

BMJ Open

BMJ Open is committed to open peer review. As part of this commitment we make the peer review history of every article we publish publicly available.

When an article is published we post the peer reviewers' comments and the authors' responses online. We also post the versions of the paper that were used during peer review. These are the versions that the peer review comments apply to.

The versions of the paper that follow are the versions that were submitted during the peer review process. They are not the versions of record or the final published versions. They should not be cited or distributed as the published version of this manuscript.

BMJ Open is an open access journal and the full, final, typeset and author-corrected version of record of the manuscript is available on our site with no access controls, subscription charges or pay-per-view fees (<http://bmjopen.bmj.com>).

If you have any questions on BMJ Open's open peer review process please email info.bmjopen@bmj.com

BMJ Open

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

Journal:	<i>BMJ Open</i>
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, Ou Hai 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 University, 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 University, 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

SCHOLARONE™
Manuscripts

1
2
3 **Low level of serum aromatic amino acid associated with elevated risk of arsenic-induced skin**
4 **lesions: data from a metabolomics study.**
5

6 Yaping Wei, MS^{a,b}, Chaonan Jia, MS^{a,b}, Yuan Lan, MS^c, Xiangqing Hou, MS^{a,b}, Jingjing Zuo, MS^c,
7 Jushuang Li^{a,b}, Tao Wang^{a,b}, Guangyun Mao, MD, PhD^{a,b,d}✉
8
9

10
11 ^a Department of Preventive Medicine, School of Public Health and Management, Wenzhou Medical
12 University, Wenzhou, China.
13

14 ^b Center on Evidence-Based Medicine & Clinical Epidemiological Research, School of Public Health,
15 Wenzhou Medical University, Wenzhou, China.
16

17 ^c School of Ophthalmology & Optometry, Wenzhou Medical University, Wenzhou, China.
18

19 ^d Center on Clinical Research, the Affiliated Eye Hospital of Wenzhou Medical University, Wenzhou,
20 China.
21
22

23
24 Yaping Wei and Chaonan Jia contributed equally to this paper.
25
26

27
28 ✉Corresponding author: Guangyun Mao, Room 7A321, School of Public Health and Management,
29 Wenzhou Medical University, Chashan University town, Ouhai district, Wenzhou, China. 325035
30

31 E-mail: maogy@wmu.edu.cn
32

33 Telephone: +86-577-86689848
34

35 Fax numbers: +86-577-86699122
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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.

1 INTRODUCTION

2 Chronic arsenic exposure via drinking water is widely believed as a global health concern,
3 affecting a large amount of people worldwide. It may give rise to several human health effects and has
4 been documented to associate with cardiovascular disease, diabetes, cancer and others^{1 2}. With the
5 industrial boom and dramatic rise of worldwide water pollution including arsenic contamination in the
6 past, the prevalence and burden of arsenic-induced health damage will continue to increase. Skin has
7 been confirmed as one of the most common and susceptible targets of arsenic-induced health lesions.
8 Cutaneous skin lesions are typical signs of arsenicosis after persistent arsenic exposure for a long term,
9 characterized by hyperkeratosis and hyperpigmentation. Considerable evidences of the prevalence of
10 arsenical skin lesions have been observed in many countries³⁻⁵.

11 As arsenic induced skin lesions (AISL) have been widely accepted as the major early
12 manifestation of arsenic toxicity⁶ and may be indicators of susceptibility to more serious
13 arsenic-induced health hazards⁷, it is particularly crucial to identify participants at risk as early as
14 possible for preventing the onset or delaying the progression of the serious health problems effectively.
15 Several possible mechanism may explain arsenic poisoning lesion, such as genetic differences⁸,
16 oxidative stress⁹ and epigenetic dysregulation¹⁰ and so on. Studies have shown that arsenic
17 methylation in vivo is tightly associated with metabolic syndrome^{11 12}, which is believed to be closely
18 related to many kinds of metabolites.

19 Amino acids (AA) are the "basic unit" that make up body's various proteins and necessary to
20 maintain the health status. Some AA are important regulators of key metabolic pathways and of great
21 importance in maximizing efficiency of food utilization, enhancing protein accretion and health
22 improvement^{13 14}. Abnormal metabolism of AA will disturbs whole body homeostasis, impairs growth
23 and development, and even causes death¹⁵. So, serum AA level may be an important implication for
24 the metabolic status and disease conditions. As a powerful tool in system biology research,
25 metabolomic approach is beneficial on unbiased monitoring changes in endogenous

1
2
3 26 metabolism-related physiological processes, providing integrative information on the distinctive
4
5 27 features across multiple functional levels, and offering a window to capture core attributes responsible
6
7 28 for various phenotypes, which is particularly important in understanding the relevant
8
9 29 pathophysiological changes in a disease state, identifying novel biomarkers for person at risk
10
11 30 screening, diagnosis, treatment and prognosis of important human diseases¹⁶⁻¹⁸.

12
13 31 Animal experiments and epidemiological study have reported obvious arsenic-related
14
15 32 metabolomics perturbations^{19 20}. All of these researches substantially suggest that the relationship
16
17 33 between specific metabolites and arsenic-induced health lesions should be investigated. However, few
18
19 34 works have been conducted to comprehensively examine the metabolic mechanism relevant to AISL,
20
21 35 especially for the AA metabolism. Based on our previous non-targeted metabolomics data, the present
22
23 36 study aims to quantitatively examine the association of several specific AA with the risk of AISL and
24
25 37 their ability to predict AISL.

28 29 30 38 **Methods**

31 32 33 39 **Patient and Public Involvement**

34
35
36 40 No patients were involved.

37 38 41 **Study Population**

39
40 42 Participants enrolled in the present study came from our previous population-based chronic
41
42 43 arsenic exposure cohort established in the Hetao plain of Inner Mongolia, China in 2010. Among 450
43
44 44 residents aged 18 to 79 years chronically exposed to arsenic via drinking water, 56 of them were
45
46 45 diagnosed as AISL and selected as the case. Another 56 participants without AISL matched by gender
47
48 46 and similar age (± 1 year) from the same population were picked out as the control. The inclusion
49
50 47 criteria were subjects having the metabolomic test. The exclusion criteria included unmatched
51
52 48 participants and those without serum metabolites data. The AISL was diagnosed as the presence of
53
54 49 arsenic induced keratosis, hyperpigmentation or depigmentation by a physician from Wenzhou
55
56
57
58
59
60

1
2
3 50 medical university strictly follow the criteria of arsenicosis²¹. Informed consent was obtained from all
4
5 51 participates and this study was approved by ethics committee of Wenzhou Medical University,
6
7 52 Wenzhou, China.

9 53 **Data Collection and Assessment**

11 54 Detailed data collection of blood and urine samples and assessment methods for clinical and
12
13 55 sociodemographic variables had been published previously²². The epuration of various urinary arsenic
14
15 56 species was conducted by means of a high-performance liquid chromatography coupled mass
16
17 57 spectrometer system for separation and detection²³. The species of arsenic in urine samples consisted
18
19 58 of inorganic arsenic (iAs, [iAs[□] plus iAs[□]]), monomethyl arsenate (MMA, [MMA[□] plus MMA[□]]) and
20
21 59 dimethyl arsenate (DMA, [DMA[□] plus DMA[□]]). All arsenic species corrected by creatinine.

24 60 **Distinct Metabolites Identification**

26 61 Detection of serum metabolites were performed using ultra-performance liquid chromatography
27
28 62 couple with tandem mass spectrometry (UPLC-MS/MS). The serum samples preparation and
29
30 63 metabolites determination were described in our previous report²⁴. The peak intensity of metabolites
31
32 64 for the 56 pairs were acquired and then imported to MetaboAnalyst 4.0 (<http://www.metaboanalyst.ca/>)
33
34 65 for statistical analysis. A partial least-squares discriminant analysis (PLS-DA), a supervised and well
35
36 66 accepted pattern recognition approach, was used for the differentiation between the cases and controls.
37
38 67 False discovery rate (FDR) adjustment p-value in univariate analysis were performed to reduce the
39
40 68 potential impact induced by false positive on the results. The criteria used for distinct metabolites
41
42 69 screening were defined as both variable importance plot (VIP) scores >1 in PLS-DA and the crude or
43
44 70 FDR adjusted p-value all < 0.05. We identified a total of 114 extracted small molecular metabolites
45
46 71 significantly contributed to the cases recognition. Among them, 4 metabolites were confirmed as
47
48 72 amino acids and obviously down-regulated in the AISL group.

52 73 **Statistical Analysis**

54 74 The normality of continuous data was assessed using Shapiro-Wilk test. The comparison of them

1
2
3 75 between the case and control was applied with paired *t*-test if they met normal or similar normal
4
5 76 distribution. Otherwise, Wilcoxon signed rank test would be used. Differences in the proportion of
6
7 77 categorical variables between the two groups were evaluated by McNemar-Bowker tests. As the
8
9 78 distinct metabolites might be high related with each other, spearman correlation analysis was used to
10
11 79 investigate relationship among the various screened AA. To avoid or reduce the influences due to
12
13 80 collinearity among the AA, a multiple stepwise regression analysis was applied to screen relevant AA
14
15 81 independently associated with AISL. The inclusion and exclusion criteria were 0.05 and 0.10,
16
17 82 respectively. Author firstly used the locally weighted regression model to estimate the relationship
18
19 83 between serum AA level and the risk to develop AISL. Then, generalized linear models (GLMs) were
20
21 84 performed to examine the association between the contributing AA levels and AISL after adjusting for
22
23 85 some potential confounding factors. The individual impacts of AA metabolites on the risk of AISL
24
25 86 were quantified separately by odds ratio (OR) and 95% confidence interval (CI) with GLMs in the
26
27 87 following two ways: with AA as a categorical variable (quartiles) and as a continuous variable [scaled
28
29 88 to interquartile range (IQR)]. The combined effects of relevant AA on AISL were also performed. In
30
31 89 addition, receiver operator characteristic (ROC) analysis was applied to evaluate the value and
32
33 90 feasibility of the AA as the potential sensitive and specific biomarker to predict AISL. Data
34
35 91 management, analysis and figure drawing were finished using R version 3.4.4 (Copyright © 2018 The
36
37 92 R Foundation for Statistical Computing). All tests were two-sides and $P \leq 0.05$ was set as significant
38
39 93 level.
40
41
42
43
44

45 94 RESULTS

46
47 95 **Table 1** summarizes the general characteristics of the study population. The comparison of
48
49 96 demographical, clinical features and urinary arsenic species in the 56 pairs of subjects were presented
50
51 97 in Table 1. The median (1st quartile, 3rd quartile) of age was 50.30 (44.70, 58.70) years for cases and
52
53 98 50.40 (44.60, 58.70) years for the controls with female proportion as 58.93% and there was no
54
55 99 statistically significant difference in urinary arsenic levels between the two groups. More than half of
56
57
58
59

1
2
3 100 them had no history of smoking or alcohol consumption. As compared to the counterpart, the serum
4
5 101 triglycerides level in AISL participants was significantly lower ($P=0.041$). While the other variables
6
7 102 were similar between AISL participants and their control ($P>0.05$), which indicated that the
8
9 103 participants in the two groups were quite comparable to some extent.

104 **Table 1** Demographic characteristics of the study population[‡].

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[‡]			
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

38 105 [‡]AISL: arsenic induced skin lesions; the variables met normal distribution was described with mean± standard deviation; otherwise,
39 106 median (1st quartile, 3rd quartile) was used to describe their features. Number of cases (percentage) was used to describe the proportion
40 107 of categorical variables between the two groups.

41 108 [‡] iAs: inorganic arsenic (iAs[□]+iAs[□]); MMA: monomethyl arsenate (MMA[□]+MMA[□]); DMA: dimethyl arsenate (DMA[□]+DMA[□]); tAs:
42 109 total arsenic (iAs[□]+iAs[□]+MMA+DMA); iAs%= iAs/tAs*100%; MMA%=MMA/tAs*100% and DMA%=DMA/tAs*100%.

43
44 110 **Table 2** shows the Four AA with FDR adjusted p-value <0.05 and VIP>1 in the cases were
45
46 111 observed significantly lower than those of their counterparts. Two of them were aromatic amino
47
48 112 acids (AAA) identified as phenylalanine and tryptophan, one is branched-chain amino acids (BCAA)
49
50 113 appraised as leucine and the last one is phenylalanyl phenylalanine. Spearman's rank correlation to
51
52 114 investigate the relationships among these four AA before the following analyzes and observed that
53
54 115 the AA were obviously correlated to each other with the correlation coefficient as 0.18 to 0.86 (**Table**

1
2
3 116 **S1)** Through a multivariable logistic stepwise regression model to obtain variables independently
4
5 117 associated with the risk to develop AISL and to avoid the potential collinearity among all
6
7 118 independent variables in the model. respectively. Finally, we found that tryptophan and
8
9 119 phenylalanine were significantly related to the AISL risk.

120 **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 ^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
Phenylalanyl phenylalanine	5.048	313.155	1.833	0.004	0.033

20 121 ^ξFrom Wilcoxon signed-rank test; ^ζAdjusted by false discovery rate (FDR).

22 122 **Figure 1** presents the individual association between AA and AISL. Based on locally weighted
23
24 123 regression curve, we observed obvious “dose-response” relationships between the levels of serum
25
26 124 tryptophan and phenylalanine with AISL. The risk to get AISL was significantly decreased in
27
28 125 participants with high serum tryptophan and phenylalanine levels. **Table 3** shows that As compared
29
30 126 to the lowest quartile, participants with the 3rd and 4th quartiles of tryptophan were significantly
31
32 127 linked to the decreased risk of AISL, with odd ratio (OR) and 95% confidence interval (CI) as
33
34 128 0.11(0.02, 0.54) for the former and 0.18(0.05, 0.70) for the latter, after adjusting for the potential
35
36 129 confounders such as age, gender, cigarette smoking status, alcohol consumption, fasting plasma
37
38 130 glucose, triglycerides, low-density lipoprotein and iAS% (percentage of inorganic arsenic by total
39
40 131 arsenic). A significant linear trend existed between the risk to develop AISL and serum tryptophan,
41
42 132 decreased by 56% (OR: 0.44; 95% CI: 0.24, 0.84) with per IQR increment of serum tryptophan
43
44 133 ($P=0.013$). Meanwhile, the similar association between serum phenylalanine level and AISL was also
45
46 134 found in the present study. After adjusting for the above-mentioned potential confounding factors,
47
48 135 the risk to get AISL for subjects in the 3rd and 4th quartiles significantly decreased by 84% (OR:
49
50 136 0.16; 95% CI: 0.04, 0.69) and 83% (OR: 0.17; 95% CI: 0.04, 0.68), respectively. There also existed a
51
52 137 linear trend between the serum phenylalanine concentration and the risk of AISL, significantly
53
54
55
56
57
58
59
60

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

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

Model	Tryptophan		Phenylalanine	
	Crude	Adjusted [§]	Crude	Adjusted [§]
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

141 [§] Values are odds ratios (95% confidence intervals) for arsenic-induced skin lesions from conditional logistic regression. IQR:
 142 interquartile range; Q1: the first quartering; Q2: the second quartering; Q3: the third quartering; Q4: the fourth quartering.

143 [§] Adjusted for: age, gender, cigarette smoking status, alcohol consumption, fasting plasma glucose, triglycerides, low-density
 144 lipoprotein and percentage of inorganic arsenic by total arsenic (iAS%).

145 **Table 4** shows the joint effect of tryptophan and phenylalanine levels on AISL We firstly
 146 classified the serum tryptophan and phenylalanine into two categories with the cut-off value of the
 147 mass spectrum peak area based on the ROC analysis as 1730858.446 for the former and
 148 1649420.815 for the latter, respectively. The higher level of these two serum AA was separately
 149 defined as equal to or over the cut-off value, while the lower categories were considered as less than
 150 the associated cut-off values. As compared to the reference group (both serum tryptophan and
 151 phenylalanine were at the lower levels), after adjusting for the impacts induced by some potential
 152 confounding factors, participants with higher level of both tryptophan and phenylalanine had the
 153 lowest risk to get AISL, significantly decreased by 94% (OR=0.06; 95%CI: 0.02, 0.22; $P<0.001$).
 154 This suggested that AISL jointly and negatively associated with serum tryptophan and phenylalanine
 155 concentration. While no significant interaction between tryptophan and phenylalanine on the
 156 occurrence of skin damage could be observed ($P=0.462$).

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

Tryptophan		Phenylalanine		N	Cases (%)	Crude		Adjusted [§]	
<cut-off value [§]	<cut-off value [§]	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

159 [‡] Cut-off value was determined by means of receiver operator characteristic (ROC) analysis.

160 [‡] Adjusted for age, gender, cigarette smoking status, alcohol consumption, fasting plasma glucose, triglycerides, low-density
161 lipoprotein and percentage of inorganic arsenic by total arsenic (iAS%).

162 **Table 5** shows that based on the ROC analysis, both serum tryptophan and phenylalanine were
163 suggested to be potential biomarkers in distinguishing AISL from a chronic arsenic exposure
164 population ($P=0.0020$ and $P=0.0017$). The area under the curve (AUC) and its related 95% CI,
165 sensitivity, specificity, positive predictive value and negative predictive value were 0.67 (0.57, 0.77),
166 69.64%, 62.50%, 65.00 and 67.31% for tryptophan, and 0.70 (0.60, 0.80), 69.64%, 69.64%, 69.64%
167 and 69.64% for phenylalanine, respectively. For the combined prediction of tryptophan and
168 phenylalanine on the AISL risk, the AUC (95% CI), sensitivity, specificity, positive predictive value
169 and negative predictive value were 0.72 (0.62, 0.81), 76.79%, 58.93%, 65.15 and 71.74%,
170 respectively. Our results suggested that these two AA could be either individually or jointly used as
171 indicators of arsenic induced skin lesions identification.

172

173 **Table 5** Combination of diagnostic indicators and ROC analysis results[§].

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 [‡]	0.72(0.62,0.81)	76.79	58.93	65.15	71.74	<0.001

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

177 [‡] Combined: tryptophan and phenylalanine. The combination is modeled according to the formula $\beta_1X_1 + \beta_2X_2$, with X_j denoting the
 178 standardized value for the j^{th} amino acid and β_j denoting the regression coefficient from the logistic regression model.

179 DISCUSSION

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

187 Participants enrolled in the current study were chronically exposed to arsenic via drinking water.
 188 The geometric mean (GM) and its related 95% CI of urinary iAs/creatinine and tAs/creatinine in this
 189 population were 17.49 (14.90, 20.53) $\mu\text{g/g}$ and 147.20 (129.00, 167.97) $\mu\text{g/g}$, respectively. They were
 190 much higher than those in the 20- $\mu\text{g/L}$ -exposed to arsenic via drinking water [GM (95% CI): 0.4
 191 (0.3,0.5) $\mu\text{g/g}$ for iAs and 9.1 (6.5,12.7) $\mu\text{g/g}$ for tAs], while obviously lower than those in the
 192 90- $\mu\text{g/L}$ -exposed group [GM (95% CI): 39.4 (31.4, 49.6) $\mu\text{g/g}$ for iAs and 248.7 (208.8, 296.3)
 193 $\mu\text{g/g}$]²⁵. The available reports show that arsenic induced health lesions can scarcely been cured though
 194 the medical technology has made great progress in the past²⁶. It is crucial to identify those who are
 195 most likely to progress to overt arsenic damages including AISL among people at risk as early as
 196 possible. Due to the arsenic methylation⁹, genetic differences⁸, and other possible metabolic
 197 mechanisms¹⁰, some residents exposed to arsenic will not develop to severe arsenic poisoning,
 198 especially for those at the early stages of arsenic exposure⁴. Metabolomics study, which mainly focus

1
2
3 199 on thoroughly assessing the variation of metabolites possibly linked to diseases occurrence and
4
5 200 development, has been widely utilized to help us understanding the pathogenesis and others because of
6
7 201 its much closer to the phenotypes as compared to other ‘OMICs’ study²⁷. Moreover, mathematical
8
9 202 modeling to assess the linkage between small molecular metabolites and arsenic toxicity has grown²⁸.
10
11 203 Developing a simple and interpretable modeling approach for the early detection of arsenic induced
12
13 204 health lesions is of great theoretical value and realistic meaning²⁹, though it might be difficult due to
14
15 205 population specific complexities and the impacts due to some potential unmeasured covariates such as
16
17 206 diet and genetic determinants.

18
19
20 207 Previous studies reported that gene-gene and gene-environment interaction were involved in
21
22 208 arsenicosis through toxicological mechanisms including genomic instability³⁰ and oxidative stress³¹.
23
24 209 Skin hyperpigmentation and palmoplantar hyperkeratosis could be biomarkers for long-term (many
25
26 210 years) internal dose. Identifying the differences in metabolites that are really associated with
27
28 211 phenotypes through metabolites analysis may promote our understanding and identification of AISL.
29
30 212 Animal study reveals that the disruption of amino acid metabolism upon arsenic exposure in rat which
31
32 213 may be beneficial on understanding the arsenic toxicity³². A recent population-based metabolomics
33
34 214 study also shows that the serum metabolites alteration significantly related to the risk of
35
36 215 arsenic-induced health damages²⁴. In the current study, three of the four amino acids are BCAA or
37
38 216 AAA are observed to be relevant to AISL occurrence. Several studies across numerous ethnic
39
40 217 backgrounds supports the usage of BCAA including leucine, isoleucine as well as valine and AAA
41
42 218 profile such as phenylalanine, tryptophan and tyrosine as biomarkers in determining metabolic
43
44 219 diseases^{27 33}. Simultaneously, Zhou *et al* report that arsenic-induced transformed cells exhibit apparent
45
46 220 alterations in metabolite profiles including down-regulated of leucine, tryptophan, and phenylalanine
47
48 221 in skin lesions group³⁴. Consistent with Zhou’s findings, two serum AAA (tryptophan, phenylalanine)
49
50 222 levels significantly associated with elevated risk to get AISL in our study.
51
52
53
54
55
56
57
58
59
60

1
2
3 223 Normal metabolism of AA are necessary for whole body homeostasis, growth and development,
4
5 224 and body health¹⁵. Studies have reported that the changes in the availability of AAA lead to a profound
6
7 225 effect on cell signaling, gene expression, brain, and neuroendocrine function³⁵. Tryptophan, an amino
8
9 226 acid metabolism related biomarker, is also a sensitive and specific indicator of oxidation. Tryptophan
10
11 227 metabolism in mammals is a physiological means of preserving immune homeostasis and associated
12
13 228 with oxidative stress, and inflammation^{36 37}. In addition, phenylalanine can be transformed into
14
15 229 specific neurotransmitters such as dopamine and adrenaline by the action of related enzymes. Wu and
16
17 230 colleagues³⁸ report that arsenic exposure will cause neurotransmitter metabolism disturbance, which
18
19 231 may explained the reduction of phenylalanine. Furthermore, as one of the peptide-bound
20
21 232 phenylalanine, phenylalanyl phenylalanine has been reported to affect protein synthesis and
22
23 233 secretion³⁹, potentially indicating the possible relation between endothelium dysfunction and
24
25 234 phenylalanine metabolism disorder.

26
27
28
29 235 The relationship between amino acid metabolism and AISL was unclear. The notable alteration of
30
31 236 tryptophan and phenylalanine in this study may well indicate the occurrence of metabolic disorders
32
33 237 due to arsenic exposure, is beneficial on intensive understanding the effects of arsenic toxicity and is of
34
35 238 great importance and realistic value in the early identification of the occurrence as well as delaying the
36
37 239 progression of various arsenic-induced health lesions including AISL. The main strength of this study
38
39 240 might be that the findings were depended on a community-based long-term arsenic exposure cohort
40
41 241 with well-designed quality assurance and quality control throughout the study. However, the findings
42
43 242 are mainly based on case-control study, which only revealed the association between amino acid
44
45 243 metabolism and the risk of AISL rather than confirming their causal relationship. The participants
46
47 244 were mainly exposed to arsenic via drinking water, which would limit the findings extrapolated to the
48
49 245 other arsenic exposure population via food and other ways. Therefore, Additional elaborate
50
51 246 population-based studies are needed to verify our discoveries.
52
53
54
55
56
57
58
59
60

1
2
3 247 Taken together, two AAA's (tryptophan, phenylalanine) reduction were closely linked to the
4
5 248 higher risk of AISL. It also suggests that tryptophan and phenylalanine are useful for distinguishing
6
7 249 AISL earlier or screening of the high-risk individuals from their counterparts in a long-term exposed to
8
9 250 low-level arsenic population via drinking water.
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For peer review only

1
2
3 **Contributors:** Guangyun Mao and Yaping Wei designed the study. Chaonan Jia participated in
4 collecting data. Yuan Lan and Chaonan Jia audited the data. Yaping Wei, Xiangqing hou, Jushuang Li,
5 Tao Wang conducted the literature search, Yaping Wei, Chaonan Jia conducted statistical analysis and
6 interpreted the results. Yaping Wei and Chaonan Jia wrote the first draft of the manuscript. Guangyun
7 Mao reviewed the final manuscript and did substantial contributions.
8

9
10 **Funding:** This work was supported by the Science Technology Department of Zhejiang Province
11 grant (2013C33169), the initial Scientific Research Fund (KYQD170301) and the major project of the
12 Eye Hospital of Wenzhou Medical University (YNZD201602).
13

14 **Acknowledgments:** We gratefully acknowledge all of the study participants of this project. We wish
15 to thank all of the dedication and hard work of the field team at the department of preventive medicine,
16 school of public health and management, Wenzhou medical university.
17

18 **Conflicts of Interest:** The authors declare no conflict of interest.
19

20 **Data Sharing Statement:** No additional unpublished data are available.
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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.

