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

Association of Arsenic Exposure with Whole Blood DNA Methylation: An Epigenome-Wide Study of Bangladeshi Adults

Kathryn Demanelis, Maria Argos, Lin Tong, Justin Shinkle, Mekala Sabarinathan, Muhammad Rakibuz-Zaman, Golam Sarwar, Hasan Shahriar, Tariqul Islam, Mahfuzar Rahman, Mohammad Yunus, Joseph H. Graziano, Karin Broberg, Karin Engström, Farzana Jasmine, Habibul Ahsan, and Brandon L. Pierce

Table of Contents

Table S1. Results from different epigenome-wide analysis methods and exposure modeling in HEALS (n=396).

Table S2. Association between estimated cell type proportions and $\log_2$-transformed urinary arsenic concentration (creatinine-adjusted) in HEALS (n=396).

Table S3. Regions of methylation associated with $\log_2$-transformed urinary arsenic concentration in HEALS (minimum FDR $< 10^{-4}$).

Table S4. Results from gene set enrichment analysis based on KEGG pathways for genes annotating to urinary arsenic-associated CpGs from HEALS discovery analysis and meta-analysis of HEALS and BEST.

Table S5. CpGs associated with $\log_2$-transformed urinary arsenic from meta-analysis ($p < 1.3 \times 10^{-7}$) and their relationship with gene expression in BEST (n=371).

Table S6. Summary of cis-mQTLs results from HEALS (n=389) for CpGs identified in meta-analysis.

Table S7. Summary of disease and trait associations with genetic variants within or near genes annotated to urinary arsenic-associated CpGs from the GWAS Catalog.

Additional File- Excel Document

References

Figure S1. QQ-plot of epigenome-wide analysis results of $\log_2$-transformed urinary arsenic concentration (creatinine-adjusted) in HEALS (n=396).
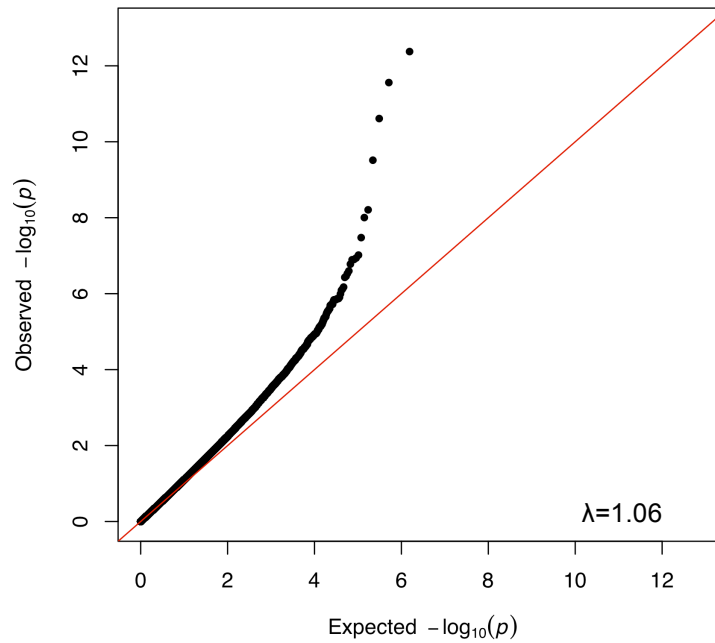
Using data from the EPIC (850K) array, associations between $\log_2$-transformed urinary arsenic concentration (creatinine-adjusted) and DNA methylation was evaluated at 771,192 CpGs and adjusted for sex, age, BMI, smoking status, and estimated SVs (n=27). The QQ-plot shows the observed p-values. The red line represents the expected distribution of p-values under the null, and λ corresponds to the genomic inflation factor (see **Methods**).

Figure S2. Urinary arsenic concentration (creatinine-adjusted) and DNA methylation at arsenic-associated CpGs (FDR < 0.05) in HEALS and reported Pearson correlations.

