsenic poisoning. Previous studies also suggested that arsenic methylation in vivo is associated with metabolic syndrome[11 12].

Amino acids (AAs) are the "basic unit" of all proteins and are necessary to maintain health. Some AAs are important regulators of key metabolic pathways and also help in maximizing food utilization, enhancing protein accretion, and improving health[13 14]. Abnormal metabolism of AAs can disturb homeostasis in the body, impair growth and development, and even cause death[15]. Thus, the levels of serum AAs may be important indicators of metabolic status and disease condition. As powerful tools in system biology research, metabolomics approaches are beneficial for unbiased monitoring of changes in endogenous metabolism–related physiological processes, providing integrative information on distinct features across multiple functional levels. These methods capture the core attributes responsible for various phenotypes, which are particularly important in

understanding the relevant pathophysiological changes of a disease and its status, and in identifying novel biomarkers for risk screening, diagnosis, treatment, and prognosis of important human diseases[16–18].

Animal experiments and epidemiological studies have reported obvious arsenic–related metabolomics perturbations[19 20]. These results suggest that the relationship between specific metabolites and arsenic–induced health lesions should be investigated. However, only a few studies have been conducted to comprehensively examine the metabolic mechanism relevant to AISL, particularly for AA metabolism. This study was conducted to quantitatively examine the association of several specific AAs with AISL and their ability to identify AISL.

METHODS

Study Population

The study data were originally obtained from a randomized, double–blind, and placebo–controlled clinical trial (NCT02235948) performed in 2010, in which all subjects were randomly selected by permuted block randomization from a single rural area in which a population was chronically exposed to low–level arsenic drinking water, had similar lifestyles, and were influenced by similar environmental factors. Information on the inclusion and exclusion criteria of the participants can be found in our previous study[21]. Strictly following the criteria of arsenicosis[22], AISL was diagnosed as the presence of arsenic–induced keratosis, hyperpigmentation, or depigmentation by a physician from the Wenzhou Medical University at the beginning of the trial. This was a matched case–control study (1:1 matching). Among 450 residents aged 18–79 years enrolled in the previous trial, 56 were diagnosed as having AISL and selected as the case group. Another 56 participants without AISL matched by gender and age (±1 year) from the same population were evaluated as controls. The inclusion criteria were subjects who underwent a metabolomic test. Unmatched participants and those without serum metabolites data were excluded.

Data Collection and Assessment

Information on age, gender, exposure year, body mass index, smoking, alcohol consumption, and education level, among other factors, was collected using a standardized questionnaire. Blood and urine samples were also collected at the time of participants' enrollment. The detailed methods for analyzing blood and urine samples and assessment methods for clinical variables including fasting plasma glucose (FPG), serum urea nitrogen, serum folate, total homocysteine, total cholesterol (TC), triglycerides (TG), high–density lipoprotein (HDL), low–density lipoprotein (LDL), and others have been published previously[21]. Various urinary arsenic species were separated and detected by high–performance liquid chromatography coupled mass spectrometry system[23]. The species of arsenic in urine samples consisted of inorganic arsenic (iAs, [$iAs^{Ⅲ}$ plus $iAs^{Ⅴ}$]), monomethyl arsenate (MMA, [$MMA^{Ⅲ}$ plus $MMA^{Ⅴ}$]), and dimethyl arsenate (DMA, [$DMA^{Ⅲ}$ plus $DMA^{Ⅴ}$]). All arsenic species were corrected by creatinine. The total arsenic (tAs) was the sum of iAs, MMA and DMA. The percentages of arsenic species were defined as: iAs% = iAs/tAs*100%, MMA% = MMA/tAs*100% and DMA% = DMA/tAs*100%, respectively.

UPLC–QTOF–MS Metabonomic Profiling

Serum samples were thawed to 4°C, and then 600 μL of a mixture of 90% acetonitrile and 10%water was added to each sample (200 μL in microcentrifuge tubes). The samples were vigorously mixed for 20 s and then centrifuged for 5 min at 12,000 g(20°C). Four hundred microliters of each supernatant were transferred to a new tube and dried in a vacuum concentrator centrifuge. The dried samples were re–suspended in 130 μL of water (containing 15% acetonitrile), mixed vigorously for 20-s, and-centrifuged as described above. Two microliters of the supernatant were collected for analysis. Serum metabolic profile was acquired using ACQUITY UPLC®/Xevo® G2 QTof/MS^E (Waters Corp., Milford, MA, USA). Chromatographic separation was performed at 50 °C using a Waters HSS T3 column (2.1 mm × 100 mm, 1.7 μm; Milford, MA, USA) at a flow rate of 0.4 mL/min. The mobile phase was a mixture of (A) H_2O with 0.1% formic acid and (B) methanol with 0.1% formic acid. Elution was performed over a linear gradient-as follows: 0 min with 100% A

and 0% B, 1 min with 100% A and 0% B, 8 min with 0%A and 100% B, and 13 min with 0% A and 100% B. The mass spectrometer was operated under both electrospray ionization positive–ion (ESI^+) mode and negative–ion (ESI^-) mode. The scan range was 50–1200 m/z. The capillary voltage was set to 3000 and 2500 V, respectively. The desolvation flow rate was 800 L/h at 350°C. Argon was used as the collision gas, and the collision energy was adjusted from 10 to 40 eV in each analysis. Quantum clustering (QC) samples were prepared by pooling aliquots of each sample and used to reflect the reliability of further metabolomics analysis. After peak deconvolution, alignment, integration, and normalization, the data including retention time (RT), mass to charge ratio (m/z), and peak intensity were extracted from raw chromatograms using Progenesis QI 2.0 (Waters Corp., Milford, MA, USA). MS/MS was performed to determine metabolite levels using MarkerLynx Applications Manager Version 4.1 (Waters Corp., Milford, MA, USA).