- 1
- 2
- 3 24. Jia C, Wei Y, Guan X, et al. Comprehensive Analysis of the Metabolomic Characteristics on the Health Lesions
- 4 Induced by Chronic Arsenic Exposure: A Metabolomics Study. (under review) .
- 5
- 6 25. Sun G, Xu Y, Li X, et al. Urinary arsenic metabolites in children and adults exposed to arsenic in drinking water
- 7 in Inner Mongolia, China. *Environmental Health Perspectives*. 2007; 115:648-52.
- 8
- 9 26. Kuo CC, Moon KA, Wang SL, et al. The Association of Arsenic Metabolism with Cancer, Cardiovascular
- 10 Disease, and Diabetes: A Systematic Review of the Epidemiological Evidence. *Environmental Health*
- 11 *Perspectives*. 2017; 125:15.
- 12
- 13 27. Wang TJ, Larson MG, Vasan RS, et al. Metabolite profiles and the risk of developing diabetes. *Nature Medicine*.
- 14 2011; 17:448-U83.
- 15
- 16 28. Lawley SD, Cinderella M, Hall MN, et al. Mathematical model insights into arsenic detoxification. *Theor Biol*
- 17 *Med Model*. 2011; 8:31.
- 18
- 19 29. Juhasz AL, Herde P, Herde C, et al. Validation of the predictive capabilities of the Sbrc-G in vitro assay for
- 20 estimating arsenic relative bioavailability in contaminated soils. *Environ Sci Technol*. 2014; 48:12962-9.
- 21
- 22 30. Kibriya MG, Jasmine F, Parvez F, et al. Association between genome-wide copy number variation and
- 23 arsenic-induced skin lesions: a prospective study. *Environ Health*. 2017; 16:13.
- 24
- 25 31. Priori R, Scrivo R, Brandt J, et al. Metabolomics in rheumatic diseases: the potential of an emerging
- 26 methodology for improved patient diagnosis, prognosis, and treatment efficacy. *Autoimmun Rev*. 2013;
- 27 12:1022-30.
- 28
- 29 32. Wang X, Mu X, Zhang J, et al. Serum metabolomics reveals that arsenic exposure disrupted lipid and amino
- 30 acid metabolism in rats: a step forward in understanding chronic arsenic toxicity. *Metallomics Integrated*
- 31 *Biometal Science*. 2015; 7:544.
- 32
- 33 33. Weng L, Quinlivan E, Gong Y, et al. Association of Branched and Aromatic Amino Acids Levels with
- 34 Metabolic Syndrome and Impaired Fasting Glucose in Hypertensive Patients. *Metabolic Syndrome And Related*
- 35 *Disorders*. 2015; 13:195-202.
- 36
- 37 34. Zhou YY, Wang YF, Su J, et al. Integration of microRNAome, proteomics and metabolomics to analyze
- 38 arsenic-induced malignant cell transformation. *Oncotarget*. 2017; 8:90879-96.
- 39
- 40 35. Lee JC, Chen MJ, Chang CH, et al. Plasma amino acid levels in patients with colorectal cancers and liver
- 41 cirrhosis with hepatocellular carcinoma. *Hepato-Gastroenterol*. 2003; 50:1269-73.
- 42
- 43 36. Munn DH, Mellor AL. Indoleamine 2,3 dioxygenase and metabolic control of immune responses. *Trends In*
- 44 *Immunology*. 2013; 34:137-43.
- 45
- 46 37. Pawlak K, Domaniewski T, Mysliwiec M, et al. The kynurenines are associated with oxidative stress,
- 47 inflammation and the prevalence of cardiovascular disease in patients with end-stage renal disease.
- 48 *Atherosclerosis*. 2009; 204:309-14.
- 49
- 50 38. Wu L-L, Gong W, Shen S-P, et al. Multiple metal exposures and their correlation with monoamine
- 51 neurotransmitter metabolism in Chinese electroplating workers. *Chemosphere*. 2017; 182:745-52.
- 52
- 53 39. Zhou MM, Wu YM, Liu HY, et al. Effects of phenylalanine and threonine oligopeptides on milk protein
- 54 synthesis in cultured bovine mammary epithelial cells. *J Anim Physiol Anim Nutr*. 2015; 99:215-20.
- 55
- 56
- 57
- 58
- 59
- 60

1
2
3 **Figure 1** Association between the peak intensity of tryptophan and phenylalanine and
4 arsenic-induced skin lesions based on multivariable locally weighted regression models. a:tryptophan;
5
6
7 b:phenylalanine.
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

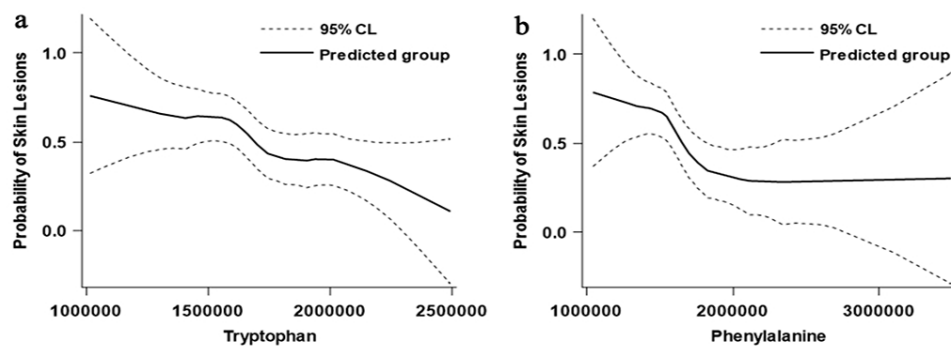


Figure 1 Association between the peak intensity of tryptophan and phenylalanine and arsenic-induced skin lesions based on multivariable locally weighted regression models. a: tryptophan; b: phenylalanine.

90x31mm (300 x 300 DPI)

Table S1Correlation 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).

For peer review only

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:	<i>BMJ Open</i>
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

SCHOLARONE™
Manuscripts

1
2
3 **Tryptophan, phenylalanine and arsenic-induced skin lesions in a chronic arsenic exposure**
4 **Chinese population via drinking water: data from a metabolomics study**
5
6
7
8
9

10 Yaping Wei^{1,2}, Chaonan Jia,^{1,2} Yuan Lan³, Xiangqing Hou^{1,2}, Jingjing Zuo³, Jushuang Li^{1,2}, Tao
11 Wang^{1,2}, Guangyun Mao^{1,2,4}✉
12
13
14

15
16 ¹ Department of Preventive Medicine, School of Public Health and Management, Wenzhou Medical
17 University, Wenzhou, China.

18
19 ² Center on Evidence-Based Medicine & Clinical Epidemiological Research, School of Public Health,
20 Wenzhou Medical University, Wenzhou, China.
21

22
23 ³ School of Ophthalmology & Optometry, Wenzhou Medical University, Wenzhou, China.

24
25 ⁴ Center on Clinical Research, the Affiliated Eye Hospital of Wenzhou Medical University, Wenzhou,
26 China.
27

28
29 Yaping Wei and Chaonan Jia contributed equally to this paper.
30

31
32
33 ✉Corresponding author: Guangyun Mao, Room 7A321, School of Public Health and Management,
34 Wenzhou Medical University, Chashan University town, Ouhai district, Wenzhou, China. 325035

35 E-mail: maogy@wmu.edu.cn

36 Telephone: +86-577-86689848

37 Fax numbers: +86-577-86699122
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For peer review only

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.

1 INTRODUCTION

2 Chronic arsenic exposure via drinking water is widely believed as a global health concern,
3 affecting a large amount of people worldwide. It may give rise to several human health issues and
4 has been documented to associate with cardiovascular disease, diabetes, cancer and others^{1 2}. With
5 the industrial boom and dramatic rise of worldwide water pollution including arsenic contamination
6 in the past, the prevalence and burden of arsenic-induced health damage will continue to increase.
7 Skin has been confirmed as one of the most common and susceptible target of arsenic-induced health
8 lesions. Cutaneous skin lesions are typical signs of arsenicosis after persistent arsenic exposure for a
9 long term which are characterized by hyperkeratosis and hyperpigmentation. Considerable evidences
10 of the prevalence of arsenical skin lesions had been observed in many countries³⁻⁵.

11 As arsenic-induced skin lesions (AISL) have been widely accepted as the major early
12 manifestation of arsenic toxicity⁶ and might be indicators of susceptibility to more serious
13 arsenic-induced health hazards⁷, it is particularly crucial to identify participants at risk as early as
14 possible for preventing the onset or delaying the progression of the serious health problems
15 effectively. Several possible mechanism such as genetic differences⁸, oxidative stress⁹ and epigenetic
16 dysregulation¹⁰ and others may explain arsenic poisoning. Previous studies also reported that arsenic
17 methylation in vivo might be associated with metabolic syndrome^{11 12}.

18 Amino acids (AAs) are the "basic unit" that make up the body's various proteins and necessary
19 to maintain the health. Some AAs are important regulators of some key metabolic pathways and have
20 great importance in maximizing efficiency of food utilization, enhancing protein accretion and health
21 improvement^{13 14}. Abnormal metabolism of AAs will disturb the homeostasis of the body, impairs
22 growth and development, and even causes death¹⁵. So, the levels of serum AAs may be an important
23 implication for the metabolic status and disease condition. As a powerful tool in system biology
24 research, metabolomics approach is beneficial on unbiased monitoring changes in endogenous
25 metabolism-related physiological processes, providing integrative information on the distinctive

1
2
3 26 features across multiple functional levels, and offering a window to capture the core attributes
4
5
6 27 responsible for various phenotypes, which are particularly important in understanding the relevant
7
8 28 pathophysiological changes of a disease and its status, identifying novel biomarkers for risk
9
10 29 screening, diagnosis, treatment and prognosis of important human diseases¹⁶⁻¹⁸.

11
12 30 Animal experiments and epidemiological study have reported obvious arsenic-related
13
14 31 metabolomics perturbations^{19 20}. All of these researches substantially suggests that the relationship
15
16 32 between specific metabolites and arsenic-induced health lesions should be investigated. However,
17
18 33 few works have been conducted to comprehensively examine the metabolic mechanism relevant to
19
20 34 AISL, especially for the AAs metabolism. The present study aims to quantitatively examine the
21
22 35 association of several specific AAs with AISL and the ability to identify AISL.
23
24
25
26

27 36 **Methods**

30 37 **Patient and Public Involvement**

31
32
33
34 38 No patients were involved.
35

36 39 **Study Population**

37
38 40 This study was originally from a randomized, double-blind, and placebo controlled clinical trial
39
40 41 (NCT02235948) in 2010, in which all subjects were randomly selected using permuted block
42
43 42 randomization from a single rural area in a population chronically exposed to low-level arsenic
44
45 43 drinking water, had similar life style and influences under similar environmental factors. Information
46
47 44 on the inclusion and exclusion criteria of the participants could be found in our previous study²¹.
48
49 45 Strictly following the criteria of arsenicosis²². AISL was diagnosed as the presence of arsenic-
50
51 46 induced keratosis, hyperpigmentation or depigmentation by a physician from Wenzhou medical
52
53 47 university at the beginning of the trial. Among 450 residents aged 18 to 79 years old enrolled in the
54
55 48 above-mentioned trial, 56 of them were diagnosed as AISL and selected as the case. Another 56
56
57 49 participants without AISL matched by gender and similar age (± 1 year) from the same population
58
59
60

1
2
3 50 were picked out as the control. The inclusion criteria were subjects having the metabolomic test.
4
5
6 51 Unmatched participants and those without serum metabolites data were excluded. Informed consent
7
8 52 was obtained from all participants and this study was approved by the ethics committee of Wenzhou
9
10 53 Medical University, Wenzhou, China.

12 54 **Data Collection and Assessment**

14
15 55 Blood and urine samples were also collected at the time of participants' enrollment. Detailed
16
17 56 data collection of blood and urine samples and assessment methods for clinical and
18
19 57 sociodemographic variables had been published previously²¹. The epuration of various urinary
20
21 58 arsenic species were conducted by means of a high-performance liquid chromatography coupled
22
23 59 mass spectrometer system for separation and detection²³. The species of arsenic in urine samples
24
25 60 consisted of inorganic arsenic (iAs, [iAs^{III} plus iAs^V]), monomethyl arsenate (MMA, [MMA^{III} plus
26
27 61 MMA^V]) and dimethyl arsenate (DMA, [DMA^{III} plus DMA^V]). All arsenic species were corrected
28
29 62 by creatinine.

33 63 **UPLC-MS/MS Metabonomic Profiling**

34
35
36 64 Serum samples (200 µL in microcentrifuge tubes) were thawed to room temperature (25°C)
37
38 65 and 600 µL mixture (90% acetonitrile - 10%water) were added to each sample. The samples were
39
40 66 vigorously mixed for 20 seconds and centrifuged for 5 min at 12000 rpm (20°C). The top 400 µL of
41
42 67 each supernatant were then transferred and dried down in a vacuum concentrator centrifuge. The
43
44 68 dried samples were re-suspended in 130 µL of water (including 15% acetonitrile), mixed vigorously
45
46 69 for 20 seconds and repeated the centrifugation method described above. Two µL of the supernatant
47
48 70 were collected as samples to be determined. Serum metabolic profile acquisition was performed by
49
50 71 using ACQUITY UPLC[®]/Xevo[®] G2 QToF/MS^E (Waters Corp., Milford, MA, USA).
51
52 72 Chromatographic separation was performed at 50°C using a WATERS HSS T3 column (2.1×100
53
54 73 mm, 1.7 µm) with a flow rate of 0.4mL/min. The mobile phase was a mixture of (A) H₂O with 0.1%
55
56 74 formic acid and (B) methanol with 0.1% formic acid. Elution was in linear gradient with the
57
58
59
60

1
2
3 75 programmed gradient at 0 min with 100% A and 0% B, 1.00min with 100%A and 0% B, 8 min with
4
5 76 0%A and 100% B, 13.00 min with 0% A and 100% B. The mass spectrometer was operated under
6
7
8 77 both positive-ion (ESI⁺) mode and negative-ion (ESI⁻) mode electrospray ionization. The scan range
9
10 78 was from 50 to 1200 m/z. Data was collected in both ESI⁺ and ESI⁻ modes. Capillary voltage was set
11
12 79 at 3000 V and 2500 V, respectively. The desolvation flow rate was 800 L/h at 350°C. Argon was
13
14
15 80 used as a collision gas, and the collision energy was adjusted from 10 eV to 40 eV for each analysis.
16
17 81 Quantum clustering (QC) samples were prepared by pooling aliquots of each sample and used to
18
19 82 reflect the reliability of further metabolomics analysis. After peak deconvolution, alignment,
20
21
22 83 integration and normalization, the data including retention time (RT), mass to charge ratio(m/z), and
23
24 84 peak intensity were extracted from raw chromatograms using Progenesis QI 2.0 (Waters Corp.,
25
26 85 Milford, MA, USA). The MS/MS mode was performed to obtain metabolites levels processed with
27
28
29 86 MarkerLynx Applications Manager Version 4.1 (Waters Corp., Milford, MA, USA).

30 87 **Distinct Metabolites Identification**

31
32
33 88 The peak intensity of metabolites for the 56 pairs were acquired and then imported to
34
35 89 MetaboAnalyst 4.0 (<http://www.metaboanalyst.ca/>) for statistical analysis. A partial least-squares
36
37 90 discriminant analysis (PLS-DA), which is a supervised and well accepted pattern recognition
38
39
40 91 approach, was used for the differentiation between the cases and controls. False discovery rate (FDR)
41
42 92 adjusted p-value in univariate analysis were performed to reduce the potential impact induced by
43
44
45 93 false positive of the results. The criteria used in the selection of metabolites include variable
46
47 94 importance in the project (VIP) scores >1 in PLS-DA and the crude or FDR adjusted p-value all <
48
49 95 0.05. We identified a total of 70 extracted small molecular metabolites that were linked to the
50
51 96 recognition of AISL. Among them, 4 metabolites were confirmed as AAs and obviously
52
53
54 97 down-regulated in the AISL group.

55 98 **Statistical Analysis**

56
57
58 99 The normality of continuous data was assessed using both QQ-plots and Shapiro-Wilk test. The
59
60

1
2
3 100 comparison between the cases and controls was performed with the paired *t*-test if they met normal
4
5
6 101 or similar normal distribution. Otherwise, Wilcoxon signed rank test would be used. Differences in
7
8 102 the proportion of categorical variables between the two groups were evaluated by McNemar-Bowker
9
10 103 tests. We firstly used locally weighted regression (Loess) models to estimate the “real” relationship
11
12 104 between serum AAs levels and the odds of AISL. Then, multivariable conditional logistic regression
13
14 105 models were performed to examine the association between the contributing AAs levels and AISL
15
16
17 106 after adjusting for some potential confounding factors. The individual impacts of AA metabolites on
18
19 107 the risk of AISL were quantified separately by odds ratio (OR) and 95% confidence interval (CI) in
20
21
22 108 the following two ways: with AA as a categorical variable (quartiles) and as a continuous variable
23
24 109 [scaled to interquartile range (IQR)]. As the distinct metabolites might be high related to each other,
25
26 110 collinearity should be well considered. To screen appropriate covariates in the logistic models, some
27
28
29 111 potential risk factors of AISL such as the duration of arsenic exposure, serum folate, cigarette
30
31 112 smoking, alcohol consumption, blood urea nitrogen and others were added in the model as covariate,
32
33 113 respectively. As too many covariates in a multiple regression model will lead to overfitting to some
34
35 114 extent²⁴, we finally select no more than 5 variables as confounding factors to decrease the potential
36
37
38 115 overfitting when assessing the association between AAs and AISL. Meanwhile, we used the
39
40 116 variance inflation factor (VIF) based on the VIF package of R software to detect potential
41
42 117 collinearity among the AAs. When the VIF is greater than 1.5, it would be considered as collinearity
43
44
45 118 existed in the model and the associated variable would be removed from the model. The combined
46
47 119 effects of relevant AAs on AISL were also performed. In addition, receiver operator characteristic
48
49 120 (ROC) analysis was applied to evaluate the value and feasibility of the AAs as the potential sensitive
50
51
52 121 and specific biomarker to recognize AISL. Data management, analysis and figure drawing were
53
54 122 finished using R version 3.4.4 (Copyright © 2018 The R Foundation for Statistical Computing). All
55
56 123 tests were two-sides and $P \leq 0.05$ was set as significant level.
57
58

59 124 **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[‡].

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[‡]			
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

[‡] 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.

[‡] 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 ^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
Phenylalanyl phenylalanine	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).

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[‡]

Models	Tryptophan		Phenylalanine		Leucine		Phenylalanyl phenylalanine	
	Crude	Adjusted [‡]	Crude	Adjusted [‡]	Crude	Adjusted [‡]	Crude	Adjusted [‡]
Amino acids as continuous variable								
Per	0.48	0.44	0.57	0.54	0.45	0.45	0.62	0.62
IQR	(0.27,0.84)	(0.24,0.80)	(0.36,0.91)	(0.32,0.89)	(0.25,0.82)	(0.24,0.83)	(0.36,1.04)	(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 (0.03,0.70)
P trend	0.007	0.007	<0.001	<0.001	0.002	0.002	0.005	0.005

[‡]Values are odds ratio (95% confidence intervals) for arsenic-induced skin lesions from conditional logistic regression. IQR: interquartile range; Q₁: the 1st quartile; Q₂: the 2nd quartile; Q₃: the 3rd quartile; Q₄: the 4th quartile.

[‡] 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 phenylalanine had the smallest VIF value (VIF=1.04) and no obvious collinearity existed (**Table S5**). 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 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

1
2
3 163 lower tryptophan and higher phenylalanine (category C), and higher levels of both tryptophan and
4
5
6 164 phenylalanine (category D), respectively. Obvious decrease trend of the probability of AISL was
7
8 165 observed among these 4 categories. As compared to the category A, adjusted OR (95% CI) for
9
10 166 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).
11
12 167 Subjects with higher levels of both tryptophan and phenylalanine had the lowest odds of AISL,
13
14
15 168 significantly decreased by 94% (OR=0.06; 95%CI: 0.02, 0.22; P<0.001), after adjusting for the
16
17 169 impacts induced by some potential confounding factors. This suggested that tryptophan and
18
19 170 phenylalanine were jointly associated with the presence of AISL. While no significant interaction
20
21
22 171 between the two AAs on the occurrence of AISL could be observed (P=0.270), which indicated that
23
24 172 each AA was independently associated with AISL though their joint impact was significant.

25
26 **Table 4** Joint association between tryptophan and phenylalanine levels with arsenic-induced skin
27 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	0.47(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

38 ^ξ Cut-off value was determined by means of receiver operator characteristic analysis.

39 ^ζ Adjusted for body mass index, serum folate, triglycerides, total arsenic.

40 173
41 174 **Table 5** shows that, based on the ROC analysis, both serum tryptophan and phenylalanine might
42
43 175 be potential biomarkers in distinguishing AISL from a chronic arsenic exposure population
44
45
46 176 (P=0.0020 and P=0.0017). The area under the curve (AUC) and its related 95% CI, sensitivity,
47
48 177 specificity, positive predictive value and negative predictive value were 0.67 (0.57, 0.77), 69.64%,
49
50 178 62.50%, 65.00 and 67.31% for tryptophan, and 0.70 (0.60, 0.80), 69.64%, 69.64%, 69.64% and
51
52 179 69.64% for phenylalanine, respectively. The AUC (95% CI), sensitivity, specificity, positive
53
54
55 180 predictive value and negative predictive value of the combination of them were 0.72 (0.62, 0.81),
56
57 181 76.79%, 58.93%, 65.15 and 71.74%, respectively. Our results suggested that these two AAs could be
58
59 182 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 X_j denoting the standardized value for the j^{th} amino acid and β_j 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) $\mu\text{g/g}$ and 147.20 (129.00, 167.97) $\mu\text{g/g}$, respectively. They were much higher than those in the 20 $\mu\text{g/L}$ exposed to arsenic via drinking water [GM (95% CI):

1
2
3 203 0.4 (0.3,0.5) $\mu\text{g/g}$ for iAs and 9.1 (6.5,12.7) $\mu\text{g/g}$ for tAs], while obviously lower than those in the 90
4
5
6 204 $\mu\text{g/L}$ exposed group [GM (95% CI): 39.4 (31.4, 49.6) $\mu\text{g/g}$ for iAs and 248.7 (208.8, 296.3) $\mu\text{g/g}$]²⁵.
7
8 205 An available report has shown hat AISL cannot be completely cured even though the medical
9
10 206 technology has already made great progress²⁶. So, it is crucial to identify those who are most likely to
11
12
13 207 progress to overt arsenic damages including AISL among people at risk as early as possible.
14
15 208 Metabolomics study, which mainly focus on thoroughly assessing the variation of metabolites
16
17 209 possibly linked to diseases occurrence and development, has been widely utilized to help us
18
19
20 210 understand pathogenesis of diseases because of its relevance to the phenotypes as compared to other
21
22 211 ‘OMICS’ study²⁷. Moreover, mathematical modeling to assess the linkage between small molecular
23
24 212 metabolites and arsenic toxicity has grown²⁸. Developing a simple and interpretable modeling
25
26 213 approach for the early detection of arsenic induced health lesions is of great theoretical value and
27
28
29 214 realistic meaning²⁹, though it might be difficult due to population specific complexities and the
30
31 215 impacts due to some potential unmeasured covariates such as diet and genetic determinants.
32

33 216 Previous studies reported that gene-gene and gene-environment interaction were involved in
34
35
36 217 arsenicosis through toxicological mechanisms including genomic instability³⁰ and oxidative stress³¹.
37
38 218 Skin hyperpigmentation and palmoplantar hyperkeratosis could be biomarkers for long-term arsenic
39
40 219 exposure identifying the differences in metabolites that are really associated with phenotypes through
41
42
43 220 metabolites analysis may promote our understanding and identification of AISL. Animal study
44
45 221 reveals that the disruption of amino acids metabolism upon arsenic exposure in rat which may be
46
47 222 beneficial on understanding arsenic toxicity³². In our previous population-based metabolomics study,
48
49
50 223 we found that serum metabolites alteration were significantly related to the risk of arsenic-induced
51
52 224 health damages. In the current study, we detected that BCAA or AAA were also significantly
53
54 225 relevant to AISL occurrence. Several studies across numerous ethnic backgrounds supports the usage
55
56 226 of BCAA including leucine, isoleucine as well as valine and AAA profile such as phenylalanine,
57
58
59 227 tryptophan and tyrosine as biomarkers in determining metabolic diseases^{27 33}. Simultaneously, Zhou
60

1
2
3 228 *et al* reports that arsenic-induced transformed cells exhibit apparent alterations in metabolite profiles
4
5 including down-regulated of leucine, tryptophan, and phenylalanine in skin lesions group³⁴.
6 229
7
8 230 Consistent with Zhou's findings, two serum AAA (tryptophan, phenylalanine) levels were also
9
10 231 significantly associated with AISL in our study.
11

12 232 Normal metabolism of amino acids are necessary for whole body homeostasis, growth and
13
14 development, and health status¹⁵. Studies have reported that the changes in the availability of AAA
15 233
16 will affect cell signaling, gene expression, brain, and neuroendocrine function³⁵. Tryptophan, an
17 234
18 amino acid metabolism related biomarker, is also a sensitive and specific indicator of oxidation.
19 235
20 Tryptophan metabolism in mammals is a physiological means of preserving immune homeostasis
21 236
22 associated with oxidative stress and inflammation^{36 37}. In addition, phenylalanine can be transformed
23 237
24 into specific neurotransmitters such as dopamine and adrenaline by the action of related enzymes.
25 238
26 Wu and colleagues³⁸ reported that arsenic exposure would lead to neurotransmitter metabolism
27 239
28 disturbance, which might explain the reduction of phenylalanine. Furthermore, as one of the
29 240
30 peptide-bound phenylalanine, phenylalanyl phenylalanine has been reported to affect protein
31 241
32 synthesis and secretion³⁹, potentially indicating the possible relation between endothelium
33 242
34 dysfunction and phenylalanine metabolism disorder. The relationship between amino acid
35 243
36 metabolism and AISL was still unclear. The notable alteration of tryptophan and phenylalanine in
37 244
38 this study may well indicate the occurrence of metabolic disorders due to arsenic exposure. It is also
39 245
40 beneficial to understand the effects of arsenic toxicity and of great importance in early identification
41 246
42 of occurrences as well as delaying the progression of various arsenic-induced health lesions
43 247
44 including AISL.
45 248
46
47
48
49
50

51 249 The current study included 56 AISL cases matched 56 non-AISL controls and the sample size might be
52
53 potentially insufficient. To estimate the impact due to this potential insufficient sample size on our conclusion,
54 250
55 the PROC POWER procedure for paired design study in SAS 9.4 (SAS Institute Inc.) was applied to assess
56 251
57 the power of the 4 AAs when assessing their associations with AISL in this study. The results showed that the
58 252
59 lowest power associated with all of these four AAs was 0.911 based on 56 pairs of participants (Figure 1S). It
60 253

1
2
3 254 suggested that with type I error as 0.05, total sample size as 56 pairs and two-sided test, the powers associated
4
5 255 with these 4 amino acids were all great than 0.8. So, we believed that the sample size for the present study, 56
6
7 256 pairs, would well balance the power of tests. Furthermore, previous metabolomics studies usually have sample
8
9
10 257 size no more than 40 cases in each group^{40 41}.

11
12 258 The main strength of this study is that the findings were depended on a community-based,
13
14 259 long-term arsenic exposure cohort with well-designed quality assurance and quality control
15
16 260 throughout the study, and the AAs were detected with non-targeted metabolomics approach through
17
18
19 261 the discovery and validation phases. However, there are also several limitations to this study. Firstly,
20
21 262 although it can assess a large amount of metabolites precisely and efficiently, untargeted
22
23 263 metabolomics approach only provide relative levels of AAs instead of their accurate quantitative
24
25
26 264 concentration. Secondly, these findings are mainly based on a case-control study, which only reveals
27
28 265 the association between amino acid metabolism and the odds of AISL rather than confirming their
29
30 266 causal relationship. Finally, the participants were mainly exposed to arsenic via drinking water,
31
32
33 267 which would limit the findings extrapolated to the other arsenic exposure population via food and
34
35 268 other ways. Therefore, additional elaborate population-based studies are needed to verify our
36
37 269 discoveries.

38
39 270 In conclusion, specific amino acids might be linked to AISL and amino acids metabolism may
40
41
42 271 play an important role in AISL early identification. Additional studies are needed to confirm our
43
44 272 findings

45
46
47 **Contributors:** Guangyun Mao and Yaping Wei designed the study. Chaonan Jia participated in
48
49
50
51
52
53
54
55
56
57
58
59
60
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).

