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

Daniel Rojas*, Julia E. Rager[†], Lisa Smeester[†], Kathryn A. Bailey[†],
Zuzana Drobná[‡], Marisela Rubio-Andrade[¶], Miroslav Stýblo^{*,‡},
Gonzalo García-Vargas[¶], and Rebecca C. Fry^{*,†,1}

*Curriculum in Toxicology, [†]Department of Environmental Sciences and Engineering, [‡]Department of Nutrition and [§]Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, North Carolina and [¶]Facultad de Medicina, Universidad Juárez del Estado de Durango, Gómez Palacio, Durango, Mexico

¹To whom correspondence should be addressed at Department of Environmental Sciences and Engineering, Gillings School of Global Public Health, 135 Dauer Drive, CB 7431, University of North Carolina, Chapel Hill, 27599, North Carolina. Fax: (919) 843-9047. E-mail: rfry@unc.edu.

ABSTRACT

Prenatal exposure to inorganic arsenic (iAs) is detrimental to the health of newborns and increases the risk of disease development later in life. Here we examined a subset of newborn cord blood leukocyte samples collected from subjects enrolled in the Biomarkers of Exposure to ARsenic (BEAR) pregnancy cohort in Gómez Palacio, Mexico, who were exposed to a range of drinking water arsenic concentrations (0.456–236 µg/l). Changes in iAs-associated DNA 5-methylcytosine methylation were assessed across 424 935 CpG sites representing 18 761 genes and compared with corresponding mRNA expression levels and birth outcomes. In the context of arsenic exposure, a total of 2919 genes were identified with iAs-associated differences in DNA methylation. Site-specific analyses identified DNA methylation changes that were most predictive of gene expression levels where CpG methylation within CpG islands positioned within the first exon, the 5' untranslated region and 200 bp upstream of the transcription start site yielded the most significant association with gene expression levels. A set of 16 genes was identified with correlated iAs-associated changes in DNA methylation and mRNA expression and all were highly enriched for binding sites of the early growth response (EGR) and CCCTC-binding factor (CTCF) transcription factors. Furthermore, DNA methylation levels of 7 of these genes were associated with differences in birth outcomes including gestational age and head circumference. These data highlight the complex interplay between DNA methylation, functional changes in gene expression and health outcomes and underscore the need for functional analyses coupled to epigenetic assessments.

Key words: arsenic; prenatal exposure; epigenetics; gene expression; DNA methylation

Exposure to inorganic arsenic (iAs) at levels that exceed the World Health Organization's (WHO) recommended limit of 10 µg/l currently impacts the health of individuals in countries around the globe (ATSDR, 2007; WHO, 2006). Chronic iAs

exposure is of considerable concern as it is associated with the development of cancers, including the liver, lung, prostate, skin, and urinary bladder as well as other chronic diseases in adults (ATSDR, 2007). In addition to the health consequences from

chronic exposure, *in utero* exposure to iAs is associated with detrimental health consequences in infancy including increased risk for infection and increased risk for both cancer and noncancer endpoints later in life (reviewed in Bailey *et al.* (2014)).

The development of iAs-associated disease likely results from the concerted action of several mechanisms of toxicity including the alteration of protein function via direct binding to sulfhydryl groups as well as the generation of oxidative stress (Jomova *et al.*, 2011). Oxidative stress, in turn, can damage cellular macromolecules such as proteins, lipids, and DNA (Jomova *et al.*, 2011). Exposure to iAs has also been shown to alter the expression of genes involved in key biological pathways such as DNA repair (Andrew *et al.*, 2003). Prenatal iAs exposure has been associated with altered gene expression in human cord blood leukocytes and various target tissues in rodents (Fry *et al.*, 2007; Liu *et al.*, 2004, 2006; Rager *et al.*, 2014a). It is probable that these changes in gene expression are regulated, at least in part, by epigenetic alterations as supported by evidence of changes in genomic 5-methylcytosine patterns associated with prenatal arsenic exposure in human cord blood leukocytes (Kile *et al.*, 2012); however, such a direct linkage has not before been carried out in the context of DNA methylation.