Pearson r: -0.250
p=4.8x10^{-7}

Pearson r: 0.218
p=1.3x10^{-5}

Pearson r: -0.146
p=3.7x10^{-3}

Pearson r: -0.126
p=1.2x10^{-2}

Pearson r: -0.167
p=8.5x10^{-4}

Pearson r: 0.210
p=2.4x10^{-5}

Pearson r: 0.216
p=1.5x10^{-5}

Pearson r: -0.106
p=3.4x10^{-2}
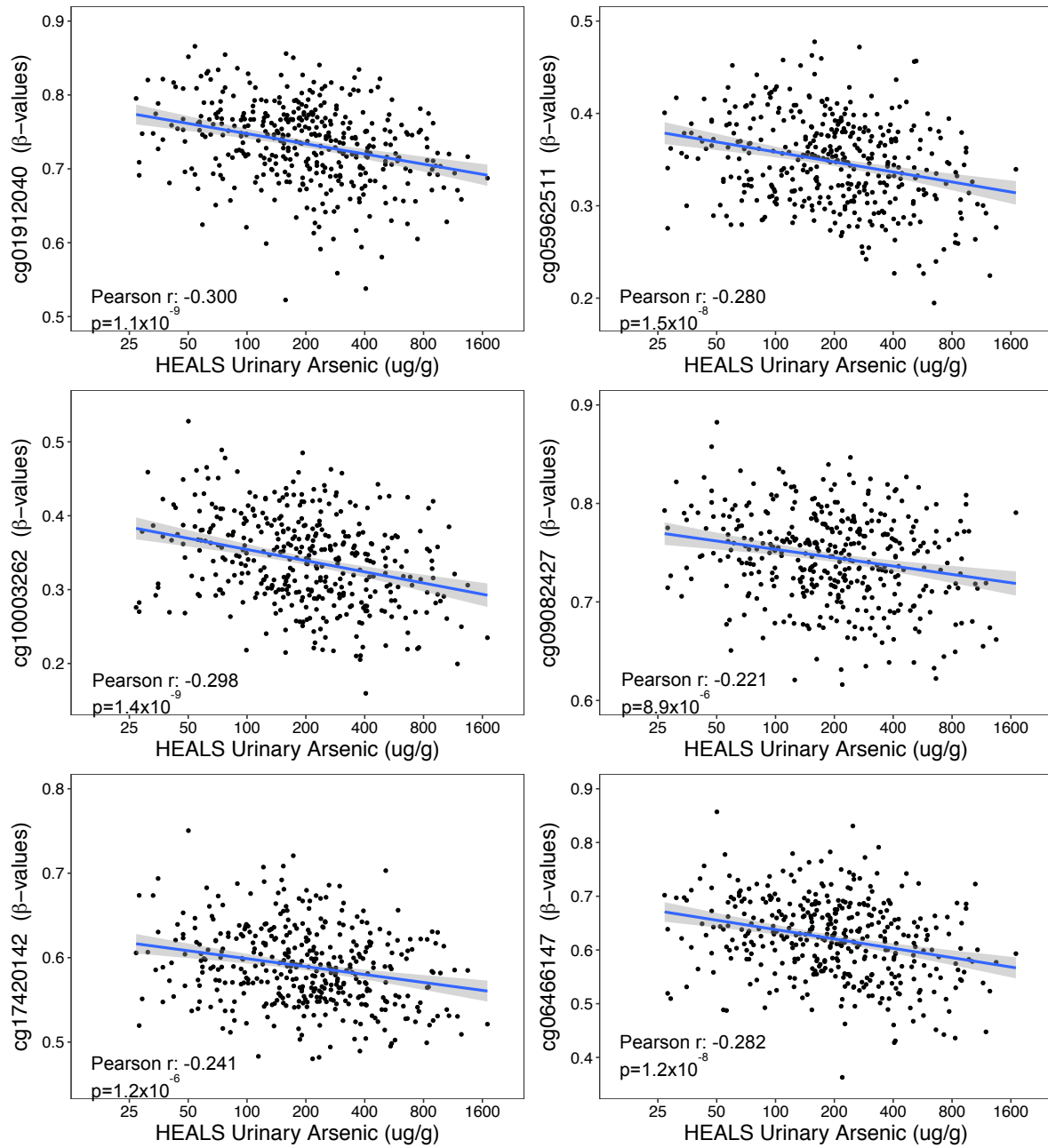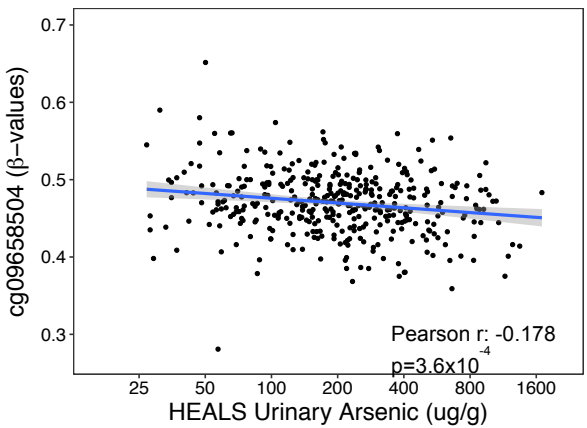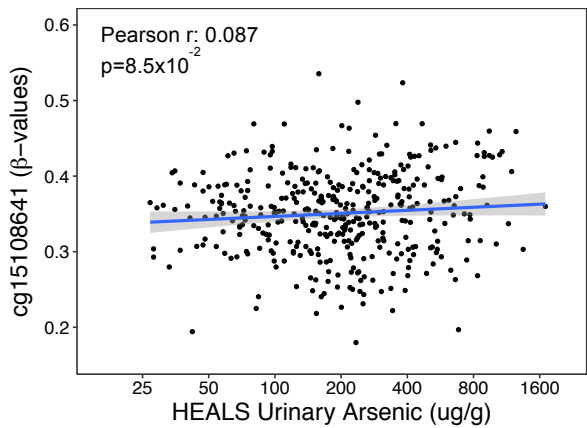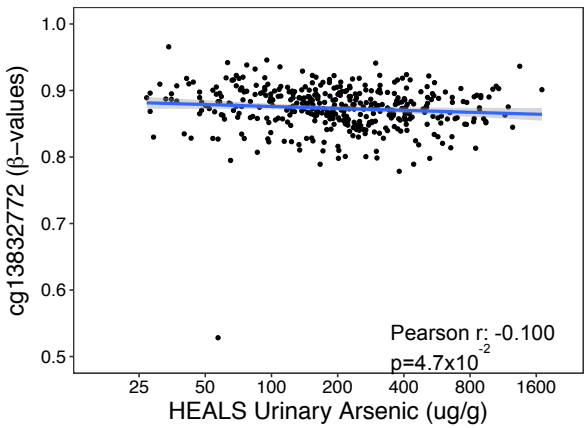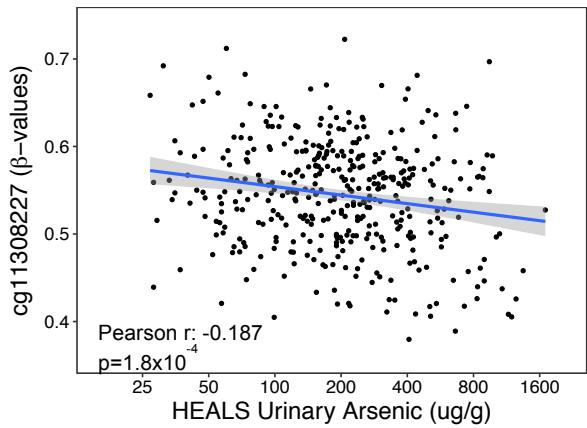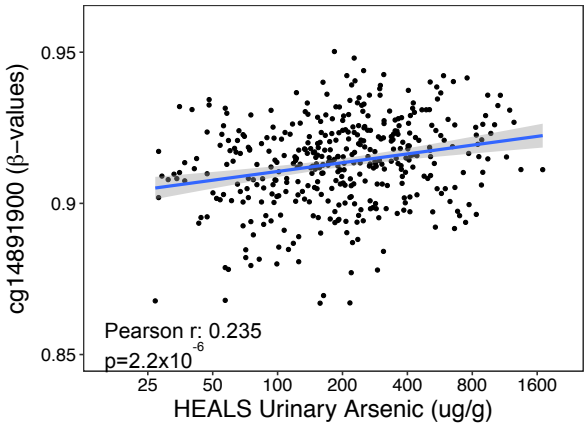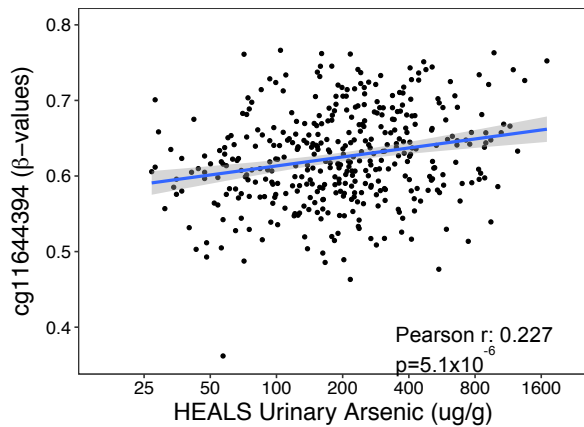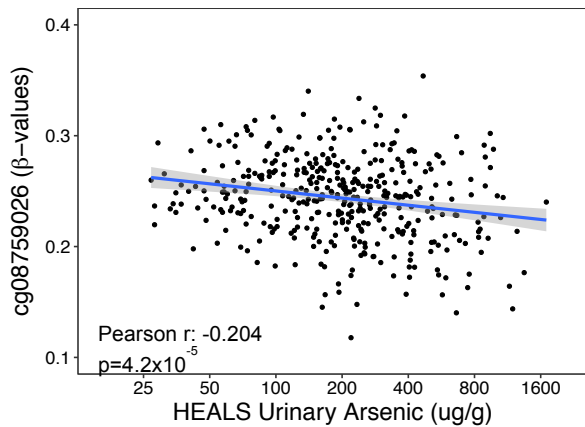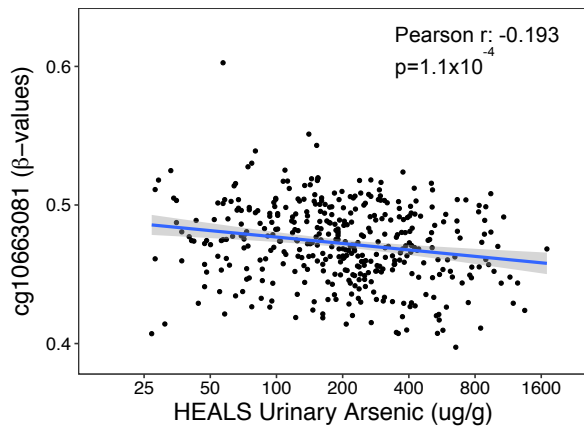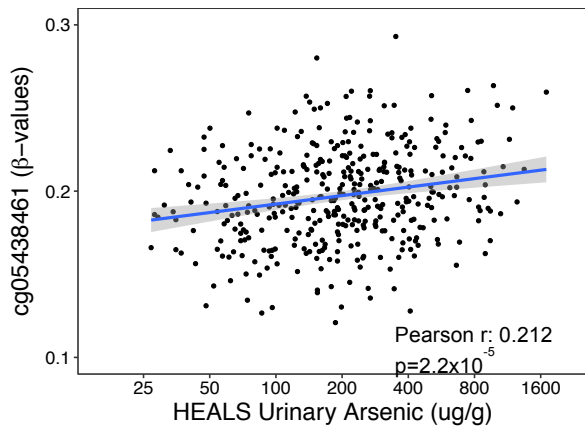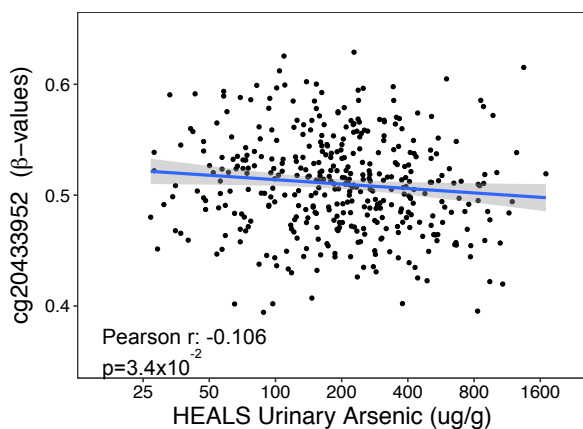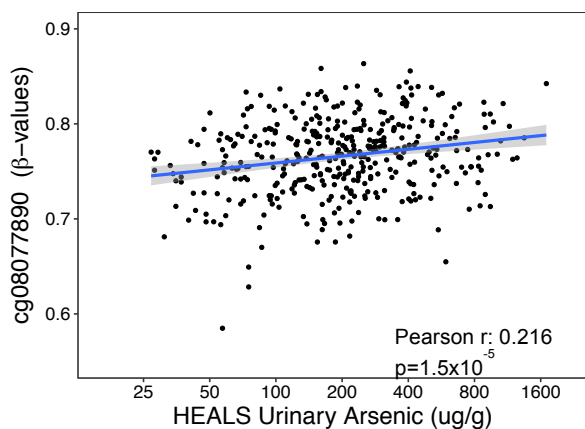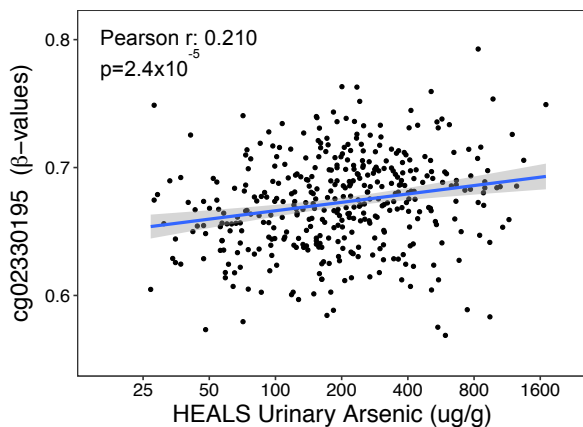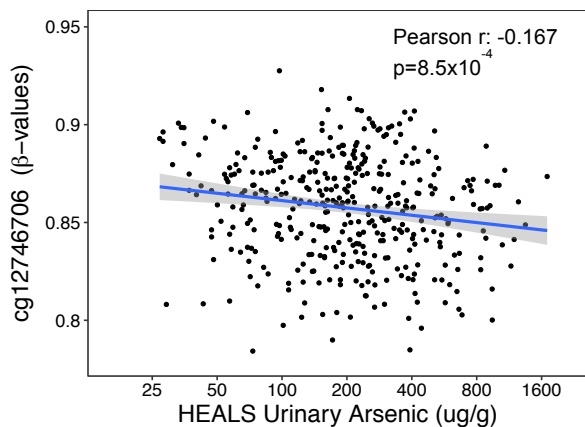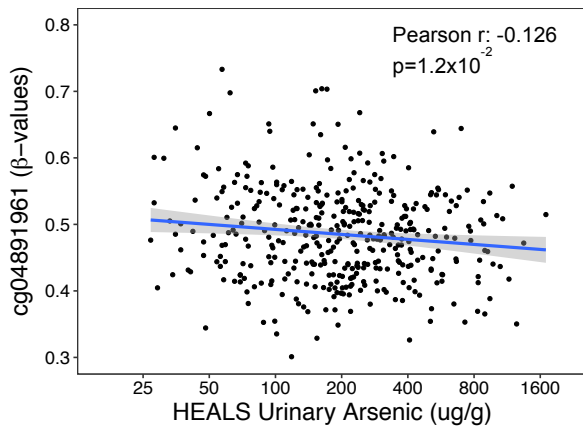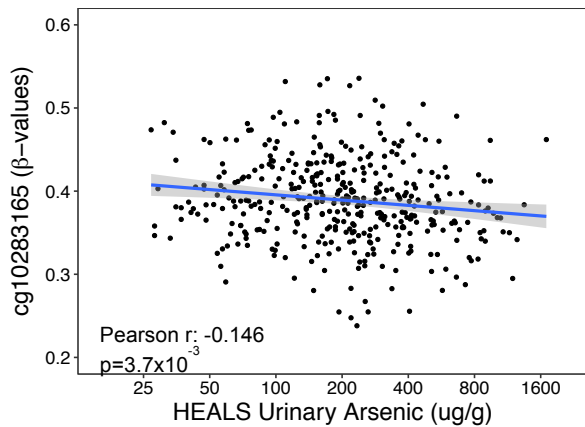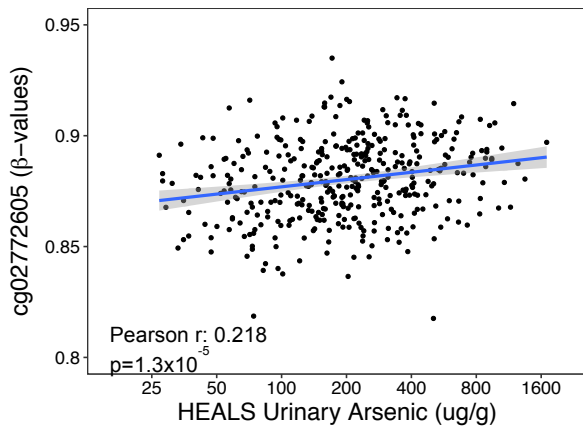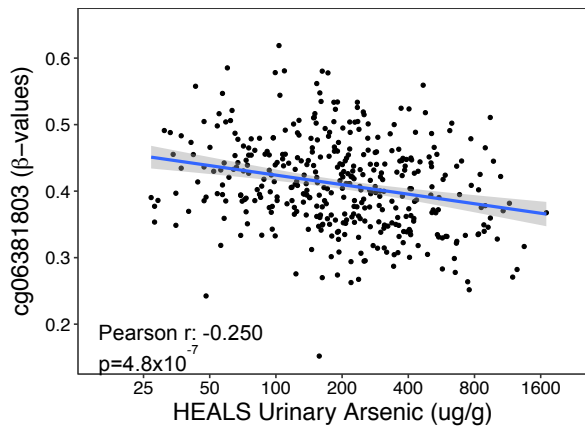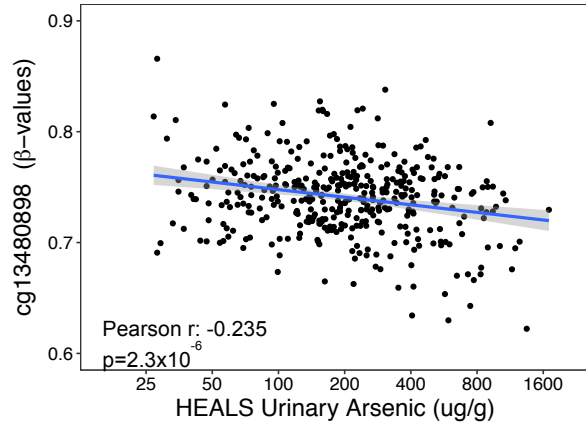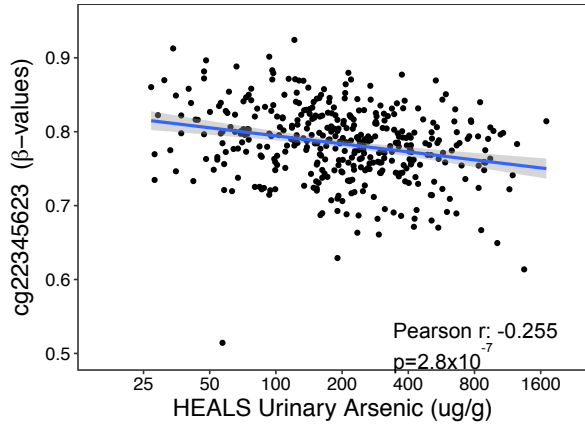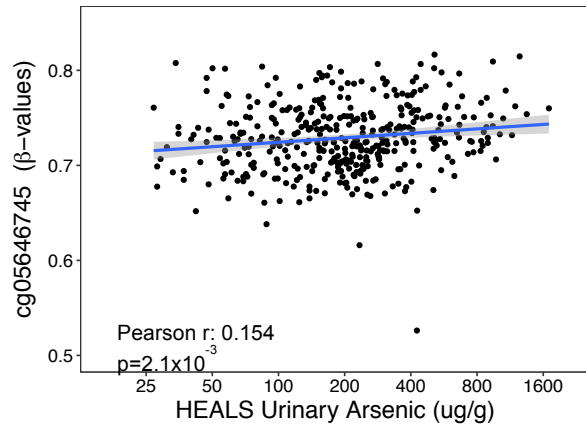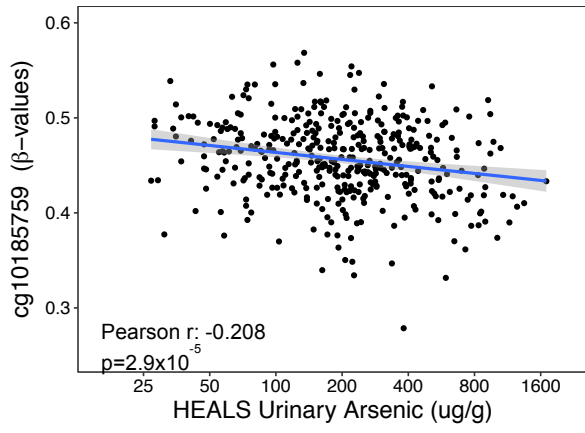
Figure S3. Relationship between water and urinary arsenic measurements in HEALS (n=396).
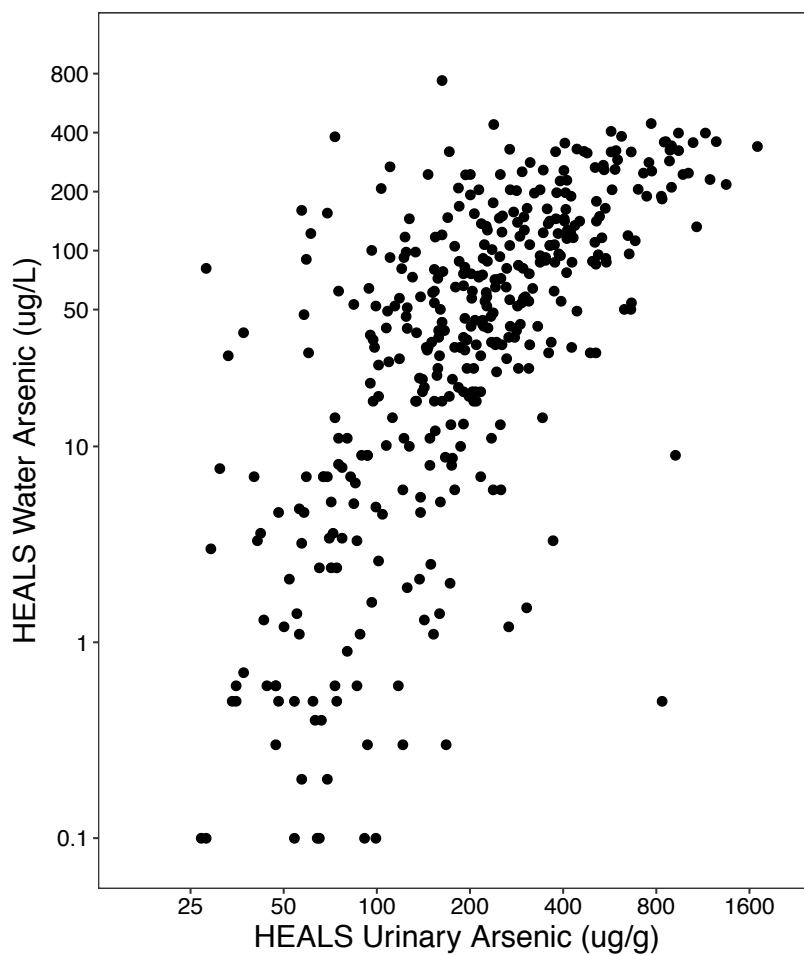
The Pearson correlation between untransformed continuous water and urinary arsenic (creatinine-adjusted) was 0.61 ($p<10^{-16}$). The correlation between $\log_2$-transformed continuous water and urinary arsenic (creatinine-adjusted) was 0.68 ($p<10^{-16}$).