1
2
3 **Acknowledgments:** We gratefully acknowledge all of the study participants of this project. We wish
4 to thank all of the dedication and hard work of the field team at the department of preventive
5 medicine, school of public health and management, Wenzhou medical university. Our thanks must
6 also go to Dr. Kendrick Hii, a native-English speaker, for his valuable suggestions and edits to the
7 manuscript.
8

9
10 **Conflicts of Interest:** The authors declare no conflict of interest.
11

12 **Data Sharing Statement:** No additional unpublished data are available.
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For peer review only

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

- 1
2
3 Randomized, Double-Blind, and Placebo Controlled Clinical Trial. *Medicine (Baltimore)*. 2015;
4 94:e1872.
5
6 22. Standard of diagnosis for endemic arsenism of China (WS/T211-2001). the Ministry of Health of
7 the People's Republic of China. Beijing 2001.
8
9 23. Chen X, Guo X, He P, et al. Interactive Influence of N6AMT1 and As3MT Genetic Variations on
10 Arsenic Metabolism in the Population of Inner Mongolia, China. *Toxicol Sci*. 2017; 155:124-34.
11
12 24. Peduzzi P, Concato J, Kemper E, et al. A simulation study of the number of events per variable in
13 logistic regression analysis. *J Clin Epidemiol*. 1996; 49:1373-9.
14
15 25. Sun G, Xu Y, Li X, et al. Urinary arsenic metabolites in children and adults exposed to arsenic in
16 drinking water in Inner Mongolia, China. *Environmental Health Perspectives*. 2007; 115:648-52.
17
18 26. Kuo CC, Moon KA, Wang SL, et al. The Association of Arsenic Metabolism with Cancer,
19 Cardiovascular Disease, and Diabetes: A Systematic Review of the Epidemiological Evidence.
20 *Environmental Health Perspectives*. 2017; 125:15.
21
22 27. Wang TJ, Larson MG, Vasan RS, et al. Metabolite profiles and the risk of developing diabetes.
23 *Nature Medicine*. 2011; 17:448-U83.
24
25 28. Lawley SD, Cinderella M, Hall MN, et al. Mathematical model insights into arsenic
26 detoxification. *Theor Biol Med Model*. 2011; 8.
27
28 29. Juhasz AL, Herde P, Herde C, et al. Validation of the predictive capabilities of the Sbrc-G in
29 vitro assay for estimating arsenic relative bioavailability in contaminated soils. *Environ Sci*
30 *Technol*. 2014; 48:12962-9.
31
32 30. Kibriya MG, Jasmine F, Parvez F, et al. Association between genome-wide copy number
33 variation and arsenic-induced skin lesions: a prospective study. *Environ Health*. 2017; 16:13.
34
35 31. Priori R, Scrivo R, Brandt J, et al. Metabolomics in rheumatic diseases: the potential of an
36 emerging methodology for improved patient diagnosis, prognosis, and treatment efficacy.
37 *Autoimmun Rev*. 2013; 12:1022-30.
38
39 32. Wang X, Mu X, Zhang J, et al. Serum metabolomics reveals that arsenic exposure disrupted lipid
40 and amino acid metabolism in rats: a step forward in understanding chronic arsenic toxicity.
41 *Metallomics Integrated Biometal Science*. 2015; 7:544.
42
43 33. Weng L, Quinlivan E, Gong Y, et al. Association of Branched and Aromatic Amino Acids Levels
44 with Metabolic Syndrome and Impaired Fasting Glucose in Hypertensive Patients. *Metabolic*
45 *Syndrome And Related Disorders*. 2015; 13:195-202.
46
47 34. Zhou YY, Wang YF, Su J, et al. Integration of microRNAome, proteomics and metabolomics to
48 analyze arsenic-induced malignant cell transformation. *Oncotarget*. 2017; 8:90879-96.
49
50 35. Lee JC, Chen MJ, Chang CH, et al. Plasma amino acid levels in patients with colorectal cancers
51 and liver cirrhosis with hepatocellular carcinoma. *Hepato-Gastroenterol*. 2003; 50:1269-73.
52
53 36. Munn DH, Mellor AL. Indoleamine 2,3 dioxygenase and metabolic control of immune responses.
54 *Trends In Immunology*. 2013; 34:137-43.
55
56 37. Pawlak K, Domaniewski T, Mysliwiec M, et al. The kynurenines are associated with oxidative
57 stress, inflammation and the prevalence of cardiovascular disease in patients with end-stage renal
58 disease. *Atherosclerosis*. 2009; 204:309-14.
59
60 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.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For peer review only

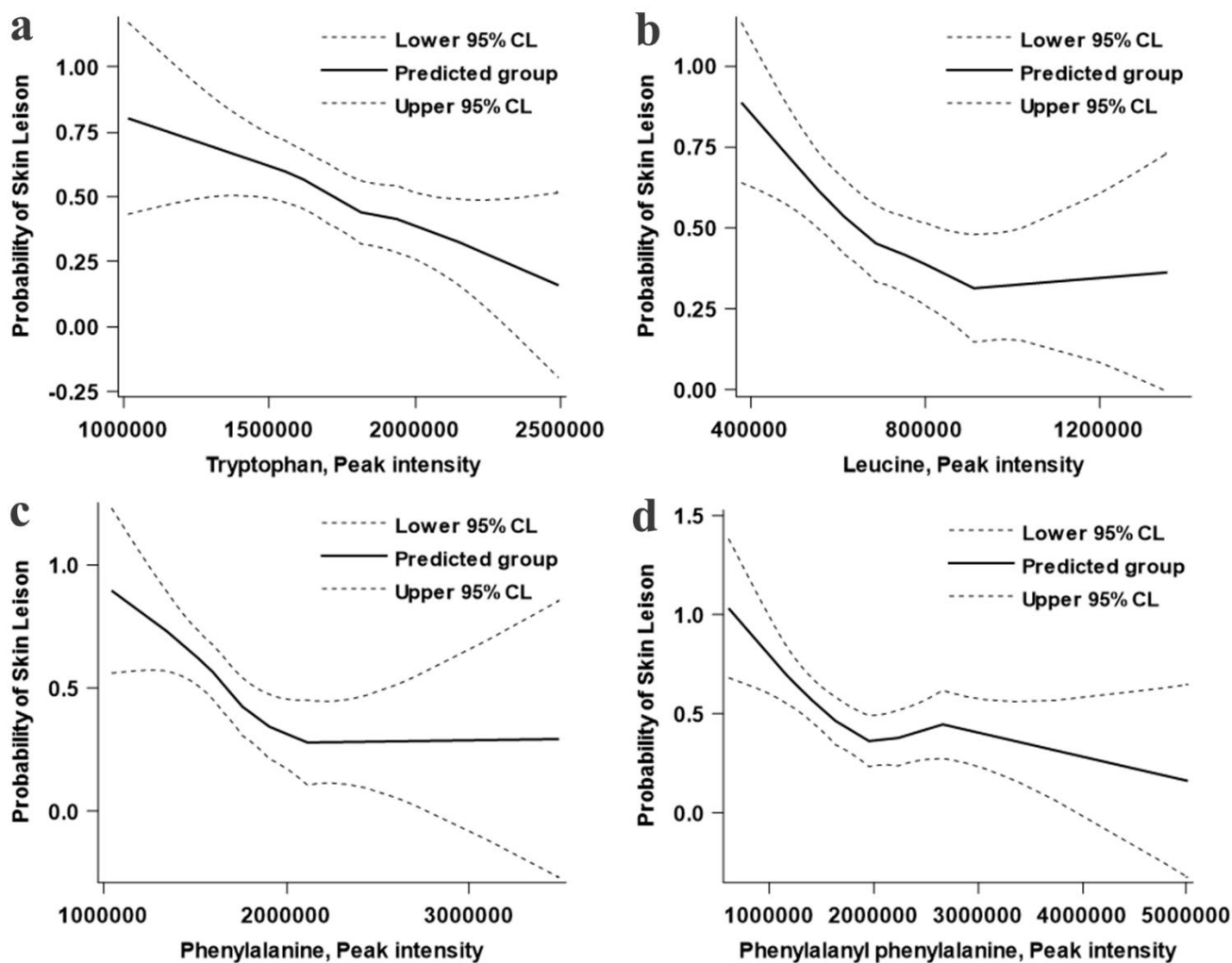


Figure 1. Association between the peak intensity of tryptophan and phenylalanine and arsenic-induced skin lesions based on multivariable locally weighted regression models. a: tryptophan; b: phenylalanine;c: leucine;d:phenylalanyl phenylalanine

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

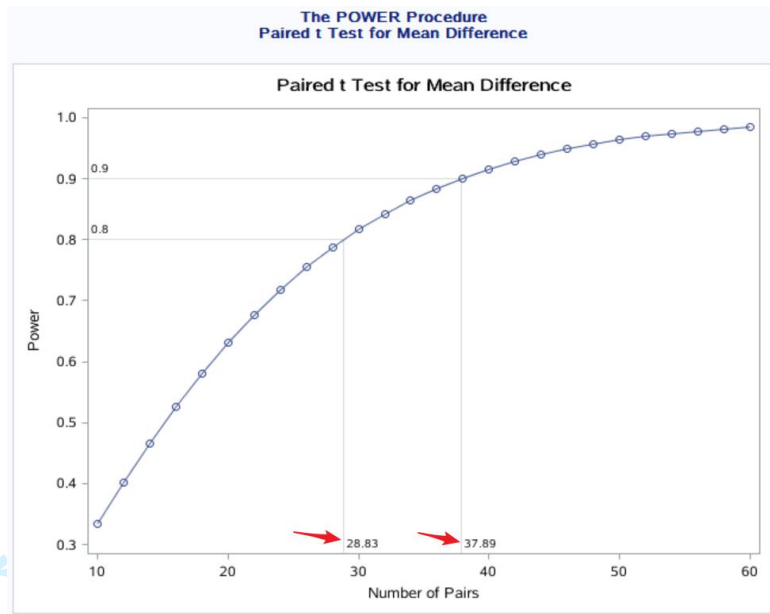


Figure 1S_A. Sample size and power estimation__Tryptophan

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

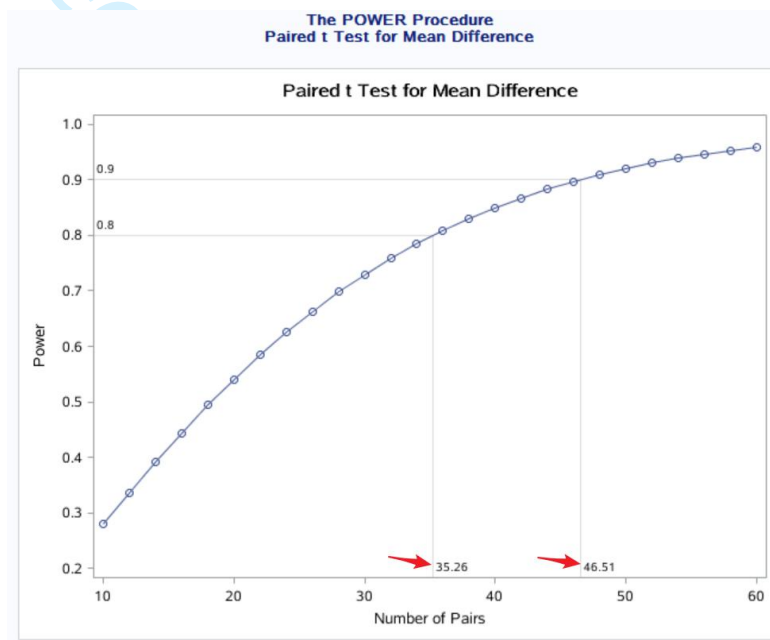


Figure 1S_B. Sample size and power estimation__Leucine

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

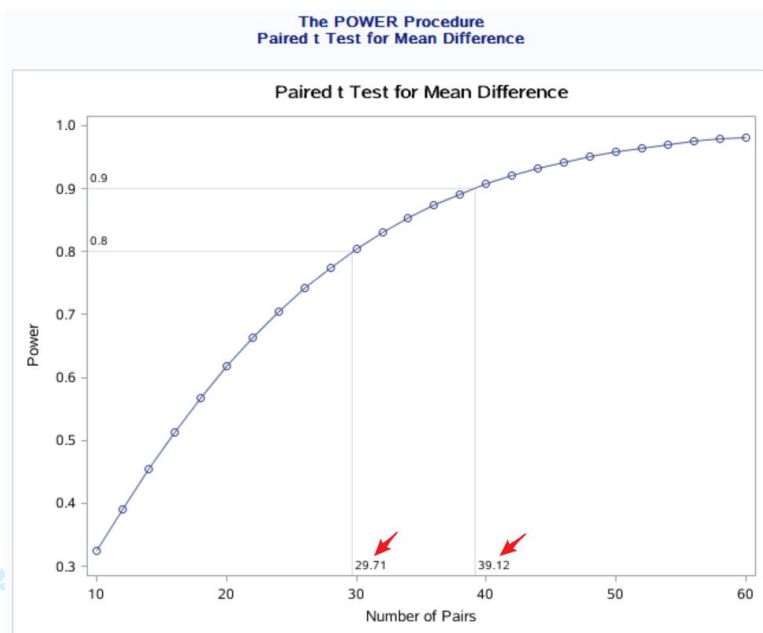


Figure 1S_C. Sample size and power estimation__Phenylalanine

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

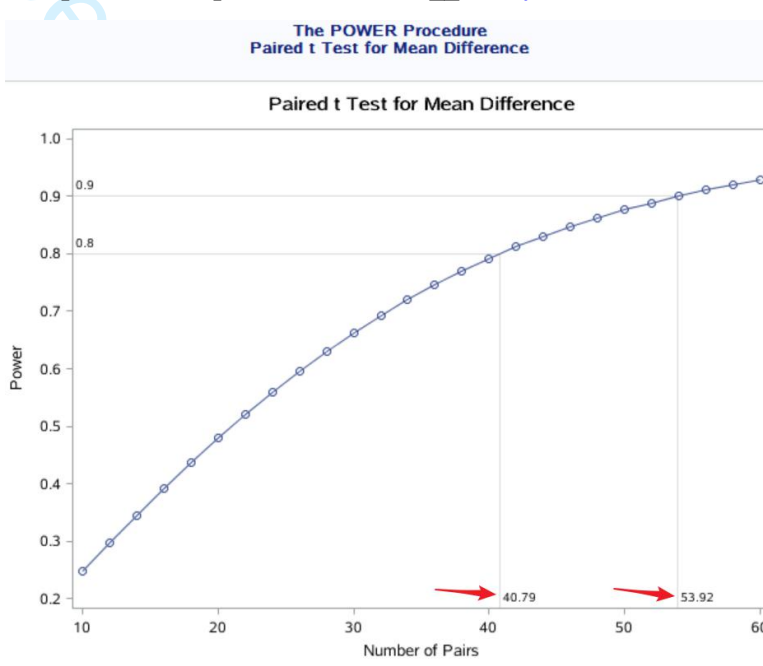


Figure 1S_D. Sample size and power estimation__Phenylalanyl phenylalanine

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

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)

[‡]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[‡].

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)

[‡]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[‡].

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)

[‡]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[‡].

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)

[‡]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	

^ξModel 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.

For peer review only

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:	<i>BMJ Open</i>
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

SCHOLARONE™
Manuscripts

1
2
3 **Tryptophan, phenylalanine and arsenic-induced skin lesions in a chronic arsenic**
4 **exposure Chinese population via drinking water: data from a metabolomics study**
5
6
7
8
9

10
11 Yaping Wei^{1,2}, Chaonan Jia,^{1,2} Yuan Lan³, Xiangqing Hou^{1,2}, Jingjing Zuo³, Jushuang
12 Li^{1,2}, Tao Wang^{1,2}, Guangyun Mao^{1,2,4}✉
13
14

15
16
17 ¹ Department of Preventive Medicine, School of Public Health and Management,
18 Wenzhou Medical University, Wenzhou, China.

19
20 ² Center on Evidence-Based Medicine & Clinical Epidemiological Research, School of
21 Public Health, Wenzhou Medical University, Wenzhou, China.

22
23 ³ School of Ophthalmology & Optometry, Wenzhou Medical University, Wenzhou,
24 China.

25
26 ⁴ Center on Clinical Research, the Affiliated Eye Hospital of Wenzhou Medical
27 University, Wenzhou, China.
28
29

30
31
32 Yaping Wei and Chaonan Jia contributed equally to this paper.
33
34

35
36 ✉Corresponding author: Guangyun Mao, Room 7A321, School of Public Health and
37 Management, Wenzhou Medical University, Chashan University town, Ouhai
38 district, Wenzhou, China. 325035
39

40
41 E-mail: maogy@wmu.edu.cn

42
43 Telephone: +86-577-86689848

44
45 Fax numbers: +86-577-86699122
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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

1
2
3 with AUC (95%CI) was 0.72 (0.62, 0.81), sensitivity of 76.79% and specificity of 58.93%
4
5 (P<0.001).
6
7

8 **Conclusions** Specific AAs might be linked to AISL and play an important role in its
9
10 early identification. Additional studies are needed to confirm our findings.
11

12 **Keywords:** Metabolomics; Chronic arsenic exposure; Skin lesions; Amino acid;
13
14 UPLC-MS/MS.
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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.

1 INTRODUCTION

2 Chronic arsenic exposure via drinking water is widely believed as a global health concern,
3 affecting a large amount of people worldwide. It may give rise to several human health issues and has
4 been documented to associate with cardiovascular disease, diabetes, cancer and others^{1 2}. With the
5 industrial boom and dramatic rise of worldwide water pollution including arsenic contamination in the
6 past, the prevalence and burden of arsenic-induced health damage will continue to increase. Skin has
7 been confirmed as one of the most common and susceptible target of arsenic-induced health lesions.
8 Cutaneous skin lesions are typical signs of arsenicosis after persistent arsenic exposure for a long term
9 which are characterized by hyperkeratosis and hyperpigmentation. Considerable evidences of the
10 prevalence of arsenical skin lesions had been observed in many countries³⁻⁵.

11 As arsenic-induced skin lesions (AISL) have been widely accepted as the major early
12 manifestation of arsenic toxicity⁶ and might be indicators of susceptibility to more serious arsenic-
13 induced health hazards⁷, it is particularly crucial to identify participants at risk as early as possible for
14 preventing the onset or delaying the progression of the serious health problems effectively. Several
15 possible mechanisms such as genetic differences⁸, oxidative stress⁹ and epigenetic dysregulation¹⁰ and
16 others may explain arsenic poisoning. Previous studies also reported that arsenic methylation in vivo
17 might be associated with metabolic syndrome^{11 12}.

18 Amino acids (AAs) are the "basic unit" that make up the body's various proteins and necessary to
19 maintain the health. Some AAs are important regulators of some key metabolic pathways and have
20 great importance in maximizing efficiency of food utilization, enhancing protein accretion and health
21 improvement^{13 14}. Abnormal metabolism of AAs will disturb the homeostasis of the body, impairs
22 growth and development, and even causes death¹⁵. So, the levels of serum AAs may be an important
23 implication for the metabolic status and disease condition. As a powerful tool in system biology
24 research, metabolomics approach is beneficial on unbiased monitoring changes in endogenous
25 metabolism-related physiological processes, providing integrative information on the distinctive

1
2
3 26 features across multiple functional levels, and offering a window to capture the core attributes
4
5 27 responsible for various phenotypes, which are particularly important in understanding the relevant
6
7
8 28 pathophysiological changes of a disease and its status, identifying novel biomarkers for risk screening,
9
10 29 diagnosis, treatment and prognosis of important human diseases¹⁶⁻¹⁸.

11
12 30 Animal experiments and epidemiological study have reported obvious arsenic-related
13
14 31 metabolomics perturbations^{19 20}. All of these researches substantially suggests that the relationship
15
16
17 32 between specific metabolites and arsenic-induced health lesions should be investigated. However, few
18
19 33 works have been conducted to comprehensively examine the metabolic mechanism relevant to AISL,
20
21 34 especially for the AAs metabolism. The present study aims to quantitatively examine the association
22
23
24 35 of several specific AAs with AISL and the ability to identify AISL.

25 26 27 36 **METHODS**

28 29 30 37 **Patient and Public Involvement**

31
32
33
34 38 No patients were involved.

35 36 39 **Study Population**

37
38 40 This study was originally from a randomized, double-blind, and placebo controlled clinical trial
39
40
41 41 (NCT02235948) in 2010, in which all subjects were randomly selected using permuted block
42
43 42 randomization from a single rural area in a population chronically exposed to low-level arsenic
44
45 43 drinking water, had similar life style and influences under similar environmental factors. Information
46
47 44 on the inclusion and exclusion criteria of the participants could be found in our previous study²¹.
48
49 45 Strictly following the criteria of arsenicosis²². AISL was diagnosed as the presence of arsenic- induced
50
51 46 keratosis, hyperpigmentation or depigmentation by a physician from Wenzhou medical university at
52
53 47 the beginning of the trial. This was a matched case-control study (1:1 matching). Among 450 residents
54
55 48 aged 18 to 79 years old enrolled in the above-mentioned trial, 56 of them were diagnosed as AISL and
56
57 49 selected as the case. Another 56 participants without AISL matched by gender and similar age (± 1
58
59
60

1
2
3 50 year) from the same population were picked out as the control. The inclusion criteria were subjects
4
5 51 having the metabolomic test. Unmatched participants and those without serum metabolites data were
6
7
8 52 excluded. Informed consent was obtained from all participants and this study was approved by the
9
10 53 ethics committee of Wenzhou Medical University, Wenzhou, China.

11 12 54 **Data Collection and Assessment**

13
14
15 55 The information on age, gender, exposure year, body mass index, smoking, alcohol consumption,
16
17 56 education level, etc. was collected with a standardized questionnaire. Blood and urine samples were
18
19 57 also collected at the time of participants' enrollment. Detailed data collection of blood and urine
20
21
22 58 samples and assessment methods for clinical variables including plasma fasting glucose (FPG), serum
23
24 59 urea nitrogen, serum folate, total homocysteine, total cholesterol (TC), triglycerides (TG), high-density
25
26 60 lipoprotein (HDL), Low-density lipoprotein (LDL) and others had been published previously²¹. The
27
28
29 61 epuration of various urinary arsenic species were conducted by means of a high-performance liquid
30
31 62 chromatography coupled mass spectrometer system for separation and detection²³. The species of
32
33 63 arsenic in urine samples consisted of inorganic arsenic (iAs, [iAs^{III} plus iAs^V]), monomethyl arsenate
34
35
36 64 (MMA, [MMA^{III} plus MMA^V]) and dimethyl arsenate (DMA, [DMA^{III} plus DMA^V]). All arsenic
37
38 65 species were corrected by creatinine. The total arsenic (tAs) was the sum of iAs, MMA and DMA. The
39
40
41 66 percentages of arsenic species were defined as: $iAs\% = iAs/tAs * 100\%$, $MMA\% = MMA/tAs * 100\%$ and
42
43 67 $DMA\% = DMA/tAs * 100\%$, respectively.

44 45 68 **UPLC-MS/MS Metabonomic Profiling**

46
47
48 69 Serum samples (200 μ L in microcentrifuge tubes) were thawed to room temperature (25°C) and
49
50 70 600 μ L mixture (90% acetonitrile - 10%water) were added to each sample. The samples were
51
52
53 71 vigorously mixed for 20 seconds and centrifuged for 5 min at 12000 rpm (20°C). The top 400 μ L of
54
55 72 each supernatant were then transferred and dried down in a vacuum concentrator centrifuge. The dried
56
57 73 samples were re-suspended in 130 μ L of water (including 15% acetonitrile), mixed vigorously for 20
58
59 74 seconds and repeated the centrifugation method described above. Two μ L of the supernatant were
60

1
2
3 75 collected as samples to be determined. Serum metabolic profile acquisition was performed by using
4
5 76 ACQUITY UPLC[®]/Xevo[®] G2 QToF/MS^E (Waters Corp., Milford, MA, USA). Chromatographic
6
7
8 77 separation was performed at 50°C using a WATERS HSS T3 column (2.1×100 mm, 1.7 μm) with a
9
10 78 flow rate of 0.4mL/min. The mobile phase was a mixture of (A) H₂O with 0.1% formic acid and (B)
11
12 79 methanol with 0.1% formic acid. Elution was in linear gradient with the programmed gradient at 0 min
13
14
15 80 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
16
17 81 with 0% A and 100% B. The mass spectrometer was operated under both positive-ion (ESI⁺) mode
18
19 82 and negative-ion (ESI⁻) mode electrospray ionization. The scan range was from 50 to 1200 m/z. Data
20
21
22 83 was collected in both ESI⁺ and ESI⁻ modes. Capillary voltage was set at 3000 V and 2500 V,
23
24 84 respectively. The desolvation flow rate was 800 L/h at 350°C. Argon was used as a collision gas, and
25
26 85 the collision energy was adjusted from 10 eV to 40 eV for each analysis. Quantum clustering (QC)
27
28 86 samples were prepared by pooling aliquots of each sample and used to reflect the reliability of further
29
30
31 87 metabolomics analysis. After peak deconvolution, alignment, integration and normalization, the data
32
33 88 including retention time (RT), mass to charge ratio(m/z), and peak intensity were extracted from raw
34
35 89 chromatograms using Progenesis QI 2.0 (Waters Corp., Milford, MA, USA). The MS/MS mode was
36
37
38 90 performed to obtain metabolites levels processed with MarkerLynx Applications Manager Version 4.1
39
40 91 (Waters Corp., Milford, MA, USA).

42 92 **Distinct Metabolites Identification**

43
44
45 93 The peak intensity of metabolites for the 56 pairs were acquired and then imported to
46
47 94 MetaboAnalyst 4.0 (<http://www.metaboanalyst.ca/>) for statistical analysis. A partial least-squares
48
49 95 discriminant analysis (PLS-DA), which is a supervised and well accepted pattern recognition approach,
50
51 96 was used for the differentiation between the cases and controls. False discovery rate (FDR) adjusted
52
53
54 97 p-value in univariate analysis were performed to reduce the potential impact induced by false positive
55
56 98 of the results. The criteria used in the selection of metabolites include variable importance in the project
57
58 99 (VIP) scores >1 in PLS-DA and the crude or FDR adjusted p-value all < 0.05 in Wilcoxon signed rank
59
60

1
2
3 100 test. We identified a total of 70 extracted small molecular metabolites that were linked to the
4
5
6 101 recognition of AISL. Among them, four amino acid metabolites (Phenylalanine, Tryptophan, Leucine,
7
8 102 Phenylalanylphenylalanine) were identified.

