

Assessing the impact of arsenic metabolism efficiency on DNA methylation using Mendelian randomization

Anthony DiGiovanni^{a,*}, Kathryn Demanelis^{a,*}, Lin Tong^a, Maria Argos^b, Justin Shinkle^a, Farzana Jasmine^a, Mekala Sabarinathan^a, Muhammad Rakibuz-Zaman^c, Golam Sarwar^c, Md. Tariqul Islam^c, Hasan Shahriar^c, Tariqul Islam^c, Mahfuzar Rahman^{c,d}, Md. Yunus^e, Joseph Graziano^f, Mary V. Gamble^f, Habibul Ahsan^{a,g,h,i}, Brandon L. Pierce^{a,g,h}

Background: Arsenic exposure affects >100 million people globally and increases risk for chronic diseases. One possible toxicity mechanism is epigenetic modification. Previous epigenome-wide association studies (EWAS) have identified associations between arsenic exposure and CpG-specific DNA methylation. To provide additional evidence that observed associations represent causal relationships, we examine the association between genetic determinants of arsenic metabolism efficiency (percent dimethylarsinic acid, DMA%, in urine) and DNA methylation among individuals from the Health Effects of Arsenic Longitudinal Study (n = 379) and Bangladesh Vitamin E and Selenium Trial (n = 393).

Methods: We used multivariate linear models to assess the association of methylation at 221 arsenic-associated CpGs with DMA% and measures of genetically predicted DMA% derived from three SNPs (rs9527, rs11191527, and rs61735836). We also conducted two-sample Mendelian randomization analyses to estimate the association between arsenic metabolism efficiency and CpG methylation.

Results: Among the associations between DMA% and methylation at each of 221 CpGs, 64% were directionally consistent with associations observed between arsenic exposure and the 221 CpGs from a prior EWAS. Similarly, among the associations between genetically predicted DMA% and each CpG, 62% were directionally consistent with the prior EWAS results. Two-sample Mendelian randomization analyses produced similar conclusions.

Conclusion: Our findings support the hypothesis that arsenic exposure effects DNA methylation at specific CpGs in whole blood. Our novel approach for assessing the impact of arsenic exposure on DNA methylation requires larger samples in order to draw more robust conclusions for specific CpG sites.

Keywords: Arsenic; Bangladesh; Epigenomics; DNA methylation; Mendelian randomization

^aDepartment of Public Health Sciences, The University of Chicago, Chicago, Illinois; ^bDivision of Epidemiology and Biostatistics, University of Illinois at Chicago, Chicago, Illinois; ^cUChicago Research Bangladesh, Mohakhali, Dhaka, Bangladesh; ^dResearch and Evaluation Division, BRAC, Dhaka, Bangladesh; ^eInternational Centre for Diarrhoeal Disease Research, Bangladesh, Dhaka, Bangladesh; ^fDepartment of Environmental Health Sciences, Mailman School of Public Health, Columbia University, New York; ^gDepartment of Human Genetics, and ^hComprehensive Cancer Center, The University of Chicago, Chicago, Illinois; and ⁱDepartment of Medicine, The University of Chicago, Chicago, Illinois.

*A.D. and K.D. are co-first authors.

This work was supported by the Department of Public Health Sciences Summer Undergraduate Fellowship and College Research Fellows Program at the University of Chicago and by grants P30CA014599 from the National Cancer Institute, P30ES027792 from the National Institute of Environmental Health Sciences, T32AG000243 from the National Institute on Aging, R01ES020506 from the National Institute of Environmental Health Sciences, R35ES028379 from the National Institute of Environmental Health Sciences, P42ES010349 from the National Institute of Environmental Health Sciences, R01CA102484 from the National Cancer Institute, and R01CA107431 from the National Cancer Institute.

SDC Supplemental digital content is available through direct URL citations in the HTML and PDF versions of this article (www.enviroepidem.com).

Scripts used for the analyses in this article are available upon request, and our results can be replicated according to the Methods section, using R version 3.5.2 with open-source packages. Individual-level genetic data cannot currently be made publicly available.

Copyright © 2020 The Authors. Published by Wolters Kluwer Health, Inc. on behalf of The Environment Epidemiology. All rights reserved. This is an open access article distributed under the Creative Commons Attribution License 4.0 (CCBY), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Environmental Epidemiology (2020) 4:e083

Received: 17 June 2019; Accepted 10 January 2020

Published online 20 March 2020

DOI: 10.1097/EE9.000000000000083

*Corresponding Author. Address: The University of Chicago, 5841 South Maryland Avenue, W264, MC2000 Chicago, IL 60637. E-mail: brandonpierce@uchicago.edu (B.L. Pierce).

Introduction

Exposure to arsenic from consumption of naturally contaminated drinking water impacts >100 million people globally, including ~50 million in Bangladesh.¹ Arsenic is a toxic and carcinogenic metal, and consumption of drinking water with arsenic concentrations above 50–100 µg/L is associated with risk for several cancer types in multiple populations.^{2,3} Arsenic exposure is also associated with increased risk of cardiovascular diseases,^{4,5} respiratory disease,⁶ diabetes,⁷ nonmalignant lung disease,⁸ and increased overall mortality.⁹ Chronically exposed

What this study adds

This article provides preliminary evidence supporting the hypothesis that many previously reported associations of arsenic exposure with DNA methylation at specific CpG sites represent causal relationships. Using a Mendelian randomization approach, we demonstrate that the directions of these previously reported associations are generally consistent with associations we estimate between genetically predicted arsenic metabolism efficiency and these arsenic-associated CpGs. Our work is unique because we apply Mendelian randomization to provide evidence supporting the hypothesis that findings from an epigenome-wide association study of an environmental exposure represent causal relationships, using genetic instruments that have well-established associations with arsenic metabolism efficiency.

individuals maintain high risks of arsenic-associated diseases and mortality for several decades.¹⁰ Thus, an understanding of arsenic toxicity mechanisms is needed to better assess risk and develop effective prevention and treatment strategies for arsenic-associated diseases and health effects.

The mechanisms of arsenic toxicity are complex and remain to be fully elucidated. Prior studies suggest that arsenic alters DNA methylation, and arsenic-associated DNA methylation alterations may be an important mechanism of arsenic toxicity.¹¹ DNA methylation is an epigenetic modification that primarily occurs at cytosine nucleotides within CpG sites (CpGs) and reflects chromatin conformation and regulation of gene expression. Prenatal arsenic exposure is associated with CpG methylation in cord blood according to numerous studies,^{12–16} and differential methylation in blood has been found in adult Bangladeshi, American, and Argentinian populations with a wide range of exposure levels (measured in urine, from ~ 10 µg/g creatinine to >1600 µg/g creatinine).^{17–20} A recent epigenome-wide association study (EWAS) of Bangladeshi adults identified associations between urinary arsenic and methylation at >200 CpGs in whole blood.²¹ These CpGs are enriched in genes belonging to cancer, aging, inflammation, and toxicant response pathways.²¹ While a mechanism of arsenic-induced carcinogenesis involving DNA methylation changes is plausible, we do not know that associations observed in EWAS represent causal effects of arsenic on the epigenome.