Distinct Metabolite Identification

The peak intensities of metabolites for the 56 pairs of subjects were acquired and imported into MetaboAnalyst 4.0 (http://www.metaboanalyst.ca/) for statistical analyses. A partial least–squares discriminant analysis (PLS–DA), which is a supervised and well–accepted pattern recognition approach, was used to differentiate between the cases and controls. The false discovery rate (FDR) adjusted P–value in univariate analysis was determined to reduce the potential effect of false–positive results. The criteria used to select metabolites included variable importance in projection (VIP) scores >1 using PLS–DA and a crude or FDR–adjusted P–value < 0.05 using Wilcoxon signed–rank test. We identified 70 extracted small molecular metabolites linked to the recognition of AISL. The Human Metabolome Database (http://www.hmdb.ca) was used to identify the metabolites, which include four amino acid metabolites (phenylalanine, tryptophan, leucine, and phenylalanylphenylalanine).

Statistical Analysis

The normality of continuous data was assessed using both QQ–plots and Shapiro–Wilk test. The

data of cases and controls were analyzed using paired *t*–test if they showed a normal or similar normal distribution. Otherwise, Wilcoxon signed–rank test was used. Differences in the proportion of categorical variables between the two groups were evaluated using McNemar–Bowker test. We first used locally weighted scatterplot smoothing (LOESS) models to estimate the "real" relationship between serum AA levels and the probability of AISL. Next, multivariable conditional logistic regression models were used to examine the association between contributing AA levels and AISL after adjusting for potential confounding factors. The individual effects of AA metabolites on the risk of AISL were quantified separately using the odds ratio (OR) and 95% confidence interval (CI) as follows: with AAs as a categorical variable (quartiles) and as a continuous variable [scaled to an interquartile range (IQR)]. Variables with a P–value less than 0.2 in the comparison between two groups were selected as potential confounders. This approach has been widely used in many studies, particularly for small sample size. The variance inflation factor (VIF) was used to examine the potential collinearity among variables. As too many covariates in a multiple regression model can lead to overfitting[24], we selected no more than four variables as confounding factors to decrease the potential of overfitting when assessing the association between AAs and AISL. Furthermore, as distinct metabolites may be highly related to each other, collinearity should be considered. Thus, we used the VIF based on VIF package of R software to detect potential collinearity among the AAs. A VIF greater than 1.5 is considered to indicate collinearity in the model, and the associated variable is removed. The combined effect of relevant AAs on AISL was also determined using a multivariable logistic regression model. A receiver operator characteristic (ROC) analysis was applied to evaluate the value and feasibility of AAs as potentially sensitive and specific biomarkers for recognizing AISL. Data management, analysis, and figure drawing were performed using R version 3.4.4 (Copyright © 2018 The R Foundation for Statistical Computing). All tests were two–sided and the results with P values < 0.05 were considered to indicate significance.

Patient and Public Involvement

The present study was designed as an observational study, and as such patients and the public were not involved in the planning, recruitment and conduct of this study. All participants were informed about the purpose of this study and signed informed consent at the beginning of the study. The results of this study have not yet been disseminated to the participants.

RESULTS

Table 1 summarizes the general characteristics of the study population. Comparisons of the demographic data, clinical features, and urinary arsenic species in the 56 pairs of subjects are presented in Table 1. The median (first and third quartile) age of AISL population was 50.30 (44.70 and 58.70) for the cases and 50.40 (44.60 and 58.70) years for the controls. Both groups contained the same proportion of females (58.93%), and there was no significant difference in urinary arsenic levels between the two groups. More than half of the subjects had no history of smoking or alcohol consumption. The serum triglyceride level in subjects with AISL was significantly lower than that in the controls ($P = 0.041$). Other variables were similar between the subjects with AISL and controls ($P > 0.05$). This indicates that the participants in the two groups were comparable.

Table 1 Demographic characteristics of the study population[ξ].

Variables	AISL (n=56)	Non–AISL (n=56)	P
Clinical Characteristics			
Age (years)	50.30(44.70,58.70)	50.40(44.60,58.70)	0.425
Exposure year (years)	48.19±11.53	47.62±10.97	0.489
Body mass index (kg/m^2)	24.12±3.14	23.91±2.86	0.697
Fasting plasma glucose (mmol/L)	4.89(4.60,5.25)	5.12(4.53,5.40)	0.137
Folate (ng/mL)	4.00(3.20,5.10)	4.25(3.35,5.40)	0.392
Total homocysteine (μmol/L)	12.30(10.32,16.50)	12.67(11.21,14.69)	0.961
Blood urea nitrogen (mmol/L)	6.45(5.42,7.69)	6.84(5.36,8.80)	0.603
Total cholesterol (mmol/L)	4.58(4.10,5.69)	4.65(3.96,5.95)	0.904
Triglycerides (mmol/L)	1.41(0.90,1.74)	1.45(1.09,2.29)	0.041
High–density lipoprotein (mmol/L)	1.19±0.34	1.16±0.31	0.675
Low–density lipoprotein (mmol/L)	3.04±0.80	3.25±0.84	0.110
Women [# (%)]	33(58.93)	33(58.93)	1.000
Cigarette smoking [# (%)]	20(35.71)	22(39.29)	0.696
Alcohol consumption [# (%)]	17(30.91)	21(37.50)	0.464
Illiteracy [# (%)]	21(37.50)	15(25.00)	0.252
Urinary arsenic species[ζ]			
iAS%	12.26(8.13,14.68)	12.31(10.04,16.54)	0.148
MMA%	24.68(20.11,29.68)	25.85(20.90,31.66)	0.420
DMA%	61.84(56.62,71.01)	61.84(47.85,64.99)	0.096
tAs (μg/g creatinine)	140.93(104.41,208.53)	186.77(80.11,217.30)	0.445