Figure S4. QQ-plot of the observed p-values from epigenome-wide analysis of $\log_2$-transformed water arsenic concentration in HEALS (n=396).
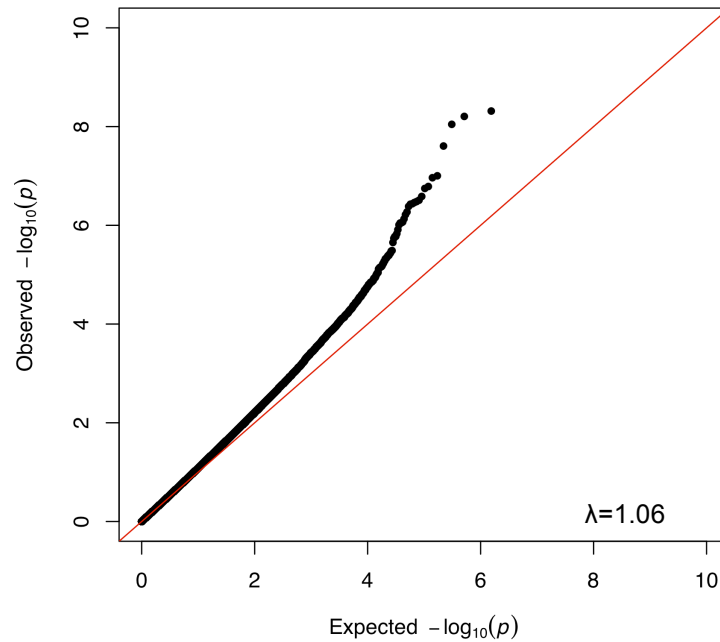
Using data from the EPIC (850K) array, the association between $\log_2$-transformed water arsenic concentration and DNA methylation was evaluated at 771,192 CpGs and adjusted for sex, age, BMI, smoking status, and estimated SVs (n=27). The QQ-plot shows the observed p-values. The red line represents expected distribution of p-values under the null, and λ corresponds to the genomic inflation factor.