10 103 **Statistical Analysis**

11
12 104 The normality of continuous data was assessed using both QQ-plots and Shapiro-Wilk test. The
13
14
15 105 comparison between the cases and controls was performed with the paired *t*-test if they met normal or
16
17 106 similar normal distribution. Otherwise, Wilcoxon signed rank test would be used. Differences in the
18
19 107 proportion of categorical variables between the two groups were evaluated by McNemar-Bowker tests.
20
21
22 108 We firstly used locally weighted scatterplot smoothing (LOESS) models to estimate the “real”
23
24 109 relationship between serum AAs levels and the odds of AISL. Then, multivariable conditional logistic
25
26 110 regression models were performed to examine the association between the contributing AAs levels
27
28
29 111 and AISL after adjusting for some potential confounding factors. The individual impacts of AA
30
31 112 metabolites on the risk of AISL were quantified separately by odds ratio (OR) and 95% confidence
32
33 113 interval (CI) in the following two ways: with AA as a categorical variable (quartiles) and as a
34
35
36 114 continuous variable [scaled to interquartile range (IQR)]. Variables with p-value less than 0.2 in the
37
38 115 comparison between two groups were selected as potential confounders because the sample size of the
39
40 116 current study was not too large. Then variance inflation factor (VIF) was used to examine the potential
41
42
43 117 collinearity among them. As too many covariates in a multiple regression model will lead to overfitting
44
45 118 to some extent²⁴, we finally select no more than 5 4 variables as confounding factors to decrease the
46
47 119 potential overfitting when assessing the association between AAs and AISL. Furthermore, as the
48
49 120 distinct metabolites might be high related to each other, collinearity should be well considered. So, we
50
51
52 121 used the variance inflation factor (VIF) based on the VIF package of R software to detect potential
53
54 122 collinearity among the AAs. When the VIF is greater than 1.5, it was considered as collinearity existed
55
56 123 in the model and the associated variable would be removed. The combined effect of relevant AAs on
57
58
59 124 AISL were also performed using a multivariable logistic regression model. In addition, receiver
60

1
2
3 125 operator characteristic (ROC) analysis was applied to evaluate the value and feasibility of the AAs as
4
5
6 126 the potential sensitive and specific biomarker to recognize AISL. Data management, analysis and
7
8 127 figure drawing were finished using R version 3.4.4 (Copyright © 2018 The R Foundation for Statistical
9
10 128 Computing). All tests were two-sides and $P \leq 0.05$ was set as significant level.
11
12

13 129 RESULTS

14
15
16 130 **Table 1** summarizes the general characteristics of the study population. The comparison of
17
18 131 demographical, clinical features and urinary arsenic species in the 56 pairs of subjects were presented
19
20 132 in Table 1. The median (1st quartile, 3rd quartile) age of AISL population was 50.30 (44.70, 58.70) for
21
22
23 133 the cases and 50.40 (44.60, 58.70) years for the controls. Both groups have the same proportion of
24
25 134 female population (58.93%), and there was no obvious statistical difference in the urinary arsenic
26
27 135 levels between the two groups. More than half of them had no history of smoking or alcohol
28
29
30 136 consumption. When compared to the controls, the serum triglycerides level in AISL participants was
31
32 137 significantly lower ($P=0.041$). While the other variables were similar between AISL participants and
33
34 138 the control ($P>0.05$). This indicates that the participants in two groups are comparable to some extent.
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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^ξ			
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.

^ξ 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 ^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 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[‡]

Amino acids	N	Cases (%)	Crude		Adjusted [‡]	
			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

[‡]Values are odds ratio (95% confidence intervals) for arsenic-induced skin lesions from conditional logistic regression. IQR: interquartile range; Q₁: the 1st quartile; Q₂: the 2nd quartile; Q₃: the 3rd quartile; Q₄: the 4th quartile.

[‡]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 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

1
2
3 159 spectrum peak area based on the ROC analysis, respectively. The higher levels of these two serum AA
4
5
6 160 were defined as equal to or over the cut-off values, while the lower categories were considered as less
7
8 161 than the associated values.
9
10

11 162 **Table 4** shows the joint impacts of tryptophan and phenylalanine levels on AISL after considering
12
13 163 the collinearity of variables in the model. The proportions of AISL were 74.3%, 60.0%, 50.0% and
14
15
16 164 18.2% for participants with lower levels of both tryptophan and phenylalanine (category A), with
17
18 165 higher tryptophan and lower phenylalanine (category B), with lower tryptophan and higher
19
20 166 phenylalanine (category C), and higher levels of both tryptophan and phenylalanine (category D),
21
22
23 167 respectively. Obvious decrease trend of the probability of AISL was observed among these 4 categories.
24
25 168 As compared to the category A, adjusted OR (95% CI) for participants in the category B, C and D
26
27 169 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
28
29
30 170 tryptophan and phenylalanine had the lowest odds of AISL, significantly decreased by 92% (OR=0.02;
31
32 171 95%CI: 0.02, 0.25; P<0.001), after adjusting for the impacts induced by some potential confounding
33
34 172 factors. This suggested that tryptophan and phenylalanine were jointly associated with the presence of
35
36 173 AISL. While no significant interaction between the two AAs on the occurrence of AISL could be
37
38
39 174 observed (P=0.419), which indicated that each AA was independently associated with AISL though
40
41 175 their joint impact was significant.
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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 ^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.
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.

^ξ 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 ^ξ	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.

^ξ 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 β_j 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

1
2
3 191 higher level of both AAs would have lowest odds of AISL. These two AAs might also be able to serve
4
5
6 192 as the indicators of AISL.

7
8 193 The probability of the initiation and development of AISL would be affected by a large number
9
10 194 of factors including age, gender, life styles, arsenic exposure, metabolism and others. These factors
11
12 195 would be important confounding factors and will largely affect our results. To adjust for the impacts
13
14
15 196 due to these cofactors, we firstly selected all participants using permuted block randomization from a
16
17 197 single rural area in which population were chronically exposed to arsenic in a same way, had similar
18
19 198 life style and environmental factors. Secondly, the cases and controls were matched by gender and age
20
21
22 199 (± 1 year). All of these may be the reason why so many potential confounders including arsenic
23
24 200 exposure do not differ significantly between the cases and controls (table 1).

25
26 201 Participants enrolled in the current study were chronically exposed to arsenic via drinking water.
27
28
29 202 The geometric mean (GM) and its related 95% CI of urinary iAs/creatinine and tAs/creatinine in this
30
31 203 population were 17.49 (14.90, 20.53) $\mu\text{g/g}$ and 147.20 (129.00, 167.97) $\mu\text{g/g}$, respectively. They were
32
33 204 much higher than those in the 20 $\mu\text{g/L}$ exposed to arsenic via drinking water [GM (95% CI): 0.4
34
35 205 (0.3,0.5) $\mu\text{g/g}$ for iAs and 9.1 (6.5,12.7) $\mu\text{g/g}$ for tAs], while obviously lower than those in the 90 $\mu\text{g/L}$
36
37
38 206 exposed group [GM (95% CI): 39.4 (31.4, 49.6) $\mu\text{g/g}$ for iAs and 248.7 (208.8, 296.3) $\mu\text{g/g}$]²⁵. An
39
40 207 available report has shown that AISL cannot be completely cured even though the medical technology
41
42
43 208 has already made great progress²⁶. So, it is crucial to identify those who are most likely to progress to
44
45 209 overt arsenic damages including AISL among people at risk as early as possible. Metabolomics study,
46
47 210 which mainly focus on thoroughly assessing the variation of metabolites possibly linked to diseases
48
49
50 211 occurrence and development, has been widely utilized to help us understand pathogenesis of diseases
51
52 212 because of its relevance to the phenotypes as compared to other 'OMICS' study²⁷. Moreover,
53
54 213 mathematical modeling to assess the linkage between small molecular metabolites and arsenic toxicity
55
56 214 has grown²⁸. Developing a simple and interpretable modeling approach for the early detection of
57
58
59 215 arsenic induced health lesions is of great theoretical value and realistic meaning²⁹, though it might be
60

1
2
3 216 difficult due to population specific complexities and the impacts due to some potential unmeasured
4
5
6 217 covariates such as diet and genetic determinants.

7
8 218 Previous studies reported that gene-gene and gene-environment interaction were involved in
9
10 219 arsenicosis through toxicological mechanisms including genomic instability³⁰ and oxidative stress³¹.
11
12 220 Skin hyperpigmentation and palmoplantar hyperkeratosis could be biomarkers for long-term arsenic
13
14
15 221 exposure identifying the differences in metabolites that are really associated with phenotypes through
16
17 222 metabolites analysis may promote our understanding and identification of AISL. Animal study reveals
18
19 223 that the disruption of amino acids metabolism upon arsenic exposure in rat which may be beneficial
20
21
22 224 on understanding arsenic toxicity³². In our previous population-based metabolomics study, we found
23
24 225 that serum metabolites alteration was significantly related to the risk of arsenic-induced health
25
26 226 damages. In the current study, we detected that BCAA or AAA were also significantly relevant to
27
28
29 227 AISL occurrence. Several studies across numerous ethnic backgrounds supports the usage of BCAA
30
31 228 including leucine, isoleucine as well as valine and AAA profile such as phenylalanine, tryptophan and
32
33 229 tyrosine as biomarkers in determining metabolic diseases^{27 33}. Simultaneously, Zhou *et al* reports that
34
35 230 arsenic-induced transformed cells exhibit apparent alterations in metabolite profiles including down-
36
37
38 231 regulated of leucine, tryptophan, and phenylalanine in skin lesions group³⁴. Consistent with Zhou's
39
40 232 findings, two serum AAA (tryptophan, phenylalanine) levels were also significantly associated with
41
42 233 AISL in our study.

43
44
45 234 Normal metabolism of amino acids are necessary for whole body homeostasis, growth and
46
47 235 development, and health status¹⁵. Studies have reported that the changes in the availability of AAA
48
49 236 will affect cell signaling, gene expression, brain, and neuroendocrine function³⁵. Tryptophan, an amino
50
51
52 237 acid metabolism related biomarker, is also a sensitive and specific indicator of oxidation. Tryptophan
53
54 238 metabolism in mammals is a physiological means of preserving immune homeostasis associated with
55
56 239 oxidative stress and inflammation^{36 37}. In addition, phenylalanine can be transformed into specific
57
58
59 240 neurotransmitters such as dopamine and adrenaline by the action of related enzymes. Wu and
60

1
2
3 241 colleagues³⁸ reported that arsenic exposure would lead to neurotransmitter metabolism disturbance,
4
5
6 242 which might explain the reduction of phenylalanine. Furthermore, as one of the peptide-bound
7
8 243 phenylalanine, phenylalanylphenylalanine has been reported to affect protein synthesis and secretion³⁹,
9
10 244 potentially indicating the possible relation between endothelium dysfunction and phenylalanine
11
12 245 metabolism disorder. The relationship between amino acid metabolism and AISL was still unclear.
13
14
15 246 The notable alteration of tryptophan and phenylalanine in this study may well indicate the occurrence
16
17 247 of metabolic disorders due to arsenic exposure. It is also beneficial to understand the effects of arsenic
18
19 248 toxicity and of great importance in early identification of occurrences as well as delaying the
20
21
22 249 progression of various arsenic-induced health lesions including AISL.

23
24 250 The current study included 56 AISL cases matched 56 non-AISL controls and the sample size
25
26 251 might be potentially insufficient. To estimate the impact due to this potential insufficient sample size
27
28
29 252 on our conclusion, the PROC POWER procedure for paired design study in SAS 9.4 (SAS Institute
30
31 253 Inc.) was applied to assess the power of the 4 AAs when assessing their associations with AISL in this
32
33 254 study. The results showed that the lowest power associated with all of these four AAs was 0.911 based
34
35 255 on 56 pairs of participants. It suggested that with type I error as 0.05, total sample size as 56 pairs and
36
37
38 256 two-sided test, the powers associated with these 4 amino acids were all great than 0.8. So, we believed
39
40 257 that the sample size for the present study, 56 pairs, would well balance the power of tests. Furthermore,
41
42 258 previous metabolomics studies usually have sample size no more than 40 cases in each group^{40 41}.

43
44
45 259 The main strength of this study is that the findings were depended on a community-based, long-
46
47 260 term arsenic exposure cohort with well-designed quality assurance and quality control throughout the
48
49 261 study, and the AAs were detected with non-targeted metabolomics approach through the discovery and
50
51
52 262 validation phases. However, there are also several limitations to this study. Firstly, although untargeted
53
54 263 metabolomics approach can assess a large amount number of metabolites precisely and efficiently, it
55
56 264 only provides relative levels of AAs instead of their accurate quantitative concentration. Secondly,
57
58 265 these findings are mainly based on a case-control study, which only reveals the association between
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

266 amino acid metabolism and the odds of AISL rather than confirming their causal relationship.
267 Furthermore, as it is suggested that the ratio of approximately 10 to 15 observations per predictor in a
268 logistic regression model will produce reasonably stable estimations²⁴, we selected only 4 covariates
269 in the models due to the small sample size and large number of predictors to avoid potential overfitting
270 and obtain a more stable estimation. Finally, the participants were mainly exposed to arsenic via
271 drinking water, which would limit the findings extrapolated to the other arsenic exposure population
272 via food and other ways. Therefore, additional elaborate population-based studies are needed to verify
273 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.

1
2
3
4 **Contributors:** Guangyun Mao and Yaping Wei designed the study. Chaonan Jia participated in
5 collecting data. Yuan Lan and Chaonan Jia audited the data. Yaping Wei, Xiangqing hou, Jushuang
6 Li, Tao Wang conducted the literature search, Yaping Wei, Chaonan Jia conducted statistical analysis
7 and interpreted the results. Yaping Wei and Chaonan Jia wrote the first draft of the manuscript. Jingjing
8 Zuo helped with copyediting. Guangyun Mao reviewed the final manuscript and did substantial
9 contributions.
10

11
12 **Funding:** This work was supported by the basic public welfare research project of Zhejiang province
13 (LGF19H260011), the initial scientific research fund (KYQD170301) and the major project of the eye
14 hospital of Wenzhou medical university (YNZD201602).
15

16 **Acknowledgments:** We gratefully acknowledge all of the study participants of this project. We wish
17 to thank all of the dedication and hard work of the field team at the department of preventive medicine,
18 school of public health and management, Wenzhou medical university. Our thanks must also go to Dr.
19 Kendrick Hii, a native-English speaker, for his valuable suggestions and edits to the manuscript.
20

21
22 **Conflicts of Interest:** The authors declare no conflict of interest.
23

24 **Data Sharing Statement:** No additional unpublished data are available.
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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

- 1
2
3 regression analysis. *J Clin Epidemiol*. 1996; 49:1373-9.
- 4 25. Sun G, Xu Y, Li X, et al. Urinary arsenic metabolites in children and adults exposed to arsenic in drinking water
5 in Inner Mongolia, China. *Environmental Health Perspectives*. 2007; 115:648-52.
- 6 26. Kuo CC, Moon KA, Wang SL, et al. The Association of Arsenic Metabolism with Cancer, Cardiovascular
7 Disease, and Diabetes: A Systematic Review of the Epidemiological Evidence. *Environmental Health*
8 *Perspectives*. 2017; 125:15.
- 9 27. Wang TJ, Larson MG, Vasan RS, et al. Metabolite profiles and the risk of developing diabetes. *Nature Medicine*.
10 2011; 17:448-U83.
- 11 28. Lawley SD, Cinderella M, Hall MN, et al. Mathematical model insights into arsenic detoxification. *Theor Biol*
12 *Med Model*. 2011; 8.
- 13 29. Juhasz AL, Herde P, Herde C, et al. Validation of the predictive capabilities of the Sbrc-G in vitro assay for
14 estimating arsenic relative bioavailability in contaminated soils. *Environ Sci Technol*. 2014; 48:12962-9.
- 15 30. Kibriya MG, Jasmine F, Parvez F, et al. Association between genome-wide copy number variation and arsenic-
16 induced skin lesions: a prospective study. *Environ Health*. 2017; 16:13.
- 17 31. Priori R, Scrivo R, Brandt J, et al. Metabolomics in rheumatic diseases: the potential of an emerging
18 methodology for improved patient diagnosis, prognosis, and treatment efficacy. *Autoimmun Rev*. 2013;
19 12:1022-30.
- 20 32. Wang X, Mu X, Zhang J, et al. Serum metabolomics reveals that arsenic exposure disrupted lipid and amino
21 acid metabolism in rats: a step forward in understanding chronic arsenic toxicity. *Metallomics Integrated*
22 *Biometal Science*. 2015; 7:544.
- 23 33. Weng L, Quinlivan E, Gong Y, et al. Association of Branched and Aromatic Amino Acids Levels with Metabolic
24 Syndrome and Impaired Fasting Glucose in Hypertensive Patients. *Metabolic Syndrome And Related Disorders*.
25 2015; 13:195-202.
- 26 34. Zhou YY, Wang YF, Su J, et al. Integration of microRNAome, proteomics and metabolomics to analyze arsenic-
27 induced malignant cell transformation. *Oncotarget*. 2017; 8:90879-96.
- 28 35. Lee JC, Chen MJ, Chang CH, et al. Plasma amino acid levels in patients with colorectal cancers and liver
29 cirrhosis with hepatocellular carcinoma. *Hepato-Gastroenterol*. 2003; 50:1269-73.
- 30 36. Munn DH, Mellor AL. Indoleamine 2,3 dioxygenase and metabolic control of immune responses. *Trends In*
31 *Immunology*. 2013; 34:137-43.
- 32 37. Pawlak K, Domaniewski T, Mysliwiec M, et al. The kynurenines are associated with oxidative stress,
33 inflammation and the prevalence of cardiovascular disease in patients with end-stage renal disease.
34 *Atherosclerosis*. 2009; 204:309-14.
- 35 38. Wu L-L, Gong W, Shen S-P, et al. Multiple metal exposures and their correlation with monoamine
36 neurotransmitter metabolism in Chinese electroplating workers. *Chemosphere*. 2017; 182:745-52.
- 37 39. Zhou MM, Wu YM, Liu HY, et al. Effects of phenylalanine and threonine oligopeptides on milk protein
38 synthesis in cultured bovine mammary epithelial cells. *J Anim Physiol Anim Nutr*. 2015; 99:215-20.
- 39 40. Chen LY, Cheng CY, Choi HW, et al. Plasma Metabonomic Profiling of Diabetic Retinopathy. *Diabetes*. 2016;
40 65:1099-108.
- 41 41. Piszcz J, Lemancewicz D, Dudzik D, et al. Differences and similarities between LC-MS derived serum
42 fingerprints of patients with B-cell malignancies. *Electrophoresis*. 2013; 34:2857-64.
- 43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

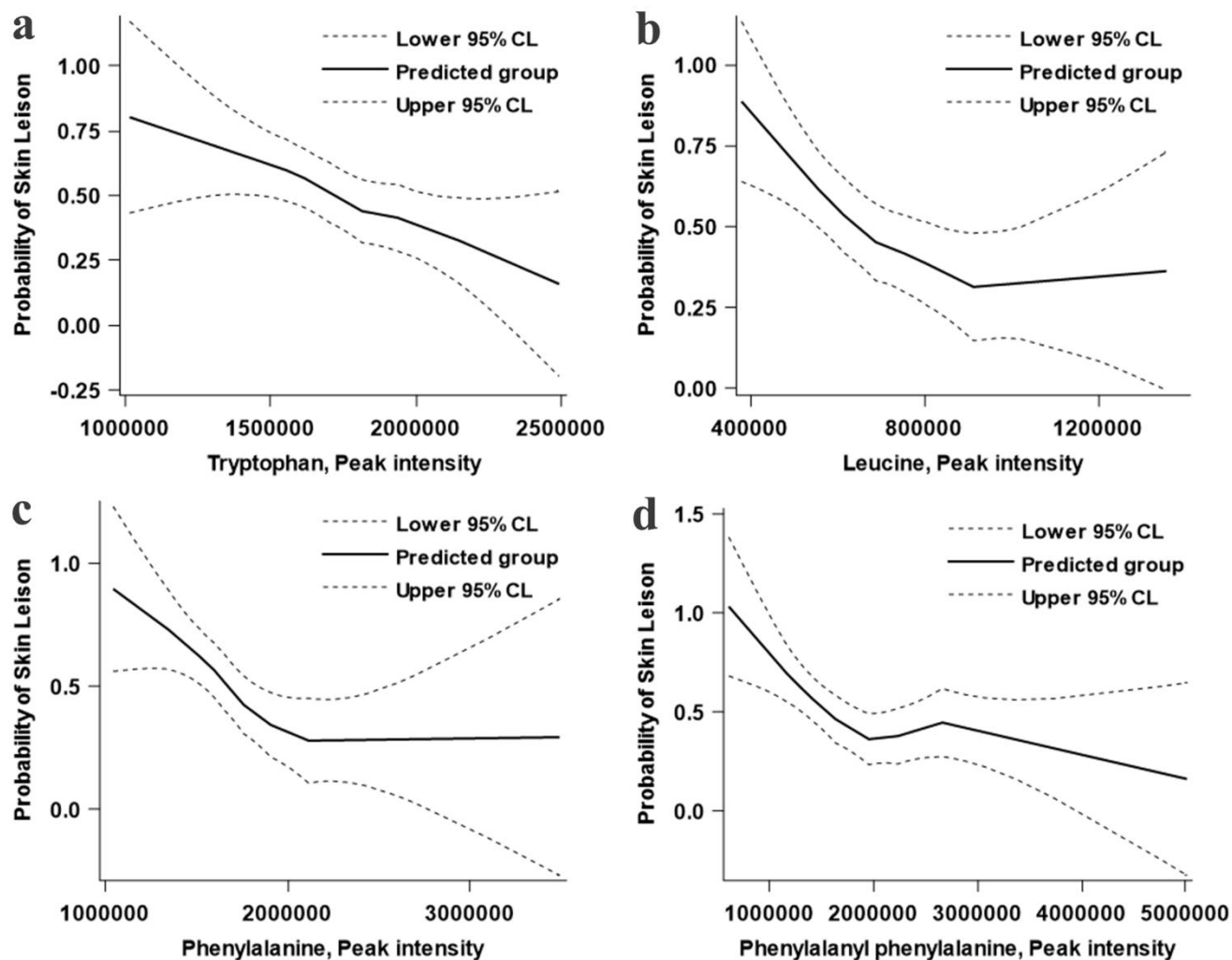


Figure 1. Association between the peak intensity of tryptophan and phenylalanine and arsenic-induced skin lesions based on multivariable locally weighted regression models. a: Tryptophan; b: Phenylalanine; c: Leucine; d: Phenylalanylphenylalanine

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[§].

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	

[§]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:	<i>BMJ Open</i>
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

SCHOLARONE™
Manuscripts

1
2
3 **Tryptophan, phenylalanine and arsenic-induced skin lesions in a chronic arsenic exposure**
4 **Chinese population via drinking water: data from a metabolomics study**
5
6

7 Yaping Wei^{1,2}, Chaonan Jia^{1,2}, Yuan Lan³, Xiangqing Hou^{1,2}, Jingjing Zuo³, Jushuang Li^{1,2}, Tao
8 Wang^{1,2}, Guangyun Mao^{1,2,4}✉
9
10

11
12
13 ¹ Department of Preventive Medicine, School of Public Health and Management, Wenzhou Medical
14 University, Wenzhou, China.
15

16 ² Center on Evidence-Based Medicine & Clinical Epidemiological Research, School of Public Health,
17 Wenzhou Medical University, Wenzhou, China.
18

19 ³ School of Ophthalmology & Optometry, Wenzhou Medical University, Wenzhou, China.
20

21 ⁴ Center on Clinical Research, the Affiliated Eye Hospital of Wenzhou Medical University, Wenzhou,
22 China.
23
24

25
26 Yaping Wei and Chaonan Jia contributed equally to this paper.
27
28

29
30 ✉Corresponding author: Guangyun Mao, Room 7A321, School of Public Health and Management,
31 Wenzhou Medical University, Chashan University town, Ouhai district, Wenzhou, China. 325035
32

33 E-mail: maogy@wmu.edu.cn
34

35 Telephone: +86-577-86689848
36

37 Fax numbers: +86-577-86699122
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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.

1 INTRODUCTION

2 Chronic arsenic exposure via drinking water is widely believed as a global health concern,
3 affecting a large amount of people worldwide. It may give rise to several human health issues and
4 has been documented to associate with cardiovascular disease, diabetes, cancer and others^{1 2}. With
5 the industrial boom and dramatic rise of worldwide water pollution including arsenic contamination
6 in the past, the prevalence and burden of arsenic-induced health damage will continue to increase.
7 Skin has been confirmed as one of the most common and susceptible target of arsenic-induced health
8 lesions. Cutaneous skin lesions are typical signs of arsenicosis after persistent arsenic exposure for a
9 long term which are characterized by hyperkeratosis and hyperpigmentation. Considerable evidences
10 of the prevalence of arsenical skin lesions had been observed in many countries³⁻⁵.

11 As arsenic-induced skin lesions (AISL) have been widely accepted as the major early
12 manifestation of arsenic toxicity⁶ and might be indicators of susceptibility to more serious
13 arsenic-induced health hazards⁷, it is particularly crucial to identify participants at risk as early as
14 possible for preventing the onset or delaying the progression of the serious health problems
15 effectively. Several possible mechanisms such as genetic differences⁸, oxidative stress⁹ and
16 epigenetic dysregulation¹⁰ and others may explain arsenic poisoning. Previous studies also reported
17 that arsenic methylation in vivo might be associated with metabolic syndrome^{11 12}.

18 Amino acids (AAs) are the "basic unit" that make up the body's various proteins and necessary
19 to maintain the health. Some AAs are important regulators of some key metabolic pathways and have
20 great importance in maximizing efficiency of food utilization, enhancing protein accretion and health
21 improvement^{13 14}. Abnormal metabolism of AAs will disturb the homeostasis of the body, impairs
22 growth and development, and even causes death¹⁵. So, the levels of serum AAs may be an important
23 implication for the metabolic status and disease condition. As a powerful tool in system biology
24 research, metabolomics approach is beneficial on unbiased monitoring changes in endogenous
25 metabolism-related physiological processes, providing integrative information on the distinctive

1
2
3 26 features across multiple functional levels, and offering a window to capture the core attributes
4
5
6 27 responsible for various phenotypes, which are particularly important in understanding the relevant
7
8 28 pathophysiological changes of a disease and its status, identifying novel biomarkers for risk
9
10 29 screening, diagnosis, treatment and prognosis of important human diseases¹⁶⁻¹⁸.

11
12 30 Animal experiments and epidemiological study have reported obvious arsenic-related
13
14 31 metabolomics perturbations^{19 20}. All of these researches substantially suggests that the relationship
15
16 32 between specific metabolites and arsenic-induced health lesions should be investigated. However,
17
18 33 few works have been conducted to comprehensively examine the metabolic mechanism relevant to
19
20 34 AISL, especially for the AAs metabolism. The present study aims to quantitatively examine the
21
22 35 association of several specific AAs with AISL and the ability to identify AISL.
23
24
25
26

27 36 **METHODS**

28 29 30 37 **Patient and Public Involvement**

31
32
33
34 38 No patients were involved.

35 36 39 **Study Population**

37
38 40 This study was originally from a randomized, double-blind, and placebo controlled clinical trial
39
40 41 (NCT02235948) in 2010, in which all subjects were randomly selected using permuted block
42
43 42 randomization from a single rural area in a population chronically exposed to low-level arsenic
44
45 43 drinking water, had similar life style and influences under similar environmental factors. Information
46
47 44 on the inclusion and exclusion criteria of the participants could be found in our previous study²¹.
48
49 45 Strictly following the criteria of arsenicosis²². AISL was diagnosed as the presence of arsenic-
50
51 46 induced keratosis, hyperpigmentation or depigmentation by a physician from Wenzhou medical
52
53 47 university at the beginning of the trial. This was a matched case-control study (1:1 matching).
54
55 48 Among 450 residents aged 18 to 79 years old enrolled in the above-mentioned trial, 56 of them were
56
57 49 diagnosed as AISL and selected as the case. Another 56 participants without AISL matched by
58
59
60

1
2
3 50 gender and similar age (± 1 year) from the same population were picked out as the control. The
4
5
6 51 inclusion criteria were subjects having the metabolomic test. Unmatched participants and those
7
8 52 without serum metabolites data were excluded. Informed consent was obtained from all participants
9
10 53 and this study was approved by the ethics committee of Wenzhou Medical University, Wenzhou,
11
12 54 China.

