Note to readers with disabilities: *EHP* strives to ensure that all journal content is accessible to all readers. However, some figures and Supplemental Material published in *EHP* articles may not conform to 508 standards due to the complexity of the information being presented. If you need assistance accessing journal content, please contact ehp508@niehs.nih.gov. Our staff will work with you to assess and meet your accessibility needs within 3 working days.

Supplemental Material

Locus-Specific Differential DNA Methylation and Urinary Arsenic: An Epigenome-Wide Association Study in Blood among Adults with Low-to-Moderate Arsenic Exposure

Anne K. Bozack, Arce Domingo-Relloso, Karin Haack, Mary V. Gamble, Maria Tellez-Plaza, Jason G. Umans, Lyle G. Best, Joseph Yracheta, Matthew O. Gribble, Andres Cardenas, Kevin A. Francesconi, Walter Goessler, Wan-Yee Tang, M. Daniele Fallin, Shelley A. Cole, and Ana Navas-Acien

Table of Contents

Table S8. Results of a linear model for the association between urinary As levels and M-values of cg14595618, annotated to *HK1*, including the interaction term ln(total urinary As) × diabetes status and adjusted age, sex, BMI, smoking status, education, study center (Arizona, Oklahoma, North and South Dakota), estimated glomerular filtration rate, and cell type proportion estimates.

Table S9. Differentially methylated CpG sites associated with ln(total urinary arsenic levels) at at $P_{FDR} < 0.05$ in our primary analyses, and results from sensitivity analyses performed on probe-type normalized data.

Table S10. Differentially methylated CpG sites associated with ln(total urinary arsenic levels) at at $P_{FDR} < 0.05$ in our primary analyses, and results from sensitivity analyses using ReFACTor, a reference-free approach to adjusted for cell type proportions.

Figure S1. Flowchart of eligible participant and participants selected for DNA methylation analysis.

Figure S2. Q-Q plot from *limma* models of M-values of ln(total urinary arsenic levels) adjusted for age, sex, BMI, self-reported smoking status, education (< high school, high school graduate or GED, > high school), study center (Arizona, Oklahoma, North and South Dakota), estimated glomerular filtration rate, and cell type proportion estimates.

Figure S3. Sensitivity analyses for probe-type normalization. Q-Q plot from *limma* models of M-values normalized using single sample Noob and RCP normalization adjusted for age, sex, BMI, self-reported smoking status, education (< high school, high school graduate or GED, > high school), study center (Arizona, Oklahoma, North and South Dakota), estimated glomerular filtration rate, and cell type proportion estimates.

Figure S4. Sensitivity analyses for reference-free cell type proportion adjustment. Q-Q plot from *limma* models of M-values adjusted for age, sex, BMI, self-reported smoking status, education (< high school, high school graduate or GED, > high school), study center (Arizona, Oklahoma, South Dakota), estimated glomerular filtration rate, and six ReFACTor components.

References

Additional File- Excel Document

Table S1: Studies identified reporting the association between As exposure and DNA methylation measured using the 450K and 850K microarrays and included in the lookup approach.

Study	Population; location	N	Exposure assessment	Mean exposure	DNA methylation assessment	Platform	Number of significant CpGs [a]	n (%) replicated [b]
Koestler et al. 2013 (1)	Birth cohort; USA	134	Maternal urinary arsenic	Mean = 4.1 µg/L	Cord blood	450K	0 after correction for multiple testing	NA
Broberg et al. 2014 (2)	Birth cohort; Bangladesh	127	Maternal urinary arsenic	Mean = 66 µg/L	Cord blood	450K	3 in boys ($p_{FDR} < 0.05$)	0 (0%)
Kile et al. 2014 (3)[c]	Birth cohort; Bangladesh	44	Maternal drinking water arsenic	Median = 12 µg/L	Cord blood	450K	1 ($p_{FDR} < 0.05$); 10 ($P < 2x10^{-6}$)	1 (10%)
Liu et al. 2014 (4)	Adults; USA	46	Toenail arsenic	Low exposure: <0.0649 µg/kg; high exposure: ≥ 0.1442 µg/kg	Whole blood cells	450K	0 after correction for multiple testing	NA
Argos et al. 2015 (5)	Adults; Bangladesh	400	Blood and urinary arsenic	Blood As, mean = 9.3 µg/L; urinary As, mean = 302 µg/g creatinine	Whole blood	450K	4 ($p_{Bonferroni} < 0.05$)	1 (25%)
Rojas et al. 2015 (6)	Birth cohort; Mexico	38	Maternal urinary arsenic	Mean = 73.9 µg/L	Cord blood	450K	4,771 ($p_{FDR} < 0.05$)	141 (3%)
Cardenas et al. 2015 (7)	Birth cohort; Bangladesh	Placenta: N = 37; cord blood: N = 45; HUVEC: N = 52	Maternal drinking water arsenic	Mean = 63.7 µg/L	Placenta, cord blood, and human umbilical vein endothelial cells (HUVEC)	450K	5 ($p_{Bonferroni} < 0.05$)	1 (20%)
Green et al. 2016 (8)	Birth cohort; USA	343	Maternal toenail arsenic and placental arsenic	Maternal toenail As, mean = 0.05 µg/kg; placental As, mean = 0.82 µg/kg	Placenta	450K	Maternal toenail arsenic: 1; placental arsenic: 163 ($p_{FDR} < 0.05$)	3 (2%)
Kaushal et al. 2017 (9)	Birth cohort; Taiwan	64	Maternal urinary arsenic	Mean = 23.2 µg/g creatinine	Cord blood	450K	579 ($p_{FDR} < 0.05$)	28 (5%0
Ameer et al. 2017 (10)	Women; Argentina	96	Urinary arsenic	Median = 185 µg/L	Whole blood	450K	6 ($p_{FDR} < 0.05$)	0 (0%)