Figure S5. Genome-wide associations between $\log_2$-transformed water arsenic concentration and CpG site-specific methylation in HEALS (n=396).
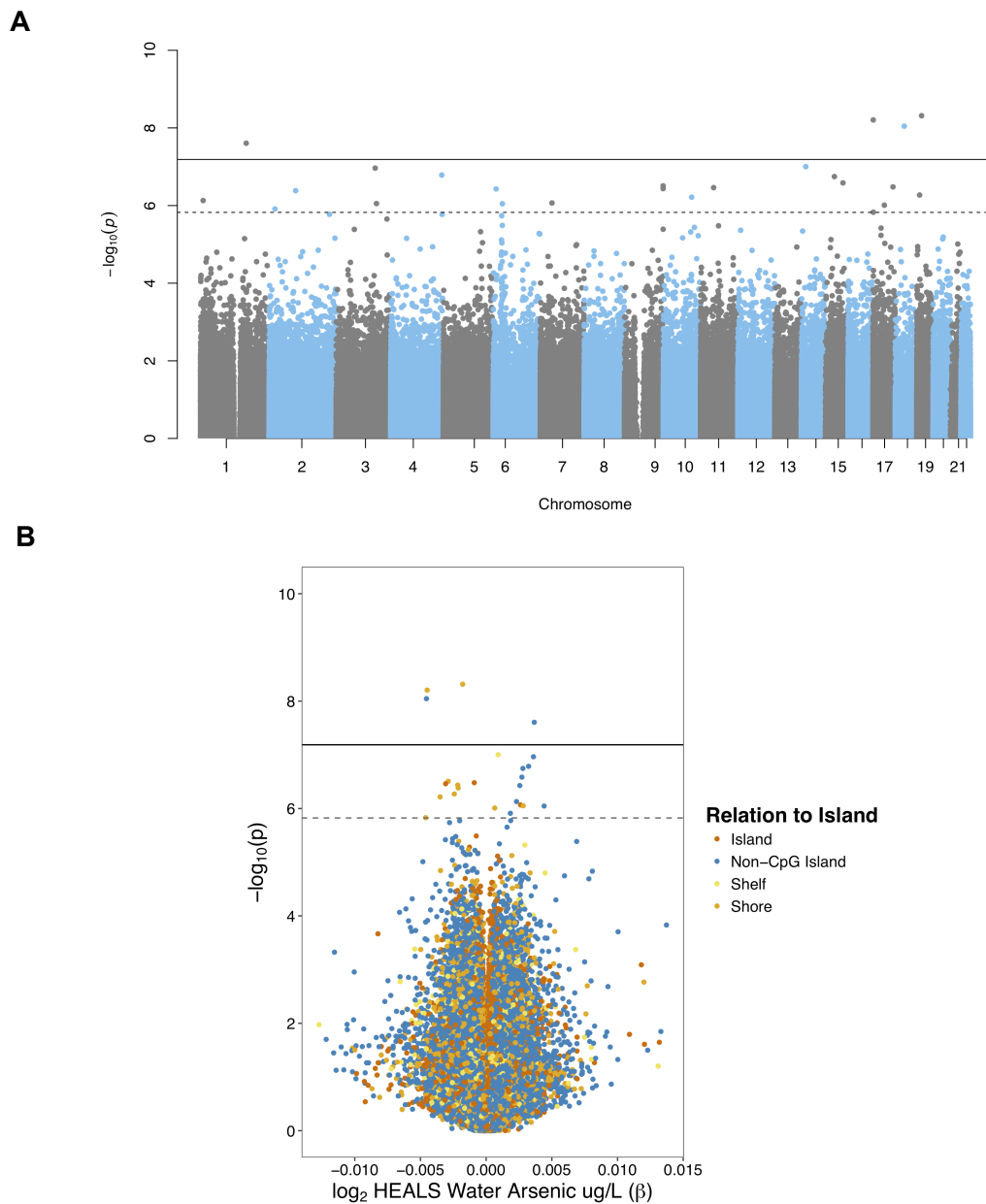
A