15 55 **Data Collection and Assessment**

16
17 56 The information on age, gender, exposure year, body mass index, smoking, alcohol
18
19 57 consumption, education level, etc. was collected with a standardized questionnaire. Blood and urine
20
21 58 samples were also collected at the time of participants' enrollment. Detailed data collection of blood
22
23 59 and urine samples and assessment methods for clinical variables including plasma fasting glucose
24
25 60 (FPG), serum urea nitrogen, serum folate, total homocysteine, total cholesterol (TC), triglycerides
26
27 61 (TG), high-density lipoprotein (HDL), low-density lipoprotein (LDL) and others had been published
28
29 62 previously²¹. The epuration of various urinary arsenic species were conducted by means of a
30
31 63 high-performance liquid chromatography coupled mass spectrometer system for separation and
32
33 64 detection²³. The species of arsenic in urine samples consisted of inorganic arsenic (iAs, [iAs^{III} plus
34
35 65 iAs^V]), monomethyl arsenate (MMA, [MMA^{III} plus MMA^V]) and dimethyl arsenate (DMA,
36
37 66 [DMA^{III} plus DMA^V]). All arsenic species were corrected by creatinine. The total arsenic (tAs) was
38
39 67 the sum of iAs, MMA and DMA. The percentages of arsenic species were defined as:
40
41 68 iAs%=iAs/tAs*100%, MMA%=MMA/tAs*100% and DMA%=DMA/tAs*100%, respectively.
42
43
44
45
46
47

48 69 **UPLC-MS/MS Metabonomic Profiling**

49
50 70 Serum samples (200 μ L in microcentrifuge tubes) were thawed to room temperature (25°C)
51
52 71 and 600 μ L mixture (90% acetonitrile - 10%water) were added to each sample. The samples were
53
54 72 vigorously mixed for 20 seconds and centrifuged for 5 min at 12000 rpm (20°C). The top 400 μ L of
55
56 73 each supernatant were then transferred and dried down in a vacuum concentrator centrifuge. The
57
58 74 dried samples were re-suspended in 130 μ L of water (including 15% acetonitrile), mixed vigorously

1
2
3 75 for 20 seconds and repeated the centrifugation method described above. Two μL of the supernatant
4
5 76 were collected as samples to be determined. Serum metabolic profile acquisition was performed by
6
7
8 77 using ACQUITY UPLC[®]/Xevo[®] G2 QToF/MS^E (Waters Corp., Milford, MA, USA).
9
10 78 Chromatographic separation was performed at 50°C using a WATERS HSS T3 column (2.1×100
11
12 79 mm, 1.7 μm) with a flow rate of 0.4mL/min. The mobile phase was a mixture of (A) H₂O with 0.1%
13
14
15 80 formic acid and (B) methanol with 0.1% formic acid. Elution was in linear gradient with the
16
17 81 programmed gradient at 0 min with 100% A and 0% B, 1.00min with 100%A and 0% B, 8 min with
18
19 82 0%A and 100% B, 13.00 min with 0% A and 100% B. The mass spectrometer was operated under
20
21
22 83 both positive-ion (ESI⁺) mode and negative-ion (ESI⁻) mode electrospray ionization. The scan range
23
24 84 was from 50 to 1200 m/z. Data was collected in both ESI⁺ and ESI⁻ modes. Capillary voltage was set
25
26 85 at 3000 V and 2500 V, respectively. The desolvation flow rate was 800 L/h at 350°C. Argon was
27
28
29 86 used as a collision gas, and the collision energy was adjusted from 10 eV to 40 eV for each analysis.
30
31 87 Quantum clustering (QC) samples were prepared by pooling aliquots of each sample and used to
32
33 88 reflect the reliability of further metabolomics analysis. After peak deconvolution, alignment,
34
35 89 integration and normalization, the data including retention time (RT), mass to charge ratio(m/z), and
36
37 90 peak intensity were extracted from raw chromatograms using Progenesis QI 2.0 (Waters Corp.,
38
39
40 91 Milford, MA, USA). The MS/MS mode was performed to obtain metabolites levels processed with
41
42 92 MarkerLynx Applications Manager Version 4.1 (Waters Corp., Milford, MA, USA).

43 44 45 93 **Distinct Metabolites Identification**

46
47 94 The peak intensity of metabolites for the 56 pairs were acquired and then imported to
48
49 95 MetaboAnalyst 4.0 (<http://www.metaboanalyst.ca/>) for statistical analysis. A partial least-squares
50
51 96 discriminant analysis (PLS-DA), which is a supervised and well accepted pattern recognition
52
53
54 97 approach, was used for the differentiation between the cases and controls. False discovery rate (FDR)
55
56 98 adjusted p-value in univariate analysis were performed to reduce the potential impact induced by
57
58 99 false positive of the results. The criteria used in the selection of metabolites include variable
59
60

1
2
3 100 importance in the project (VIP) scores >1 in PLS-DA and the crude or FDR adjusted p-value all <
4
5
6 101 0.05 in Wilcoxon signed rank test. We identified a total of 70 extracted small molecular metabolites
7
8 102 that were linked to the recognition of AISL. The Human Metabolome Database (<http://www.hmdb.ca>)
9
10 103 was used to identify the name of metabolites. Among them, there were four amino acid metabolites
11
12 104 (Phenylalanine, Tryptophan, Leucine, Phenylalanylphenylalanine).

15 105 **Statistical Analysis**

16
17 106 The normality of continuous data was assessed using both QQ-plots and Shapiro-Wilk test. The
18
19 107 comparison between the cases and controls was performed with the paired *t*-test if they met normal
20
21
22 108 or similar normal distribution. Otherwise, Wilcoxon signed rank test would be used. Differences in
23
24 109 the proportion of categorical variables between the two groups were evaluated by McNemar-Bowker
25
26 110 tests. We firstly used locally weighted scatterplot smoothing (LOESS) models to estimate the “real”
27
28 111 relationship between serum AAs levels and the probability of AISL. Then, multivariable conditional
29
30 112 logistic regression models were performed to examine the association between the contributing AAs
31
32
33 113 levels and AISL after adjusting for some potential confounding factors. The individual impacts of
34
35 114 AA metabolites on the risk of AISL were quantified separately by odds ratio (OR) and 95%
36
37
38 115 confidence interval (CI) in the following two ways: with AA as a categorical variable (quartiles) and
39
40 116 as a continuous variable [scaled to interquartile range (IQR)]. Variables with p-value less than 0.2 in
41
42 117 the comparison between two groups were selected as potential confounders, which had been widely
43
44
45 118 performed in many studies especially when the sample size of the current study was not too large.
46
47 119 Then variance inflation factor (VIF) was used to examine the potential collinearity among them. As
48
49 120 too many covariates in a multiple regression model will lead to overfitting to some extent²⁴, we
50
51 121 finally select no more than 4 variables as confounding factors to decrease the potential overfitting
52
53
54 122 when assessing the association between AAs and AISL. Furthermore, as the distinct metabolites
55
56 123 might be high related to each other, collinearity should be well considered. So, we used the VIF
57
58 124 based on the VIF package of R software to detect potential collinearity among the AAs. When the
59
60

1
2
3 125 VIF is greater than 1.5, it was considered as collinearity existed in the model and the associated
4
5
6 126 variable would be removed. The combined effect of relevant AAs on AISL were also performed
7
8 127 using a multivariable logistic regression model. In addition, receiver operator characteristic (ROC)
9
10 128 analysis was applied to evaluate the value and feasibility of the AAs as the potential sensitive and
11
12 129 specific biomarker to recognize AISL. Data management, analysis and figure drawing were finished
13
14
15 130 using R version 3.4.4 (Copyright © 2018 The R Foundation for Statistical Computing). All tests
16
17 131 were two-sides and $P \leq 0.05$ was set as significant level.
18
19

20 132 RESULTS

21
22
23 133 **Table 1** summarizes the general characteristics of the study population. The comparison of
24
25 134 demographical, clinical features and urinary arsenic species in the 56 pairs of subjects were presented
26
27 135 in Table 1. The median (1st quartile, 3rd quartile) age of AISL population was 50.30 (44.70, 58.70)
28
29
30 136 for the cases and 50.40 (44.60, 58.70) years for the controls. Both groups have the same proportion
31
32 137 of female population (58.93%), and there was no obvious statistical difference in the urinary arsenic
33
34 138 levels between the two groups. More than half of them had no history of smoking or alcohol
35
36 139 consumption. When compared to the controls, the serum triglycerides level in AISL participants was
37
38
39 140 significantly lower ($P=0.041$). While the other variables were similar between AISL participants and
40
41 141 the control ($P>0.05$). This indicates that the participants in two groups are comparable to some
42
43 142 extent.
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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

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[§]			
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

[‡] 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.

[§] 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 figure 1, 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 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 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[‡]

Amino acids	N	Cases (%)	Crude		Adjusted [‡]	
			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

[‡]Values are odds ratio (95% confidence intervals) for arsenic-induced skin lesions from conditional logistic regression. IQR: interquartile range; Q₁: the 1st quartile; Q₂: the 2nd quartile; Q₃: the 3rd quartile; Q₄: the 4th quartile.

[‡]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 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

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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

1
2
3 195 higher level of both AAs would have lowest odds of AISL. These two AAs might also be able to
4
5
6 196 serve as the indicators of AISL.

7
8 197 The probability of the initiation and development of AISL would be affected by a large number
9
10 198 of factors including age, gender, life styles, arsenic exposure, metabolism and others. These factors
11
12 199 would be important confounding factors and will largely affect our results. To adjust for the impacts
13
14 200 due to these cofactors, we firstly selected all participants using permuted block randomization from a
15
16
17 201 single rural area in which population were chronically exposed to arsenic in a same way, had similar
18
19 202 life style and environmental factors. Secondly, the cases and controls were matched by gender and
20
21
22 203 age (± 1 year). All of these may be the reason why so many potential confounders including arsenic
23
24 204 exposure do not differ significantly between the cases and controls (table 1).

25
26 205 Participants enrolled in the current study were chronically exposed to arsenic via drinking water.
27
28 206 The geometric mean (GM) and its related 95% CI of urinary iAs/creatinine and tAs/creatinine in this
29
30 207 population were 17.49 (14.90, 20.53) $\mu\text{g/g}$ and 147.20 (129.00, 167.97) $\mu\text{g/g}$, respectively. They
31
32
33 208 were much higher than those in the 20 $\mu\text{g/L}$ exposed to arsenic via drinking water [GM (95% CI):
34
35 209 0.4 (0.3,0.5) $\mu\text{g/g}$ for iAs and 9.1 (6.5,12.7) $\mu\text{g/g}$ for tAs], while obviously lower than those in the 90
36
37
38 210 $\mu\text{g/L}$ exposed group [GM (95% CI): 39.4 (31.4, 49.6) $\mu\text{g/g}$ for iAs and 248.7 (208.8, 296.3) $\mu\text{g/g}$]²⁵.
39
40 211 An available report has shown that AISL cannot be completely cured even though the medical
41
42 212 technology has already made great progress²⁶. So, it is crucial to identify those who are most likely to
43
44
45 213 progress to overt arsenic damages including AISL among people at risk as early as possible.
46
47 214 Metabolomics study, which mainly focus on thoroughly assessing the variation of metabolites
48
49 215 possibly linked to diseases occurrence and development, has been widely utilized to help us
50
51
52 216 understand pathogenesis of diseases because of its relevance to the phenotypes as compared to other
53
54 217 ‘OMICS’ study²⁷. Moreover, mathematical modeling to assess the linkage between small molecular
55
56 218 metabolites and arsenic toxicity has grown²⁸. Developing a simple and interpretable modeling
57
58
59 219 approach for the early detection of arsenic induced health lesions is of great theoretical value and
60

1
2
3 220 realistic meaning²⁹, though it might be difficult due to population specific complexities and the
4
5
6 221 impacts due to some potential unmeasured covariates such as diet and genetic determinants.

7
8 222 Previous studies reported that gene-gene and gene-environment interaction were involved in
9
10 223 arsenicosis through toxicological mechanisms including genomic instability³⁰ and oxidative stress³¹.
11
12 224 Skin hyperpigmentation and palmoplantar hyperkeratosis could be biomarkers for long-term arsenic
13
14
15 225 exposure identifying the differences in metabolites that are really associated with phenotypes through
16
17 226 metabolites analysis may promote our understanding and identification of AISL. Animal study
18
19 227 reveals that the disruption of amino acids metabolism upon arsenic exposure in rat which may be
20
21
22 228 beneficial on understanding arsenic toxicity³². In our previous population-based metabolomics study,
23
24 229 we found that serum metabolites alteration was significantly related to the risk of arsenic-induced
25
26 230 health damages. In the current study, we detected that BCAA or AAA were also significantly
27
28
29 231 relevant to AISL occurrence. Several studies across numerous ethnic backgrounds supports the usage
30
31 232 of BCAA including leucine, isoleucine as well as valine and AAA profile such as phenylalanine,
32
33 233 tryptophan and tyrosine as biomarkers in determining metabolic diseases^{27 33}. Simultaneously, Zhou
34
35 234 *et al* reports that arsenic-induced transformed cells exhibit apparent alterations in metabolite profiles
36
37
38 235 including down-regulated of leucine, tryptophan, and phenylalanine in skin lesions group³⁴.
39
40 236 Consistent with Zhou's findings, two serum AAA (tryptophan, phenylalanine) levels were also
41
42 237 significantly associated with AISL in our study.

43
44
45 238 Normal metabolism of amino acids are necessary for whole body homeostasis, growth and
46
47 239 development, and health status¹⁵. Studies have reported that the changes in the availability of AAA
48
49 240 will affect cell signaling, gene expression, brain, and neuroendocrine function³⁵. Tryptophan, an
50
51
52 241 amino acid metabolism related biomarker, is also a sensitive and specific indicator of oxidation.
53
54 242 Tryptophan metabolism in mammals is a physiological means of preserving immune homeostasis
55
56 243 associated with oxidative stress and inflammation^{36 37}. In addition, phenylalanine can be transformed
57
58 244 into specific neurotransmitters such as dopamine and adrenaline by the action of related enzymes.
59
60

1
2
3 245 Wu and colleagues³⁸ reported that arsenic exposure would lead to neurotransmitter metabolism
4
5
6 246 disturbance, which might explain the reduction of phenylalanine. Furthermore, as one of the
7
8 247 peptide-bound phenylalanine, phenylalanylphenylalanine has been reported to affect protein
9
10 248 synthesis and secretion³⁹, potentially indicating the possible relation between endothelium
11
12 249 dysfunction and phenylalanine metabolism disorder. The relationship between amino acid
13
14
15 250 metabolism and AISL was still unclear. The notable alteration of tryptophan and phenylalanine in
16
17 251 this study may well indicate the occurrence of metabolic disorders due to arsenic exposure. It is also
18
19 252 beneficial to understand the effects of arsenic toxicity and of great importance in early identification
20
21
22 253 of occurrences as well as delaying the progression of various arsenic-induced health lesions
23
24 254 including AISL.

25
26 255 The current study included 56 AISL cases matched 56 non-AISL controls and the sample size
27
28
29 256 might be potentially insufficient. To estimate the impact due to this potential insufficient sample size
30
31 257 on our conclusion, the PROC POWER procedure for paired design study in SAS 9.4 (SAS Institute
32
33 258 Inc.) was applied to assess the power of the 4 AAs when assessing their associations with AISL in
34
35 259 this study. The results showed that the lowest power associated with all of these four AAs was 0.911
36
37
38 260 based on 56 pairs of participants (Figure 1S). It suggested that with type I error as 0.05, total sample
39
40 261 size as 56 pairs and two-sided test, the powers associated with these 4 amino acids were all great than
41
42 262 0.8. So, we believed that the sample size for the present study, 56 pairs, would well balance the
43
44
45 263 power of tests. Furthermore, previous metabolomics studies usually have sample size no more than
46
47 264 40 cases in each group^{40 41}.

48
49 265 The main strength of this study is that the findings were depended on a community-based,
50
51
52 266 long-term arsenic exposure cohort with well-designed quality assurance and quality control
53
54 267 throughout the study, and the AAs were detected with non-targeted metabolomics approach through
55
56 268 the discovery and validation phases. However, there are also several limitations to this study. Firstly,
57
58 269 although untargeted metabolomics approach can assess a large amount number of metabolites
59
60

1
2
3 270 precisely and efficiently, it only provides relative levels of AAs instead of their accurate quantitative
4
5
6 271 concentration. Secondly, these findings are mainly based on a case-control study, which only reveals
7
8 272 the association between amino acid metabolism and the odds of AISL rather than confirming their
9
10 273 causal relationship. Furthermore, as it is suggested that the ratio of approximately 10 to 15
11
12 274 observations per predictor in a logistic regression model will produce reasonably stable estimations²⁴,
13
14
15 275 we selected only 4 covariates in the models due to the small sample size and these results need to be
16
17 276 confirmed in new studies. Finally, the participants were mainly exposed to arsenic via drinking
18
19 277 water, which would limit the findings extrapolated to the other arsenic exposure population via food
20
21
22 278 and other ways. Therefore, additional elaborate population-based studies are needed to verify our
23
24 279 discoveries.

25
26 280 In conclusion, specific AAs might be linked to AISL and amino acids metabolism may play an
27
28
29 281 important role in AISL early identification. Additional studies may be needed to confirm our
30
31 282 findings.

1
2
3
4 **Contributors:** Guangyun Mao and Yaping Wei designed the study. Chaonan Jia participated in
5 collecting data. Yuan Lan and Chaonan Jia audited the data. Yaping Wei, Xiangqing hou, Jushuang
6 Li, Tao Wang conducted the literature search, Yaping Wei, Chaonan Jia conducted statistical
7 analysis and interpreted the results. Yaping Wei and Chaonan Jia wrote the first draft of the
8 manuscript. Jingjing Zuo helped with copyediting. Guangyun Mao reviewed the final manuscript and
9 did substantial contributions.
10

11
12 **Funding:** This work was supported by the basic public welfare research project of Zhejiang province
13 (LGF19H260011), the initial scientific research fund (KYQD170301) and the major project of the
14 eye hospital of Wenzhou medical university (YNZD201602).
15

16 **Acknowledgments:** We gratefully acknowledge all of the study participants of this project. We wish
17 to thank all of the dedication and hard work of the field team at the department of preventive
18 medicine, school of public health and management, Wenzhou medical university. Our thanks must
19 also go to Dr. Kendrick Hii, a native-English speaker, for his valuable suggestions and edits to the
20 manuscript.
21

22
23 **Conflicts of Interest:** The authors declare no conflict of interest.
24

25 **Data Sharing Statement:** No additional unpublished data are available.
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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

- 1
2
3 regression analysis. *J Clin Epidemiol*. 1996; 49:1373-9.
- 4 25. Sun G, Xu Y, Li X, et al. Urinary arsenic metabolites in children and adults exposed to arsenic in drinking
5 water in Inner Mongolia, China. *Environmental Health Perspectives*. 2007; 115:648-52.
- 6 26. Kuo CC, Moon KA, Wang SL, et al. The Association of Arsenic Metabolism with Cancer, Cardiovascular
7 Disease, and Diabetes: A Systematic Review of the Epidemiological Evidence. *Environmental Health*
8 *Perspectives*. 2017; 125:15.
- 9 27. Wang TJ, Larson MG, Vasan RS, et al. Metabolite profiles and the risk of developing diabetes. *Nature*
10 *Medicine*. 2011; 17:448-U83.
- 11 28. Lawley SD, Cinderella M, Hall MN, et al. Mathematical model insights into arsenic detoxification. *Theor Biol*
12 *Med Model*. 2011; 8.
- 13 29. Juhasz AL, Herde P, Herde C, et al. Validation of the predictive capabilities of the Sbrc-G in vitro assay for
14 estimating arsenic relative bioavailability in contaminated soils. *Environ Sci Technol*. 2014; 48:12962-9.
- 15 30. Kibriya MG, Jasmine F, Parvez F, et al. Association between genome-wide copy number variation and
16 arsenic-induced skin lesions: a prospective study. *Environ Health*. 2017; 16:13.
- 17 31. Priori R, Scrivo R, Brandt J, et al. Metabolomics in rheumatic diseases: the potential of an emerging
18 methodology for improved patient diagnosis, prognosis, and treatment efficacy. *Autoimmun Rev*. 2013;
19 12:1022-30.
- 20 32. Wang X, Mu X, Zhang J, et al. Serum metabolomics reveals that arsenic exposure disrupted lipid and amino
21 acid metabolism in rats: a step forward in understanding chronic arsenic toxicity. *Metallomics Integrated*
22 *Biometal Science*. 2015; 7:544.
- 23 33. Weng L, Quinlivan E, Gong Y, et al. Association of Branched and Aromatic Amino Acids Levels with
24 Metabolic Syndrome and Impaired Fasting Glucose in Hypertensive Patients. *Metabolic Syndrome And*
25 *Related Disorders*. 2015; 13:195-202.
- 26 34. Zhou YY, Wang YF, Su J, et al. Integration of microRNAome, proteomics and metabolomics to analyze
27 arsenic-induced malignant cell transformation. *Oncotarget*. 2017; 8:90879-96.
- 28 35. Lee JC, Chen MJ, Chang CH, et al. Plasma amino acid levels in patients with colorectal cancers and liver
29 cirrhosis with hepatocellular carcinoma. *Hepato-Gastroenterol*. 2003; 50:1269-73.
- 30 36. Munn DH, Mellor AL. Indoleamine 2,3 dioxygenase and metabolic control of immune responses. *Trends In*
31 *Immunology*. 2013; 34:137-43.
- 32 37. Pawlak K, Domaniewski T, Mysliwiec M, et al. The kynurenines are associated with oxidative stress,
33 inflammation and the prevalence of cardiovascular disease in patients with end-stage renal disease.
34 *Atherosclerosis*. 2009; 204:309-14.
- 35 38. Wu L-L, Gong W, Shen S-P, et al. Multiple metal exposures and their correlation with monoamine
36 neurotransmitter metabolism in Chinese electroplating workers. *Chemosphere*. 2017; 182:745-52.
- 37 39. Zhou MM, Wu YM, Liu HY, et al. Effects of phenylalanine and threonine oligopeptides on milk protein
38 synthesis in cultured bovine mammary epithelial cells. *J Anim Physiol Anim Nutr*. 2015; 99:215-20.
- 39 40. Chen LY, Cheng CY, Choi HW, et al. Plasma Metabonomic Profiling of Diabetic Retinopathy. *Diabetes*.
40 2016; 65:1099-108.
- 41 41. Piszcz J, Lemancewicz D, Dudzik D, et al. Differences and similarities between LC-MS derived serum
42 fingerprints of patients with B-cell malignancies. *Electrophoresis*. 2013; 34:2857-64.
- 43
44
45
46
47
48
49
50
51
52
53

54 **Figure 1.** Association between the peak intensity of tryptophan and phenylalanine and
55 arsenic-induced skin lesions based on multivariable locally weighted regression models. a:
56 Tryptophan; b: Phenylalanine; c: Leucine; d: Phenylalanylphenylalanine
57
58
59
60

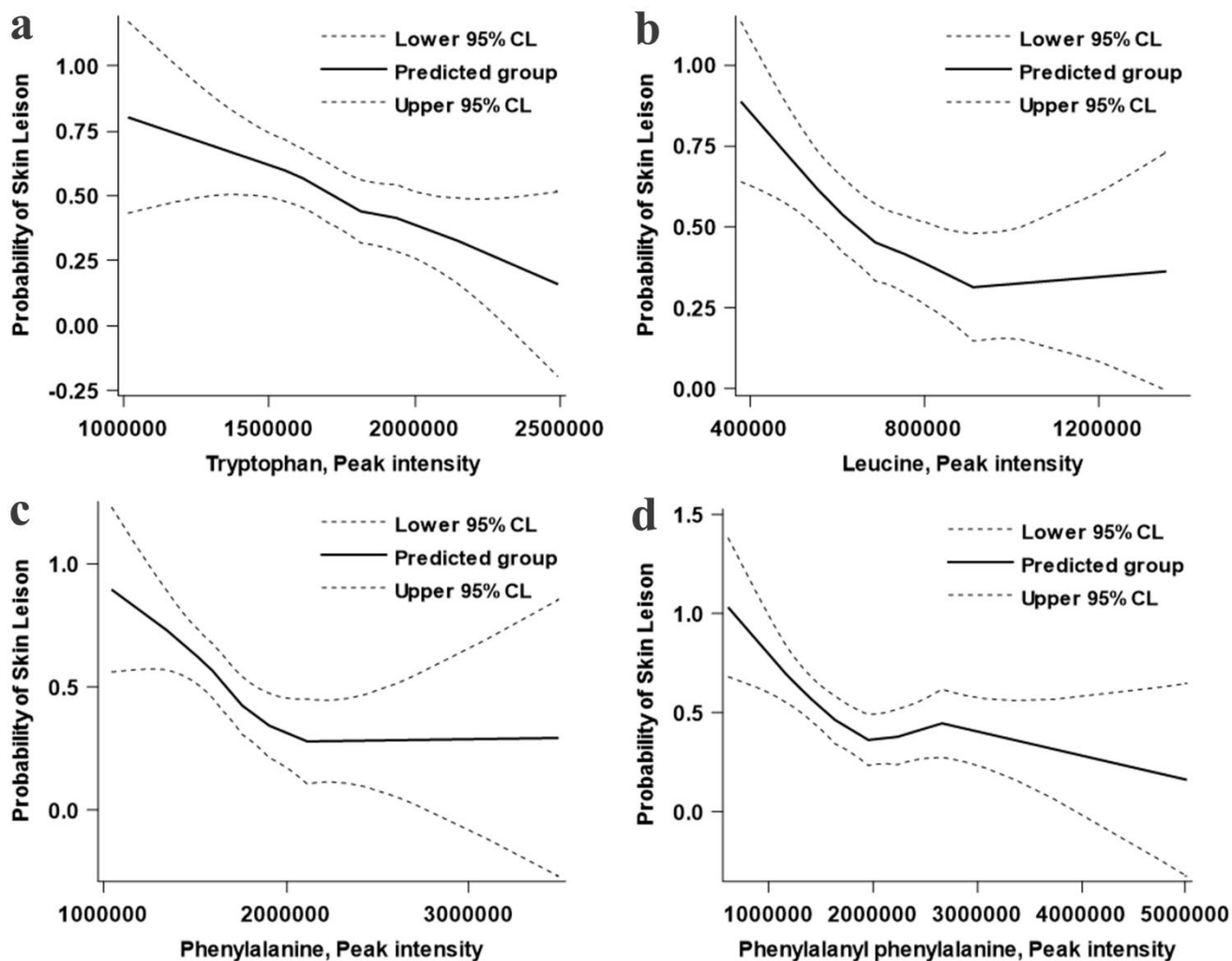


Figure 1. Association between the peak intensity of tryptophan and phenylalanine and arsenic-induced skin lesions based on multivariable locally weighted regression models. a: Tryptophan; b: Phenylalanine;c: Leucine;d: Phenylalanylphenylalanine

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[§].