Bozack et al. 2018 (11)[c]	Birth cohort; Bangladesh	44	Maternal drinking water arsenic	Mean = 63.7 µg/L	Cord blood	450K	380 ($p < 1.10 \times 10^{-6}$; |β regression|>0.10); top 10 reported	0 (0%)
Guo et al. 2018 (12)	Families (adults and children); China	102	Village drinking water arsenic > or < 10µg/L	Among exposed group, mean = 266.4 ppb	Buffy coat	450K	85 ($p_{FDR} < 0.05$)	7 (8%)
Gliga et al. 2018 (13)	Birth cohort; Bangladesh	113	Maternal urinary arsenic	Mean = 68 µg/L	Blood mononuclear cells collected at 9 years	450K	9 in all children, 57 in boys, 15 in girls ($p_{FDR} < 0.05$)	5 (7%)
Demanelis et al. 2019 (14)	Adults; Bangladesh	396	Urinary As; drinking water As	Urinary As, median = 201.5 µg/g creatinine; water As median: 50.5 µg/L	Whole blood	850K	Urinary As: 34 ($p_{FDR} < 0.05$); water As: 24 ($p_{FDR} < 0.05$)	4 (5%)

a. A total of 4,612 differentially methylated positions, and 4,595 unique 4,612 differentially methylated positions were identified across studies. b. In the current study, nominal $p < 0.05$ for ln(urinary As levels) in *limma* models adjusted for age, sex, BMI, self-reported smoking status, education (< high school, high school graduate or GED, > high school), study center (Arizona, Oklahoma, North and South Dakota), estimated glomerular filtration rate, and cell type proportion estimates. c. Analyses of the same birth cohort recruited in the Sirajdikhan and Pabna Sadar upazilas of Bangladesh.

Table S2: Descriptive characteristics for eligible participants and participants for DNA methylation analysis.

	Included (N=2,325)	Eligible (N=2,731)
Age, median (IQR)	55 (49, 62)	55 (49, 62)
Sex (% male)	41.46	40.75
Smoking status		
% Current	38.41	37.79
% Former	32.17	32.95
BMI, median (IQR)	29.59 (26.22, 33.63)	29.67 (26.29, 33.69)
Education		
No high school	17.51	17.18
Some high school	23.91	23.43
Completed high school	58.58	59.39
Total urinary arsenic (µg/g creatinine)	8.56 (5.24, 14.42)	8.43 (5.15, 14.32)
IQR = interquartile range		

Table S3: Linear models for the association between ln(total urinary arsenic levels) and imputed cell type proportions.

Cell type	Mean proportion	B (95% CI)	p
CD8+ T cells	0.07	0.001 (-0.004, 0.002)	0.41
CD4+ T cells	0.18	-0.002 (-0.006, 0.002)	0.35
NK	0.12	0.005 (0.002, 0.009)	0.004
B cells	0.08	-0.005 (-0.007, -0.002)	<0.001
Monocytes	0.05	0.002 (0.000, 0.004)	0.06
Granulocytes	0.51	0.001(-0.007, 0.008)	0.81
Models were adjusted for age, BMI, smoking status, and sex.			

Table S4: Differentially methylated CpG sites associated with ln(total urinary arsenic levels) at $P_{FDR} < 0.05$, and effects sizes and p-values stratified by sex.

CpG	Chr	Position [a]	Gene	Feature category	Unstratified p	Female (N = 1,361) Effect estimate [b]	p	Male (N = 964) Effect estimate [b]	p
cg06690548	4	139162808	SLC7A11	Body	9.56E-10	-0.11	2.08E-07	-0.09	2.81E-04
cg03497652	16	4751569	ANKS3	Body	1.49E-09	0.09	1.04E-06	0.08	7.22E-04
cg22294740	19	2294961	LINGO3	5'UTR	8.78E-09	0.09	2.47E-06	0.08	7.51E-04
cg20493718	17	80202979	CSNK1D	Body, 3'UTR	1.04E-08	0.13	3.24E-07	0.09	0.008
cg18616702	1	150523808	ADAMTSL4; MIR4257	5'UTR; TSS1500	2.35E-08	0.05	5.04E-05	0.05	1.66E-04
cg12116137	17	1576449	PRPF8	Body	1.22E-07	0.07	2.28E-04	0.09	7.45E-04
cg12106731	1	36023279	NCDN; KIAA0319L	TSS200; TSS1500	1.24E-07	0.04	2.62E-04	0.06	2.09E-04
cg00500428	16	3074502	HCFC1R1; THOC6	TSS1500; Body	1.83E-07	-0.06	3.78E-05	-0.06	0.002
cg04940901	3	52418250	DNAH1	Body	1.96E-07	0.04	0.001	0.06	5.22E-05
cg14827056	8	141550539	EIF2C2	Body	2.14E-07	0.06	4.12E-04	0.08	4.54E-04
cg21369801	17	80202961	CSNK1D	Body, 3'UTR	3.96E-07	0.07	5.91E-05	0.07	0.005
cg09280971	9	73036509	Intergenic		4.63E-07	0.02	0.002	0.04	2.40E-05
cg27178850	11	9039431	Intergenic		4.66E-07	0.03	0.001	0.05	2.45E-04
cg08059112	19	2294887	LINGO3	5'UTR	6.19E-07	0.04	3.56E-05	0.03	0.011
cg20509831	19	58866362	A1BG-AS1; A1BG; ZNF497	Body; TSS1500; 3'UTR	6.45E-07	0.04	7.034E-04	0.05	4.68E-04
cg01538969	6	30624636	DHX16	Body	7.09E-07	0.04	0.004	0.07	6.01E-05
cg14595618	10	71135446	HK1	Body	7.63E-07	0.06	3.35E-05	0.05	0.010
cg07021906	16	87866833	SLC7A5	Body	9.19E-07	0.05	0.002	0.08	3.18E-04
cg07317306	1	110314824	Intergenic		9.79E-07	0.02	0.004	0.04	2.62E-04
cg03036214	15	63640658	CA12	Body	1.23E-06	0.05	9.90E-04	0.06	0.002

Chr = chromosome. a. GRCh37/hg19 assembly. b. Effect estimate from *limma* models adjusted for age, BMI, self-reported smoking status, education (< high school, high school graduate or GED, > high school), study center (Arizona, Oklahoma, North and South Dakota), estimated glomerular filtration rate, and cell type proportion estimates.

Table S5: Differentially methylated CpG sites associated with ln(total urinary arsenic concentration) at $P_{FDR} < 0.05$, and effects sizes and p-values stratified by smoking status.