Arsenic undergoes biotransformation by arsenite methyltransferase (*AS3MT*), and arsenic metabolism efficiency modulates internal dose of arsenic by facilitating excretion of arsenic in urine. *AS3MT* converts inorganic arsenic (iAs) via methylation reactions into monomethylarsonic acid (MMA) and then dimethylarsinic acid (DMA).^{22,23} The proportion of DMA among all arsenic species in urine (DMA%) is a measure of arsenic metabolism efficiency.²⁴ These methylated species, especially DMA, are more readily excreted in urine than iAs, and higher DMA relative to MMA in urine is associated with lower arsenic toxicity risk.^{25–28} Therefore, arsenic metabolism efficiency reflects variability in the internal arsenic dose for individuals at a given level of exposure. Because DMA% has known genetic determinants, it is possible to use a Mendelian randomization (MR) approach to estimate associations between DMA% and methylation at specific CpGs that are not biased due to confounding by environmental or lifestyle factors. Genetic determinants of arsenic metabolism can be utilized as instrumental variables (IVs); and given that these IVs (1) affect arsenic metabolism and (2) are associated with CpG methylation exclusively through their effect on arsenic metabolism, MR provides

accurate effect estimates.²⁹ If these conditions are satisfied, MR estimates avoid biases that may affect associations with directly measured arsenic metabolism, such as confounding or reverse causation. Genome-wide association (GWA) and candidate gene studies have identified independent associations of variants in the 10q24.32/*AS3MT* region, rs9527 and rs11191527, with DMA% among arsenic-exposed Bangladeshi individuals.^{24,30,31} DMA% is also associated with rs61735836 in exon 3 of *FTCD* (Formiminotransferase cyclodeaminase), which has a catalytic role in the one-carbon/folate cycle, a source of methyl groups for arsenic methylation.³²

In this study, we use an MR approach to evaluate the relationship between arsenic metabolism efficiency and methylation of arsenic-associated CpGs identified in prior EWAS.²¹ Efficient metabolism of arsenic (i.e., high DMA%) should reduce the internal dose of arsenic; thus, we hypothesize that DMA% will be associated with the CpGs discovered in EWAS in a direction opposite to that of arsenic exposure. Using data on 379 Bangladeshi participants in the Health Effects of Arsenic Longitudinal Study (HEALS) and 393 participants in the Bangladesh Vitamin E and Selenium Trial (BEST),^{33,34} we obtain MR-based estimates (Figure 1) of the association between arsenic metabolism efficiency and each arsenic-associated CpG, using genetic determinants of DMA% as IVs.³⁵ Our approach represents a novel strategy to examine the toxicological relevance of molecular features showing association with environmental exposures, such as DNA methylation, in observational studies.

Methods

Study participants

Subjects analyzed in this article were participants in one of the two following studies: the HEALS and the BEST. DNA samples were obtained at baseline from participants in both cohorts. HEALS (described previously in detail) is a prospective longitudinal study of health outcomes associated with chronic arsenic exposure in Araihaazar, Bangladesh, with 11,746 adults (11,224 with arsenic measurements) enrolled at the original baseline visit (age 18–75 years).³³ Trained study physicians (blinded to arsenic measurements) conducted in-person interviews, clinical evaluations, and collection of urine and blood samples using structured protocols.³⁶ BEST is a 2 × 2 factorial randomized chemoprevention trial assessing the effects of vitamin E and selenium dietary supplements on skin cancer risk among Bangladeshi individuals (n = 7,000, age 25–65 years) with arsenical skin lesions.³⁷ Many of the study protocols in BEST, including sample collection and

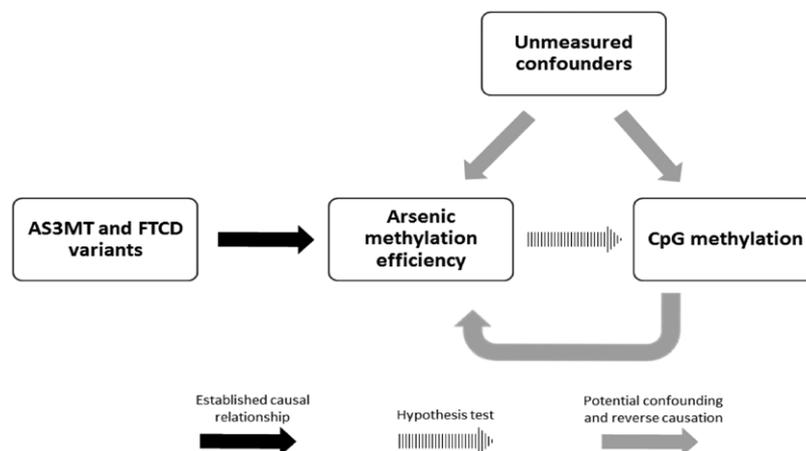


Figure 1. Causal diagram depicting the relationships among 10q24.32 and *FTCD* genetic variants, arsenic methylation efficiency, and methylation levels of CpG sites in an arsenic-exposed population. The causal relationship between the 10q24.32 and *FTCD* variants and arsenic methylation permits the use of these variants as potential IVs in our MR analysis.

exposure assessment, were identical to those in HEALS.²⁴ This article uses DNA methylation data from a recent EWAS of 396 randomly selected adults from HEALS, and from a prior study of 400 BEST participants.^{19,21} Study protocols were approved by the Institutional Review Boards of The University of Chicago and Columbia University, the Ethical Review Committee of the International Center for Diarrheal Disease Research, Bangladesh, and the Bangladesh Medical Research Council. Informed consent was obtained from all participants.

Exposure assessment

Arsenic was measured in urine at baseline for HEALS and BEST participants and in drinking water for HEALS participants. At baseline, each HEALS participant identified the well used as their primary source of drinking water. Urinary and water arsenic were measured using graphite furnace atomic absorption spectrometry in a single laboratory.³⁸ Total urinary arsenic concentration was divided by creatinine to compute creatinine-adjusted arsenic (micrograms/gram creatinine).³⁹ In HEALS participants, urinary arsenic metabolites were distinguished via high-performance liquid chromatography, followed by detection using inductively coupled plasma-mass spectrometry with dynamic reaction cell.²⁶ The percentages of iAs, MMA, and DMA among all arsenic species were calculated after subtracting arsenobetaine and arsenocholine (nontoxic organic arsenic from dietary sources) from total arsenic.