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	

[§]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.

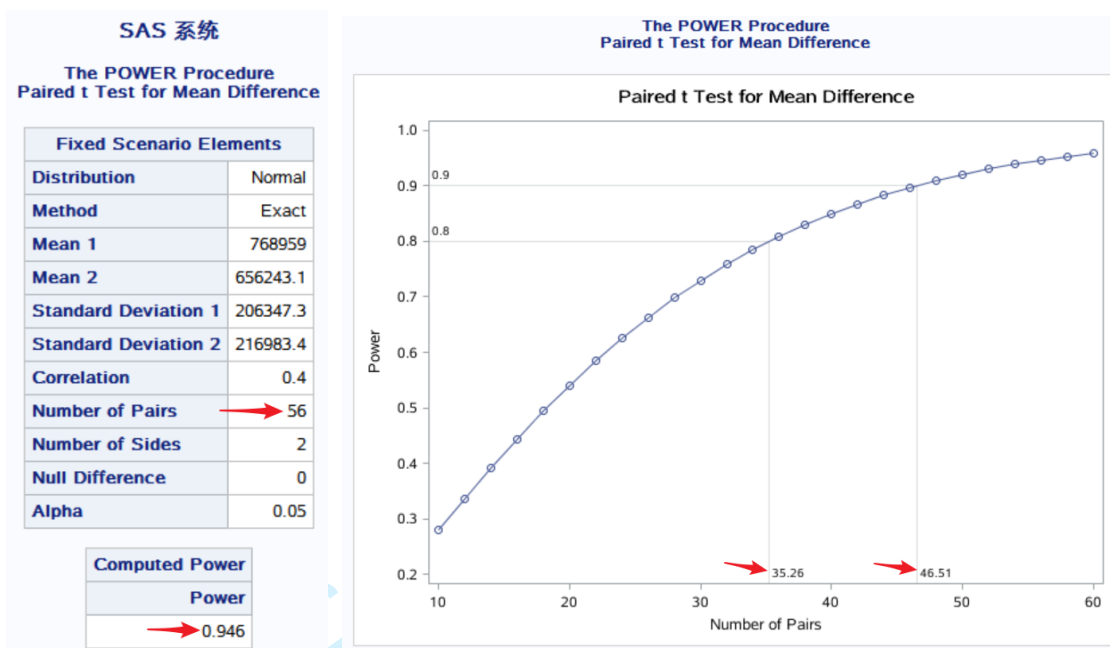


Figure 1S_A. Sample size and power estimation__Leucine

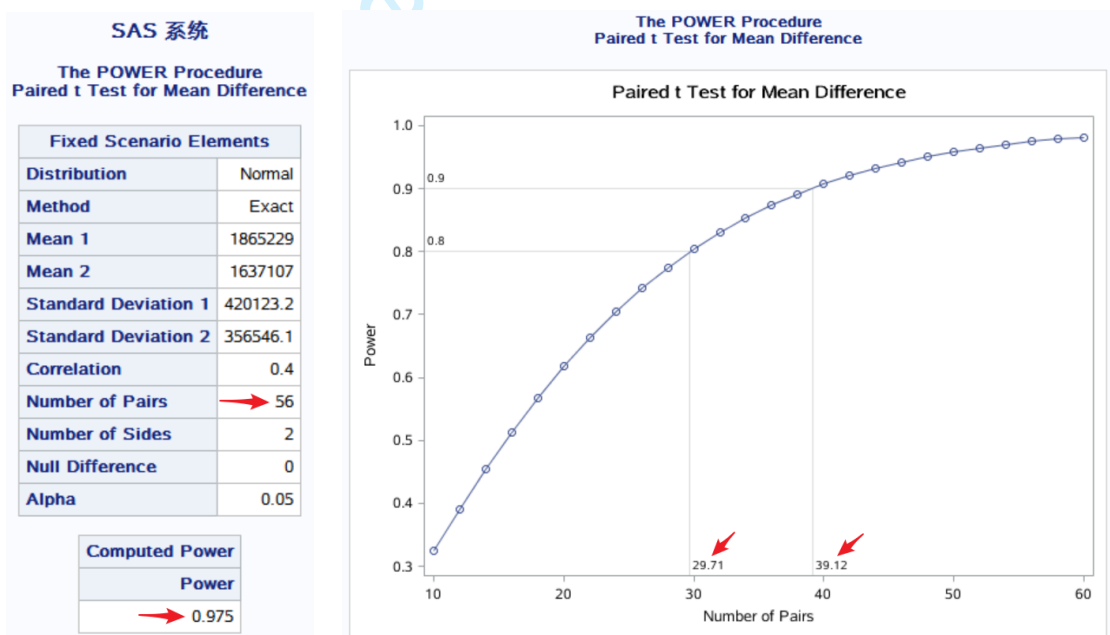


Figure 1S_B. Sample size and power estimation__Phenylalanine

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

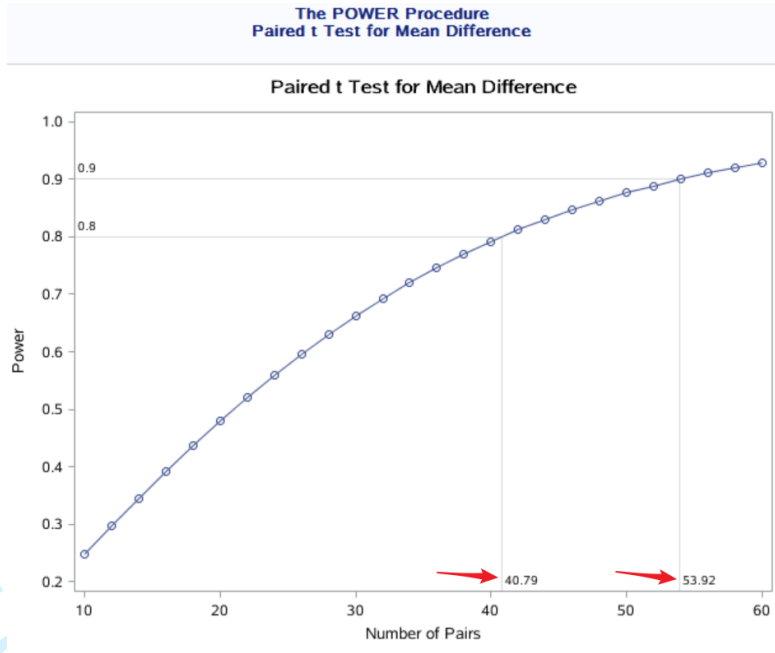


Figure 1S_C. Sample size and power estimation__Phenylalanylphenylalanine

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

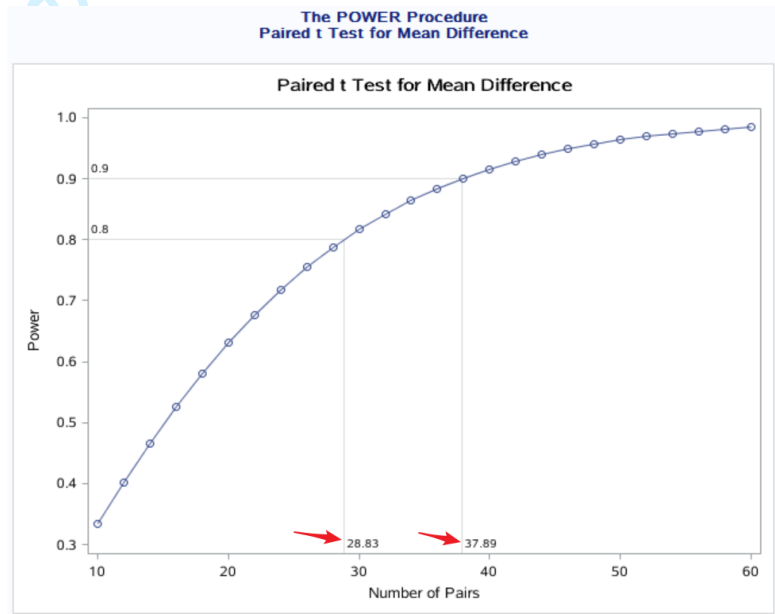


Figure 1S_D. Sample size and power estimation__Tryptophan

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:	<i>BMJ Open</i>
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, Ou Hai 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 University, 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 University, 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

SCHOLARONE™
Manuscripts

1
2
3 **Tryptophan, phenylalanine and arsenic-induced skin lesions in a chronic arsenic exposure**
4 **Chinese population via the drinking water: a case-control study.**
5
6

7 Yaping Wei^{1,2}, Chaonan Jia^{1,2}, Yuan Lan³, Xiangqing Hou^{1,2}, Jingjing Zuo³, Jushuang Li^{1,2}, Tao
8 Wang^{1,2}, Guangyun Mao^{1,2,4}✉
9
10

11
12
13 ¹ Department of Preventive Medicine, School of Public Health and Management, Wenzhou Medical
14 University, Wenzhou, China.

15
16 ² Center on Evidence-Based Medicine & Clinical Epidemiological Research, School of Public Health,
17 Wenzhou Medical University, Wenzhou, China.

18
19 ³ School of Ophthalmology & Optometry, Wenzhou Medical University, Wenzhou, China.

20
21 ⁴ Center on Clinical Research, the Affiliated Eye Hospital of Wenzhou Medical University, Wenzhou,
22 China.
23
24

25
26 Yaping Wei and Chaonan Jia contributed equally to this paper.
27
28

29
30 ✉Corresponding author: Guangyun Mao, Room 7A321, School of Public Health and Management,
31 Wenzhou Medical University, Chashan University Town, Ouhai District, Wenzhou, China. 325035

32
33 E-mail: maogy@wmu.edu.cn

34
35 Telephone: +86-577-86689848

36
37 Fax numbers: +86-577-86699122
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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% ($P < 0.001$).

1
2
3 **Conclusions** Specific AAs may be linked to AISL and play important roles in its early identification.
4

5 Additional studies are needed to confirm our findings.
6

7 **Keywords:** Chronic arsenic exposure; Skin lesions; Metabolomics; Amino acid.
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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.

1 INTRODUCTION

2 Chronic arsenic exposure via drinking water is widely believed as a global health concern,
3 affecting a large number of people worldwide. It may give rise to several human health issues and
4 has been documented to associate with cardiovascular disease, diabetes, cancer and others^{1 2}. With
5 the industrial boom and dramatic rise of worldwide water pollution including arsenic contamination
6 in the past, the prevalence and burden of arsenic-induced health damage will continue to increase.
7 The skin has been confirmed as one of the most common and susceptible targets of arsenic-induced
8 health lesions. Cutaneous skin lesions are typical signs of arsenicosis after persistent arsenic
9 exposure for the long term which are characterized by hyperkeratosis and hyperpigmentation.
10 Considerable evidence of the prevalence of arsenical skin lesions had been observed in many
11 countries³⁻⁵.

12 As arsenic-induced skin lesions (AISL) have been widely accepted as the major early
13 manifestation of arsenic toxicity⁶ and might be indicators of susceptibility to more serious
14 arsenic-induced health hazards⁷, it is particularly crucial to identify participants at risk as early as
15 possible for preventing the onset or delaying the progression of the serious health problems
16 effectively. Several possible mechanisms such as genetic differences⁸, oxidative stress⁹, and
17 epigenetic dysregulation¹⁰ and others may explain arsenic poisoning. Previous studies also reported
18 that arsenic methylation in vivo might be associated with metabolic syndrome^{11 12}.

19 Amino acids (AAs) are the "basic unit" that make up the body's various proteins and necessary
20 to maintain the health. Some AAs are important regulators of some key metabolic pathways and have
21 great importance in maximizing the efficiency of food utilization, enhancing protein accretion and
22 health improvement^{13 14}. Abnormal metabolism of AAs will disturb the homeostasis of the body,
23 impairs growth and development, and even causes death¹⁵. So, the levels of serum AAs may be an
24 important implication for the metabolic status and disease condition. As a powerful tool in system
25 biology research, metabolomics approach is beneficial on unbiased monitoring changes in

1
2
3 26 endogenous metabolism-related physiological processes, providing integrative information on the
4
5 27 distinctive features across multiple functional levels, offering a window to capture the core attributes
6
7
8 28 responsible for various phenotypes, which are particularly important in understanding the relevant
9
10 29 pathophysiological changes of a disease and its status, and identifying novel biomarkers for risk
11
12 30 screening, diagnosis, treatment and prognosis of important human diseases¹⁶⁻¹⁸.

13
14
15 31 Animal experiments and epidemiological studies have reported obvious arsenic-related
16
17 32 metabolomics perturbations^{19 20}. All of these researches substantially suggest that the relationship
18
19 33 between specific metabolites and arsenic-induced health lesions should be investigated. However,
20
21
22 34 few works have been conducted to comprehensively examine the metabolic mechanism relevant to
23
24 35 AISL, especially for amino acid's metabolism. The present study aims to quantitatively examine the
25
26 36 association of several specific AAs with AISL and the ability to identify AISL.

27 28 29 37 **METHODS**

30 31 32 38 **Study Population**

33
34 39 This study was originally from a randomized, double-blind, and placebo-controlled clinical trial
35
36 40 (NCT02235948) in 2010, in which all subjects were randomly selected using permuted block
37
38
39 41 randomization from a single rural area in a population chronically exposed to low-level arsenic
40
41 42 drinking water, had similar lifestyle and influences under similar environmental factors. Information
42
43 43 on the inclusion and exclusion criteria of the participants could be found in our previous study²¹.
44
45 44 Strictly following the criteria of arsenicosis²². AISL was diagnosed as the presence of arsenic-
46
47 45 induced keratosis, hyperpigmentation or depigmentation by a physician from Wenzhou medical
48
49 46 university at the beginning of the trial. This was a matched case-control study (1:1 matching).
50
51 47 Among 450 residents aged 18 to 79 years old enrolled in the trial mentioned above, 56 of them were
52
53 48 diagnosed as AISL and selected as the case. Another 56 participants without AISL matched by
54
55 49 gender and similar age (± 1 year) from the same population were picked out as the control. The
56
57
58
59
60

1
2
3 50 inclusion criteria were subjects having the metabolomic test. Unmatched participants and those
4
5
6 51 without serum metabolites data were excluded.

8 52 **Data Collection and Assessment**

9
10 53 The information on age, gender, exposure year, body mass index, smoking, alcohol
11
12 54 consumption, education level, etc. was collected with a standardized questionnaire. Blood and urine
13
14 55 samples were also collected at the time of participants' enrollment. Detailed data collection of blood
15
16
17 56 and urine samples and assessment methods for clinical variables including plasma fasting glucose
18
19 57 (FPG), serum urea nitrogen, serum folate, total homocysteine, total cholesterol (TC), triglycerides
20
21 58 (TG), high-density lipoprotein (HDL), low-density lipoprotein (LDL) and others had been published
22
23
24 59 previously²¹. The epuration of various urinary arsenic species was conducted utilizing
25
26 60 high-performance liquid chromatography coupled mass spectrometer system for separation and
27
28
29 61 detection²³. The species of arsenic in urine samples consisted of inorganic arsenic (iAs, [iAs^{III} plus
30
31 62 iAs^V]), monomethyl arsenate (MMA, [MMA^{III} plus MMA^V]) and dimethyl arsenate (DMA,
32
33
34 63 [DMA^{III} plus DMA^V]). All arsenic species were corrected by creatinine. The total arsenic (tAs) was
35
36 64 the sum of iAs, MMA and DMA. The percentages of arsenic species were defined as:
37
38
39 65 $iAs\% = iAs/tAs * 100\%$, $MMA\% = MMA/tAs * 100\%$ and $DMA\% = DMA/tAs * 100\%$, respectively.

41 66 **UPLC-QTOF-MS Metabonomic Profiling**

42
43 67 Serum samples were thawed to a temperature of 4°C and 600 µL mixture (90% acetonitrile -
44
45 68 10%water) were added to each sample (200 µL in microcentrifuge tubes). The samples were
46
47
48 69 vigorously mixed for 20 seconds and centrifuged for 5 min at 12000 rpm (20°C). The top 400 µL of
49
50 70 each supernatant was then transferred and dried down in a vacuum concentrator centrifuge. The dried
51
52 71 samples were re-suspended in 130 µL of water (including 15% acetonitrile), mixed vigorously for 20
53
54
55 72 seconds and repeat the centrifugation method described above. Two µL of the supernatant was
56
57 73 collected as samples to be determined. Serum metabolic profile acquisition was performed by using
58
59 74 ACQUITY UPLC[®]/Xevo[®] G2 QTof/MS^E (Waters Corp., Milford, MA, USA). Chromatographic
60

1
2
3 75 separation was performed at 50°C using a WATERS HSS T3 column (2.1×100 mm, 1.7 µm) with a
4
5
6 76 flow rate of 0.4mL/min. The mobile phase was a mixture of (A) H₂O with 0.1% formic acid and (B)
7
8 77 methanol with 0.1% formic acid. Elution was in linear gradient with the programmed gradient at 0
9
10 78 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
11
12 79 min with 0% A and 100% B. The mass spectrometer was operated under both positive-ion (ESI⁺)
13
14
15 80 mode and negative-ion (ESI⁻) mode electrospray ionization. The scan range was from 50 to 1200 m/z.
16
17 81 Data were collected in both ESI⁺ and ESI⁻ modes. The capillary voltage was set at 3000 V and 2500
18
19 82 V, respectively. The desolvation flow rate was 800 L/h at 350°C. Argon was used as collision gas,
20
21
22 83 and the collision energy was adjusted from 10 eV to 40 eV for each analysis. Quantum clustering
23
24 84 (QC) samples were prepared by pooling aliquots of each sample and used to reflect the reliability of
25
26 85 further metabolomics analysis. After peak deconvolution, alignment, integration and normalization,
27
28 86 the data including retention time (RT), mass to charge ratio(m/z), and peak intensity were extracted
29
30
31 87 from raw chromatograms using Progenesis QI 2.0 (Waters Corp., Milford, MA, USA). The MS/MS
32
33 88 mode was performed to obtain metabolites levels processed with MarkerLynx Applications Manager
34
35 89 Version 4.1 (Waters Corp., Milford, MA, USA).

37 38 90 **Distinct Metabolites Identification**

39
40 91 The peak intensity of metabolites for the 56 pairs was acquired and then imported to
41
42 92 MetaboAnalyst 4.0 (<http://www.metaboanalyst.ca/>) for statistical analysis. A partial least-squares
43
44 93 discriminant analysis (PLS-DA), which is a supervised and well-accepted pattern recognition
45
46 94 approach, was used for the differentiation between the cases and controls. False discovery rate (FDR)
47
48 95 adjusted p-value in univariate analysis was performed to reduce the potential impact induced by the
49
50 96 false positive of the results. The criteria used in the selection of metabolites include variable
51
52 97 importance in the project (VIP) scores >1 in PLS-DA and the crude or FDR adjusted p-value all <
53
54 98 0.05 in Wilcoxon signed-rank test. We identified a total of 70 extracted small molecular metabolites
55
56 99 that were linked to the recognition of AISL. The Human Metabolome Database (<http://www.hmdb.ca>)
57
58
59
60

1
2
3 100 was used to identify the name of metabolites. Among them, there were four amino acid metabolites
4
5
6 101 (phenylalanine, tryptophan, leucine, phenylalanylphenylalanine).

8 102 **Statistical Analysis**

9
10 103 The normality of continuous data was assessed using both QQ-plots and Shapiro-Wilk test. The
11
12 104 comparison between the cases and controls was performed with the paired *t*-test if they met normal
13
14
15 105 or similar normal distribution. Otherwise, the Wilcoxon signed-rank test would be used. Differences
16
17 106 in the proportion of categorical variables between the two groups were evaluated by
18
19 107 McNemar-Bowker tests. We firstly used locally weighted scatterplot smoothing (LOESS) models to
20
21
22 108 estimate the “real” relationship between serum AAs levels and the probability of AISL. Then,
23
24 109 multivariable conditional logistic regression models were performed to examine the association
25
26 110 between the contributing AAs levels and AISL after adjusting for some potentially confounding
27
28
29 111 factors. The individual impacts of AA metabolites on the risk of AISL were quantified separately by
30
31 112 odds ratio (OR) and 95% confidence interval (CI) in the following two ways: with AAs as a
32
33 113 categorical variable (quartiles) and as a continuous variable [scaled to an interquartile range (IQR)].
34
35
36 114 Variables with a p-value less than 0.2 in the comparison between two groups were selected as
37
38 115 potential confounders, which had been widely performed in many studies especially when the sample
39
40 116 size of the current study was not too large. The variance inflation factor (VIF) was used to examine
41
42
43 117 the potential collinearity among them. As too many covariates in a multiple regression model will
44
45 118 lead to overfitting to some extent²⁴, we finally select no more than 4 variables as confounding factors
46
47 119 to decrease the potential overfitting when assessing the association between AAs and AISL.
48
49 120 Furthermore, as the distinct metabolites might be highly related to each other, collinearity should be
50
51
52 121 well considered. So, we used the VIF based on the VIF package of R software to detect potential
53
54 122 collinearity among the AAs. When the VIF is greater than 1.5, it was considered as collinearity
55
56 123 existed in the model and the associated variable would be removed. The combined effect of relevant
57
58 124 AAs on AISL was also performed using a multivariable logistic regression model. Besides, receiver
59
60

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 $P \leq 0.05$ 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^ξ			
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.

^ξ 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 figure 1, 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

1
2
3 153 compared to their lowest quartiles, respectively. The category boundaries of the quartiles were shown
4
5
6 154 in Table S1. Significant linear trends existed between AISL and those four serum AAs. Meanwhile,
7
8 155 the same linear negative association between AISL and per IQR rise of the four serum AAs was
9
10 156 observed when these AAs were considered as continuous variables in the present study.
11

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

Amino acids	N	Cases (%)	Crude		Adjusted [‡]	
			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

48
49 [‡]Values are odds ratio (95% confidence intervals) for arsenic-induced skin lesions from conditional logistic regression. IQR: interquartile range; Q₁: the 1st quartile; Q₂: the 2nd quartile; Q₃: the 3rd quartile; Q₄: the 4th quartile.

50
51 [‡]Adjusted for plasma glucose, low-density lipoprotein, triglyceride, and urinary dimethyl arsenate.

52
53 157 As these four specific AAs are significantly or marginal significantly associated with the odds
54
55 158 of AISL, so it is needed to examine the joint impacts among them on AISL. However, the results of
56
57 159 potential collinearity examination revealed that among these four specific AAs, both tryptophan and
58
59 160 phenylalanine had the smallest VIF value (VIF=1.04) and no obvious collinearity existed (**Table S2**).

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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 β_j 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

1
2
3 198 phenylalanine levels in a chronic arsenic exposure population via drinking water. Participants with a
4
5
6 199 higher level of both AAs would have the lowest odds of AISL. These two AAs might also be able to
7
8 200 serve as the indicators of AISL.

9
10 201 The probability of the initiation and development of AISL would be affected by a large number
11
12 202 of factors including age, gender, lifestyles, arsenic exposure, metabolism, and others. These factors
13
14
15 203 would be important confounding factors and will largely affect our results. To adjust for the impacts
16
17 204 due to these co-factors, we firstly selected all participants using permuted block randomization from
18
19 205 a single rural area in which the population were chronically exposed to arsenic in the same way, had
20
21
22 206 a similar lifestyle and environmental factors. Secondly, the cases and controls were matched by
23
24 207 gender and age (± 1 year). All of these may be the reason why so many potential confounders
25
26 208 including arsenic exposure do not differ significantly between the cases and controls.

27
28
29 209 Participants enrolled in the current study were chronically exposed to arsenic via drinking water.
30
31 210 The geometric mean (GM) and its related 95% CI of urinary iAs/creatinine and tAs/creatinine in this
32
33 211 population were 17.49 (14.90, 20.53) $\mu\text{g/g}$ and 147.20 (129.00, 167.97) $\mu\text{g/g}$, respectively. They
34
35 212 were much higher than those in the 20 $\mu\text{g/L}$ exposed to arsenic via drinking water [GM (95% CI):
36
37
38 213 0.4 (0.3,0.5) $\mu\text{g/g}$ for iAs and 9.1 (6.5,12.7) $\mu\text{g/g}$ for tAs], while obviously lower than those in the 90
39
40 214 $\mu\text{g/L}$ exposed group [GM (95% CI): 39.4 (31.4, 49.6) $\mu\text{g/g}$ for iAs and 248.7 (208.8, 296.3) $\mu\text{g/g}$ for
41
42 215 tAs]²⁵. An available report has shown that AISL cannot be completely cured even though medical
43
44
45 216 technology has already made great progress²⁶. So, it is crucial to identify those who are most likely to
46
47 217 progress to overt arsenic damages including AISL among people at risk as early as possible.
48
49 218 Metabolomics study, which mainly focuses on thoroughly assessing the variation of metabolites
50
51
52 219 possibly linked to diseases occurrence and development, has been widely utilized to help us
53
54 220 understand the pathogenesis of diseases because of its relevance to the phenotypes as compared to
55
56 221 other 'OMICs' study²⁷. Moreover, mathematical modeling to assess the linkage between small
57
58
59 222 molecular metabolites and arsenic toxicity has grown²⁸. Developing a simple and interpretable
60

1
2
3 223 modeling approach for the early detection of arsenic-induced health lesions is of great theoretical
4
5 value and realistic meaning²⁹, though it might be difficult due to population-specific complexities
6 224
7
8 225 and the impacts due to some potential unmeasured covariates such as diet and genetic determinants.
9

10 226 Previous studies reported that gene-gene and gene-environment interaction were involved in
11
12 arsenicosis through toxicological mechanisms including genomic instability³⁰ and oxidative stress³¹.
13 227
14
15 228 Skin hyperpigmentation and palmoplantar hyperkeratosis could be biomarkers for long-term arsenic
16
17 229 exposure identifying the differences in metabolites that are associated with phenotypes through
18
19 230 metabolites analysis may promote our understanding and identification of AISL. Animal study
20
21 reveals that the disruption of amino acids metabolism upon arsenic exposure in the rat which may be
22 231
23
24 232 beneficial in understanding arsenic toxicity³². In our previous population-based metabolomics study,
25
26 233 we found that serum metabolites alteration was significantly related to the risk of arsenic-induced
27
28 health damages. In the current study, we detected that some BCAAs or AAAs were also significantly
29 234
30
31 235 relevant to AISL occurrence. Several studies across numerous ethnic backgrounds support the usage
32
33 236 of BCAAs including leucine, isoleucine as well as valine and AAAs profile such as phenylalanine,
34
35 237 tryptophan, and tyrosine as biomarkers in determining metabolic diseases^{27 33}. Simultaneously, Zhou
36
37 *et al* reported that arsenic-induced transformed cells exhibit apparent alterations in metabolite
38 238
39 profiles including down-regulated of leucine, tryptophan, and phenylalanine in skin lesions group³⁴.
40 239
41
42 240 Consistent with Zhou's findings, two serum AAAs (tryptophan, phenylalanine) levels were also
43
44 significantly associated with AISL in our study.
45 241
46

47 242 Normal metabolism of amino acids is necessary for whole-body homeostasis, growth and
48
49 243 development, and health status¹⁵. Studies have reported that the changes in the availability of AAAs
50
51 244 will affect cell signaling, gene expression, brain, and neuroendocrine function³⁵. Tryptophan, an
52
53 amino acid metabolism-related biomarker, is also a sensitive and specific indicator of oxidation.
54 245
55
56 246 Tryptophan metabolism in mammals is a physiological means of preserving immune homeostasis
57
58 247 associated with oxidative stress and inflammation^{36 37}. Besides, phenylalanine can be transformed
59
60

1
2
3 248 into specific neurotransmitters such as dopamine and adrenaline by the action of related enzymes.
4
5
6 249 Wu and colleagues³⁸ reported that arsenic exposure would lead to neurotransmitter metabolism
7
8 250 disturbance, which might explain the reduction of phenylalanine. Furthermore, as one of the
9
10 251 peptide-bound phenylalanine, phenylalanylphenylalanine has been reported to affect protein
11
12 252 synthesis and secretion³⁹, potentially indicating the possible relation between endothelium
13
14
15 253 dysfunction and phenylalanine metabolism disorder. The relationship between amino acid
16
17 254 metabolism and AISL was still unclear. The notable alteration of tryptophan and phenylalanine in
18
19 255 this study may well indicate the occurrence of metabolic disorders due to arsenic exposure. It is also
20
21
22 256 beneficial to understand the effects of arsenic toxicity and of great importance in early identification
23
24 257 of occurrences as well as delaying the progression of various arsenic-induced health lesions
25
26 258 including AISL.