CpG	Chr	Position [b]	Gene	Feature category	Unstratified p	Nonsmoker [a] (N = 684) Effect estimate [d]	p	Smoker [b] (N = 1,641) Effect estimate [d]	p
cg06690548	4	139162808	SLC7A11	Body	9.56E-10	-0.09	0.004	-0.10	5.41E-08
cg03497652	16	4751569	ANKS3	Body	1.49E-09	0.11	2.15E-04	0.08	1.07E-06
cg22294740	19	2294961	LINGO3	5'UTR	8.78E-09	0.11	5.74E-05	0.07	1.76E-05
cg20493718	17	80202979	CSNK1D	Body, 3'UTR	1.04E-08	0.14	7.82E-04	0.12	1.65E-06
cg18616702	1	150523808	ADAMTSL4; MIR4257	5'UTR; TSS1500	2.35E-08	0.05	0.002	0.05	8.27E-07
cg12116137	17	1576449	PRPF8	Body	1.22E-07	0.11	4.10E-04	0.07	1.08E-04
cg12106731	1	36023279	NCDN; KIAA0319L	TSS200; TSS1500	1.24E-07	0.05	0.002	0.05	1.34E-05
cg00500428	16	3074502	HCFC1R1; THOC6	TSS1500; Body	1.83E-07	-0.05	0.017	-0.06	6.20E-06
cg04940901	3	52418250	DNAH1	Body	1.96E-07	0.05	0.003	0.05	1.03E-05
cg14827056	8	141550539	EIF2C2	Body	2.14E-07	0.08	0.002	0.06	2.34E-05
cg21369801	17	80202961	CSNK1D	Body, 3'UTR	3.96E-07	0.09	0.002	0.07	2.89E-05
cg09280971	9	73036509	Intergenic		4.63E-07	0.03	0.011	0.03	2.98E-06
cg27178850	11	9039431	Intergenic		4.66E-07	0.06	2.11E-04	0.03	3.51E-04
cg08059112	19	2294887	LINGO3	5'UTR	6.19E-07	0.06	3.09E-04	0.03	3.24E-04
cg20509831	19	58866362	A1BG-AS1; A1BG; ZNF497	Body; TSS1500; 3'UTR	6.45E-07	0.06	0.001	0.04	6.08E-05
cg01538969	6	30624636	DHX16	Body	7.09E-07	0.07	6.22E-04	0.05	1.44E-04
cg14595618	10	71135446	HK1	Body	7.63E-07	0.06	0.008	0.06	3.36E-05
cg07021906	16	87866833	SLC7A5	Body	9.19E-07	0.08	6.89E-04	0.06	2.18E-04
cg07317306	1	110314824	Intergenic		9.79E-07	0.04	0.007	0.03	1.74E-05
cg03036214	15	63640658	CA12	Body	1.23E-06	0.05	0.018	0.05	3.43E-05

Chr = chromosome. a. Never smoker. b. Former smoker or current smoker. c. GRCh37/hg19 assembly. d. Effect estimate from *limma* models adjusted for age, sex, BMI, education (< high school, high school graduate or GED, > high school), study center (Arizona, Oklahoma, North and South Dakota), estimated glomerular filtration rate, and cell type proportion estimates.

Table S6: Differentially methylated CpG sites associated with ln(total urinary arsenic concentration) at $P_{FDR} < 0.05$, and effects sizes and p-values stratified by study center.

					Unstratified	Arizona (N = 312)		Oklahoma (N = 981)		North & South Dakota (N = 1,032)	
CpG	Chr	Position [a]	Gene	Feature category	p	Effect estimate [b]	p	Effect estimate [b]	p	Effect estimate [b]	p
cg06690548	4	139162808	SLC7A11	Body	9.56E-10	-0.11	0.026	-0.14	3.42E-08	-0.067	0.004
cg03497652	16	4751569	ANKS3	Body	1.49E-09	0.15	0.001	0.09	1.63E-04	0.071	8.41E-04
cg22294740	19	2294961	LINGO3	5'UTR	8.78E-09	0.17	4.73E-04	0.09	4.27E-05	0.050	0.013
cg20493718	17	80202979	CSNK1D	Body, 3'UTR	1.04E-08	0.20	0.002	0.13	1.08E-04	0.086	0.004
cg18616702	1	150523808	ADAMTSL4; MIR4257	5'UTR; TSS1500	2.35E-08	0.05	0.066	0.05	1.91E-04	0.036	0.004
cg12116137	17	1576449	PRPF8	Body	1.22E-07	0.12	0.018	0.07	0.003	0.074	0.001
cg12106731	1	36023279	NCDN; KIAA0319L	TSS200; TSS1500	1.24E-07	0.06	0.033	0.06	8.62E-05	0.038	0.005
cg00500428	16	3074502	HCFC1R1; THOC6	TSS1500; Body	1.83E-07	-0.11	0.003	-0.04	0.020	-0.057	3.96E-04
cg04940901	3	52418250	DNAH1	Body	1.96E-07	0.07	0.026	0.04	0.010	0.045	7.17E-04
cg14827056	8	141550539	EIF2C2	Body	2.14E-07	0.10	0.027	0.06	0.003	0.051	0.004
cg21369801	17	80202961	CSNK1D	Body, 3'UTR	3.96E-07	0.15	0.002	0.06	0.004	0.055	0.010
cg09280971	9	73036509	Intergenic		4.63E-07	0.04	0.055	0.04	9.22E-05	0.020	0.018
cg27178850	11	9039431	Intergenic		4.66E-07	0.07	0.008	0.05	2.25E-04	0.024	0.028
cg08059112	19	2294887	LINGO3	5'UTR	6.19E-07	0.08	0.002	0.03	0.005	0.030	0.006
cg20509831	19	58866362	A1BG-AS1; A1BG; ZNF497	Body; TSS1500; 3'UTR	6.45E-07	0.01	0.839	0.05	0.002	0.050	6.78E-05
cg01538969	6	30624636	DHX16	Body	7.09E-07	0.12	0.002	0.05	0.006	0.040	0.016
cg14595618	10	71135446	HK1	Body	7.63E-07	0.11	0.011	0.05	0.011	0.050	0.003
cg07021906	16	87866833	SLC7A5	Body	9.19E-07	0.15	4.07E-04	0.04	0.030	0.052	0.006
cg07317306	1	110314824	Intergenic		9.79E-07	0.04	0.075	0.03	0.021	0.038	1.21E-04
cg03036214	15	63640658	CA12	Body	1.23E-06	0.09	0.020	0.05	0.010	0.041	0.008

Chr = chromosome. a. GRCh37/hg19 assembly. b. Effect estimate from *limma* models adjusted for age, sex, BMI, self-reported smoking status, education (< high school, high school graduate or GED, > high school), estimated glomerular filtration rate, and cell type proportion estimates.