Genotyping

DNA extraction and quality control for HEALS and BEST blood samples have been described in the respective studies in which the 10q24.32 and *FTCD* variant associations with

arsenic metabolism efficiency were discovered.^{24,32} Genotypes for the 10q24.32 and *FTCD* variants were obtained from Illumina HumanCytoSNP-12 v2.1 chips and Illumina's exome array v1.1, respectively. Genotypes for rs9527 and rs11191527 were available for 379 individuals in HEALS and 393 in BEST, and genotypes for rs61735836 were available for 340 individuals in HEALS and 383 in BEST (Table 1).

DNA methylation

DNA methylation was measured on the Illumina EPIC array for 396 HEALS participants and 450K array for 400 BEST participants.^{21,37} Preprocessing and normalization of these arrays has been described previously.²¹ Briefly, beta mixture quantile (BMIQ) normalization was applied to reduce type I/II probe bias, and an EWAS of log₂-transformed urinary arsenic was conducted in both HEALS and BEST using models adjusted for age, sex, smoking status, BMI, and surrogate variables.⁴⁰ The METAL software was used to conduct a meta-analysis using inputted summary statistics from 390,810 CpGs measured in both HEALS and BEST.⁴¹ Analyses of the EPIC array data from HEALS participants identified associations between log₂-transformed urinary arsenic and 34 CpGs at FDR = 0.05 (eTable 1; <http://links.lww.com/EE/A77>). From the meta-analysis of both HEALS and BEST, we identified 221 CpGs associated with log₂-transformed urinary arsenic at FDR = 0.05 (eTable 2; <http://links.lww.com/EE/A77>).

Estimation of GP-DMA%

We generated a weighted SNP score representing genetically predicted DMA% (GP-DMA%) as a measure of arsenic metabolism efficiency. β coefficients for associations of rs9527, rs11191527,

Table 1. Descriptive statistics (median [25th,75th percentiles] or n [%]) for HEALS and BEST cohorts and subcohorts used for this analysis

	HEALS		BEST	
	Cohort	Subcohort	Cohort	Subcohort
N	11,224	379	7,000	393
Age (years)	36 (29, 45)	37 (30, 43)	41 (33, 50)	44 (35, 51)
Water arsenic ($\mu\text{g/L}$)	60 (12, 147)	51 (11, 123.5)		
Urinary arsenic ($\mu\text{g/g creatinine}$)	199 (106, 352)	200 (109.5, 344)	484 (145, 1,066)	137 (76, 395)
DMA%	72.1 (66.3, 77.2)	72.4 (66.8, 77.6)		
Weighted GP-DMA% ^a		17.4 (16.0, 17.4)		17.4 (14.5, 17.4)
Sex				
	Male	4,855 (43.3%)	160 (42.2%)	2,840 (40.6%)
	Female	6,369 (56.7%)	219 (57.8%)	4,160 (59.4%)
Cigarette smoking				
	Never	7,204 (64.2%)	230 (60.7%)	5,086 (72.7%)
	Former	743 (6.6%)	21 (5.5%)	495 (7.1%)
	Current	3,271 (29.2%)	128 (33.8%)	1,419 (20.3%)
BMI				
	<18.5	4,421 (39.7%)	149 (39.3%)	2,894 (41.3%)
	18.5–25	5,931 (53.2%)	197 (52.0%)	3,677 (52.5%)
	>25	791 (7.1%)	33 (8.7%)	429 (6.1%)
Years of education				
	0	4,944 (44.1%)	168 (44.3%)	2,923 (41.8%)
	1–5	3,335 (29.7%)	123 (32.5%)	1,972 (28.2%)
	>5	2,587 (26.2%)	88 (23.2%)	2,105 (30.1%)
Genotypes				
	rs9527		379 (100%)	393 (100%)
	rs11191527		379 (100%)	393 (100%)
	rs61735836		340 (89.7%)	383 (97.5%)
Weighted GP-DMA% ^a		17.4 (16.0, 17.4)		17.4 (14.5, 17.4)
Binary SNP Score ^b				
	0		135 (35.6%)	138 (35.1%)
	1		244 (64.4%)	255 (64.9%)

^aSum of effect allele counts weighted by strength of associations (see Methods).

^bIndicator variable that equals 1 when individuals are homozygotes for the high efficiency allele at both rs9527 and rs61735836, otherwise zero (see Methods).

and rs61735836 with DMA% were obtained from prior GWA study in HEALS.^{24,32} The β coefficient for the association of each SNP with DMA% (Table 2) was multiplied by its high-efficiency allele (effect allele) count, and the score was computed as the sum of these products. GP-DMA% captured the increase in DMA% relative to an arbitrary baseline DMA%. Let X_j be the high-efficiency allele count of each j th SNP, and β_{X_j} be the effect size of the association between SNP j and DMA%. Then:

$$\text{GP-DMA\%} = \sum_j \beta_{X_j} X_j.$$

In a supplementary analysis, we defined a binary score indicating whether each individual carried only high-efficiency alleles at both rs9527 and rs61735836 (or if they carried at least one low-efficiency allele at one or both of these SNPs). This strategy avoided specifying weights, rather relying on the two SNPs with (1) large effect sizes (compared with rs11191527) and (2) low frequency of the low-efficiency allele (Table 2).

Statistical methods

Associations of CpG methylation with DMA%, 10q24.32 and *FTCD* SNPs, and GP-DMA% were estimated by linear regression. Because of small sample sizes of minor allele homozygotes among our SNPs (at most 2.3% of the sample), these individuals were combined with heterozygotes and classified as minor allele carriers for analyses of single SNPs. Each SNP genotype was coded as a binary variable for higher effect allele count. Because of moderate LD between rs9527 and rs11191527, these SNPs were included together as covariates in individual SNP analyses. All regressions included age, sex, methylation batch, and cigarette smoking status (never, former, and current smoker) as covariates. Analyses in which HEALS and BEST were combined were adjusted for cohort (HEALS or BEST). Regressions with DMA% as a predictor (restricted to HEALS) also included BMI, years of education, and log-transformed water arsenic as covariates, to adjust for potential confounders of this non-genetic variable. As an additional adjustment for confounding due to arsenic exposure levels, associations between log-transformed DMA and CpG methylation were adjusted for log-transformed urinary arsenic and log-transformed urinary creatinine in addition to the same covariates as in DMA% regressions.

Due to low power for testing the association of SNPs (or DMA%) with any specific CpG site, we tested the hypothesis that the set of arsenic-associated CpGs tended to associate with DMA% (and with genotypes that impact DMA%) in directions consistent with the CpGs' associations with arsenic. We applied binomial tests to binary variables representing the directional consistency of association for each CpG site. For each CpG, the estimate of association with DMA% or GP-DMA% (or the MR estimate) was considered consistent with prior EWAS if the β coefficient had a direction opposite to that of the previously

reported association between arsenic exposure and CpG methylation. In other words, we hypothesized that if arsenic metabolism efficiency impacts DNA methylation, increased DMA% (or GP-DMA%) would lower the internal arsenic dose, thus the direction of association for each arsenic-associated CpG site would be opposite to that of arsenic exposure (Figure 2). Under the null hypothesis, the number of CpGs showing a direction of association consistent with the prior EWAS is equal (or less than) to the number of CpGs with a direction of association that is inconsistent with the prior EWAS. We report one-sided P -values (from a binomial test) that correspond to the alternative hypothesis the number of CpG showing an association directionally consistent with prior EWAS is greater than the number of CpGs showing a directionally inconsistent association.