[ξ] AISL: arsenic –induced skin lesions; the variables met normal distribution was described with mean± standard deviation; otherwise, median (1st quartile, 3rd quartile) was used to describe their features. Number of cases (percentage) was used to describe the proportion of categorical variables between the two groups.
*[ζ] iAS: inorganic arsenic (iAs^{III}+iAs^{V}); MMA: monomethyl arsenate (MMA^{III}+MMA^{V}); DMA: dimethyl arsenate (DMA^{III}+DMA^{V}); tAs: total arsenic (iAs^{III}+iAs^{V}+MMA+DMA); iAs%= iAS/tAs*100%; MMA%=MMA/tAs*100% and DMA%=DMA/tAs*100%.*

Table 2 shows that the four AAs, with an FDR–adjusted P–value of < 0.05 and VIP of > 1, were significantly lower in the cases than in the controls. Two were aromatic amino acids (AAAs) identified as phenylalanine and tryptophan, one was a branched–chain amino acid (BCAA) identified as leucine, and the last AA was phenylalanylphenylalanine. The individual associations of AAs with AISL are presented in Figure 1, which shows obvious "dose–response" relationships.

Table 2 Distinct metabolites in population with arsenic – induced skin lesions and their counterparts.

Serum amino acid metabolites	Retention time (min)	Mass–to–Charge Ratio	VIP value	p–values[ξ]	Adjusted p–values[ζ]
Phenylalanine	3.402	166.087	1.508	<0.001	0.009
Tryptophan	3.886	203.082	1.046	0.003	0.014
Leucine	2.642	132.102	1.014	0.001	0.020
Phenylalanylphenylalanine	5.048	313.155	1.833	0.004	0.033

VIP: variable importance in the project; ξ Wilcoxon signed–rank test; ζ Adjusted by false discovery rate (FDR).

Table 3 shows that participants in the third and fourth quartiles of the four specific AAs were significantly linked to decreased odds of AISL after adjusting for FPG, LDL, TG, and DMA%, as compared with that of their lowest quartiles. The category boundaries of the quartiles are shown in

Table S1. Significant linear trends were detected between AISL and the four serum AAs. The same linear negative association between AISL and per IQR increase in the four serum AAs was observed when these AAs were considered as continuous variables in the present study.

Table 3 Relationship of amino acid levels with the odds of arsenic–induced skin lesions[ξ]

Amino acids	N	Cases (%)	Crude		Adjusted[ζ]	
			OR (95% CI)	P	OR (95% CI)	P
Tryptophan						
Per IQR	112	56(50)	0.48(0.27,0.84)	0.011	0.48(0.27,0.86)	0.013
Quartiles						
Q_1	28	20(71.40)	1.00(1.00,1.00)	Ref.	1.00(1.00,1.00)	Ref.
Q_2	28	15(53.60)	0.50(0.16,1.54)	0.225	0.56(0.16,1.98)	0.370
Q_3	28	10(35.70)	0.12(0.03,0.53)	0.005	0.12(0.02,0.60)	0.010
Q_4	28	11(39.30)	0.19(0.05,0.71)	0.014	0.21(0.05,0.84)	0.028
P for trend				0.008		0.012
Phenylalanine						
Per IQR	112	56(50)	0.57(0.36,0.91)	0.019	0.56(0.33,0.94)	0.028
Quartiles						
Q_1	28	20(71.40)	1.00(1.00,1.00)	Ref.	1.00(1.00,1.00)	Ref.
Q_2	28	19(67.90)	0.79(0.22,2.82)	0.712	0.70(0.18,2.76)	0.609
Q_3	28	8(28.60)	0.18(0.05,0.66)	0.010	0.20(0.05,0.77)	0.019
Q_4	28	9(32.10)	0.25(0.08,0.79)	0.018	0.20(0.05,0.75)	0.017
P for trend				<0.001		0.001
Leucine						
Per IQR	112	56(50)	0.45(0.25,0.82)	0.019	0.43(0.21,0.86)	0.016
Quartiles						
Q_1	28	21(75.00)	1.00(1.00,1.00)	Ref.	1.00(1.00,1.00)	Ref.
Q_2	28	14(50.00)	0.33(0.11,1.03)	0.057	0.31(0.10,1.01)	0.052
Q_3	28	11(39.30)	0.22(0.07,0.68)	0.009	0.22(0.07,0.73)	0.014
Q_4	28	10(35.70)	0.19(0.06,0.59)	0.004	0.19(0.06,0.65)	0.008
P for trend				0.003		0.007
Phenylalanylphenylalanine						
Per IQR	112	56(50)	0.62(0.36,1.04)	0.070	0.71(0.41,1.24)	0.227
Quartiles						
Q_1	27	21(77.80)	1.00(1.00,1.00)	Ref.	1.00(1.00,1.00)	Ref.
Q_2	29	14(48.30)	0.16(0.03,0.75)	0.021	0.14(0.03,0.73)	0.019
Q_3	28	9(32.10)	0.08(0.02,0.39)	0.002	0.09(0.02,0.52)	0.007
Q_4	28	12(42.90)	0.12(0.02,0.57)	0.008	0.11(0.02,0.66)	0.016
P for trend				0.006		0.023

[ξ]Values are odds ratio (95% confidence intervals) for arsenic–induced skin lesions from conditional logistic regression. IQR: interquartile range; Q_1: the 1st quartile; Q_2: the 2nd quartile; Q_3: the 3rd quartile; Q_4: the 4th quartile.

[ζ] Adjusted for plasma glucose, low–density lipoprotein, triglyceride, and urinary dimethyl arsenate.

As these four specific AAs were significantly or marginally significantly associated with the odds of AISL, we examined their combined effects on AISL. However, analysis of potential collinearity revealed that among the four specific AAs, both tryptophan and phenylalanine had the smallest VIF value (VIF = 1.04), and no obvious collinearity existed (**Table S2**). Thus, we mainly focused on tryptophan and phenylalanine to assess the combined effects of AAs on AISL and only

presented the results associated with these two AAs in the present study. To avoid the effects of insufficient power because of unreasonable grouping of the results, we classified both tryptophan and phenylalanine into two categories, according to the cut–off values of their mass spectrum peak area based on the ROC analysis, respectively. The higher levels of these two serum AAs were defined as equal to or higher than the cut–off values, while the lower categories were considered as less than the associated values.