Table S7: Differentially methylated CpG sites associated with ln(total urinary arsenic levels) at $P_{FDR} < 0.05$, and effects sizes and p-values stratified by diabetes status.

CpG	Chr	Position [a]	Gene	Feature category	Unstratified p	Diabetic (N = 968) Effect estimate [b]	p	Non-diabetic, creatinine adjusted [c] (N = 1,357) Effect estimate [b]	p	Non-diabetic, SG adjusted [d] (N = 1,357) Effect estimate [b]	p
cg06690548	4	139162808	SLC7A11	Body	9.56E-10	-0.06	0.028	-0.12	3.63E-09	-0.10	1.26E-07
cg03497652	16	4751569	ANKS3	Body	1.49E-09	0.11	1.24E-05	0.07	1.64E-04	0.06	2.96E-04
cg22294740	19	2294961	LINGO3	5'UTR	8.78E-09	0.10	5.78E-05	0.06	2.54E-04	0.05	0.004
cg20493718	17	80202979	CSNK1D	Body, 3'UTR	1.04E-08	0.13	6.39E-05	0.11	7.53E-05	0.10	1.03E-04
cg18616702	1	150523808	ADAMTSL4; MIR4257	5'UTR; TSS1500	2.35E-08	0.04	0.011	0.05	3.64E-06	0.05	1.95E-06
cg12116137	17	1576449	PRPF8	Body	1.22E-07	0.12	1.63E-05	0.06	0.002	0.06	0.002
cg12106731	1	36023279	NCDN; KIAA0319L	TSS200; TSS1500	1.24E-07	0.05	0.001	-0.06	6.24E-05	0.04	9.20E-04
cg00500428	16	3074502	HCFC1R1; THOC6	TSS1500; Body	1.83E-07	-0.05	0.007	0.05	1.78E-05	-0.05	8.40E-05
cg04940901	3	52418250	DNAH1	Body	1.96E-07	0.04	0.010	0.05	4.74E-05	0.04	1.15E-04
cg14827056	8	141550539	EIF2C2	Body	2.14E-07	0.09	6.15E-05	0.05	0.003	0.03	0.062
cg21369801	17	80202961	CSNK1D	Body, 3'UTR	3.96E-07	0.09	3.32E-04	0.06	7.07E-04	0.06	0.0014
cg09280971	9	73036509	Intergenic		4.63E-07	0.03	8.84E-04	0.02	0.001	0.02	0.003
cg27178850	11	9039431	Intergenic		4.66E-07	0.06	2.11E-06	0.02	0.027	0.03	0.001
cg08059112	19	2294887	LINGO3	5'UTR	6.19E-07	0.05	4.17E-04	0.03	0.002	0.02	0.008
cg20509831	19	58866362	A1BG-AS1; A1BG; ZNF497	Body; TSS1500; 3'UTR	6.45E-07	0.04	0.007	0.05	6.53E-05	0.04	4.73E-04
cg01538969	6	30624636	DHX16	Body	7.09E-07	0.07	9.09E-04	0.05	0.001	0.04	6.89E-04
cg14595618	10	71135446	HK1	Body	7.63E-07	0.10	2.72E-07	0.02	0.088	0.01	0.28
cg07021906	16	87866833	SLC7A5	Body	9.19E-07	0.09	1.41E-04	0.04	0.011	0.05	0.001
cg07317306	1	110314824	Intergenic		9.79E-07	0.03	0.004	0.03	5.75E-04	0.02	0.003
cg03036214	15	63640658	CA12	Body	1.23E-06	0.09	1.38E-06	0.03	0.036	0.03	0.028

Chr = chromosome. a. GRCh37/hg19 assembly. b. Effect estimate from *limma* models adjusted for age, sex, BMI, self-reported smoking status, education (< high school, high school graduate or GED, > high school), study center (Arizona, Oklahoma, North and South Dakota), estimated glomerular filtration rate, and cell type proportion estimates. c. Urinary creatinine used to adjust for urine dilution (i.e., total urinary As concentration modeled in units of μg/g creatinine). d. Specific gravity (SG) used to adjust for urine dilution.

Table S8: Results of a linear model for the association between urinary As levels and M-values of cg14595618, annotated to *HK1*, including the interaction term ln(total urinary As) × diabetes status and adjusted age, sex, BMI, smoking status, education, study center (Arizona, Oklahoma, North and South Dakota), estimated glomerular filtration rate, and cell type proportion estimates.

Variable	*B*	*P*
ln(total urinary As)	0.03	0.017
Age	-0.003	0.003
Sex[a]	-0.09	<0.001
Smoking status[b]		
Former smoker	-0.04	0.022
Current smoker	0.09	<0.001
BMI	0.00	0.85
Diabetes	-0.07	0.15
Study center[c]		
Oklahoma	0.11	<0.001
North and South Dakota	0.07	0.003
eGFR	0.001	0.027
Cell type proportion		
CD8+ T	-3.32	<0.001
CD4+ T	-4.38	<0.001
NK	-3.11	<0.001
B cells	-5.30	<0.001
Monocytes	-5.16	<0.001
Granulocytes	-5.60	<0.001
ln(total urinary As) × diabetes	0.06	0.027

a. Reference = female. b. Reference = never smoker. c. Reference = Arizona.

Table S9: Differentially methylated CpG sites associated with ln(total urinary arsenic levels) at at $P_{FDR} < 0.05$ in our primary analyses, and results from sensitivity analyses performed on probe-type normalized data.