Mendelian randomization

MR estimates were computed using the “Mendelian randomization” R package,⁴² using rs9527, rs11191527, and rs61735836 as IVs in the inverse-variance weighted (IVW) MR method. Maximum likelihood MR was also conducted as a sensitivity analysis. Effect sizes (β_{X_j}) and standard errors ($se(\beta_{X_j})$) for associations between each j th SNP and DMA% were derived from prior publications from HEALS (Table 2).^{24,32} For SNP ~ CpG associations, effect estimates (β_{Y_j}) and standard errors ($se(\beta_{Y_j})$) were obtained from the regression analyses conducted in this study.

Results

Participant characteristics

Genotyping and methylation data were available for 379 HEALS participants and 393 BEST participants. The median age in HEALS was 37 (IQR: 30, 43) and 44 years in BEST (IQR: 35, 51), with 57.8% women in HEALS and 46.8% women in BEST (Table 1). In HEALS, 33.8% were current smokers and 5.5% former smokers; and in BEST, 27.2% were current smokers and 10.2% former. The median urinary arsenic (adjusted for creatinine) was 200 $\mu\text{g/g}$ (IQR: 109.5, 344) in HEALS and 137 $\mu\text{g/g}$ (IQR: 76, 395) in BEST. Water arsenic and DMA% data are not available for BEST participants.

Associations between CpG methylation and arsenic metabolism

Median DMA% in the HEALS participants was 72.4% (IQR: 66.8%, 77.6%) (Table 1). The correlation between DMA% and log-transformed creatinine-adjusted urinary arsenic was -0.16 ($n = 383$, $P = 1.3 \times 10^{-3}$), and between log-transformed DMA and log-transformed creatinine-adjusted urinary arsenic was 0.52 ($n = 383$, $P = 1.8 \times 10^{-27}$) (eFigure 1; <http://links.lww.com/EE/A77>).

After adjusting for covariates (age, sex, methylation batch, cigarette smoking status, BMI, and education) and log-transformed water arsenic, the directions of DMA% associations were consistent with arsenic exposure associations from prior EWAS (i.e., opposite in sign; see Methods) for 142 out of 221 CpGs (binomial $P = 1.4 \times 10^{-5}$) (Table 3; Figure 3A). β coefficients for log-transformed DMA had directions consistent with arsenic exposure associations in 146 CpGs ($P = 1.0 \times 10^{-6}$) (eTable 3; <http://links.lww.com/EE/A77>, eFigure 2; <http://links.lww.com/EE/A77>). This pattern confirms the expectation that increasing arsenic exposure and lower DMA% (i.e., higher internal dose of arsenic due to lower metabolism efficiency) should have the same directions of associations with CpGs that are influenced by arsenic exposure (Figure 2).

Table 2.

SNPs used as IVs for arsenic metabolism efficiency were selected based on their association with DMA% in prior GWA studies

SNP	Chromosome	Position	Frequency of	
			DMA%-increasing allele	β^a (SE)
rs9527 ^b	10	104613568	0.92	3.60 (0.47)
rs11191527 ^c	10	104785124	0.16	2.22 (0.34)
rs61735836 ^b	21	47572637	0.93	5.09 (0.51)

^a β represents the percentage point change in DMA% per one-allele increase in effect allele count, derived from a prior GWA study that identified the association of these three SNPs with DMA%.^{24,30,32}

^bEffect allele is major (G).

^cEffect allele is minor (A).

T3,F3

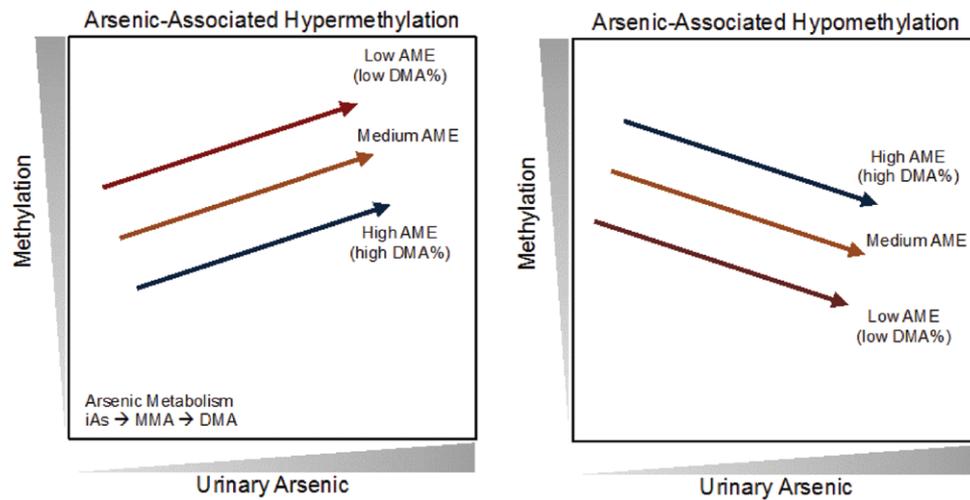


Figure 2. Hypothesized relationships among arsenic metabolism efficiency (AME), urinary arsenic, and CpG methylation. For CpGs that show a positive association with urinary arsenic (hypermethylated CpGs, left panel), higher AME should reduce the internal dose of arsenic and thus show a negative association with CpG methylation. For CpGs negatively associated with urinary arsenic (hypomethylated CpGs, right panel), higher AME should show a positive association with CpG methylation.

Associations between CpG methylation and genetically predicted DMA%

Using genotype data for 340 individuals from HEALS and 383 from BEST, we analyzed associations between CpG methylation and genetically predicted DMA% (GP-DMA%) in this combined cohort. Stratified results for HEALS and BEST are available in eTable 4; <http://links.lww.com/EE/A77>. Median weighted GP-DMA% in the HEALS participants was 17.4 (IQR: 16.0, 17.4) and 17.4 in BEST participants (IQR: 14.5, 17.4) (Table 1), relative to an individual carrying zero high-efficiency alleles. After adjusting for sex, age, methylation batch, and cigarette smoking status, weighted GP-DMA% associations with CpGs were consistent with arsenic exposure associations in 137 of the 221 arsenic-associated CpGs ($P = 0.0002$) (Table 3; Figure 3B). Among 41 CpGs that were associated with arsenic exposure based on a Bonferroni P -value threshold ($P < 1.3 \times 10^{-7}$), 29

were associated with weighted GP-DMA% in a consistent direction ($P = 0.006$) (eTable 5; <http://links.lww.com/EE/A77>). As a sensitivity analysis, we also tested a binary SNP score in 379 individuals from HEALS and 393 from BEST (see Methods). Binary SNP score and CpG associations were consistent with arsenic exposure associations in 148 CpGs ($P = 2.5 \times 10^{-7}$), and 32 CpGs of Bonferroni P -value threshold in EWAS ($P = 0.0002$) (eTable 3; <http://links.lww.com/EE/A77>, eTable 5; <http://links.lww.com/EE/A77>, eFigure 3; <http://links.lww.com/EE/A77>).