Table 4 shows the combined effects of tryptophan and phenylalanine levels on AISL after considering the collinearity of variables in the model. The proportions of AISL were 74.3%, 60.0%, 50.0%, and 18.2% for participants with lower levels of both tryptophan and phenylalanine (category A), with higher level of tryptophan and lower level of phenylalanine (category B), with lower level of tryptophan and higher level of phenylalanine (category C), and higher levels of both tryptophan and phenylalanine (category D), respectively. An obvious decrease in the probability of AISL was observed among these four categories. Compared with that of category A, the adjusted OR (95% CI) for participants in categories B, C, and D was 0.49 (0.15–1.63), 0.32 (0.10–1.02), and 0.08 (0.02–0.25), respectively. Subjects with higher levels of both tryptophan and phenylalanine had the lowest odds of AISL, which significantly decreased by 92% (OR=0.08; 95% CI 0.02–0.25; P<0.001), after adjusting for the effects of some potential confounding factors. This suggests that tryptophan and phenylalanine are jointly associated with AISL, although no significant interaction between the two AAs and occurrence of AISL was observed (P = 0.419).

Table 4 Joint association between tryptophan and phenylalanine levels with arsenic–induced skin lesions.

Tryptophan <cut–off value[ξ]	Phenylalanine <cut–off value[ξ]	N	Cases (%)	Crude		Adjusted[ζ]	
				OR (95% CI)	P	OR (95% CI)	P
Yes	Yes	35	26(74.3)	1.00(1.00,1.00)	Ref.	1.00(1.00,1.00)	Ref.
No	Yes	20	12(60.0)	0.52(0.16,1.68)	0.273	0.49(0.15,1.63)	0.244
Yes	No	24	12(50.0)	0.35(0.12,1.04)	0.059	0.32(0.10,1.02)	0.053
No	No	33	6(18.2)	0.08(0.02,0.25)	<0.001	0.08(0.02,0.25)	<0.001
Interaction					0.320	0.49(0.09,2.78)	0.419

[ξ] *Cut–off value was determined using receiver operator characteristic analysis.*
[ζ] *Adjusted for plasma glucose, low–density lipoprotein, triglyceride, and urinary dimethyl arsenate.*

Table 5 shows that, based on the ROC analysis, both serum tryptophan and phenylalanine levels might be potential biomarkers in distinguishing AISL from a chronic arsenic exposure population (P = 0.0020 and P = 0.0017). The area under the curve (AUC) and its related 95% CI, sensitivity, specificity, positive predictive value and negative predictive value were 0.67 (0.57–0.77), 69.64%, 62.50%, 65.00 and 67.31% for tryptophan, and 0.70 (0.60–0.80), 69.64%, 69.64%, 69.64% and 69.64% for phenylalanine, respectively. The AUC (95% CI), sensitivity, specificity, positive predictive value and negative predictive value of the combination of them were 0.72 (0.62–0.81), 76.79%, 58.93%, 65.15 and 71.74%, respectively. Our results suggest that these two AAs could be either individually or jointly used as indicators of AISL identification.

Table 5 Combination of diagnostic indicators and ROC analysis results[ξ].

Indicators	AUC (95% CI)	Sensitivity, %	Specificity, %	Predict^{+}, %	Predict^{-}, %	P
Tryptophan	0.67(0.57,0.77)	69.64	62.50	65.00	67.31	0.002
Phenylalanine	0.70(0.60,0.80)	69.64	69.64	69.64	69.64	0.002
Combined[ζ]	0.72(0.62,0.81)	76.79	58.93	65.15	71.74	<0.001

[ξ] ROC: a receiver operator characteristic; AUC: area under the roc curve; CI: confidence interval; The sensitivities, specificity, positive predictive value, and negative predictive value were calculated at their best cut–off points; Predict+: positive predictive value; Predict–: negative predictive value.

[ζ] Combined: tryptophan and phenylalanine. The combination is modeled according to the formula $\beta_1 X_1 + \beta_2 X_2$, with X_j denoting the standardized value for the j^{th} amino acid and βj denoting the regression coefficient from the logistic regression model.

DISCUSSION

In the present study, the association of serum tryptophan and phenylalanine, screened in our previous non–targeted metabolomics study using UPLC–MS/MS, with AISL and their ability to indicate AISL occurrence were quantitatively evaluated in individual and joint modes. Our results clearly showed that AISL was significantly and negatively associated with serum tryptophan and phenylalanine levels in a chronically arsenic–exposed population via drinking water. Participants with a higher level of both AAs showed the lowest odds of AISL. These two AAs may be useful as indicators of AISL.

The probability of initiation and development of AISL is affected by numerous factors including age, gender, lifestyles, arsenic exposure, and metabolism. These factors may be important confounding factors and largely affect our results. To adjust for the effects of these co–factors, we

first selected all participants using permuted block randomization from a single rural area in which the population was chronically exposed to arsenic in the same route, had a similar lifestyle, and was influenced by similar environmental factors. Secondly, the cases and controls were matched by gender and age (±1 year). All of these may be the reason for the so many potential confounders including arsenic exposure, which did not differ significantly between the cases and controls.