CpG	Chr	Position[c]	Gene	Feature category	Primary analyses[a]			Probe-type normalized[b]		
					Effect estimate	P	P_{FDR}	Effect estimate	P	P_{FDR}
cg06690548	4	139162808	SLC7A11	Body	-0.10	9.56E-10	5.88E-04	-0.11	3.01E-09	0.001
cg03497652	16	4751569	ANKS3	Body	0.09	1.49E-09	5.88E-04	0.09	1.47E-09	0.001
cg22294740	19	2294961	LINGO3	5'UTR	0.08	8.78E-09	0.002	0.08	9.72E-09	0.002
cg20493718	17	80202979	CSNK1D	Body, 3'UTR	0.12	1.04E-08	0.002	0.11	1.27E-08	0.002
cg18616702	1	150523808	ADAMTSL4; MIR4257	5'UTR; TSS1500	0.05	2.35E-08	0.004	0.06	2.63E-08	0.004
cg12116137	17	1576449	PRPF8	Body	0.08	1.22E-07	0.014	0.08	1.32E-07	0.017
cg12106731	1	36023279	NCDN; KIAA0319L	TSS200; TSS1500	0.05	1.24E-07	0.014	0.06	1.22E-06	0.050
cg00500428	16	3074502	HCFC1R1; THOC6	TSS1500; Body	-0.06	1.83E-07	0.017	-0.08	1.56E-07	0.017
cg04940901	3	52418250	DNAH1	Body	0.05	1.96E-07	0.017	0.06	1.75E-07	0.017
cg14827056	8	141550539	EIF2C2	Body	0.07	2.14E-07	0.017	0.06	2.40E-07	0.021
cg21369801	17	80202961	CSNK1D	Body, 3'UTR	0.07	3.96E-07	0.028	0.07	4.34E-07	0.031
cg09280971	9	73036509	Intergenic		0.03	4.63E-07	0.028	0.04	3.28E-07	0.026
cg27178850	11	9039431	Intergenic		0.04	4.66E-07	0.028	0.05	9.38E-07	0.044
cg08059112	19	2294887	LINGO3	5'UTR	0.04	6.19E-07	0.034	0.05	5.24E-07	0.033
cg20509831	19	58866362	A1BG-AS1; A1BG; ZNF497	Body; TSS1500; 3'UTR	0.05	6.45E-07	0.034	0.06	5.39E-07	0.033
cg01538969	6	30624636	DHX16	Body	0.06	7.09E-07	0.035	0.07	9.59E-07	0.044
cg14595618	10	71135446	HK1	Body	0.06	7.63E-07	0.035	0.07	8.64E-07	0.044
cg07021906	16	87866833	SLC7A5	Body	0.06	9.19E-07	0.040	0.06	1.02E-06	0.045
cg07317306	1	110314824	Intergenic		0.03	9.79E-07	0.041	0.04	1.45E-06	0.056
cg03036214	15	63640658	CA12	Body	0.05	1.23E-06	0.048	0.07	1.57E-06	0.056

Chr = chromosome. a. Effect estimate and *P*-values from *limma* models performed on data normalized using single sample Noob and adjusted for age, sex, BMI, self-reported smoking status, education (< high school, high school graduate or GED, > high school), study center (Arizona, Oklahoma, South Dakota), estimated glomerular filtration rate, and cell type proportion estimates. b. Effect estimate and *P*-values from *limma* models performed on data normalized using single sample Noob and RCP normalization and adjusted for age, sex, BMI, self-reported smoking status, education (< high school, high school graduate or GED, > high school), study center (Arizona, Oklahoma, South Dakota), estimated glomerular filtration rate, and cell type proportion estimates. c. GRCh37/hg19 assembly.

Table S10: Differentially methylated CpG sites associated with ln(total urinary arsenic levels) at at $P_{FDR} < 0.05$ in our primary analyses, and results from sensitivity analyses using ReFACTor, a reference-free approach to adjusted for cell type proportions.

CpG	Chr	Position[c]	Gene	Feature category	Reference-based adjustment for cell type proportions (Houseman regression calibration)[a]			Reference-free adjustment for cell type proportions (ReFACTor)[b]		
					Effect estimate	P	P_{FDR}	Effect estimate	P	P_{FDR}
cg06690548	4	139162808	SLC7A11	Body	-0.10	9.56E-10	5.88E-04	-0.10	3.51E-10	2.77E-04
cg03497652	16	4751569	ANKS3	Body	0.09	1.49E-09	5.88E-04	0.09	1.08E-09	4.25E-04
cg22294740	19	2294961	LINGO3	5'UTR	0.08	8.78E-09	0.002	0.08	2.56E-09	5.06E-04
cg20493718	17	80202979	CSNK1D	Body, 3'UTR	0.12	1.04E-08	0.002	0.12	1.90E-09	5.00E-04
cg18616702	1	150523808	ADAMTSL4; MIR4257	5'UTR; TSS1500	0.05	2.35E-08	0.004	0.05	2.69E-08	0.004
cg12116137	17	1576449	PRPF8	Body	0.08	1.22E-07	0.014	0.08	2.92E-07	0.017
cg12106731	1	36023279	NCDN; KIAA0319L	TSS200; TSS1500	0.05	1.24E-07	0.014	0.05	2.10E-07	0.015
cg00500428	16	3074502	HCFC1R1; THOC6	TSS1500; Body	-0.06	1.83E-07	0.017	-0.06	1.11E-07	0.012
cg04940901	3	52418250	DNAH1	Body	0.05	1.96E-07	0.017	0.04	1.68E-06	0.044
cg14827056	8	141550539	EIF2C2	Body	0.07	2.14E-07	0.017	0.05	7.60E-06	0.105
cg21369801	17	80202961	CSNK1D	Body, 3'UTR	0.07	3.96E-07	0.028	0.07	1.36E-07	0.013
cg09280971	9	73036509	Intergenic		0.03	4.63E-07	0.028	0.03	3.83E-07	0.020
cg27178850	11	9039431	Intergenic		0.04	4.66E-07	0.028	0.03	4.04E-05	0.210
cg08059112	19	2294887	LINGO3	5'UTR	0.04	6.19E-07	0.034	0.04	3.35E-08	0.004
cg20509831	19	58866362	A1BG-AS1; A1BG; ZNF497	Body; TSS1500; 3'UTR	0.05	6.45E-07	0.034	0.04	2.17E-05	0.171
cg01538969	6	30624636	DHX16	Body	0.06	7.09E-07	0.035	0.05	2.04E-07	0.015
cg14595618	10	71135446	HK1	Body	0.06	7.63E-07	0.035	0.06	4.68E-07	0.023
cg07021906	16	87866833	SLC7A5	Body	0.06	9.19E-07	0.040	0.06	2.75E-07	0.017
cg07317306	1	110314824	Intergenic		0.03	9.79E-07	0.041	0.03	1.04E-04	0.284
cg03036214	15	63640658	CA12	Body	0.05	1.23E-06	0.048	0.05	3.05E-07	0.0172