Mendelian randomization

We used a two-sample MR approach based on summary statistics to obtain an MR-based estimate of the association between DMA% and CpG methylation using the 10q24.32 and *FTCD* variants as IVs for DMA%, in the combined HEALS and BEST cohort ($n = 772$). IVW-MR association estimates for 134 out of 221 CpGs were consistent in direction with arsenic exposure associations (binomial $P = 0.0010$) (Table 3; Figure 3C). When we used the maximum-likelihood MR method as a sensitivity analysis, the effect estimates for 134 CpGs were consistent in direction with arsenic associations ($P = 0.0010$) (eTable 3; <http://links.lww.com/EE/A77>). Plots for CpGs with the strongest MR effect estimates are included in the Supplement (eFigure 4; <http://links.lww.com/EE/A77>). When analyses were restricted to 41 CpGs whose associations passed a Bonferroni P -value threshold in EWAS, 27 were consistent ($P = 0.03$) (eTable 5; <http://links.lww.com/EE/A77>). To assess the sensitivity of our MR results to the method of analysis via summary statistics, we also tested associations of CpG methylation levels with individual effect allele counts of 10q24.32 and *FTCD* SNPs (see Methods). Among 221 CpGs, associations between allele scores and CpG methylation were consistent with arsenic exposure associations in 132 CpGs for rs9527 ($P = 0.0023$), 120 for rs11191527 ($P = 0.11$), and 131 for rs61735836 ($P = 0.0035$) (eTable 6; <http://links.lww.com/EE/A77>, eFigure 5; <http://links.lww.com/EE/A77>).

Table 3.

Summary of associations between arsenic metabolism efficiency phenotypes and 221 arsenic-associated CpG sites (compared with results from prior EWAS meta-analysis of urinary arsenic)

	Cohort	Consistent ^a	Inconsistent ^b	One-sided P ^c	% Consistent
DMA% ^d	HEALS	142	79	1.4×10^{-5}	64.3
Weighted	Combined	137	84	2.2×10^{-4}	62.0
GP-DMA% ^e					
MR-IVW ^f	Combined	134	87	9.5×10^{-4}	60.6

^aNumber of associations with methylation at CpG sites that are opposite in sign from arsenic exposure associations observed in prior EWAS (i.e., consistent with prior EWAS).

^bNumber of associations with methylation at CpG sites that are the same in sign compared with arsenic exposure associations observed in prior EWAS (i.e., inconsistent with prior EWAS).

^cCorresponding to the alternative hypothesis that signs of associations among the set of 221 CpGs tend to be consistent with arsenic exposure associations. Reported one-sided P -values.

^dCpG methylation ~ DMA% + age + sex + methylation batch + smoking status + BMI + education + ln(water As).

^eCpG methylation ~ weighted GP-DMA% + age + sex + methylation batch + smoking status.

^fInverse-variance weighted MR.

Sensitivity analysis of arsenic-associated CpGs discovered in HEALS

We repeated the analyses above, restricting to the set of 34 CpGs measured on the EPIC array and found to be associated (at FDR = 0.05) with urinary arsenic among HEALS participants