Participants enrolled in the current study were chronically exposed to arsenic via drinking water. The geometric mean (GM) and its related 95% CI of urinary iAs/creatinine and tAs/creatinine in this population were 17.49 (14.90–20.53) μg/g and 147.20 (129.00–167.97) μg/g, respectively. They were much higher than those in the 20 μg/L exposed to arsenic via drinking water [GM (95% CI): 0.4 (0.3–0.5) μg/g for iAs and 9.1 (6.5–12.7) μg/g for tAs], while obviously lower than those in the 90 μg/L exposed group [GM (95% CI): 39.4 (31.4–49.6) μg/g for iAs and 248.7 (208.8–296.3) μg/g for tAs][25]. An available report has shown that AISL cannot be completely cured even though medical technology has already made great progress[26]. Therefore, it is crucial to identify those who are most likely to progress to overt arsenic damages including AISL among people at risk as early as possible. Metabolomics study, which mainly focuses on thoroughly assessing the variation in metabolites possibly linked to disease occurrence and development, has been widely utilized to help understand the pathogenesis of diseases because of its relevance to the phenotypes compared to other "OMICs" study[27]. Moreover, mathematical modeling to assess the linkage between small molecular metabolites and arsenic toxicity has advanced [28]. Developing a simple and interpretable modeling approach for the early detection of arsenic–induced lesions is of great theoretical value and realistic meaning[29], although it might be difficult due to population–specific complexities and effect of some potential unmeasured covariates such as diet and genetic determinants.

Previous studies have reported that gene–gene and gene–environment interactions are involved in arsenicosis through toxicological mechanisms including genomic instability[30] and oxidative stress[31]. Skin hyperpigmentation and palmoplantar hyperkeratosis may be biomarkers for long–term

arsenic exposure and useful for identifying differences in metabolites associated with phenotypes. Metabolite analysis may improve the understanding and identification of AISL. An animal study revealed that examining the disruption in AA metabolism upon arsenic exposure in rats–may be beneficial for understanding arsenic toxicity[32]. In our previous population–based metabolomics study, we found that serum metabolite alterations were significantly related to the risk of arsenic–induced health damages. In the present study, we found that some BCAAs or AAAs were significantly related to AISL occurrence. Several studies across numerous ethnic backgrounds support the use of BCAAs including leucine, isoleucine, valine, and AAA profiles such as those of phenylalanine, tryptophan, and tyrosine as biomarkers for identifying metabolic diseases[27 33]. Zhou et al. reported that arsenic–induced transformed cells had altered metabolite profiles including downregulation of leucine, tryptophan, and phenylalanine in the skin lesion group[34]. Consistent with these findings, the levels of two serum AAAs (tryptophan and phenylalanine) were significantly associated with AISL in our study.

Normal metabolism of AAs is necessary for whole‐body homeostasis, growth, and development, and health[15]. Studies have reported that changes in the availability of AAAs affect cell signaling, gene expression, brain function, and neuroendocrine function[35]. Tryptophan, an AA metabolism‐related biomarker, is a sensitive and specific indicator of oxidation. Tryptophan metabolism in mammals is a physiological means of preserving immune homeostasis associated with oxidative stress and inflammation[36 37]. Additionally, phenylalanine can be transformed into specific neurotransmitters such as dopamine and adrenaline via the action of related enzymes. Wu et al.[38] reported that arsenic exposure can lead to neurotransmitter metabolism, which may explain the reduction in phenylalanine. Furthermore, as a peptide–bound phenylalanine, phenylalanylphenylalanine has been reported to affect protein synthesis and secretion[39], indicating a relationship between endothelium dysfunction and phenylalanine metabolism disorder. The relationship between AA metabolism and AISL remains unclear. The notable alteration of tryptophan

and phenylalanine in this study indicates the occurrence of metabolic disorders due to arsenic exposure. These results are also beneficial for understanding the effects of arsenic toxicity and importance of early identification of exposure for delaying the progression of various arsenic–induced lesions, including AISL.

The main strength of this study is that the findings were based on a community–based, long–term arsenic exposure cohort with well–designed quality assurance and quality control throughout the study. However, there were some limitations to this study. Firstly, although the untargeted metabolomics approach can assess a large amount number of metabolites precisely and efficiently, it only provided relative levels of AAs instead of their accurate quantitative concentration. Secondly, these findings were mainly based on a case–control study, which only reveals the association between AA metabolism and the odds of AISL rather than confirming their causal relationship. Furthermore, the study included 56 AISL cases–matched and 56 non–AISL controls and the sample size might be potentially insufficient. Although metabolomics studies usually have a sample size of no more than 40 cases in each group[40 41], our study may be under–powered and thus larger studies are needed in the future. Besides, as it is suggested that the ratio of approximately 10–15 observations per predictor in a logistic regression model will produce reasonably stable estimations[24], we selected only four covariates in the models due to the small sample size and these results need to be confirmed in further studies. Finally, the participants were mainly exposed to arsenic via drinking water, which would limit the findings from being extrapolated to other arsenic exposure populations via food and other routes. Therefore, additional elaborate population–based studies are needed to verify our findings.

In conclusion, specific AAs may be linked to AISL and AA metabolism may play an important role in early AISL identification. Additional studies are needed to confirm our findings.

Contributors: Guangyun Mao and Yaping Wei designed the study. Chaonan Jia participated in collecting data. Yuan Lan and Chaonan Jia audited the data. Yaping Wei, Xiangqing hou, Jushuang Li, Tao Wang conducted the literature search, Yaping Wei, Chaonan Jia conducted statistical analysis and interpreted the results. Yaping Wei and Chaonan Jia wrote the first draft of the manuscript. Jingjing Zuo helped with copyediting. Guangyun Mao reviewed the final manuscript and did substantial contributions.

Funding: This work was supported by the basic public welfare research project of Zhejiang province (LGF19H260011), the initial scientific research fund (KYQD170301) and the major project of the eye hospital of Wenzhou medical university (YNZD201602).

Acknowledgments: We gratefully acknowledge all of the study participants of this project. We wish to thank all of the dedication and hard work of the field team at the department of preventive medicine, school of public health and management, Wenzhou medical university. Our thanks must also go to Dr. Kendrick Hii, a native–English speaker, for his valuable suggestions and edits to the manuscript.

Competing Interest: None declared.