Chr = chromosome. a. Effect estimate and *P*-values from *limma* models adjusted for age, sex, BMI, self-reported smoking status, education (< high school, high school graduate or GED, > high school), study center (Arizona, Oklahoma, South Dakota), estimated glomerular filtration rate, and cell type proportion estimates from Houseman regression callibration. b. Effect estimate and *P*-values from *limma* models adjusted for age, sex, BMI, self-reported smoking status, education (< high school, high school graduate or GED, > high school), study center (Arizona, Oklahoma, South Dakota), estimated glomerular filtration rate, and six ReFACTor components. c. GRCh37/hg19 assembly.
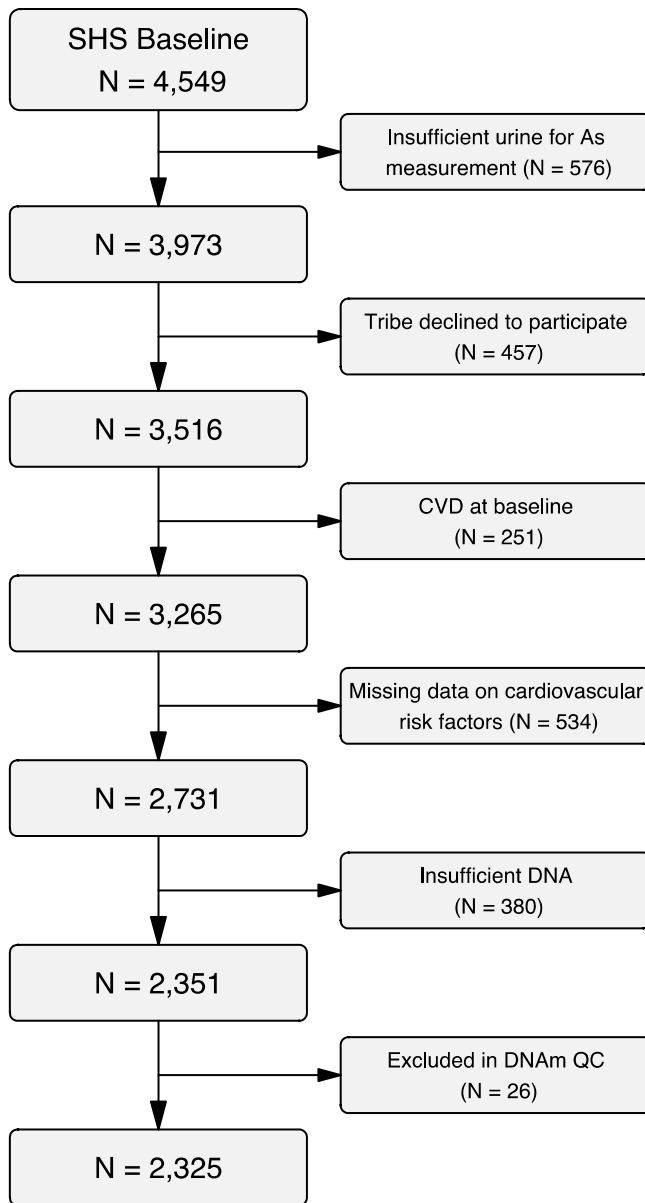
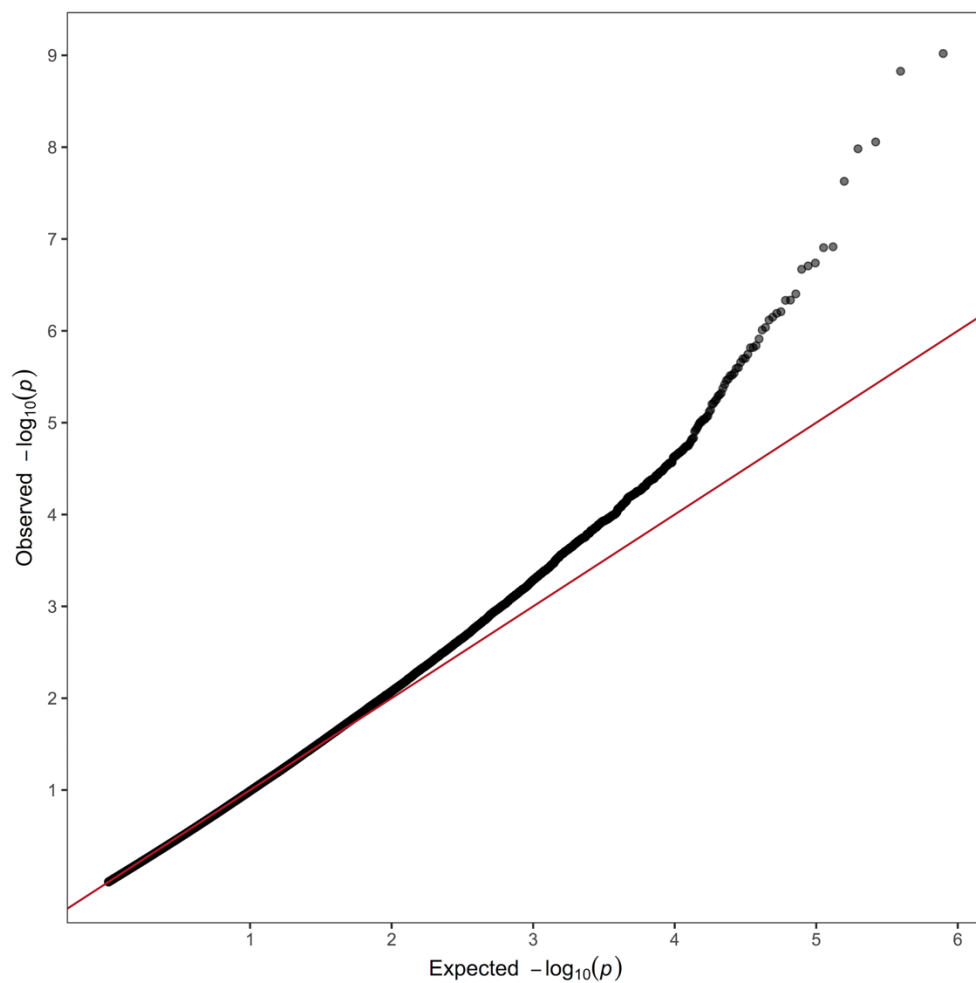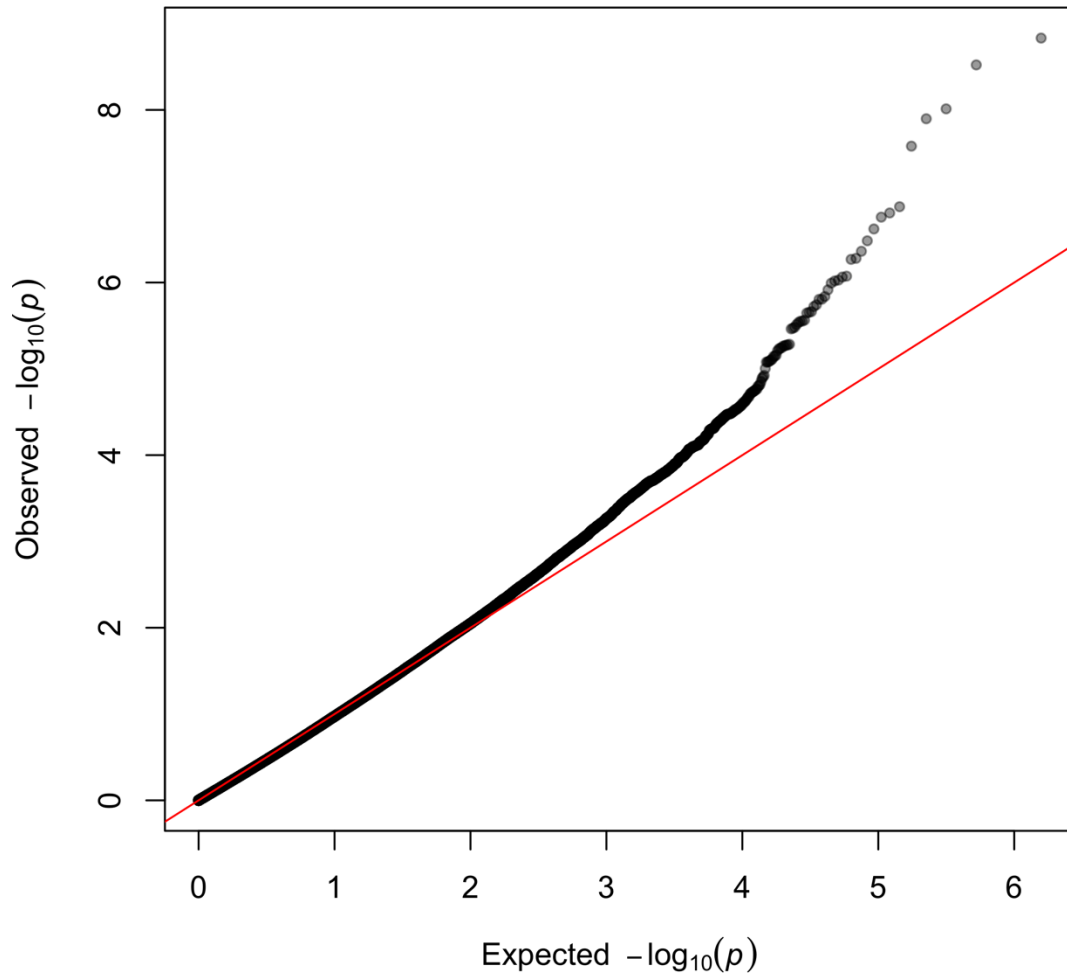**Figure S1:** Flowchart of eligible participant and participants selected for DNA methylation analysis.