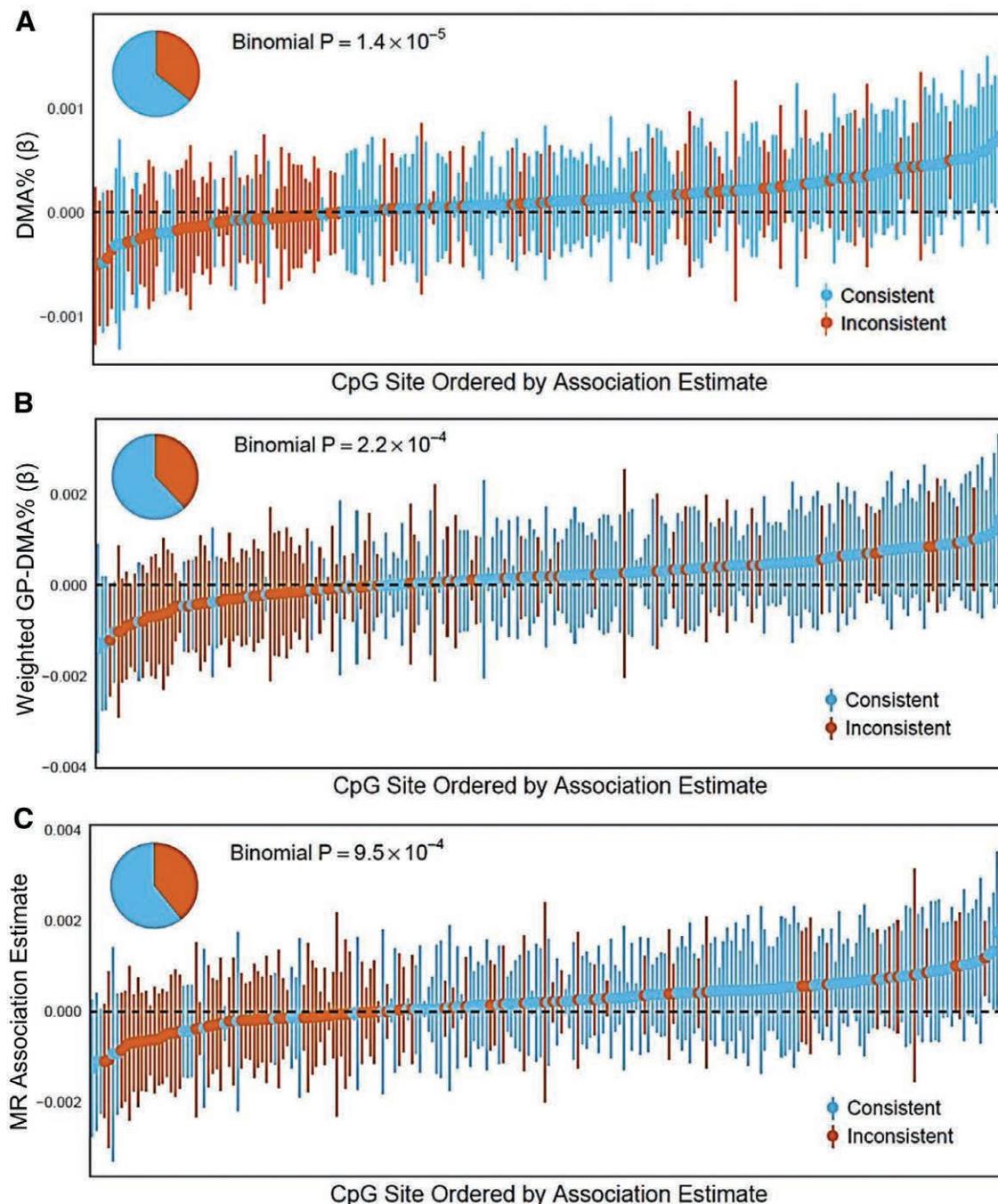


Figure 3. Associations of DMA% and genetically predicted DMA% with DNA methylation at 221 arsenic-associated CpGs discovered in meta-analysis. β coefficients for associations of (A) DMA% and (B) weighted GP-DMA% with CpG sites, as well as (C) MR-based association estimates for each CpG site, were considered consistent with associations between $\ln(\text{creatinine-adjusted urinary As})$ and CpG sites (from prior EWAS) if their signs were opposite. Study populations were HEALS ($n = 379$) for DMA% analyses, a subset of the combined cohort with full genotypic data ($n = 723$) for weighted GP-DMA%, and a combined cohort of HEALS and BEST ($n = 772$) for MR.

in the prior EWAS ($n = 379$ for all analyses except weighted GP-DMA%, where $n = 340$). Associations with DMA% and log-transformed DMA were consistent with arsenic exposure associations in 22 CpGs ($P = 0.061$) and 24 CpGs ($P = 0.012$), respectively (eTable 7; <http://links.lww.com/EE/A77>, Figure 4A). Weighted GP-DMA% and binary SNP score associations were consistent with arsenic exposure associations in 23 CpGs ($P = 0.029$) and 27 CpGs ($P = 0.0004$), respectively (eTable 7; <http://links.lww.com/EE/A77>, Figure 4B). MR estimates were consistent with arsenic exposure associations in 25 CpGs ($P = 0.0045$) (eTable 7; <http://links.lww.com/EE/A77>, Figure 4C).

Discussion

In this study, we used an MR approach to provide evidence that supports the hypothesis that arsenic metabolism efficiency is causally related to DNA methylation measured in whole blood at specific CpG sites previously reported to be associated with urinary arsenic. Using a sample of 379 Bangladeshi adults from the HEALS cohort, we demonstrate that the directions of association between DMA% and DNA methylation levels at many arsenic-associated CpG sites tend to be in the opposite direction of the association between urinary arsenic and the same CpGs. Additionally, genetically predicted DMA% (combining

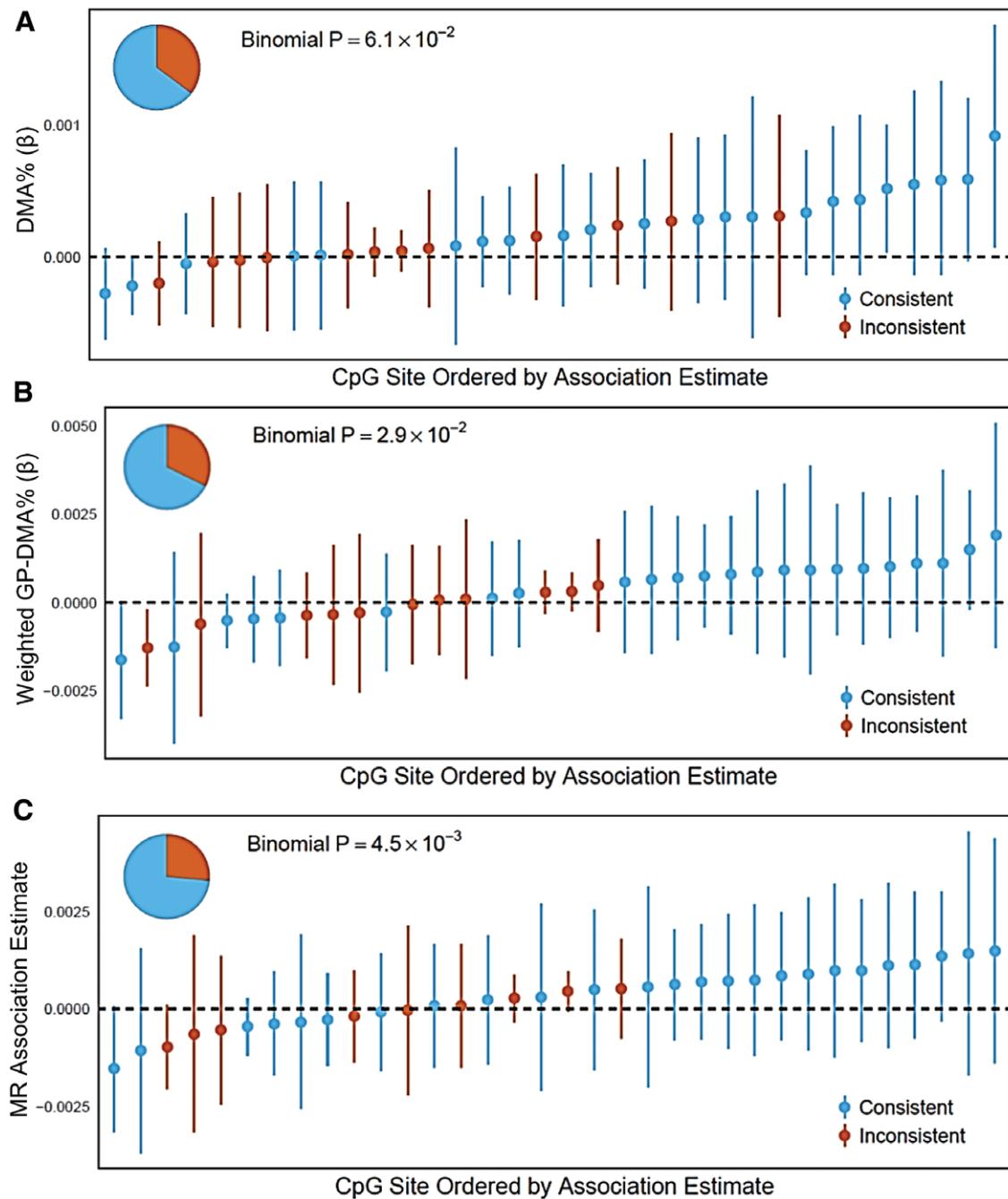


Figure 4. Associations of DMA% and genetically predicted DMA% with DNA methylation at 34 arsenic-associated CpGs discovered in HEALS. (A) DMA%, (B) weighted GP-DMA%, (C) MR. Study populations were HEALS ($n = 379$) for DMA% and MR analyses, and a subset of HEALS with full genotypic data ($n = 340$) for weighted GP-DMA%.

information across multiple SNPs) tends to be associated with these CpGs in directions opposite to that of arsenic exposure. Finally, the MR estimates for arsenic-associated CpGs tend to be consistent with the associations between these CpGs and arsenic exposure. These observations support the hypothesis that efficient metabolism of arsenic (i.e., high DMA%) reduces an individual's internal arsenic dose and should reduce the impact of arsenic on CpGs to which arsenic is causally related.