27
28
29 259 The current study included 56 AISL cases matched 56 non-AISL controls and the sample size
30
31 260 might be potentially insufficient. To estimate the impact due to this potential insufficient sample size
32
33 261 on our conclusion, the PROC POWER procedure for paired design study in SAS 9.4 (SAS Institute
34
35 262 Inc.) was applied to assess the power of the four AAs when assessing their associations with AISL in
36
37
38 263 this study. The results showed that the lowest power associated with all of these four AAs was 0.911
39
40 264 based on 56 pairs of participants (Figure 1S). It suggested that with type I error as 0.05, the total
41
42 265 sample size as 56 pairs and two-sided test, the powers associated with these four AAs were all great
43
44
45 266 than 0.8. So, we believed that the sample size for the present study, 56 pairs, would well balance the
46
47 267 power of tests. Furthermore, previous metabolomics studies usually have a sample size of no more
48
49 268 than 40 cases in each group^{40 41}.

50
51
52 269 The main strength of this study is that the findings were based on a community-based, long-term
53
54 270 arsenic exposure cohort with well-designed quality assurance and quality control throughout the
55
56 271 study, and the AAs were detected with non-targeted metabolomics approach through the discovery
57
58 272 and validation phases. However, there are also several limitations to this study. Firstly, although the
59
60

1
2
3 273 untargeted metabolomics approach can assess a large amount number of metabolites precisely and
4
5
6 274 efficiently, it only provided relative levels of AAs instead of their accurate quantitative concentration.
7
8 275 Secondly, these findings were mainly based on a case-control study, which only reveals the
9
10 276 association between amino acid metabolism and the odds of AISL rather than confirming their causal
11
12 277 relationship. Furthermore, as it is suggested that the ratio of approximately 10 to 15 observations per
13
14
15 278 predictor in a logistic regression model will produce reasonably stable estimations²⁴, we selected
16
17 279 only 4 covariates in the models due to the small sample size and these results need to be confirmed in
18
19 280 new studies. Finally, the participants were mainly exposed to arsenic via drinking water, which
20
21
22 281 would limit the findings extrapolated to the other arsenic exposure population via food and other
23
24 282 ways. Therefore, additional elaborate population-based studies are needed to verify our discoveries.
25

26 283 In conclusion, specific AAs might be linked to AISL and amino acid metabolism may play an
27
28
29 284 important role in AISL early identification. Additional studies may be needed to confirm our
30
31 285 findings.
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4 **Contributors:** Guangyun Mao and Yaping Wei designed the study. Chaonan Jia participated in
5 collecting data. Yuan Lan and Chaonan Jia audited the data. Yaping Wei, Xiangqing hou, Jushuang
6 Li, Tao Wang conducted the literature search, Yaping Wei, Chaonan Jia conducted statistical
7 analysis and interpreted the results. Yaping Wei and Chaonan Jia wrote the first draft of the
8 manuscript. Jingjing Zuo helped with copyediting. Guangyun Mao reviewed the final manuscript and
9 did substantial contributions.
10
11

12 **Funding:** This work was supported by the basic public welfare research project of Zhejiang province
13 (LGF19H260011), the initial scientific research fund (KYQD170301) and the major project of the
14 eye hospital of Wenzhou medical university (YNZD201602).
15

16 **Acknowledgments:** We gratefully acknowledge all of the study participants of this project. We wish
17 to thank all of the dedication and hard work of the field team at the department of preventive
18 medicine, school of public health and management, Wenzhou medical university. Our thanks must
19 also go to Dr. Kendrick Hii, a native-English speaker, for his valuable suggestions and edits to the
20 manuscript.
21
22

23 **Competing Interest:** None declared.
24

25 **Ethics approval:** Ethical approval was received from the ethics committee of Wenzhou Medical
26 University, Wenzhou, China.
27

28 **Provenance and peer review:** Not commissioned; externally peer reviewed.
29

30 **Data Sharing Statement:** No additional unpublished data are available.
31

32 **Patient consent for publication:** Not required.
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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

- 1
2
3 regression analysis. *J Clin Epidemiol*. 1996; 49:1373-9.
- 4 25. Sun G, Xu Y, Li X, et al. Urinary arsenic metabolites in children and adults exposed to arsenic in drinking
5 water in Inner Mongolia, China. *Environmental Health Perspectives*. 2007; 115:648-52.
- 6 26. Kuo CC, Moon KA, Wang SL, et al. The Association of Arsenic Metabolism with Cancer, Cardiovascular
7 Disease, and Diabetes: A Systematic Review of the Epidemiological Evidence. *Environmental Health*
8 *Perspectives*. 2017; 125:15.
- 9 27. Wang TJ, Larson MG, Vasan RS, et al. Metabolite profiles and the risk of developing diabetes. *Nature*
10 *Medicine*. 2011; 17:448-U83.
- 11 28. Lawley SD, Cinderella M, Hall MN, et al. Mathematical model insights into arsenic detoxification. *Theor Biol*
12 *Med Model*. 2011; 8.
- 13 29. Juhasz AL, Herde P, Herde C, et al. Validation of the predictive capabilities of the Sbrc-G in vitro assay for
14 estimating arsenic relative bioavailability in contaminated soils. *Environ Sci Technol*. 2014; 48:12962-9.
- 15 30. Kibriya MG, Jasmine F, Parvez F, et al. Association between genome-wide copy number variation and
16 arsenic-induced skin lesions: a prospective study. *Environ Health*. 2017; 16:13.
- 17 31. Priori R, Scrivo R, Brandt J, et al. Metabolomics in rheumatic diseases: the potential of an emerging
18 methodology for improved patient diagnosis, prognosis, and treatment efficacy. *Autoimmun Rev*. 2013;
19 12:1022-30.
- 20 32. Wang X, Mu X, Zhang J, et al. Serum metabolomics reveals that arsenic exposure disrupted lipid and amino
21 acid metabolism in rats: a step forward in understanding chronic arsenic toxicity. *Metallomics Integrated*
22 *Biometal Science*. 2015; 7:544.
- 23 33. Weng L, Quinlivan E, Gong Y, et al. Association of Branched and Aromatic Amino Acids Levels with
24 Metabolic Syndrome and Impaired Fasting Glucose in Hypertensive Patients. *Metabolic Syndrome And*
25 *Related Disorders*. 2015; 13:195-202.
- 26 34. Zhou YY, Wang YF, Su J, et al. Integration of microRNAome, proteomics and metabolomics to analyze
27 arsenic-induced malignant cell transformation. *Oncotarget*. 2017; 8:90879-96.
- 28 35. Lee JC, Chen MJ, Chang CH, et al. Plasma amino acid levels in patients with colorectal cancers and liver
29 cirrhosis with hepatocellular carcinoma. *Hepato-Gastroenterol*. 2003; 50:1269-73.
- 30 36. Munn DH, Mellor AL. Indoleamine 2,3 dioxygenase and metabolic control of immune responses. *Trends In*
31 *Immunology*. 2013; 34:137-43.
- 32 37. Pawlak K, Domaniewski T, Mysliwiec M, et al. The kynurenines are associated with oxidative stress,
33 inflammation and the prevalence of cardiovascular disease in patients with end-stage renal disease.
34 *Atherosclerosis*. 2009; 204:309-14.
- 35 38. Wu L-L, Gong W, Shen S-P, et al. Multiple metal exposures and their correlation with monoamine
36 neurotransmitter metabolism in Chinese electroplating workers. *Chemosphere*. 2017; 182:745-52.
- 37 39. Zhou MM, Wu YM, Liu HY, et al. Effects of phenylalanine and threonine oligopeptides on milk protein
38 synthesis in cultured bovine mammary epithelial cells. *J Anim Physiol Anim Nutr*. 2015; 99:215-20.
- 39 40. Chen LY, Cheng CY, Choi HW, et al. Plasma Metabonomic Profiling of Diabetic Retinopathy. *Diabetes*.
40 2016; 65:1099-108.
- 41 41. Piszcz J, Lemancewicz D, Dudzik D, et al. Differences and similarities between LC-MS derived serum
42 fingerprints of patients with B-cell malignancies. *Electrophoresis*. 2013; 34:2857-64.
- 43
44
45
46
47
48
49
50
51
52
53

54 **Figure 1.** Association between the peak intensity of tryptophan and phenylalanine and
55 arsenic-induced skin lesions based on multivariable locally weighted regression models. a:
56 Tryptophan; b: Phenylalanine; c: Leucine; d: Phenylalanylphenylalanine
57
58
59
60

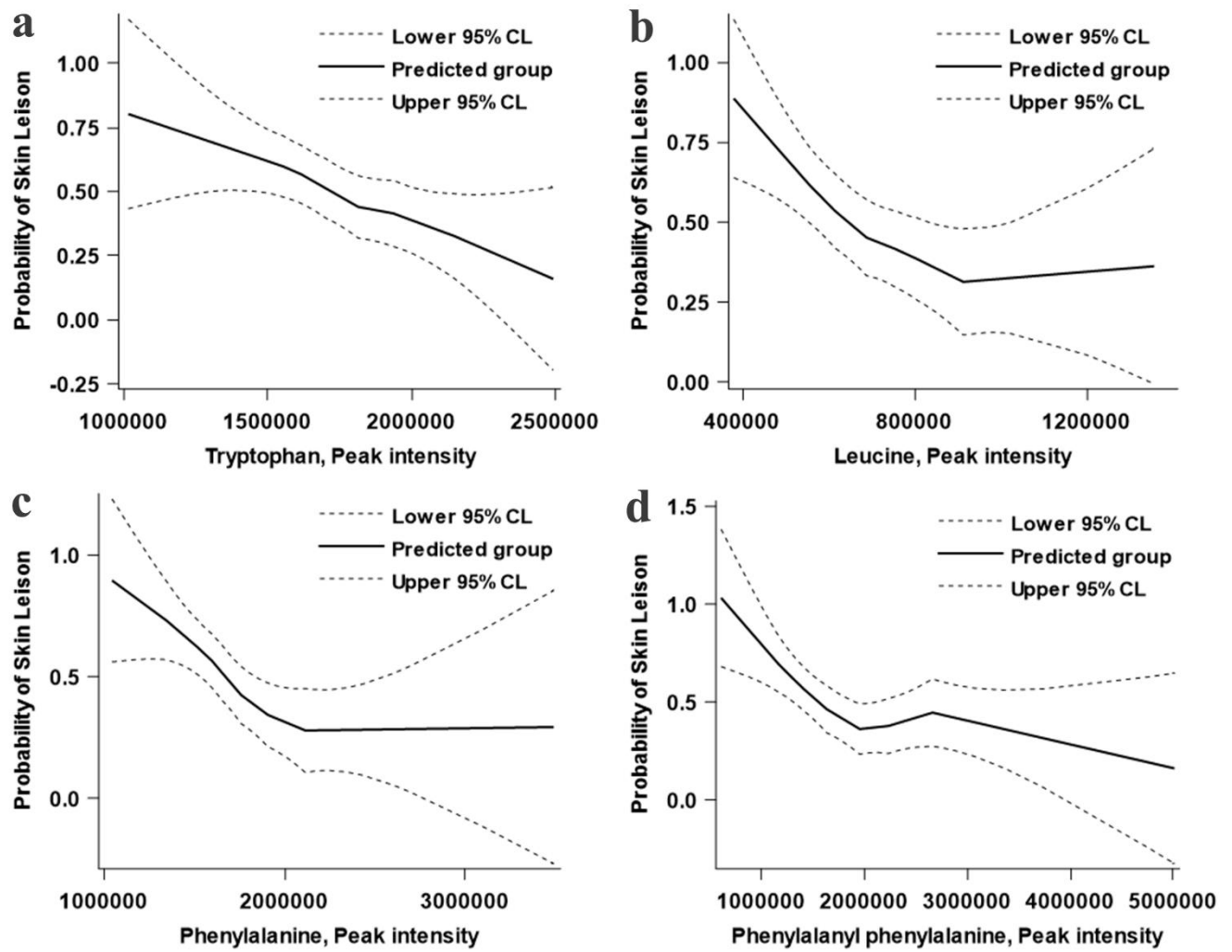


Figure 1. Association between the peak intensity of tryptophan and phenylalanine and arsenic-induced skin lesions based on multivariable locally weighted regression models. a: Tryptophan; b: Phenylalanine; c: Leucine; d: Phenylalanylphenylalanine

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[§].

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	

[§]Model 1: Tryptophan, phenylalanine, leucine and phenylalanylphenylalanine;

Model 2: Tryptophan, leucine and phenylalanylphenylalanine;

Model 3: Tryptophan, phenylalanine and phenylalanylphenylalanine;

Model 4: Tryptophan and phenylalanine.

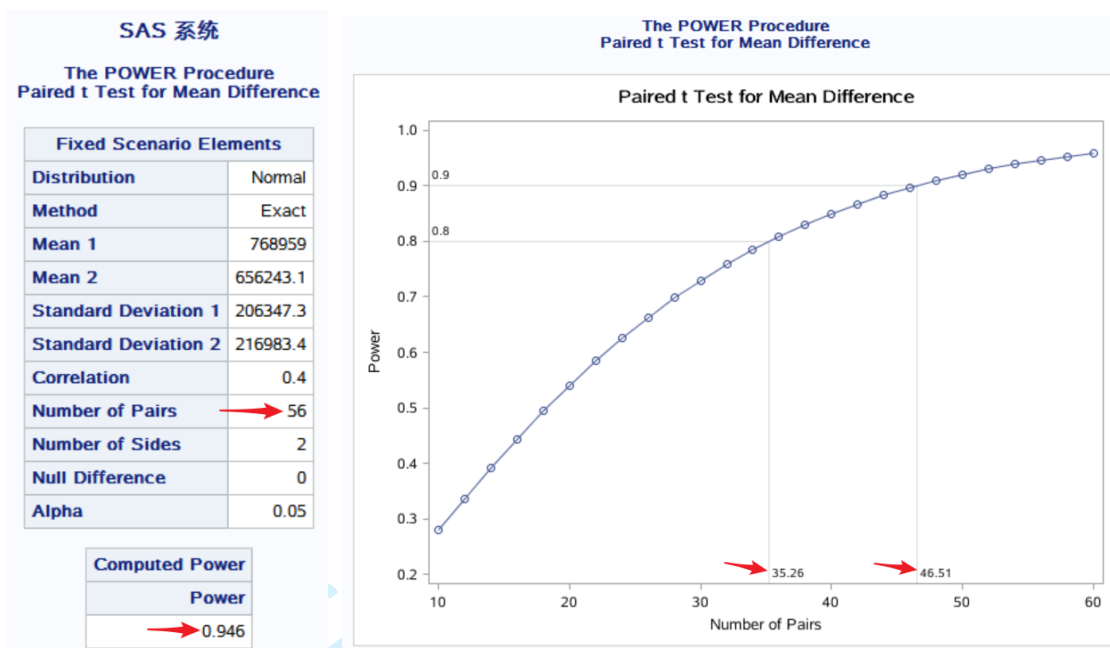


Figure 1S_A. Sample size and power estimation__Leucine

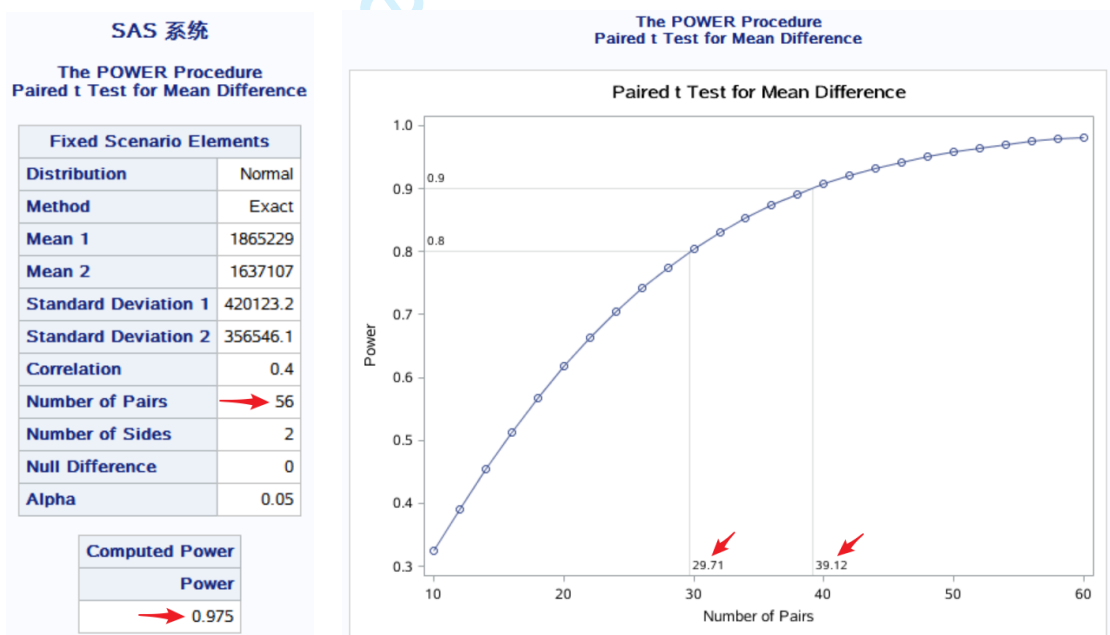


Figure 1S_B. Sample size and power estimation__Phenylalanine

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

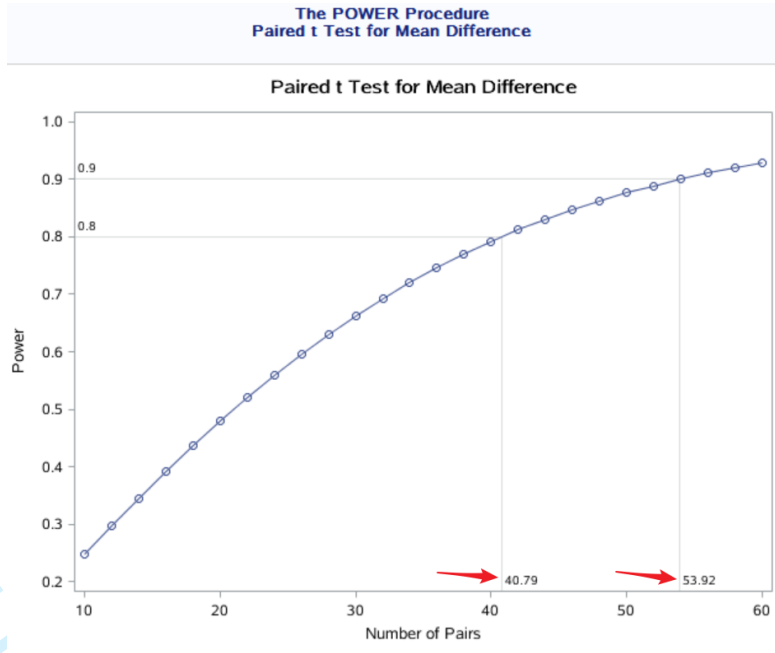


Figure 1S_C. Sample size and power estimation__Phenylalanylphenylalanine

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

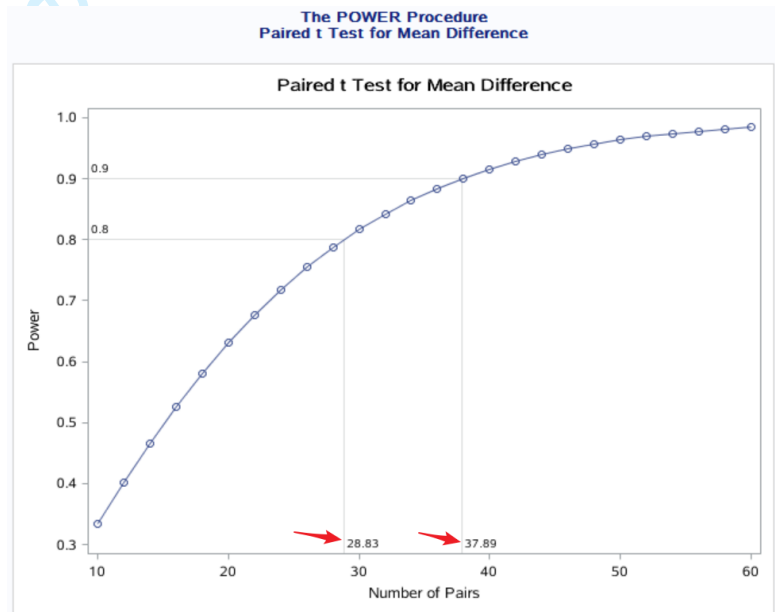


Figure 1S_D. Sample size and power estimation__Tryptophan

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:	<i>BMJ Open</i>
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

SCHOLARONE™
Manuscripts

1
2
3 **Tryptophan and phenylalanine are associated with arsenic-induced skin lesions in a Chinese**
4 **population chronically exposed to arsenic via drinking water: a case-control study**
5
6

7 Yaping Wei^{1,2}, Chaonan Jia^{1,2}, Yuan Lan³, Xiangqing Hou^{1,2}, Jingjing Zuo³, Jushuang Li^{1,2}, Tao
8 Wang^{1,2}, Guangyun Mao^{1,2,4}✉
9
10

11
12
13 ¹ Department of Preventive Medicine, School of Public Health and Management, Wenzhou Medical
14 University, Wenzhou, China.

15
16 ² Center on Evidence-Based Medicine & Clinical Epidemiological Research, School of Public
17 Health, Wenzhou Medical University, Wenzhou, China.

18
19 ³ School of Ophthalmology & Optometry, Wenzhou Medical University, Wenzhou, China.

20
21 ⁴ Center on Clinical Research, the Affiliated Eye Hospital of Wenzhou Medical University, Wenzhou,
22 China.
23
24

25
26 Yaping Wei and Chaonan Jia contributed equally to this paper.
27
28

29
30 ✉Corresponding author: Guangyun Mao, Room 7A321, School of Public Health and Management,
31 Wenzhou Medical University, Chashan University Town, Ouhai District, Wenzhou, China. 325035

32
33 E-mail: maogy@wmu.edu.cn

34
35 Telephone: +86 - 577 - 86689848

36
37 Fax numbers: +86 - 577 - 86699122
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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 ($P <$

1
2
3 0.05). The combined pattern with AUC (95% CI) was 0.72 (0.62–0.81), showing a sensitivity of
4
5 76.79% and specificity of 58.93% ($P < 0.001$).
6

7
8 **Conclusions** Specific AAs may be linked to AISL and play important roles in early AISL
9
10 identification.
11

12 **Keywords:** Chronic arsenic exposure; Skin lesions; Metabolomics; Amino acid.
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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.

1 INTRODUCTION

2 Chronic arsenic exposure via drinking water is widely considered a global health concern
3 affecting several people worldwide. This exposure may cause various human health issues such as
4 cardiovascular diseases, diabetes, and cancer^{1 2}. With the industrial boom and considerable increase
5 in global water pollution including arsenic contamination in the past, the prevalence and burden of
6 arsenic-induced health damage will continue to increase. The skin has been confirmed as one of the
7 most common and susceptible targets of arsenic-induced health effects. Cutaneous skin lesions are
8 typical signs of arsenicosis after persistent and long-term arsenic exposure. These lesions are
9 characterized by hyperkeratosis and hyperpigmentation. Considerable evidence of the prevalence of
10 arsenic-induced skin lesions (AISL) has been reported in several countries³⁻⁵.

11 As AISL are widely accepted as the major early manifestation of arsenic toxicity⁶ and may be an
12 indicator of susceptibility to more serious arsenic-induced health hazards⁷, it is critical to identify
13 people who are at risk as early as possible to prevent the onset or delay the progression of serious
14 health problems. Several possible mechanisms such as genetic differences⁸, oxidative stress⁹, and
15 epigenetic dysregulation¹⁰ may explain arsenic poisoning. Previous studies also suggested that
16 arsenic methylation in vivo is associated with metabolic syndrome^{11 12}.

17 Amino acids (AAs) are the "basic unit" of all proteins and are necessary to maintain health.
18 Some AAs are important regulators of key metabolic pathways and also help in maximizing food
19 utilization, enhancing protein accretion, and improving health^{13 14}. Abnormal metabolism of AAs can
20 disturb homeostasis in the body, impair growth and development, and even cause death¹⁵. Thus, the
21 levels of serum AAs may be important indicators of metabolic status and disease condition. As
22 powerful tools in system biology research, metabolomics approaches are beneficial for unbiased
23 monitoring of changes in endogenous metabolism-related physiological processes, providing
24 integrative information on distinct features across multiple functional levels. These methods capture
25 the core attributes responsible for various phenotypes, which are particularly important in

1
2
3 26 understanding the relevant pathophysiological changes of a disease and its status, and in identifying
4
5 27 novel biomarkers for risk screening, diagnosis, treatment, and prognosis of important human
6
7
8 28 diseases^{16–18}.