Figure S2: Q-Q plot from *limma* models of M-values of ln(total urinary arsenic levels) adjusted for age, sex, BMI, self-reported smoking status, education (< high school, high school graduate or GED, > high school), study center (Arizona, Oklahoma, North and South Dakota), estimated glomerular filtration rate, and cell type proportion estimates.
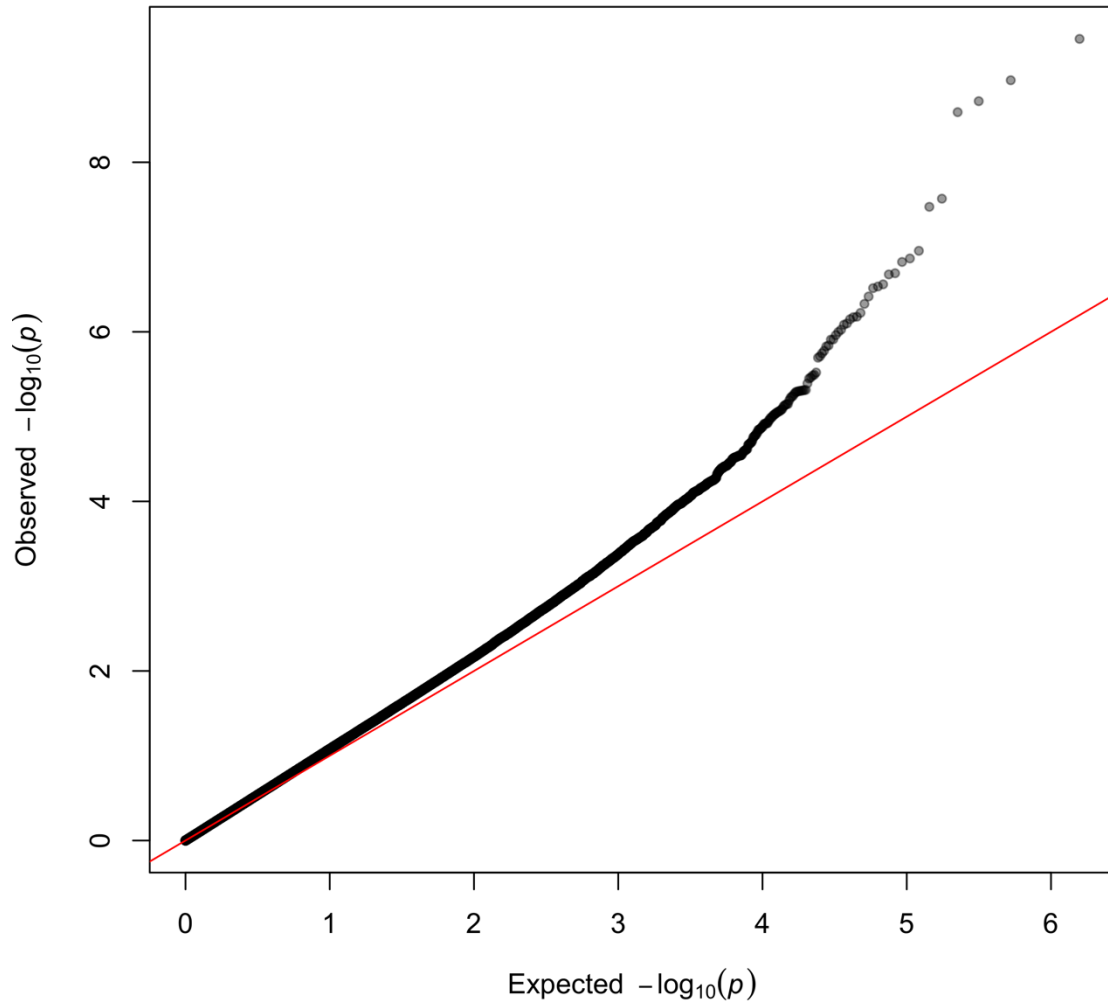
Figure S3: Sensitivity analyses for probe-type normalization. Q-Q plot from *limma* models of M-values normalized using single sample Noob and RCP normalization adjusted for age, sex, BMI, self-reported smoking status, education (< high school, high school graduate or GED, > high school), study center (Arizona, Oklahoma, North and South Dakota), estimated glomerular filtration rate, and cell type proportion estimates.