Ethics approval: Ethical approval was received from the ethics committee of Wenzhou Medical University, Wenzhou, China.

Provenance and peer review: Not commissioned; externally peer reviewed.

Data Sharing Statement: No additional unpublished data are available.

Patient consent for publication: Not required.

REFERENCES

1. Mandal BK, Suzuki KT. Arsenic round the world: a review. Talanta. 2002; 58:201–35.
2. Smith AH, Steinmaus CM. Health Effects of Arsenic and Chromium in Drinking Water: Recent Human Findings. Annual Review Of Public Health. Palo Alto: Annual Reviews, 2009:107–22.
3. Guha Mazumder DN, Haque R, Ghosh N, et al. Arsenic levels in drinking water and the prevalence of skin lesions in West Bengal, India. International journal of epidemiology. 1998; 27:871–7.
4. Guo XJ, Fujino Y, Kaneko S, et al. Arsenic contamination of groundwater and prevalence of arsenical dermatosis in the Hetao plain area, Inner Mongolia, China. Mol Cell Biochem. 2001; 222:137–40.
5. Lindberg A–L, Rahman M, Persson L–A, et al. The risk of arsenic induced skin lesions in Bangladeshi men and women is affected by arsenic metabolism and the age at first exposure. Toxicology and Applied Pharmacology. 2008; 230:9–16.
6. Tondel M, Rahman M, Magnuson A, et al. The relationship of arsenic levels in drinking water and the prevalence rate of skin lesions in Bangladesh. Environmental Health Perspectives. 1999; 107:727–29.
7. Hsu LI, Chen GS, Lee CH, et al. Use of Arsenic–Induced Palmoplantar Hyperkeratosis and Skin Cancers to Predict Risk of Subsequent Internal Malignancy. Am J Epidemiol. 2013; 177:202–12.
8. Luo LR, Li YY, Gao YH, et al. Association between arsenic metabolism gene polymorphisms and arsenic–induced skin lesions in individuals exposed to high–dose inorganic arsenic in northwest China. Sci Rep. 2018; 8:12.
9. Hughes MF, Beck BD, Chen Y, et al. Arsenic Exposure and Toxicology: A Historical Perspective. Toxicol Sci. 2011; 123:305–32.
10. Ren X, McHale CM, Skibola CF, et al. An Emerging Role for Epigenetic Dysregulation in Arsenic Toxicity and Carcinogenesis. Environmental Health Perspectives. 2011; 119:11–19.
11. Chen J–W, Wang S–L, Wang Y–H, et al. Arsenic methylation, GSTO1 polymorphisms, and metabolic syndrome in an arseniasis endemic area of southwestern Taiwan. Chemosphere. 2012; 88:432–38.
12. Wang S–L, Chang F–H, Liou S–H, et al. Inorganic arsenic exposure and its relation to metabolic syndrome in an industrial area of Taiwan. Environment International. 2007; 33:805–11.
13. Liu R, Li H, Fan WJ, et al. Leucine Supplementation Differently Modulates Branched–Chain Amino Acid Catabolism, Mitochondrial Function and Metabolic Profiles at the Different Stage of Insulin Resistance in Rats on High–Fat Diet. Nutrients. 2017; 9:20.
14. Platten M, Wick W, Van den Eynde BJ. Tryptophan Catabolism in Cancer: Beyond IDO and Tryptophan Depletion. Cancer Res. 2012; 72:5435–40.
15. Wu G. Amino acids: metabolism, functions, and nutrition. Amino Acids. 2009; 37:1–17.
16. Casadei L, Valerio M, Manetti C. Metabolomics: Challenges and Opportunities in Systems Biology Studies. Methods Mol Biol. 2018; 1702:327–36.
17. Johnson CH, Gonzalez FJ. Challenges and opportunities of metabolomics. J Cell Physiol. 2012; 227:2975–81.
18. Hui AY, McCarty WJ, Masuda K, et al. A systems biology approach to synovial joint lubrication in health, injury, and disease. Wiley Interdiscip Rev Syst Biol Med. 2012; 4:15–37.
19. Lu K, Abo RP, Schlieper KA, et al. Arsenic Exposure Perturbs the Gut Microbiome and Its Metabolic Profile in Mice: An Integrated Metagenomics and Metabolomics Analysis. Environmental Health Perspectives. 2014; 122:284–91.
20. Zhang J, Shen H, Xu W, et al. Urinary Metabolomics Revealed Arsenic Internal Dose–Related Metabolic Alterations: A Proof–of–Concept Study in a Chinese Male Cohort. Environmental Science & Technology. 2014; 48:12265–74.
21. Guo X, Cui H, Zhang H, et al. Protective Effect of Folic Acid on Oxidative DNA Damage: A Randomized, Double–Blind, and Placebo Controlled Clinical Trial. Medicine (Baltimore). 2015; 94:e1872.
22. Standard of diagnosis for endemic arsenism of China (WS/T211–2001). the Ministry of Health of the People's Republic of China. Beijing 2001.
23. Chen X, Guo X, He P, et al. Interactive Influence of N6AMT1 and As3MT Genetic Variations on Arsenic Metabolism in the Population of Inner Mongolia, China. Toxicol Sci. 2017; 155:124–34.
24. Peduzzi P, Concato J, Kemper E, et al. A simulation study of the number of events per variable in logistic

regression analysis. J Clin Epidemiol. 1996; 49:1373–9.