9
10 29 Animal experiments and epidemiological studies have reported obvious arsenic-related
11
12 30 metabolomics perturbations^{19 20}. These results suggest that the relationship between specific
13
14 31 metabolites and arsenic-induced health lesions should be investigated. However, only a few studies
15
16 32 have been conducted to comprehensively examine the metabolic mechanism relevant to AISL,
17
18 33 particularly for AA metabolism. This study was conducted to quantitatively examine the association
19
20 34 of several specific AAs with AISL and their ability to identify AISL.
21
22
23
24

25 35 **METHODS**

26 27 36 **Study Population**

28
29 37 The study data were originally obtained from a randomized, double-blind, and
30
31 38 placebo-controlled clinical trial (NCT02235948) performed in 2010, in which all subjects were
32
33 39 randomly selected by permuted block randomization from a single rural area in which a population
34
35 40 was chronically exposed to low-level arsenic drinking water, had similar lifestyles, and were
36
37 41 influenced by similar environmental factors. Information on the inclusion and exclusion criteria of
38
39 42 the participants can be found in our previous study²¹. Strictly following the criteria of arsenicosis²²,
40
41 43 AISL was diagnosed as the presence of arsenic-induced keratosis, hyperpigmentation, or
42
43 44 depigmentation by a physician from the Wenzhou Medical University at the beginning of the trial.
44
45 46 This was a matched case-control study (1:1 matching). Among 450 residents aged 18–79 years
46
47 48 enrolled in the previous trial, 56 were diagnosed as having AISL and selected as the case group.
48
49 49 Another 56 participants without AISL matched by gender and age (± 1 year) from the same
50
51 52 population were evaluated as controls. The inclusion criteria were subjects who underwent a
52
53 54 metabolomic test. Unmatched participants and those without serum metabolites data were excluded.
54
55 55
56
57 49
58
59

59 50 **Data Collection and Assessment**

60

1
2
3 51 Information on age, gender, exposure year, body mass index, smoking, alcohol consumption,
4
5
6 52 and education level, among other factors, was collected using a standardized questionnaire. Blood
7
8 53 and urine samples were also collected at the time of participants' enrollment. The detailed methods
9
10 54 for analyzing blood and urine samples and assessment methods for clinical variables including
11
12 55 fasting plasma glucose (FPG), serum urea nitrogen, serum folate, total homocysteine, total
13
14 56 cholesterol (TC), triglycerides (TG), high-density lipoprotein (HDL), low-density lipoprotein (LDL),
15
16
17 57 and others have been published previously²¹. Various urinary arsenic species were separated and
18
19 58 detected by high-performance liquid chromatography coupled mass spectrometry system²³. The
20
21
22 59 species of arsenic in urine samples consisted of inorganic arsenic (iAs, [iAs^{III} plus iAs^V]),
23
24 60 monomethyl arsenate (MMA, [MMA^{III} plus MMA^V]), and dimethyl arsenate (DMA, [DMA^{III} plus
25
26
27 61 DMA^V]). All arsenic species were corrected by creatinine. The total arsenic (tAs) was the sum of
28
29 62 iAs, MMA and DMA. The percentages of arsenic species were defined as: iAs% = iAs/tAs*100%,
30
31
32 63 MMA% = MMA/tAs*100% and DMA% = DMA/tAs*100%, respectively.

33 34 64 **UPLC-QTOF-MS Metabonomic Profiling**

35
36 65 Serum samples were thawed to 4°C, and then 600 µL of a mixture of 90% acetonitrile and
37
38 66 10% water was added to each sample (200 µL in microcentrifuge tubes). The samples were
39
40
41 67 vigorously mixed for 20 s and then centrifuged for 5 min at 12,000 g (20°C). Four hundred
42
43 68 microliters of each supernatant were transferred to a new tube and dried in a vacuum concentrator
44
45 69 centrifuge. The dried samples were re-suspended in 130 µL of water (containing 15% acetonitrile),
46
47
48 70 mixed vigorously for 20-s, and centrifuged as described above. Two microliters of the supernatant
49
50 71 were collected for analysis. Serum metabolic profile was acquired using ACQUITY UPLC[®]/Xevo[®]
51
52 72 G2 QToF/MS^E (Waters Corp., Milford, MA, USA). Chromatographic separation was performed at 50
53
54 73 °C using a Waters HSS T3 column (2.1 mm × 100 mm, 1.7 µm; Milford, MA, USA) at a flow rate of
55
56
57 74 0.4 mL/min. The mobile phase was a mixture of (A) H₂O with 0.1% formic acid and (B) methanol
58
59 75 with 0.1% formic acid. Elution was performed over a linear gradient-as follows: 0 min with 100% A
60

1
2
3 76 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
4
5 77 100% B. The mass spectrometer was operated under both electrospray ionization positive-ion (ESI⁺)
6
7 78 mode and negative-ion (ESI⁻) mode. The scan range was 50–1200 m/z. The capillary voltage was set
8
9
10 79 to 3000 and 2500 V, respectively. The desolvation flow rate was 800 L/h at 350°C. Argon was used
11
12 80 as the collision gas, and the collision energy was adjusted from 10 to 40 eV in each analysis.
13
14
15 81 Quantum clustering (QC) samples were prepared by pooling aliquots of each sample and used to
16
17 82 reflect the reliability of further metabolomics analysis. After peak deconvolution, alignment,
18
19 83 integration, and normalization, the data including retention time (RT), mass to charge ratio (m/z),
20
21
22 84 and peak intensity were extracted from raw chromatograms using Progenesis QI 2.0 (Waters Corp.,
23
24 85 Milford, MA, USA). MS/MS was performed to determine metabolite levels using MarkerLynx
25
26 86 Applications Manager Version 4.1 (Waters Corp., Milford, MA, USA).

28 87 **Distinct Metabolite Identification**

30
31 88 The peak intensities of metabolites for the 56 pairs of subjects were acquired and imported into
32
33 89 MetaboAnalyst 4.0 (<http://www.metaboanalyst.ca/>) for statistical analyses. A partial least-squares
34
35 90 discriminant analysis (PLS-DA), which is a supervised and well-accepted pattern recognition
36
37
38 91 approach, was used to differentiate between the cases and controls. The false discovery rate (FDR)
39
40 92 adjusted P-value in univariate analysis was determined to reduce the potential effect of
41
42 93 false-positive results. The criteria used to select metabolites included variable importance in
43
44 94 projection (VIP) scores >1 using PLS-DA and a crude or FDR-adjusted P-value < 0.05 using
45
46
47 95 Wilcoxon signed-rank test. We identified 70 extracted small molecular metabolites linked to the
48
49 96 recognition of AISL. The Human Metabolome Database (<http://www.hmdb.ca>) was used to identify
50
51
52 97 the metabolites, which include four amino acid metabolites (phenylalanine, tryptophan, leucine, and
53
54 98 phenylalanylphenylalanine).

56 99 **Statistical Analysis**

57
58 100 The normality of continuous data was assessed using both QQ-plots and Shapiro-Wilk test. The
59
60

1
2
3 101 data of cases and controls were analyzed using paired *t*-test if they showed a normal or similar
4
5
6 102 normal distribution. Otherwise, Wilcoxon signed-rank test was used. Differences in the proportion of
7
8 103 categorical variables between the two groups were evaluated using McNemar-Bowker test. We first
9
10 104 used locally weighted scatterplot smoothing (LOESS) models to estimate the “real” relationship
11
12 105 between serum AA levels and the probability of AISL. Next, multivariable conditional logistic
13
14 106 regression models were used to examine the association between contributing AA levels and AISL
15
16
17 107 after adjusting for potential confounding factors. The individual effects of AA metabolites on the risk
18
19 108 of AISL were quantified separately using the odds ratio (OR) and 95% confidence interval (CI) as
20
21
22 109 follows: with AAs as a categorical variable (quartiles) and as a continuous variable [scaled to an
23
24 110 interquartile range (IQR)]. Variables with a P-value less than 0.2 in the comparison between two
25
26 111 groups were selected as potential confounders. This approach has been widely used in many studies,
27
28
29 112 particularly for small sample size. The variance inflation factor (VIF) was used to examine the
30
31 113 potential collinearity among variables. As too many covariates in a multiple regression model can
32
33 114 lead to overfitting²⁴, we selected no more than four variables as confounding factors to decrease the
34
35
36 115 potential of overfitting when assessing the association between AAs and AISL. Furthermore, as
37
38 116 distinct metabolites may be highly related to each other, collinearity should be considered. Thus, we
39
40 117 used the VIF based on VIF package of R software to detect potential collinearity among the AAs. A
41
42 118 VIF greater than 1.5 is considered to indicate collinearity in the model, and the associated variable is
43
44
45 119 removed. The combined effect of relevant AAs on AISL was also determined using a multivariable
46
47 120 logistic regression model. A receiver operator characteristic (ROC) analysis was applied to evaluate
48
49 121 the value and feasibility of AAs as potentially sensitive and specific biomarkers for recognizing
50
51
52 122 AISL. Data management, analysis, and figure drawing were performed using R version 3.4.4
53
54 123 (Copyright © 2018 The R Foundation for Statistical Computing). All tests were two-sided and the
55
56 124 results with P values < 0.05 were considered to indicate significance.

58 125 **Patient and Public Involvement**

60

1
2
3 126 The present study was designed as an observational study, and as such patients and the public
4
5
6 127 were not involved in the planning, recruitment and conduct of this study. All participants were
7
8 128 informed about the purpose of this study and signed informed consent at the beginning of the study.
9
10 129 The results of this study have not yet been disseminated to the participants.
11
12

13 130 RESULTS

15
16 131 **Table 1** summarizes the general characteristics of the study population. Comparisons of the
17
18 132 demographic data, clinical features, and urinary arsenic species in the 56 pairs of subjects are
19
20 133 presented in Table 1. The median (first and third quartile) age of AISL population was 50.30 (44.70
21
22 and 58.70) for the cases and 50.40 (44.60 and 58.70) years for the controls. Both groups contained
23 134 the same proportion of females (58.93%), and there was no significant difference in urinary arsenic
24
25 135 levels between the two groups. More than half of the subjects had no history of smoking or alcohol
26
27 136 consumption. The serum triglyceride level in subjects with AISL was significantly lower than that in
28
29 137 the controls ($P = 0.041$). Other variables were similar between the subjects with AISL and controls
30
31 ($P > 0.05$). This indicates that the participants in the two groups were comparable.
32 138
33
34 139
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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^ξ			
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.

^ξ 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

1
2
3 148 Table S1. Significant linear trends were detected between AISL and the four serum AAs. The same
4
5
6 149 linear negative association between AISL and per IQR increase in the four serum AAs was observed
7
8 150 when these AAs were considered as continuous variables in the present study.
9

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

Amino acids	N	Cases (%)	Crude		Adjusted [§]	
			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

46 [§]Values are odds ratio (95% confidence intervals) for arsenic-induced skin lesions from conditional logistic regression. IQR:
47 interquartile range; Q₁: the 1st quartile; Q₂: the 2nd quartile; Q₃: the 3rd quartile; Q₄: the 4th quartile.

48 [§] Adjusted for plasma glucose, low-density lipoprotein, triglyceride, and urinary dimethyl arsenate.
49

50 151 As these four specific AAs were significantly or marginally significantly associated with the
51
52 152 odds of AISL, we examined their combined effects on AISL. However, analysis of potential
53
54
55 153 collinearity revealed that among the four specific AAs, both tryptophan and phenylalanine had the
56
57 154 smallest VIF value (VIF = 1.04), and no obvious collinearity existed (**Table S2**). Thus, we mainly
58
59 155 focused on tryptophan and phenylalanine to assess the combined effects of AAs on AISL and only
60

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 [‡]	Phenylalanine <cut-off value [‡]	N	Cases (%)	Crude		Adjusted [§]	
				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

[‡] Cut-off value was determined using receiver operator characteristic analysis.

[§] 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 β_j 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

1
2
3 197 first selected all participants using permuted block randomization from a single rural area in which
4
5
6 198 the population was chronically exposed to arsenic in the same route, had a similar lifestyle, and was
7
8 199 influenced by similar environmental factors. Secondly, the cases and controls were matched by
9
10 200 gender and age (± 1 year). All of these may be the reason for the so many potential confounders
11
12 201 including arsenic exposure, which did not differ significantly between the cases and controls.

14
15 202 Participants enrolled in the current study were chronically exposed to arsenic via drinking water.
16
17 203 The geometric mean (GM) and its related 95% CI of urinary iAs/creatinine and tAs/creatinine in this
18
19 204 population were 17.49 (14.90–20.53) $\mu\text{g/g}$ and 147.20 (129.00–167.97) $\mu\text{g/g}$, respectively. They
20
21
22 205 were much higher than those in the 20 $\mu\text{g/L}$ exposed to arsenic via drinking water [GM (95% CI):
23
24 206 0.4 (0.3–0.5) $\mu\text{g/g}$ for iAs and 9.1 (6.5–12.7) $\mu\text{g/g}$ for tAs], while obviously lower than those in the
25
26 207 90 $\mu\text{g/L}$ exposed group [GM (95% CI): 39.4 (31.4–49.6) $\mu\text{g/g}$ for iAs and 248.7 (208.8–296.3) $\mu\text{g/g}$
27
28
29 208 for tAs]²⁵. An available report has shown that AISL cannot be completely cured even though
30
31 209 medical technology has already made great progress²⁶. Therefore, it is crucial to identify those who
32
33 210 are most likely to progress to overt arsenic damages including AISL among people at risk as early as
34
35
36 211 possible. Metabolomics study, which mainly focuses on thoroughly assessing the variation in
37
38 212 metabolites possibly linked to disease occurrence and development, has been widely utilized to help
39
40 213 understand the pathogenesis of diseases because of its relevance to the phenotypes compared to other
41
42 214 “OMICS” study²⁷. Moreover, mathematical modeling to assess the linkage between small molecular
43
44
45 215 metabolites and arsenic toxicity has advanced²⁸. Developing a simple and interpretable modeling
46
47 216 approach for the early detection of arsenic-induced lesions is of great theoretical value and realistic
48
49 217 meaning²⁹, although it might be difficult due to population-specific complexities and effect of some
50
51
52 218 potential unmeasured covariates such as diet and genetic determinants.

53
54 219 Previous studies have reported that gene–gene and gene–environment interactions are involved
55
56 220 in arsenicosis through toxicological mechanisms including genomic instability³⁰ and oxidative
57
58
59 221 stress³¹. Skin hyperpigmentation and palmoplantar hyperkeratosis may be biomarkers for long-term
60

1
2
3 222 arsenic exposure and useful for identifying differences in metabolites associated with phenotypes.
4
5 223 Metabolite analysis may improve the understanding and identification of AISL. An animal study
6
7
8 224 revealed that examining the disruption in AA metabolism upon arsenic exposure in rats—may be
9
10 225 beneficial for understanding arsenic toxicity³². In our previous population-based metabolomics
11
12 226 study, we found that serum metabolite alterations were significantly related to the risk of
13
14
15 227 arsenic-induced health damages. In the present study, we found that some BCAAs or AAAs were
16
17 228 significantly related to AISL occurrence. Several studies across numerous ethnic backgrounds
18
19 229 support the use of BCAAs including leucine, isoleucine, valine, and AAA profiles such as those of
20
21
22 230 phenylalanine, tryptophan, and tyrosine as biomarkers for identifying metabolic diseases^{27 33}. Zhou et
23
24 231 al. reported that arsenic-induced transformed cells had altered metabolite profiles including
25
26 232 downregulation of leucine, tryptophan, and phenylalanine in the skin lesion group³⁴. Consistent with
27
28
29 233 these findings, the levels of two serum AAAs (tryptophan and phenylalanine) were significantly
30
31 234 associated with AISL in our study.

32
33 235 Normal metabolism of AAs is necessary for whole - body homeostasis, growth, and
34
35
36 236 development, and health¹⁵. Studies have reported that changes in the availability of AAAs affect cell
37
38 237 signaling, gene expression, brain function, and neuroendocrine function³⁵. Tryptophan, an AA
39
40 238 metabolism - related biomarker, is a sensitive and specific indicator of oxidation. Tryptophan
41
42
43 239 metabolism in mammals is a physiological means of preserving immune homeostasis associated with
44
45 240 oxidative stress and inflammation^{36 37}. Additionally, phenylalanine can be transformed into specific
46
47
48 241 neurotransmitters such as dopamine and adrenaline via the action of related enzymes. Wu et al.³⁸
49
50 242 reported that arsenic exposure can lead to neurotransmitter metabolism, which may explain the
51
52 243 reduction in phenylalanine. Furthermore, as a peptide-bound phenylalanine,
53
54 244 phenylalanylphenylalanine has been reported to affect protein synthesis and secretion³⁹, indicating a
55
56
57 245 relationship between endothelium dysfunction and phenylalanine metabolism disorder. The
58
59 246 relationship between AA metabolism and AISL remains unclear. The notable alteration of tryptophan
60

1
2
3 247 and phenylalanine in this study indicates the occurrence of metabolic disorders due to arsenic
4
5
6 248 exposure. These results are also beneficial for understanding the effects of arsenic toxicity and
7
8 249 importance of early identification of exposure for delaying the progression of various
9
10 250 arsenic-induced lesions, including AISL.

11
12 251 The main strength of this study is that the findings were based on a community-based,
13
14
15 252 long-term arsenic exposure cohort with well-designed quality assurance and quality control
16
17 253 throughout the study. However, there were some limitations to this study. Firstly, although the
18
19 254 untargeted metabolomics approach can assess a large amount number of metabolites precisely and
20
21
22 255 efficiently, it only provided relative levels of AAs instead of their accurate quantitative concentration.
23
24 256 Secondly, these findings were mainly based on a case-control study, which only reveals the
25
26 257 association between AA metabolism and the odds of AISL rather than confirming their causal
27
28
29 258 relationship. Furthermore, the study included 56 AISL cases-matched and 56 non-AISL controls and
30
31 259 the sample size might be potentially insufficient. Although metabolomics studies usually have a
32
33 260 sample size of no more than 40 cases in each group^{40 41}, our study may be under-powered and thus
34
35 261 larger studies are needed in the future. Besides, as it is suggested that the ratio of approximately
36
37
38 262 10-15 observations per predictor in a logistic regression model will produce reasonably stable
39
40 263 estimations²⁴, we selected only four covariates in the models due to the small sample size and these
41
42 264 results need to be confirmed in further studies. Finally, the participants were mainly exposed to
43
44
45 265 arsenic via drinking water, which would limit the findings from being extrapolated to other arsenic
46
47 266 exposure populations via food and other routes. Therefore, additional elaborate population-based
48
49 267 studies are needed to verify our findings.

50
51 268 In conclusion, specific AAs may be linked to AISL and AA metabolism may play an important
52
53
54 269 role in early AISL identification. Additional studies are needed to confirm our findings.
55
56
57
58
59
60

1
2
3
4 **Contributors:** Guangyun Mao and Yaping Wei designed the study. Chaonan Jia participated in
5 collecting data. Yuan Lan and Chaonan Jia audited the data. Yaping Wei, Xiangqing hou, Jushuang
6 Li, Tao Wang conducted the literature search, Yaping Wei, Chaonan Jia conducted statistical
7 analysis and interpreted the results. Yaping Wei and Chaonan Jia wrote the first draft of the
8 manuscript. Jingjing Zuo helped with copyediting. Guangyun Mao reviewed the final manuscript and
9 did substantial contributions.
10
11

12 **Funding:** This work was supported by the basic public welfare research project of Zhejiang province
13 (LGF19H260011), the initial scientific research fund (KYQD170301) and the major project of the
14 eye hospital of Wenzhou medical university (YNZD201602).
15

16 **Acknowledgments:** We gratefully acknowledge all of the study participants of this project. We wish
17 to thank all of the dedication and hard work of the field team at the department of preventive
18 medicine, school of public health and management, Wenzhou medical university. Our thanks must
19 also go to Dr. Kendrick Hii, a native-English speaker, for his valuable suggestions and edits to the
20 manuscript.
21
22

23 **Competing Interest:** None declared.
24

25 **Ethics approval:** Ethical approval was received from the ethics committee of Wenzhou Medical
26 University, Wenzhou, China.
27

28 **Provenance and peer review:** Not commissioned; externally peer reviewed.
29

30 **Data Sharing Statement:** No additional unpublished data are available.
31
32

33 **Patient consent for publication:** Not required.
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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

- 1
2
3 regression analysis. *J Clin Epidemiol*. 1996; 49:1373–9.
- 4 25. Sun G, Xu Y, Li X, et al. Urinary arsenic metabolites in children and adults exposed to arsenic in drinking
5 water in Inner Mongolia, China. *Environmental Health Perspectives*. 2007; 115:648–52.
- 6 26. Kuo CC, Moon KA, Wang SL, et al. The Association of Arsenic Metabolism with Cancer, Cardiovascular
7 Disease, and Diabetes: A Systematic Review of the Epidemiological Evidence. *Environmental Health*
8 *Perspectives*. 2017; 125:15.
- 9 27. Wang TJ, Larson MG, Vasan RS, et al. Metabolite profiles and the risk of developing diabetes. *Nature*
10 *Medicine*. 2011; 17:448–U83.
- 11 28. Lawley SD, Cinderella M, Hall MN, et al. Mathematical model insights into arsenic detoxification. *Theor Biol*
12 *Med Model*. 2011; 8.
- 13 29. Juhasz AL, Herde P, Herde C, et al. Validation of the predictive capabilities of the Sbrc–G in vitro assay for
14 estimating arsenic relative bioavailability in contaminated soils. *Environ Sci Technol*. 2014; 48:12962–9.
- 15 30. Kibriya MG, Jasmine F, Parvez F, et al. Association between genome–wide copy number variation and
16 arsenic–induced skin lesions: a prospective study. *Environ Health*. 2017; 16:13.
- 17 31. Priori R, Scrivo R, Brandt J, et al. Metabolomics in rheumatic diseases: the potential of an emerging
18 methodology for improved patient diagnosis, prognosis, and treatment efficacy. *Autoimmun Rev*. 2013;
19 12:1022–30.
- 20 32. Wang X, Mu X, Zhang J, et al. Serum metabolomics reveals that arsenic exposure disrupted lipid and amino
21 acid metabolism in rats: a step forward in understanding chronic arsenic toxicity. *Metallomics Integrated*
22 *Biometal Science*. 2015; 7:544.
- 23 33. Weng L, Quinlivan E, Gong Y, et al. Association of Branched and Aromatic Amino Acids Levels with
24 Metabolic Syndrome and Impaired Fasting Glucose in Hypertensive Patients. *Metabolic Syndrome And*
25 *Related Disorders*. 2015; 13:195–202.
- 26 34. Zhou YY, Wang YF, Su J, et al. Integration of microRNAome, proteomics and metabolomics to analyze
27 arsenic–induced malignant cell transformation. *Oncotarget*. 2017; 8:90879–96.
- 28 35. Lee JC, Chen MJ, Chang CH, et al. Plasma amino acid levels in patients with colorectal cancers and liver
29 cirrhosis with hepatocellular carcinoma. *Hepato–Gastroenterol*. 2003; 50:1269–73.
- 30 36. Munn DH, Mellor AL. Indoleamine 2,3 dioxygenase and metabolic control of immune responses. *Trends In*
31 *Immunology*. 2013; 34:137–43.
- 32 37. Pawlak K, Domaniewski T, Mysliwiec M, et al. The kynurenines are associated with oxidative stress,
33 inflammation and the prevalence of cardiovascular disease in patients with end–stage renal disease.
34 *Atherosclerosis*. 2009; 204:309–14.
- 35 38. Wu L–L, Gong W, Shen S–P, et al. Multiple metal exposures and their correlation with monoamine
36 neurotransmitter metabolism in Chinese electroplating workers. *Chemosphere*. 2017; 182:745–52.
- 37 39. Zhou MM, Wu YM, Liu HY, et al. Effects of phenylalanine and threonine oligopeptides on milk protein
38 synthesis in cultured bovine mammary epithelial cells. *J Anim Physiol Anim Nutr*. 2015; 99:215–20.
- 39 40. Chen LY, Cheng CY, Choi HW, et al. Plasma Metabonomic Profiling of Diabetic Retinopathy. *Diabetes*.
40 2016; 65:1099–108.
- 41 41. Piszcz J, Lemancewicz D, Dudzik D, et al. Differences and similarities between LC–MS derived serum
42 fingerprints of patients with B–cell malignancies. *Electrophoresis*. 2013; 34:2857–64.
- 43
44
45
46
47
48
49
50
51
52
53

54 **Figure 1.** Association between the peak intensity of tryptophan and phenylalanine and
55 arsenic–induced skin lesions based on multivariable locally weighted regression models. a:
56 Tryptophan; b: Phenylalanine; c: Leucine; d: Phenylalanylphenylalanine
57
58
59
60

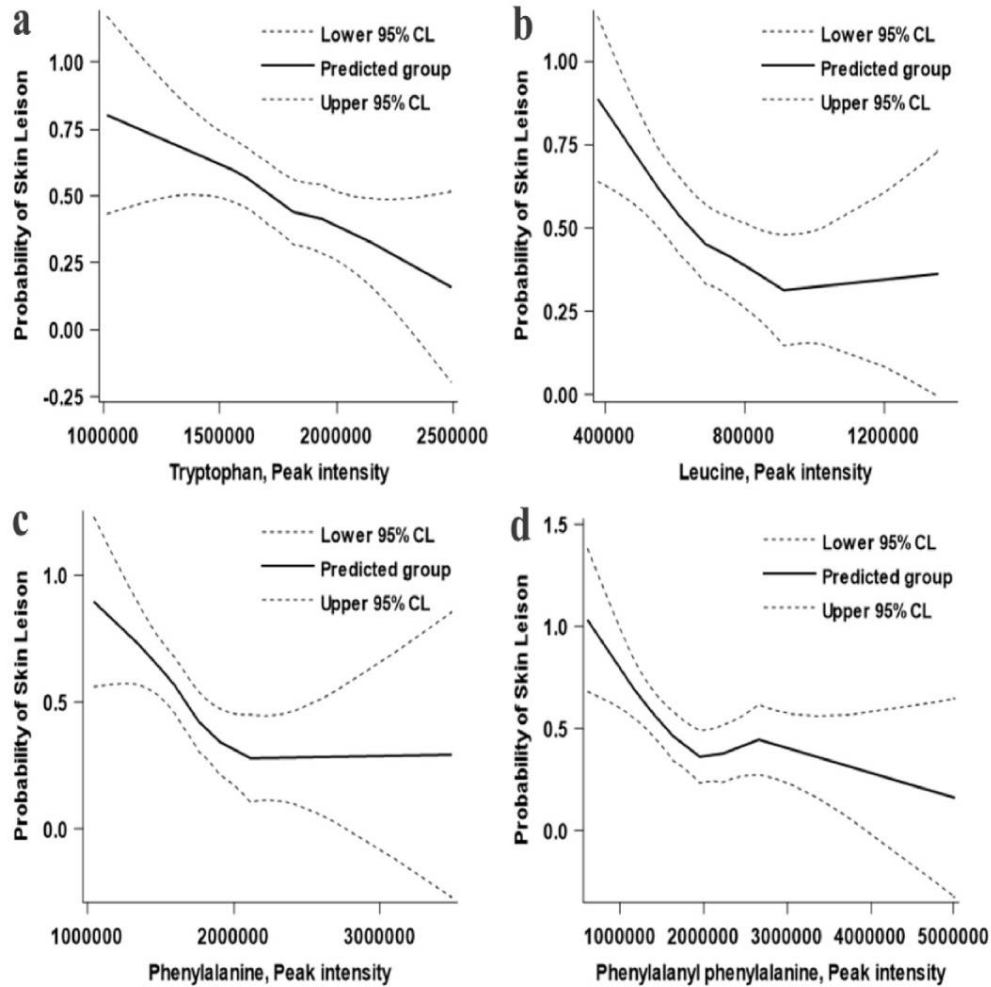


Figure 1. Association between the peak intensity of tryptophan and phenylalanine and arsenic-induced skin lesions based on multivariable locally weighted regression models. a: Tryptophan; b: Phenylalanine; c: Leucine; d: Phenylalanylphenylalanine

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[§].

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	

[§]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.