Figure S4: Sensitivity analyses for reference-free cell type proportion adjustment. Q-Q plot from *limma* models of M-values adjusted for age, sex, BMI, self-reported smoking status, education (< high school, high school graduate or GED, > high school), study center (Arizona, Oklahoma, South Dakota), estimated glomerular filtration rate, and six ReFACTor components.

References

1. 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 [Internet]. 2013 [cited 2016 Mar 22];121:971–7. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3733676&tool=pmcentrez&rendertype=abstract

2. Broberg K, Ahmed S, Engström K, Hossain MB, Jurkovic Mlakar S, Bottai M, Grandér M, Raqib R, Vahter M. Arsenic exposure in early pregnancy alters genome-wide DNA methylation in cord blood, particularly in boys. J Dev Orig Health Dis [Internet]. 2014 [cited 2016 Jun 1];5:288–98. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24965135

3. Kile ML, Houseman EA, Baccarelli AA, Quamruzzaman Q, Rahman M, Mostofa G, Cardenas A, Wright RO, Christiani DC. Effect of prenatal arsenic exposure on DNA methylation and leukocyte subpopulations in cord blood. Epigenetics [Internet]. 2014 [cited 2016 Mar 30];9:774–82. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24525453

4. Liu X, Zheng Y, Zhang W, Zhang X, LIoyd-Jones DM, Baccarelli AA, Ning H, Fornage M, He K, Liu K, et al. Blood methylomics in response to arsenic exposure in a low-exposed US population. J Expo Sci Environ Epidemiol [Internet]. 2014 [cited 2019 Jan 11];24:145–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24368509

5. Argos M, Chen L, Jasmine F, Tong L, Pierce BL, Roy S, Paul-Brutus R, Gamble M V., Harper KN, Parvez 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 [Internet]. 2015 [cited 2016 Jun 9];123:64–71. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25325195

6. Rojas D, Rager JE, Smeester L, Bailey KA, Drobná Z, Rubio-Andrade M, Stýblo M, García-Vargas G, Fry RC. Prenatal arsenic exposure and the epigenome: identifying sites of 5-methylcytosine alterations that predict functional changes in gene expression in newborn cord blood and subsequent birth outcomes. Toxicol Sci [Internet]. Oxford University Press; 2015 [cited 2019 Jan 11];143:97–106. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25304211

7. Cardenas A, Houseman EA, Baccarelli AA, Quamruzzaman Q, Rahman M, Mostofa G, Wright RO, Christiani DC, Kile ML. *In utero* arsenic exposure and epigenome-wide associations in placenta, umbilical artery, and human umbilical vein endothelial cells. Epigenetics [Internet]. 2015 [cited 2017 Feb 11];10:1054–63. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26646901

8. Green BB, Karagas MR, Punshon T, Jackson BP, Robbins DJ, Houseman EA, Marsit CJ. Epigenome-wide assessment of DNA methylation in the placenta and arsenic exposure in the New Hampshire Birth Cohort Study. Environ Health Perspect [Internet]. 2016 [cited 2019 Feb 11];124:1253–60. Available from: https://ehp.niehs.nih.gov/doi/10.1289/ehp.1510437

9. Kaushal A, Zhang H, Karmaus WJJ, Everson TM, Marsit CJ, Karagas MR, Tsai S-F, Wen H-J, Wang S-L. Genome-wide DNA methylation at birth in relation to *in utero* arsenic

exposure and the associated health in later life. Environ Heal [Internet]. BioMed Central; 2017 [cited 2019 Jan 11];16:50. Available from: http://ehjournal.biomedcentral.com/articles/10.1186/s12940-017-0262-0

10. 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 [Internet]. Academic Press; 2017 [cited 2019 Jan 11];321:57–66. Available from: https://www.sciencedirect.com/science/article/pii/S0041008X17300868?via%3Dihub#t0010

11. Bozack AK, Cardenas A, Quamruzzaman Q, Rahman M, Mostofa G, Christiani DC, Kile ML. DNA methylation in cord blood as mediator of the association between prenatal arsenic exposure and gestational age. Epigenetics [Internet]. 2018 [cited 2019 Feb 11];13:923–40. Available from: http://www.ncbi.nlm.nih.gov/pubmed/30175652

12. Guo X, Chen X, Wang J, Liu Z, Gaile D, Wu H, Yu G, Mao G, Yang Z, Di Z, et al. Multi-generational impacts of arsenic exposure on genome-wide DNA methylation and the implications for arsenic-induced skin lesions. Environ Int [Internet]. Pergamon; 2018 [cited 2019 Feb 11];119:250–63. Available from: https://www.sciencedirect.com/science/article/pii/S0160412018301521?via%3Dihub#s0055

13. Gliga AR, Engström K, Kippler M, Skröder H, Ahmed S, Vahter M, Raqib R, Broberg K. Prenatal arsenic exposure is associated with increased plasma IGFBP3 concentrations in 9-year-old children partly via changes in DNA methylation. Arch Toxicol [Internet]. 2018 [cited 2019 Feb 11];92:2487–500. Available from: http://www.ncbi.nlm.nih.gov/pubmed/29947889

14. Demanelis K, Argos M, Tong L, Shinkle J, Sabarinathan M, Rakibuz-Zaman M, Sarwar G, Shahriar H, Islam T, Rahman M, et al. Association of arsenic exposure with whole blood DNA methylation: an epigenome-wide study of Bangladeshi adults. Environ Health Perspect [Internet]. Environmental Health Perspectives; 2019 [cited 2019 Jun 3];127. Available from: https://ehp.niehs.nih.gov/doi/10.1289/EHP3849