B

Using data from the EPIC (850K) array, associations between $\log_2$-transformed water arsenic concentration and DNA methylation were evaluated at 771,192 CpGs. (**A**) Manhattan plot of the location of each CpG and p-value for each arsenic-CpG association. (**B**) Volcano plot presenting association estimate and p-value for each arsenic-CpG association. Colors correspond to each CpG's relationship to CpG islands. In **A** and **B**, solid and dashed lines designate epigenome-wide significance (Bonferroni threshold $p=6.5\times10^{-8}$) and FDR < 0.05 ($p=1.5\times10^{-6}$), respectively.
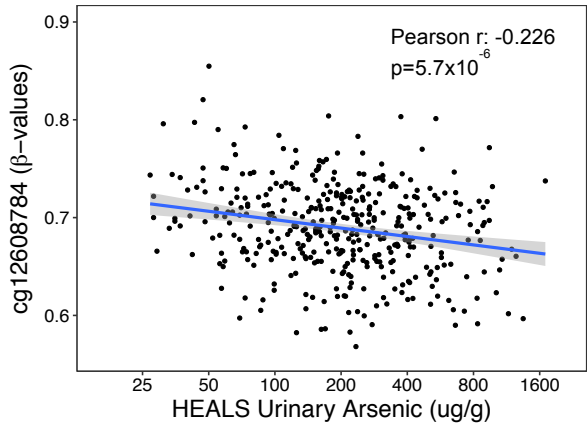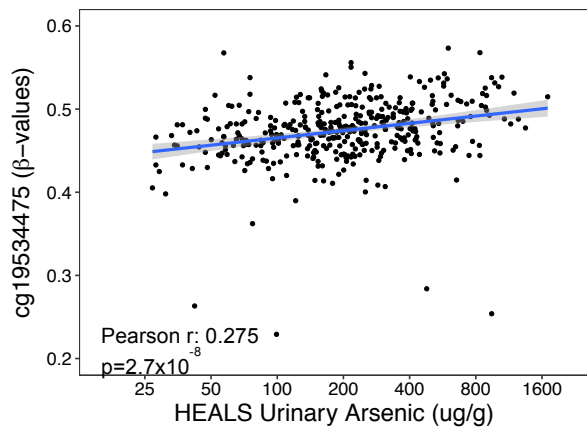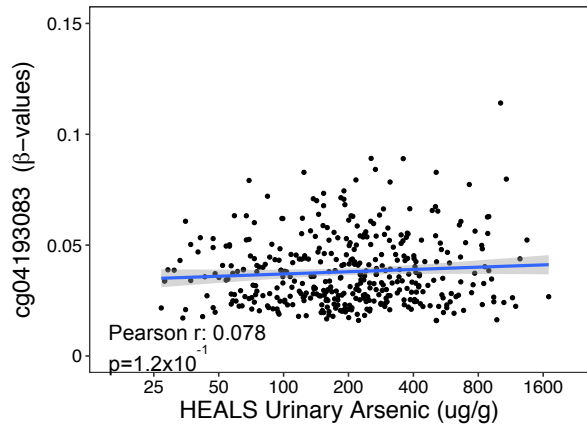
Figure S6. QQ-plot of epigenome-wide analysis results for log$_2$-transformed urinary arsenic concentration (creatinine-adjusted) in BEST (n=400).
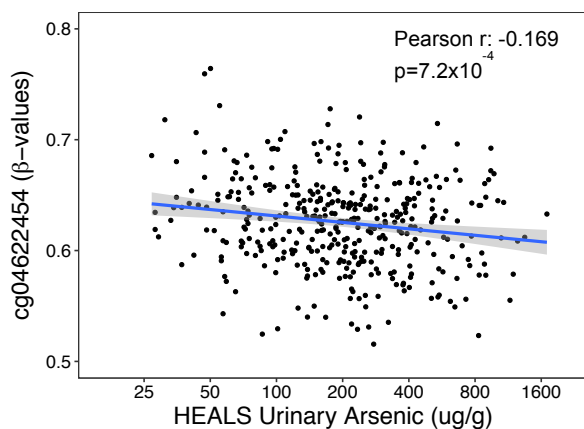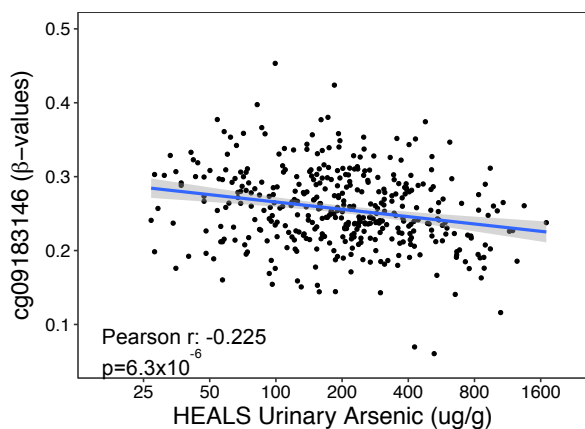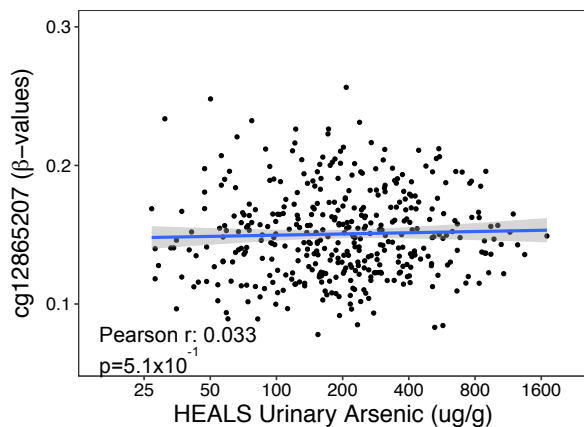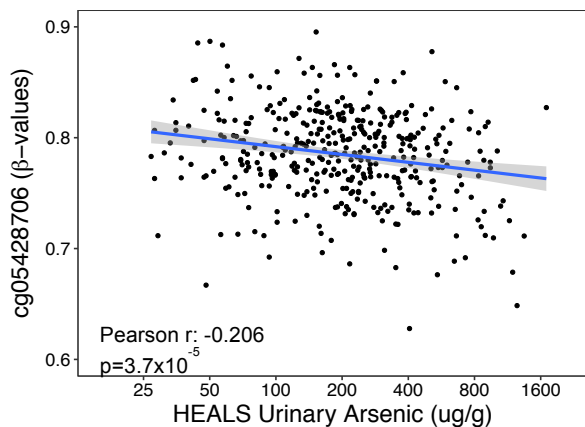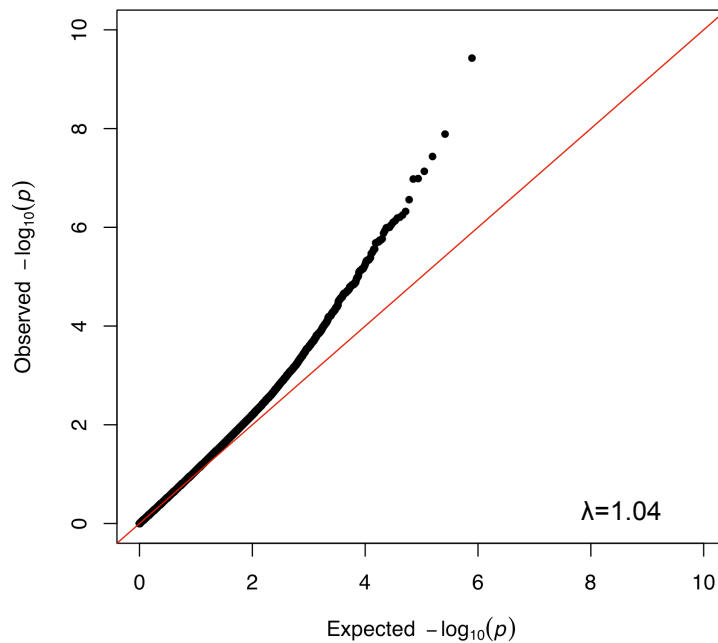

Using data from the 450K array, association between log$_2$-transformed urinary arsenic concentration (creatinine-adjusted) and DNA methylation was evaluated at 390,810 CpGs and adjusted for sex, age, BMI, smoking status, and estimated SVs (n=24). The QQ-plot shows the observed p-values. The red line represents expected distribution of p-values under the null, and λ corresponds to the genomic inflation factor.

A

B

Using data from the 450K array, association between $\log_2$-transformed urinary arsenic and methylation was evaluated at 390,810 CpGs. (**A**) Manhattan plot of the location of each CpG and p-value for each arsenic-CpG association. (**B**) Volcano plot presents association estimate and p-value for each arsenic-CpG association. Colors correspond to CpG's relationship to island. In **A** and **B**, solid and dashed lines designate epigenome-wide significance (Bonferroni threshold $p=1.3 \times 10^{-7}$) and FDR < 0.05 ($p=4.7 \times 10^{-6}$), respectively.

Figure S8. Enrichment of urinary arsenic-associated CpGs identified in meta-analysis of HEALS (n=396) and BEST (n=400) within genomic features stratified by direction of association (FDR < 0.05).
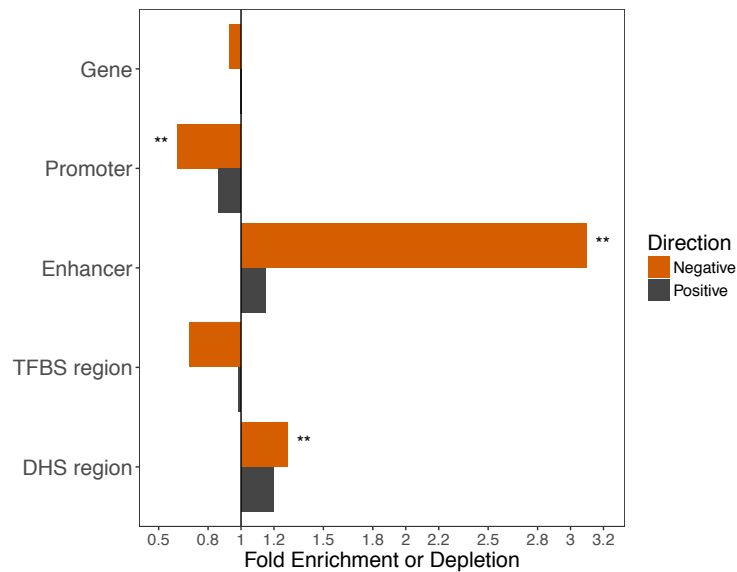
Arsenic-associated CpGs (FDR < 0.05) from meta-analysis of HEALS and BEST (n=221) were stratified by direction of association with urinary arsenic (negative [n=170] and positive [n=51]). Fold enrichment (FE) are plotted for each genomic feature comparing urinary arsenic-associated CpGs with FDR of 0.05 to the remaining CpGs below FDR of 0.05. Abbreviations correspond to transcription factor binding site (TFBS) and DNase I hypersensitive site (DHS). P-values are obtained from X^2-test comparing distribution of CpGs above and below FDR of 0.05 for each genomic feature (p<0.05 [*], p<10^{-3} [**], and p<10^{-6} [***]).

Table S1. Results from different epigenome-wide analysis methods and exposure modeling in HEALS (n=396).