While the genetics of arsenic metabolism and the associations of arsenic exposure with DNA methylation have been studied previously, in this study, we use the unique approach of applying MR to estimate the association between arsenic metabolism efficiency and methylation at specific CpG sites. The random

assignment of study participants' genotypes minimizes the risk of bias in the association of arsenic metabolism with CpG methylation due to confounding (under the assumption that *AS3MT* and *FTCD* genotypes are associated with CpG methylation only through their effect on arsenic metabolism). This article does not provide strong evidence that arsenic metabolism efficiency affects methylation at any specific CpG site (largely due to lack of statistical power for tests of individual CpGs). These results may not generalize to tissue types other than blood or to other arsenic-exposed populations.

Although the binomial tests of directional consistency are consistent with a causal influence of arsenic metabolism on some fraction of the CpG sites tested, it is unclear if the

assumptions of MR are fully satisfied for any individual CpG. MR provides valid evidence supporting causal effects if IVs are (1) associated with the risk factor (DMA%), (2) not associated with confounders of the association between the risk factor and outcome (CpG methylation), and (3) independent of the outcome conditional on the risk factor and confounders.²⁹ The third assumption (exclusion restriction) requires that the impact of 10q24.32 and *FTCD* variants on methylation levels of the CpG site be completely mediated by arsenic metabolism efficiency, and this assumption cannot be proven.²⁹ Because only three genetic variants are known to be independently associated with DMA%, the number of IVs is insufficient to conduct Egger regression and test for violations of the exclusion restriction. However, the role of *AS3MT* in arsenic metabolism is well-established, providing strong biologic plausibility for a specific effect of these SNPs on arsenic metabolism, even though the SNPs' biologic mechanisms are not completely understood. The mechanism linking *FTCD* to arsenic metabolism is less clear, but it is potentially related to folate metabolism and the availability of methyl groups for arsenic metabolism.³² Given the biologic plausibility of these IVs with respect to arsenic metabolism and the consistency of our results between *AS3MT* SNP rs9527 and *FTCD* SNP rs61735836, we feel that violations of the MR assumptions in this work are unlikely.³¹ Finally, the directions of association for rs11191527 with CpGs were not skewed either positively or negatively. This result may call into question the validity of rs11191527 as an IV. However, the power of our study for detecting associations between rs11191527 and each CpG is limited by the relatively weak association of rs11191527 with DMA%.

An additional limitation of our MR approach is that this approach cannot identify critical periods of this life course in which arsenic exposure affects DNA methylation, nor can we estimate the latency period of such effects. Applying MR only allows us to estimate the impact of lifelong differences in arsenic metabolism efficiency on DNA methylation.

The results of this article build upon those of the EWAS in which these arsenic-associated CpGs were identified, providing evidence for a potential mechanism for arsenic-induced carcinogenesis. Enrichment analyses in the prior EWAS have shown that CpGs annotated to CpG shores, DNase I hypersensitive sites, and enhancers are over-represented among arsenic-associated CpGs identified in the meta-analysis of HEALS and BEST.²¹ These genomic regions have important roles in transcription regulation, and DNA methylation in these regions is more variable and may be more sensitive to environmental exposures.⁴³ This results from this article provide additional evidence to supporting studies to further investigate arsenic-associated DNA methylation as a mediator between arsenic exposure and associated diseases and health effects.

This study provides evidence suggesting that a substantial fraction of the associations between arsenic exposure and whole blood DNA methylation observed in prior EWAS represent a causal impact of arsenic, mediated by arsenic metabolism efficiency, on the epigenome. Our MR approach in the context of EWAS is particularly novel, providing an example that can inform future attempts at causal inference in this field and studies of potential molecular effects of environmental exposures. Future studies using larger sample sizes obtained from multiple arsenic-exposed populations should provide more robust MR evidence supporting causal effects of arsenic on specific CpG sites and a more comprehensive understanding of the mechanisms of arsenic-associated disease risk.

Conflict of interest statement

The authors declare that they have no financial conflict of interest with regard to the content of this report.

Acknowledgments

We would like to acknowledge the HEALS and BEST study participants and research staff for their contributions to these cohorts and Dr. M. Kibriya for his contributions to generating the DNA methylation data utilized in this analysis.