25. Sun G, Xu Y, Li X, et al. Urinary arsenic metabolites in children and adults exposed to arsenic in drinking water in Inner Mongolia, China. Environmental Health Perspectives. 2007; 115:648–52.
26. Kuo CC, Moon KA, Wang SL, et al. The Association of Arsenic Metabolism with Cancer, Cardiovascular Disease, and Diabetes: A Systematic Review of the Epidemiological Evidence. Environmental Health Perspectives. 2017; 125:15.
27. Wang TJ, Larson MG, Vasan RS, et al. Metabolite profiles and the risk of developing diabetes. Nature Medicine. 2011; 17:448–U83.
28. Lawley SD, Cinderella M, Hall MN, et al. Mathematical model insights into arsenic detoxification. Theor Biol Med Model. 2011; 8.
29. Juhasz AL, Herde P, Herde C, et al. Validation of the predictive capabilities of the Sbrc–G in vitro assay for estimating arsenic relative bioavailability in contaminated soils. Environ Sci Technol. 2014; 48:12962–9.
30. Kibriya MG, Jasmine F, Parvez F, et al. Association between genome–wide copy number variation and arsenic–induced skin lesions: a prospective study. Environ Health. 2017; 16:13.
31. Priori R, Scrivo R, Brandt J, et al. Metabolomics in rheumatic diseases: the potential of an emerging methodology for improved patient diagnosis, prognosis, and treatment efficacy. Autoimmun Rev. 2013; 12:1022–30.
32. Wang X, Mu X, Zhang J, et al. Serum metabolomics reveals that arsenic exposure disrupted lipid and amino acid metabolism in rats: a step forward in understanding chronic arsenic toxicity. Metallomics Integrated Biometal Science. 2015; 7:544.
33. Weng L, Quinlivan E, Gong Y, et al. Association of Branched and Aromatic Amino Acids Levels with Metabolic Syndrome and Impaired Fasting Glucose in Hypertensive Patients. Metabolic Syndrome And Related Disorders. 2015; 13:195–202.
34. Zhou YY, Wang YF, Su J, et al. Integration of microRNAome, proteomics and metabolomics to analyze arsenic–induced malignant cell transformation. Oncotarget. 2017; 8:90879–96.
35. Lee JC, Chen MJ, Chang CH, et al. Plasma amino acid levels in patients with colorectal cancers and liver cirrhosis with hepatocellular carcinoma. Hepato–Gastroenterol. 2003; 50:1269–73.
36. Munn DH, Mellor AL. Indoleamine 2,3 dioxygenase and metabolic control of immune responses. Trends In Immunology. 2013; 34:137–43.
37. Pawlak K, Domaniewski T, Mysliwiec M, et al. The kynurenines are associated with oxidative stress, inflammation and the prevalence of cardiovascular disease in patients with end–stage renal disease. Atherosclerosis. 2009; 204:309–14.
38. Wu L–L, Gong W, Shen S–P, et al. Multiple metal exposures and their correlation with monoamine neurotransmitter metabolism in Chinese electroplating workers. Chemosphere. 2017; 182:745–52.
39. Zhou MM, Wu YM, Liu HY, et al. Effects of phenylalanine and threonine oligopeptides on milk protein synthesis in cultured bovine mammary epithelial cells. J Anim Physiol Anim Nutr. 2015; 99:215–20.
40. Chen LY, Cheng CY, Choi HW, et al. Plasma Metabonomic Profiling of Diabetic Retinopathy. Diabetes. 2016; 65:1099–108.
41. Piszcz J, Lemancewicz D, Dudzik D, et al. Differences and similarities between LC–MS derived serum fingerprints of patients with B–cell malignancies. Electrophoresis. 2013; 34:2857–64.

Figure 1. Association between the peak intensity of tryptophan and phenylalanine and arsenic–induced skin lesions based on multivariable locally weighted regression models. a: Tryptophan; b: Phenylalanine;c: Leucine;d: Phenylalanylphenylalanine