Analysis Method	Modeled Exposure	λ[d]	Arsenic-associated CpGs identified		Overlap with log$_2$-transformed arsenic-associated CpGs identified from SV adjusted model (FDR < 0.05, n=34)		
			Bonferroni (p < 6.5x10^{-8})	FDR < 0.05	FDR < 0.05	p < 10^{-4}	p < 0.05
SV adjusted[b]	log2-transformed continuous	1.06	7	34	NA	NA	NA
SV adjusted[b]	untransformed continuous	1.05	9	37	16	29	34
SV adjusted[b]	ordinal quartiles[a]	1.05	6	11	11	32	34
Cell-type adjusted[c]	log2-transformed continuous	1.33	14	2191	23	23	33
Unadjusted (no covariates)	log2-transformed continuous	1.17	10	382	18	21	31

All models adjusted for sex, age, smoking status, and BMI (except for unadjusted model that contains only log$_2$-transformed urinary arsenic). Overlap was determined by examining the number of arsenic-associated CpGs identified in each analysis at different significance threshold (FDR < 0.05, p < 10^{-4} (suggestive epigenome-wide significance), and nominal p < 0.05) that were identified among the log$_2$-transformed arsenic-associated CpGs (FDR < 0.05, n=34) from covariate and SV adjusted analysis.

[a] An ordinal variable across integer-coded quartiles [1st (<113.5 ug/g), 2nd (113.5-201.5 ug/g), 3rd (201.5-350 ug/g), 4th(> 350 ug/g)].

[b] SV-adjusted model adjusts for covariates and 27 estimated surrogate variables based on Leek et al 2012.

[c] Cell-type adjusted model adjusts for covariates and estimated cell type proportions (monocytes, B-cells, CD4T, CD8T, granulocytes, and natural killer cells) based on Houseman et al 2012.

[d] λ corresponds to genomic inflation factor.

Table S2. Association between estimated cell type proportions and $\log_2$-transformed urinary arsenic concentration (creatinine-adjusted) in HEALS (n=396).

Estimate Cell Type (%)	mean (sd)	Pearson Correlation	p (correlation)	β (se)	p
Bcell	7.9 (2.9)	-0.110	0.0293	-0.2 (0.1)	0.0588
CD4T	13.9 (4.4)	-0.105	0.0371	-0.5 (0.2)	0.0136
CD8T	4.3 (4.3)	0.024	0.6299	0.0 (0.2)	0.8694
Gran	60.9 (9.2)	0.096	0.0556	0.8 (0.4)	0.0482
Mono	5.7 (2.4)	-0.007	0.8850	0.0 (0.1)	0.6207
NK	7.3 (4.6)	-0.043	0.3906	-0.2 (0.2)	0.3772

Reported Pearson correlation between each cell type proportion and $\log_2$-transformed urinary arsenic concentration (creatinine-adjusted) in HEALS. Estimated cell types expressed as percentages. Results from linear model were adjusted for age, sex, smoking status, and BMI.

Table S3. Regions of methylation associated with log$_2$-transformed urinary arsenic concentration in HEALS (minimum FDR < 10^{-4}).

Chromosome	Start	End	Gene	Location	Region Width (basepairs)	minimum FDR	# CpGs (CpGs p<0.05)	Mean Effect Size (β)
17	1106553	1107460			908	6.4E-26	6(6)	-0.008
9	140348828	140349969	NSMF	Body	1142	6.7E-20	10(4)	-0.003
10	102729375	102730130			756	6.7E-20	4(3)	-0.005
17	79201972	79203106	TEPSIN	Body/3'UTR	1135	1.7E-16	9(7)	-0.005
16	68000829	68002044	SLC12A4	Promoter, Body	1216	7.8E-12	8(6)	-0.002
3	101443258	101443992	CEP97	Promoter, 1st Exon, Body	735	1.3E-10	13(10)	0.002
10	135172031	135172491	FUOM	Promoter	461	1.9E-10	3(2)	0.003
19	46998383	46999757	PNMAL2	Promoter, 5'UTR, 1st Exon	1375	1.5E-09	15(9)	-0.003
1	161171116	161172291	NDUFS2	Promoter, 5'UTR	1176	2.0E-09	17(7)	0.002
6	31870476	31870990	ZBTB12	Promoter	515	1.4E-08	7(5)	0.004
17	17109800	17110641	PLD6	Promoter	842	3.0E-08	9(6)	-0.007
4	4763259	4763753			495	4.3E-08	7(6)	0.007
22	50739843	50740499	PLXNB2	5'UTR	657	5.3E-08	5(5)	-0.004
11	2919689	2920735	SLC22A18	Promoter	1047	6.5E-08	15(9)	-0.003
1	84326185	84326856			672	1.4E-07	9(8)	-0.003
19	49843160	49843922	CD37	Body, 3'UTR	763	1.4E-07	5(4)	-0.005
11	64980819	64981596	SLC22A20	Promoter, 1st Exon	778	1.4E-07	9(9)	-0.005
19	40871421	40871823	PLD3	5'UTR	403	1.6E-07	4(4)	-0.007
10	77164720	77165618	C10orf41	Body	899	5.1E-07	7(6)	0.006
9	139648911	139649039	LCN8	Body, 3'UTR	129	6.0E-07	2(2)	-0.008
1	9375862	9376492	SPSB1	5'UTR	631	7.8E-07	3(3)	-0.005
6	32063459	32064258	TNXB	Body	800	9.0E-07	24(12)	0.006
6	32847441	32847845	PPP1R2P1	Body	405	1.3E-06	13(10)	0.005
5	171986065	171986331			267	1.8E-06	4(4)	0.005
19	18496756	18497201	GDF15	Promoter, 5'UTR, 1st Exon	446	1.8E-06	7(7)	-0.004
1	155188904	155188982	GBAP1	Body	79	2.0E-06	2(1)	-0.009
17	42081594	42082036	NAGS/PYY	Promoter, 1st Exon/Promoter, 1st Exon, 5'UTR	443	2.1E-06	10(6)	-0.002
17	41323431	41323562	NBR1	5'UTR	132	2.2E-06	2(2)	0.001
17	27939837	27939900	ANKRD13B	Body, Exon Binding	64	2.4E-06	2(2)	-0.008
10	102279330	102280155	SEC31B	Promoter, 5'UTR	826	6.4E-06	14(9)	0.003
20	48185517	48186174	PTGIS	Promoter	658	6.4E-06	5(5)	-0.008
7	73476337	73476675	ELN	Body	339	6.9E-06	4(3)	0.004
13	27999177	27999546	GTF3A	Body	370	7.0E-06	4(4)	-0.011
14	63671356	63671737	RHOJ	5'UTR, 1st exon	382	1.0E-05	4(4)	-0.021
10	134254845	134255097			253	1.1E-05	4(2)	-0.002
19	17375523	17375799	USHBP1	Promoter	277	1.3E-05	4(2)	-0.002
4	165877875	165878219	C4orf39/TRIM61	Promoter/Body	345	1.8E-05	7(6)	-0.004
2	209224225	209224437	PTH2R	Promoter	213	1.9E-05	5(4)	-0.005
6	31607340	31607648	BAT3	Body	309	2.6E-05	5(3)	0.004
2	241856000	241856143			144	3.4E-05	3(3)	0.008
17	39968838	39969397	FKBP10/P3H4	Promoter,5'UTR, 1st Exon / Promoter	560	3.4E-05	10(4)	-0.002
19	7764345	7764649	FCER2	Body	305	4.2E-05	3(2)	-0.005
8	39172020	39172120	ADAM5P	Promoter	101	5.5E-05	6(4)	0.008
16	4421445	4421486	VASN/CORO7	Promoter/Body	42	7.3E-05	2(2)	-0.003
12	4417142	4417170			29	9.2E-05	2(2)	0.011

Analyzed using DMRcate (Peters et al. 2015) and tested for regions with λ=1,000 basepairs and C=2. Gene assignments and genic locations are based on UCSC gene annotation. Mean effect size (β) is average change in methylation (expressed as beta-values) associated with log$_2$-transformed urinary arsenic concentration across the differentially methylated region (DMR).