References

1. Naujokas MF, Anderson B, Ahsan H, et al. The broad scope of health effects from chronic arsenic exposure: update on a worldwide public health problem. *Environ Health Perspect.* 2013;121:295–302.
2. Martinez VD, Vucic EA, Becker-Santos DD, Gil L, Lam WL. Arsenic exposure and the induction of human cancers. *J Toxicol.* 2011;2011:431287.
3. Saint-Jacques N, Parker L, Brown P, Dummer TJ. Arsenic in drinking water and urinary tract cancers: a systematic review of 30 years of epidemiological evidence. *Environ Health.* 2014;13:44.
4. Moon K, Guallar E, Navas-Acien A. Arsenic exposure and cardiovascular disease: an updated systematic review. *Curr Atheroscler Rep.* 2012;14:542–555.
5. Navas-Acien A, Silbergeld EK, Streeter RA, Clark JM, Burke TA, Guallar E. Arsenic exposure and type 2 diabetes: a systematic review of the experimental and epidemiological evidence. *Environ Health Perspect.* 2006;114:641–648.
6. Sanchez TR, Powers M, Perzanowski M, George CM, Graziano JH, Navas-Acien A. A meta-analysis of arsenic exposure and lung function: is there evidence of restrictive or obstructive lung disease? *Curr Environ Health Rep.* 2018;5:244–254.
7. Sung TC, Huang JW, Guo HR. Association between arsenic exposure and diabetes: a meta-analysis. *Biomed Res Int.* 2015;2015:368087.
8. Argos M, Parvez F, Rahman M, et al. Arsenic and lung disease mortality in Bangladeshi adults. *Epidemiology.* 2014;25:536–543.
9. Argos M, Kalra T, Rathouz PJ, et al. Arsenic exposure from drinking water, and all-cause and chronic-disease mortalities in Bangladesh (HEALS): a prospective cohort study. *Lancet.* 2010;376:252–258.
10. Smith AH, Marshall G, Roh T, Ferreccio C, Liaw J, Steinmaus C. Lung, bladder, and kidney cancer mortality 40 years after arsenic exposure reduction. *J Natl Cancer Inst.* 2018;110:241–249.
11. Martin EM, Fry RC. Environmental influences on the epigenome: exposure-associated DNA methylation in human populations. *Annu Rev Public Health.* 2018;39:309–333.
12. Koestler DC, Avissar-Whiting M, Houseman EA, Karagas MR, Marsit CJ. Differential DNA methylation in umbilical cord blood of infants exposed to low levels of arsenic in utero. *Environ Health Perspect.* 2013;121:971–977.
13. Cardenas A, Houseman EA, Baccarelli AA, et al. In utero arsenic exposure and epigenome-wide associations in placenta, umbilical artery, and human umbilical vein endothelial cells. *Epigenetics.* 2015;10:1054–1063.
14. Cardenas A, Koestler DC, Houseman EA, et al. Differential DNA methylation in umbilical cord blood of infants exposed to mercury and arsenic in utero. *Epigenetics.* 2015;10:508–515.
15. Green BB, Karagas MR, Punshon T, et al. Epigenome-wide assessment of DNA methylation in the placenta and arsenic exposure in the New Hampshire Birth Cohort Study (USA). *Environ Health Perspect.* 2016;124:1253–1260.
16. Kile ML, Houseman EA, Baccarelli AA, et al. Effect of prenatal arsenic exposure on DNA methylation and leukocyte subpopulations in cord blood. *Epigenetics.* 2014;9:774–782.
17. Seow WJ, Kile ML, Baccarelli AA, et al. Epigenome-wide DNA methylation changes with development of arsenic-induced skin lesions in Bangladesh: a case-control follow-up study. *Environ Mol Mutagen.* 2014;55:449–456.
18. Liu X, Zheng Y, Zhang W, et al. Blood methylomics in response to arsenic exposure in a low-exposed US population. *J Expo Sci Environ Epidemiol.* 2014;24:145–149.
19. Argos M, Chen L, Jasmine F, et al. Gene-specific differential DNA methylation and chronic arsenic exposure in an epigenome-wide association study of adults in Bangladesh. *Environ Health Perspect.* 2015;123:64–71.
20. Ameer SS, Engström K, Hossain MB, Concha G, Vahter M, Broberg K. Arsenic exposure from drinking water is associated with decreased gene expression and increased DNA methylation in peripheral blood. *Toxicol Appl Pharmacol.* 2017;321:57–66.
21. Demanelis K, Argos M, Tong L, et al. The impact of arsenic exposure on whole blood DNA methylation: an epigenome-wide study of Bangladeshi adults. *Environ Health Perspect.* 2019;127:57011.
22. Drobna Z, Styblo M, Thomas DJ. An Overview of Arsenic Metabolism and Toxicity. *Curr Protoc Toxicol.* 2009;42: 4 31 31–34 31 36.

23. Drobná Z, Walton FS, Paul DS, Xing W, Thomas DJ, Stýblo M. Metabolism of arsenic in human liver: the role of membrane transporters. *Arch Toxicol.* 2010;84:3–16.
24. Pierce BL, Kibriya MG, Tong L, et al. Genome-wide association study identifies chromosome 10q24.32 variants associated with arsenic metabolism and toxicity phenotypes in Bangladesh. *PLoS Genet.* 2012;8:e1002522.
25. Valenzuela OL, Borja-Aburto VH, Garcia-Vargas GG, et al. Urinary trivalent methylated arsenic species in a population chronically exposed to inorganic arsenic. *Environ Health Perspect.* 2005;113:250–254.
26. Ahsan H, Chen Y, Kibriya MG, et al. Arsenic metabolism, genetic susceptibility, and risk of premalignant skin lesions in Bangladesh. *Cancer Epidemiol Biomarkers Prev.* 2007;16:1270–1278.
27. Kile ML, Hoffman E, Rodrigues EG, et al. A pathway-based analysis of urinary arsenic metabolites and skin lesions. *Am J Epidemiol.* 2011;173:778–786.
28. Lindberg AL, Rahman M, Persson LA, Vahter M. The risk of arsenic induced skin lesions in Bangladeshi men and women is affected by arsenic metabolism and the age at first exposure. *Toxicol Appl Pharmacol.* 2008;230:9–16.
29. Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet Epidemiol.* 2013;37:658–665.
30. Pierce BL, Tong L, Argos M, et al. Arsenic metabolism efficiency has a causal role in arsenic toxicity: Mendelian randomization and gene-environment interaction. *Int J Epidemiol.* 2013;42:1862–1871.
31. Agusa T, Fujihara J, Takeshita H, Iwata H. Individual variations in inorganic arsenic metabolism associated with AS3MT genetic polymorphisms. *Int J Mol Sci.* 2011;12:2351–2382.
32. Pierce BL, Tong L, Dean S, et al. A missense variant in FTCD affects arsenic metabolism and toxicity phenotypes in Bangladesh. *PLOS Genetics.* 2019;15:e1007984.
33. Ahsan H, Chen Y, Parvez F, et al. Health Effects of Arsenic Longitudinal Study (HEALS): description of a multidisciplinary epidemiologic investigation. *J Expo Sci Environ Epidemiol.* 2006;16:191–205.
34. Verret WJ, Chen Y, Ahmed A, et al. A randomized, double-blind placebo-controlled trial evaluating the effects of vitamin E and selenium on arsenic-induced skin lesions in Bangladesh. *J Occup Environ Med.* 2005;47:1026–1035.
35. Lawlor DA, Harbord RM, Sterne JA, Timpson N, Davey Smith G. Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. *Stat Med.* 2008;27:1133–1163.
36. Ahsan H, Chen Y, Parvez F, et al. Arsenic exposure from drinking water and risk of premalignant skin lesions in Bangladesh: baseline results from the Health Effects of Arsenic Longitudinal Study. *Am J Epidemiol.* 2006;163:1138–1148.
37. Argos M, Rahman M, Parvez F, et al. Baseline comorbidities in a skin cancer prevention trial in Bangladesh. *Eur J Clin Invest.* 2013;43:579–588.
38. Nixon DE, Mussmann GV, Eckdahl SJ, Moyer TP. Total arsenic in urine: palladium-persulfate vs nickel as a matrix modifier for graphite furnace atomic absorption spectrophotometry. *Clin Chem.* 1991;37:1575–1579.
39. Nermell B, Lindberg AL, Rahman M, et al. Urinary arsenic concentration adjustment factors and malnutrition. *Environ Res.* 2008;106:212–218.
40. Teschendorff AE, Marabita F, Lechner M, et al. A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. *Bioinformatics.* 2013;29:189–196.
41. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics.* 2010;26:2190–2191.
42. Yavorska O. MendelianRandomization: Mendelian Randomization Package. R package version 0.3.0. 2018. <https://CRAN.R-project.org/package=MendelianRandomization>. Accessed July 2018.
43. Jones PA. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat Rev Genet.* 2012;13:484–492.