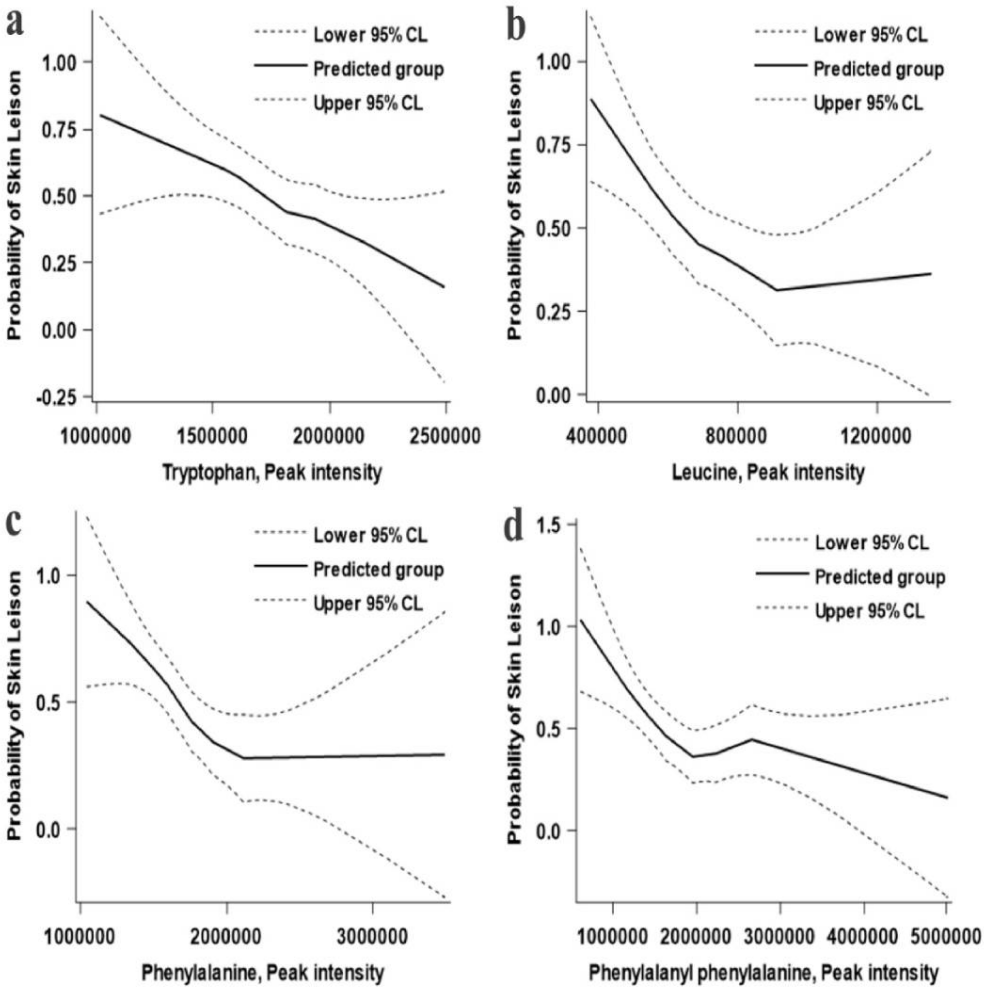

Figure 1. Association between the peak intensity of tryptophan and phenylalanine and arsenic–induced skin lesions based on multivariable locally weighted regression models. a: Tryptophan; b: Phenylalanine;c: Leucine;d: Phenylalanylphenylalanine

90x90mm (300 x 300 DPI)

Table S1 Amino acids metabolites in serum.

Amino acids	Quartile	Quartile Range	Number
Tryptophan	1	1015821.54-1555040.97	28
	2	1555370.88-1701256.07	28
	3	1704298.84-1937437.56	28
	4	1945755.37-2492319.05	28
Phenylalanine	1	1044964.46-1516708.76	28
	2	1519993.60-1649420.82	28
	3	1659818.85-1908125.28	28
	4	1929946.22-3489918.62	28
Leucine	1	379957.04-556314.46	28
	2	558744.08-688759.83	28
	3	689470.83-806033.16	28
	4	821492.90-1356991.53	28
Phenylalanylphenylalanine	1	608200.15-1291853.09	27
	2	1303597.59-1632678.66	29
	3	1638204.36-2234283.97	28
	4	2235250.29-5011775.70	28

Table S2 Variance inflation factor of amino acids in different models[ξ].

Amino acids	Model 1	Model 2	Model 3	Model 4
Tryptophan	1.04	1.04	1.04	1.04
Phenylalanine	4.56		1.50	1.04
Leucine	4.21	1.39		
Phenylalanylphenylalanine	1.49	1.38	1.48	

[ξ]*Model 1: Tryptophan, Phenylalanine, Leucine and Phenylalanyl Phenylalanine;*
Model 2: Tryptophan, Leucine and Phenylalanyl Phenylalanine;
Model 3: Tryptophan, Phenylalanine and Phenylalanylphenylalanine;
Model 4: Tryptophan and Phenylalanine.

Table1 The STROBE checklist in this study.

	Item No	Recommendation
Title and abstract	1	(a) Indicate the study's design with a commonly used term in the title or the abstract (b) Provide in the abstract an informative and balanced summary of what was done and what was found **Please see detail in the "ABSTRACT" section in the manuscript.**
Introduction		
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported **Please see detail in the second, third, fourth paragraphs of the "INTRODUCTION" section in the manuscript.**
Objectives	3	State specific objectives, including any prespecified hypotheses **Please see detail in the fourth paragraph of the "INTRODUCTION" section in the manuscript.**
Methods		
Study design	4	Present key elements of study design early in the paper **Please see detail in the "Study Population "of the "METHODS" section in the manuscript.**
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection **Please see detail in the "Study Population "of the "METHODS" section in the manuscript.**
Participants	6	(a) Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls (*b*) For matched studies, give matching criteria and the number of controls per case **Please see detail in the "Study Population "of the "METHODS" section in the manuscript.**
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable **Please see detail in the "Study Population ", "Data Collection and Assessment" and "Distinct Metabolites Identification" of the "METHODS" section in the manuscript.**
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group **Please see detail in the "Study Population ", "Data Collection and**

		Assessment" and "Distinct Metabolites Identification" of the "METHODS" section in the manuscript.
Bias	9	Describe any efforts to address potential sources of bias **Please see detail in the "Study Population "of the "METHODS" section in the manuscript.**
Study size	10	Explain how the study size was arrived at **Please see detail in the sixth paragraph of the of the "DISCUSSION" section in the manuscript.**
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why **Please see detail in the "Statistical Analysis" of the "METHODS" section in the manuscript.**
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding (*b*) Describe any methods used to examine subgroups and interactions (*c*) Explain how missing data were addressed (*d*) If applicable, explain how matching of cases and controls was addressed (*e*) Describe any sensitivity analyses **No missing values were observed in our database. Others please see detail in the "Statistical Analysis" of the "METHODS" section in the manuscript.**
Results		
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed; (b) Give reasons for non-participation at each stage; (c) Consider use of a flow diagram **Please see detail in the "Study Population "of the "METHODS" section in the manuscript.**
Descriptive data	**14***	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders (b) Indicate number of participants with missing data for each variable of interest **See Table 1 in the manuscript.**
Outcome data	15*	Report numbers in each exposure category, or summary measures of exposure **Please see Table 3 in the manuscript.**
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were

		included **Please see Table 3 in the manuscript.**
		(*b*) Report category boundaries when continuous variables were categorized **Please see Table S1 in the manuscript.**
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses **Please see Table 4 in the manuscript.**
Discussion		
Key results	18	Summarise key results with reference to study objectives **Please see detail in the first paragraph of the of the "DISCUSSION" section in the manuscript.**
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias **Please see detail in the seventh paragraph of the of the "DISCUSSION" section in the manuscript.**
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence **Please see detail in the fifth paragraph of the of the "DISCUSSION" section in the manuscript.**
Generalisability	21	Discuss the generalisability (external validity) of the study results **Please see detail in the seventh paragraph of the "DISCUSSION" section in the manuscript.**
Other information		
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based **Please see detail in the "Funding" section in the manuscript.**

*Give information separately for cases and controls.