Table S4. Results from gene set enrichment analysis based on KEGG pathways for genes annotating to urinary arsenic-associated CpGs from HEALS discovery analysis and meta-analysis of HEALS and BEST.

		top 500 CpGs[a]		FDR < 0.05[b]	
	n	genes	p	genes	p
HEALS					
Hematopoietic cell lineage	90	4	3.7×10^{-2}		
META-ANALYSIS					
African trypanosomiasis	34	4	2.2×10^{-3}	2	2.6×10^{-2}
Malaria	44	4	7.0×10^{-3}	2	4.7×10^{-2}
Pancreatic secretion	91	6	1.2×10^{-2}	4	1.1×10^{-2}
Ether lipid metabolism	47	3	3.9×10^{-2}	2	4.5×10^{-2}
Linoleic acid metabolism	29	2	4.0×10^{-2}	2	9.4×10^{-3}
alpha-Linolenic acid metabolism	25	2	4.5×10^{-2}	2	1.1×10^{-2}
Ras signaling pathway	223	10	4.8×10^{-2}	6	3.7×10^{-2}
Rheumatoid arthritis	85	6	2.8×10^{-3}		
Cell adhesion molecules (CAMs)	135	9	2.9×10^{-3}		
Morphine addiction	88	8	3.9×10^{-3}		
T cell receptor signaling pathway	97	7	5.0×10^{-3}		
Osteoclast differentiation	118	7	7.4×10^{-3}		
Epithelial cell signaling in Helicobacter pylori infection	66	5	8.1×10^{-3}		
Rap1 signaling pathway	203	11	1.3×10^{-2}		
MAPK signaling pathway	283	14	1.3×10^{-2}		
Chagas disease (American trypanosomiasis)	98	6	1.4×10^{-2}		
Collecting duct acid secretion	27	3	1.4×10^{-2}		
Estrogen signaling pathway	137	8	1.5×10^{-2}		
Th17 cell differentiation	102	6	2.0×10^{-2}		
Vibrio cholerae infection	49	4	2.1×10^{-2}		
NF-kappa B signaling pathway	89	5	2.9×10^{-2}		
Intestinal immune network for IgA production	44	3	3.0×10^{-2}		
Alcoholism	165	7	3.5×10^{-2}		
Calcium signaling pathway	178	9	3.8×10^{-2}		
Hematopoietic cell lineage	90	4	3.8×10^{-2}		
Staphylococcus aureus infection	56	3	4.1×10^{-2}		
Amoebiasis	92	5	4.2×10^{-2}		
Influenza A	143	7	4.3×10^{-2}		
Neuroactive ligand-receptor interaction	315	10	4.7×10^{-2}		
Cysteine and methionine metabolism	45	3	4.8×10^{-2}		
Metabolic pathways	1182			16	2.4×10^{-2}
Fat digestion and absorption	41			2	3.3×10^{-2}
Acute myeloid leukemia	63			3	3.6×10^{-2}
Homologous recombination	39			2	4.0×10^{-2}
Retrograde endocannabinoid signaling	133			4	4.5×10^{-2}
Carbohydrate digestion and absorption	37			2	5.0×10^{-2}

Gene annotations for arsenic-associated CpGs are provided as gene set input for the *gometh* function. Results are presented for KEGG pathways (n=324 sets) with at least two genes from arsenic-associated CpG gene set and p < 0.05. Table reports total number of genes in pathway (n), genes identified in pathway from CpG gene set, and p-value from CpG bias-corrected hypergeometric test. Background gene set for 450K is 19,246 genes and EPIC is 23,234 genes.
[a]Top 500 arsenic-associated CpGs, annotated to 348 unique genes in the HEALS discovery analysis (EPIC array), 338 unique genes in the meta-analysis (450K array).
[b]Arsenic-associated CpGs (FDR < 0.05), annotated to 21 unique genes in the HEALS discovery analysis, 153 unique genes in the meta-analysis

Table S5. CpGs associated with $\log_2$-transformed urinary arsenic from meta-analysis ($p < 1.3 \times 10^{-7}$) and their relationship with gene expression in BEST (n=371).

Name	Chr	Position	Nearest Gene	Feature	Distance	Direction of Association	Illumina ID	Pearson correlation	$p_{correlation}$	$p_{adjusted}$
							Gene Expression Analysis			
cg13480898	19	10195914	C19orf66	Promoter	892	↓↓	ILMN_1750400	0.163	1.6×10^{-3}	5.9×10^{-3}
cg07207669	1	155102388	EFNA1	Body	2039	↓↓	ILMN_2371055	0.070	1.8×10^{-1}	1.4×10^{-1}
							ILMN_2371053	-0.021	6.9×10^{-1}	6.3×10^{-1}
cg06381803	19	46119475	EML2	Body	29300	↓↓	ILMN_3240541	-0.107	3.9×10^{-2}	1.8×10^{-2}
cg09183146	16	1429863	UNKL	Promoter	34842	↓↓	ILMN_1692826	-0.044	4.0×10^{-1}	9.0×10^{-1}
							ILMN_2126408	0.013	8.1×10^{-1}	6.2×10^{-1}
cg08759026	11	69061454	MYEOV	Promoter	159	↓↓	ILMN_2198413	-0.083	1.1×10^{-1}	9.5×10^{-2}
cg17489312	1	9376039	SPSB1	5'UTR	23098	↓↓	ILMN_1714170	0.001	9.8×10^{-1}	9.3×10^{-1}
cg00472758	16	2552820	TBC1D24	3'UTR	4581	↓↓	ILMN_2060212	-0.087	9.3×10^{-2}	9.7×10^{-2}
							ILMN_1780197	0.001	9.8×10^{-1}	9.7×10^{-1}
cg05425326	16	58439361	GINS3	3'UTR	13063	↑↑	ILMN_1754272	-0.001	9.8×10^{-1}	8.3×10^{-1}
cg26435149	3	55605611	ERC2	3'UTR	896780	↓↓	ILMN_1769195	0.041	4.3×10^{-1}	5.3×10^{-1}
cg17393635	19	49843565	CD37	Body	4888	↓↓	ILMN_1786176	-0.075	1.5×10^{-1}	7.4×10^{-2}
							ILMN_2375825	-0.058	2.6×10^{-1}	1.4×10^{-1}
cg19240637	2	7172297	RNF144A	Body	114774	↓↓	ILMN_3238326	0.220	1.9×10^{-5}	2.4×10^{-5}
cg07782285	19	13085442	DAND5	3'UTR	5118	↓↓	ILMN_2230178	-0.029	5.8×10^{-1}	4.9×10^{-1}
							ILMN_2378654	-0.109	3.5×10^{-2}	6.2×10^{-2}
cg26390598	21	41032396	B3GALT5	5'UTR	3142	↑↑	ILMN_1698756	-0.056	2.8×10^{-1}	1.3×10^{-1}
							ILMN_1800713	-0.005	9.3×10^{-1}	4.8×10^{-1}
cg04622454	9	140349128	NELF	Body	4658	↓↓	ILMN_1665095	0.112	3.1×10^{-2}	5.6×10^{-2}
cg12261095	19	7764345	FCER2	Body	2687	↓↓	ILMN_1662451	-0.133	1.0×10^{-2}	8.9×10^{-3}
cg04459545	19	17375685	USHBP1	Promoter	61	↓↓	ILMN_1693274	-0.034	5.1×10^{-1}	5.6×10^{-1}
cg14145338	9	139649039	LCN8	Body	3692	↓↓	ILMN_1761917	-0.175	7.0×10^{-4}	5.3×10^{-4}
cg04920032	12	50262986	FAIM2	3'UTR	34774	↓↓	ILMN_1803855	-0.081	1.2×10^{-1}	1.8×10^{-1}
cg18413900	12	58160989	CYP27B1	Promoter	13	↑↑	ILMN_1740418	-0.003	9.5×10^{-1}	9.2×10^{-1}
cg01757312	13	112720565	SOX1	Promoter	1348	↑↑	ILMN_1716070	-0.008	8.8×10^{-1}	8.4×10^{-1}
cg05816193	6	26018127	HIST1H1A	Promoter	87	↓↓	ILMN_1692831	-0.098	5.9×10^{-2}	6.7×10^{-2}
cg02306995	3	122635049	SEMA5B	Body	111610	↓↓	ILMN_1794370	0.110	3.5×10^{-2}	8.6×10^{-2}
							ILMN_1736300	0.053	3.1×10^{-1}	4.1×10^{-1}
						↓↓	ILMN_1659565	0.029	5.8×10^{-1}	8.7×10^{-1}
cg04875062	1	17305562	MFAP2	5'UTR	2519		ILMN_1787981	0.011	8.3×10^{-1}	5.3×10^{-1}
							ILMN_2245087	0.000	9.9×10^{-1}	3.9×10^{-1}
cg13764516	9	139648911	LCN8	3'UTR	3820	↓↓	ILMN_1761917	-0.142	6.4×10^{-3}	5.4×10^{-3}
						↓↓	ILMN_1655387	-0.024	6.5×10^{-1}	6.1×10^{-1}
cg23050300	1	3281321	PRDM16	Body	295579		ILMN_1784094	-0.022	6.7×10^{-1}	6.0×10^{-1}
							ILMN_1697631	0.013	8.0×10^{-1}	7.4×10^{-1}
							ILMN_2343264	0.008	8.7×10^{-1}	7.1×10^{-1}
							ILMN_2364272	0.032	5.3×10^{-1}	3.2×10^{-1}
cg18050715	13	97996992	MBNL2	Body	69106	↑↑	ILMN_1684316	-0.022	6.7×10^{-1}	7.4×10^{-1}
							ILMN_1658015	-0.009	8.6×10^{-1}	9.5×10^{-1}
cg27092191	16	31884699	ZNF267	Promoter	380	↓↓	ILMN_1671010	0.039	4.5×10^{-1}	5.2×10^{-1}

We restricted to CpGs that annotated to promoter or gene region from results obtained from the meta-analysis ($p < 1.3 \times 10^{-7}$) (**Table 3**). Illumina probes annotated to gene containing CpG, and results are presented for all probes for a gene. Direction of association indicated by arrows from HEALS and BEST, and downward and upward arrows correspond to inverse and positive association with CpG methylation and increasing urinary arsenic exposure. Pearson correlations were computed between methylation and expression and p-value from adjusted analysis (age and sex) is reported.

Table S6. Summary of cis-mQTLs results from HEALS (n=389) for CpGs identified in meta-analysis.

CpG	CpG Location		SNPs within 1 megabase window		Lead cis-mQTL pair			
	Chr	Position	Total Tested	cis-mQTLs (FDR < 0.01)	Lead SNP	Lead SNP Position	Distance from CpG	p
cg17489312	1	9376039	6063	103	rs9726646	9375309	730	3.4×10^{-30}
cg05962511	10	102730022	4658	194	rs1570171	102731759	-1737	3.5×10^{-29}
cg19240637	2	7172297	5932	232	rs541772948	7166631	5666	6.2×10^{-23}
cg26435149	3	55605611	5494	58	rs79875865	55678578	-72967	3.3×10^{-22}
cg13764516	9	139648911	4896	231	rs1444333281	139657374	-8463	3.3×10^{-17}
cg01912040	17	1106553	5883	104	rs4968184	1113656	-7103	1.2×10^{-14}
cg26390598	21	41032396	5583	21	rs3746887	41032740	-344	5.7×10^{-13}
cg05428706	10	102730130	4658	160	rs10509748	102691412	38718	9.0×10^{-13}
cg14718533	10	33355576	5895	675	rs6481817	33299247	56329	7.0×10^{-12}
cg14145338	9	139649039	4895	225	rs1444333281	139657374	-8335	1.3×10^{-11}
cg07207669	1	155102388	2720	269	rs4390169	155106054	-3666	2.7×10^{-11}
cg09183146	16	1429863	7196	37	rs149214129	1430914	-1051	1.5×10^{-10}
cg04622454	9	140349128	4022	29	rs116914963	140349161	-33	3.9×10^{-9}
cg10003262	17	1106589	5883	36	rs78571244	1106909	-320	1.6×10^{-8}
cg01225779	5	179238472	7774	188	rs201589495	179163966	74506	4.0×10^{-8}
cg04875062	1	17305562	3263	43	rs112384534	17297705	7857	6.8×10^{-8}
cg07367302	1	19967428	5950	151	rs3020595	19970727	-3299	9.0×10^{-8}
cg04459545	19	17375685	5479	7	rs8112782	17374726	959	1.0×10^{-7}
cg22959742	10	13913931	6995	24	rs867425490	13739152	174779	2.6×10^{-7}
cg18050715	13	97996992	4399	6	rs9516919	98021849	-24857	6.7×10^{-7}
cg08596618	1	24275885	4461	8	rs142675257	24305397	-29512	5.0×10^{-6}
cg00281776	2	209224225	4523	8	rs7578859	208994927	229298	1.9×10^{-5}
cg03871754	17	79320652	6487	1	rs142808556	79391360	-70708	2.6×10^{-5}
cg13223043	1	26492308	4106	10	rs10751733	26486880	5428	3.9×10^{-5}
cg02306995	3	122635049	5827	11	rs148907356	122246027	389022	4.6×10^{-5}
cg00472758	16	2552820	3094					
cg01757312	13	112720565	6153					
cg04826368	6	27130208	2705					
cg04920032	12	50262986	2145					
cg05425326	16	58439361	6013					
cg05816193	6	26018127	4197					
cg06381803	19	46119475	3547					
cg07782285	19	13085442	3679					
cg08759026	11	69061454	5339					
cg12261095	19	7764345	5407					
cg13480898	19	10195914	4576					
cg17393635	19	49843565	2493					
cg18413900	12	58160989	3303					
cg23050300	1	3281321	6698					
cg24318728	17	39649283	2667					
cg27092191	16	31884699	1430					

Analyzed all cis-mQTL pairs within a 1 megabase window around each meta-analysis arsenic-associated CpG (n=41, $p < 1.3 \times 10^{-7}$) among HEALS participants (n=389). MatrixeQTL software was used for the analysis, and model adjusted for age, sex, and first four methylation principal components (PCs). The total SNPs tested for each CpG and number of cis-mQTLs that reached FDR of 0.01 were summarized, and we reported the genomic information for the lead SNP (smallest p-value) and distance (in basepairs) from CpG.

Table S7. Summary of disease and trait associations with genetic variants within or near genes annotated to urinary arsenic-associated CpGs from the GWAS Catalog.

Gene	Variants	Associated Health Effects and Diseases
ABR	2	heel bone mineral density
B3GALT5	0	
C19orf66	1	blood protein levels
CD37	0	
CYP27B1	1	multiple sclerosis
DAND5	2	Highest math class taken and waist-hip ratio
EFNA1	3	hair color, 1,5-anhydroglucitol levels, urinary albumin excretion
EML2	1	Type 2 diabetes
ERC2	16	heel bone mineral density, lobe attachment, lung function (FVC, FEV1/FVC), educational attainment, highest math class taken, self-reported math ability, adolescent idiopathic scoliosis, cognitive performance, smoking initiation
FAIM2	2	BMI (adult and childhood), extreme obesity, fat free mass
FAM110D	2	heel bone mineral density, white blood cell count
FCER2	8	blood protein levels, cerebrospinal fluid biomarker levels
FRMD4A	9	red blood cell distribution width, adolescent idiopathic scoliosis, lung function (FEV1/FVC), red blood cell count, Alzheimer's disease, monoclonal gammopathy
GBAP1	5	gastric cancer, blood urea nitrogen levels, creatinine levels, renal function-related traits (BUN), glomerular filtration rate
GINS3	0	
HIST1H1A	0	
HIST1H2AH	0	
ITGB1	0	
KRT36	0	
LCN8	0	
MAPRE2	1	adolescent idiopathic scoliosis
MBNL2	7	Diastolic blood pressure, self-reported math ability, highest math class taken, neuroticism, cognitive performance, menarche (age at onset)
MFAP2	4	height, lung function (FVC, FEV1/FVC), body fat distribution, pulmonary function, blood protein levels
MINOS-NBL1	0	
MYEOV	17	prostate cancer, systolic blood pressure, macular thickness, balding type 1, waist-hip ratio, male-pattern baldness, pulse pressure, hair color, diastolic blood pressure, breast size, intracranial aneurysm, cancer (pleiotropy)
NBR1	0	
NSMF	0	
NELF	0	
PRDM16	14	migraine, red blood cell count, hematocrit, mean corpuscular hemoglobin, plateletcrit, mean platelet volume, platelet distribution width, Eczema, QRS duration, systolic blood pressure, coronary artery disease, headache, motion sickness
PTH2R	2	menarche (age at onset), adolescent idiopathic scoliosis
RNF144A	1	intracranial aneurysm
SEMA4G	0	
SEMA5B	1	age-related diseases and mortality
SOX1	0	
SPSB1	9	waist-hip ratio, pulse pressure, eosinophil counts, systolic blood pressure, lung function (FVC), Eczema, neutrophil percentage of granulocytes, eosinophil percentage of granulocytes, sum eosinophil basophil counts
SQSTM1	0	
SRSF10	1	red cell distribution width
TBC1D24	0	
TMEM105	0	
UNKL	0	
USHBP1	2	high-grade serous ovarian cancer, serous invasive ovarian cancer, breast cancer, epithelial ovarian cancer, invasive epithelial ovarian cancer, breast cancer in BRCA1 mutation carriers, breast cancer (ER-negative), cancer (pleiotropy)
ZNF267	0	

We extracted gene annotations from arsenic-associated CpGs identified in HEALS discovery analysis ($p < 6.5 \times 10^{-8}$, n=6) and meta-analysis ($p < 1.3 \times 10^{-7}$, n=38). Genes were queried in the GWAS Catalog (accessed 4/22/2019), and we summarized unique SNP count among SNPs with $p < 10^{-8}$. Disease and health traits associated with the SNPs were reported.

REFERENCES

Ameer SS, Engstrom K, Hossain MB, Concha G, Vahter M, Broberg K. 2017. Arsenic exposure from drinking water is associated with decreased gene expression and increased dna methylation in peripheral blood. Toxicol Appl Pharmacol 321:57-66.

Argos M, Chen L, Jasmine F, Tong L, Pierce BL, Roy S, et al. 2015. Gene-specific differential dna methylation and chronic arsenic exposure in an epigenome-wide association study of adults in bangladesh. Environ Health Perspect 123:64-71.

Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, et al. 2012. Dna methylation arrays as surrogate measures of cell mixture distribution. BMC Bioinformatics 13:86.

Kaushal A, Zhang H, Karmaus WJJ, Everson TM, Marsit CJ, Karagas MR, et al. 2017. Genome-wide dna methylation at birth in relation to in utero arsenic exposure and the associated health in later life. Environ Health 16:50.

Kile ML, Houseman EA, Baccarelli AA, Quamruzzaman Q, Rahman M, Mostofa G, et al. 2014. Effect of prenatal arsenic exposure on dna methylation and leukocyte subpopulations in cord blood. Epigenetics 9:774-782.

Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. 2012. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. Bioinformatics 28:882-883.

Peters TJ, Buckley MJ, Statham AL, Pidsley R, Samaras K, R VL, et al. 2015. De novo identification of differentially methylated regions in the human genome. Epigenetics Chromatin 8:6.

Rojas D, Rager JE, Smeester L, Bailey KA, Drobna Z, Rubio-Andrade M, et al. 2015. 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 143:97-106.