During the DNA methylation process in mammals, a methyl group is enzymatically added to the 5' position of cytosines mainly in the context of CpG dinucleotides (Smith *et al.*, 2012). Alterations in DNA methylation across the genome can lead to changes in genomic integrity and the silencing or expression of genes or entire chromosomes (Smith *et al.*, 2012). DNA methylation patterns are highly dynamic during embryonic development, and changes during this stage may lead to permanent reprogramming and lifelong effects (Tobi *et al.*, 2009). Although it is generally accepted that CpG-based methylation can lead to a repression in gene expression, evidence shows these relationships are far more complex. For example, DNA methylation does not invariably lead to gene silencing, and in multiple cases it has been observed to result in increased expression or to have no relationship with gene expression (Bock *et al.*, 2012; Boellmann *et al.*, 2010; Dosunmu *et al.*, 2012). Previously published research has demonstrated that both chronic exposure in adults as well as prenatal exposure to iAs is associated with DNA methylation changes in humans (Kile *et al.*, 2012, 2014; Smeester *et al.*, 2011). However, the relationship between altered levels of DNA methylation, gene expression, and health outcomes has not been previously examined in newborns exposed to arsenic *in utero*.

To assess these relationships, we utilized samples and data obtained through the Biomarkers of Exposure to ARsenic (BEAR) prospective pregnancy cohort. This cohort includes women from Gómez Palacio, in the state of Durango, Mexico (Rager *et al.*, 2014a). In this area, iAs in drinking water often exceeds 50 µg/l, and adverse health effects associated with iAs exposure, including skin lesions and diabetes mellitus, have been previously identified (as described in Rager *et al.* (2014a)). We recently assessed the impact of prenatal exposure to arsenic on genome-wide mRNA expression profiles in blood leukocytes of a nested set of newborns within the cohort (Rager *et al.*, 2014a) and identified a role of microRNAs (miRNAs) as mediators of this response. In the present study, we expand upon this research to examine the relationship between DNA methylation levels and mRNA expression in a gene-specific manner. At baseline (ie, independent of iAs), the predicted negative correlation between DNA methylation and gene expression was generally observed. In the context of iAs-associated changes in DNA methylation, we demonstrate that DNA methylation at

individual CpG sites and/or methylation averaged across CpG sites for a given gene was not necessarily predictive of gene expression change. A genome position-specific analysis was used to identify the sites that were most predictive of functional change. Furthermore, a subset of genes with altered DNA methylation levels was associated with gene expression as well as birth outcomes.

MATERIALS AND METHODS

Study subject description. This study was approved by the University of North Carolina at Chapel Hill's Institutional Review Board (No. 10-1583) and at the Universidad Juárez del Estado de Durango (UJED), del Estado de Durango Gómez Palacio, Durango, Mexico. BEAR participants were recruited near the time of delivery at the General Hospital of Gómez Palacio. Detailed information on participation requirements and maternal characteristics for the larger study population are as described previously (Rager *et al.*, 2014a).

Subcohort selection and determination of DW-iAs and U-tAs. The present study focuses on a comparative analysis of mRNA expression and DNA methylation profiles from 38 cord blood samples obtained from mother-newborn pairs from the larger BEAR cohort of 200 mother-newborn pairs. The samples were selected to include newborns exposed to varying levels of arsenic. Exposure levels were determined by iAs levels in drinking water (DW-iAs; µg/l) and the levels of total maternal urinary arsenic (U-tAs; µg/l) and subjects were selected randomly within the highest and lowest exposure quartiles. U-tAs is defined as the sum of the levels of iAs and its monomethylated (MMAs) and dimethylated metabolites (DMAs; Rager *et al.*, 2014a).

Cord blood genomic and epigenomic assessment. Cord blood samples were collected using Paxgene Blood DNA tubes and extracted using the PaxGene Blood DNA kit (Qiagen, Valencia, CA) as per the manufacturer's specifications. Isolated DNA was first bisulfite-converted using the EZ DNA methylation kit (Zymo Research, Irvine, CA) and converted DNA was then hybridized onto the Illumina HumanMethylation450 BeadChip (Illumina, Inc, San Diego, CA). This platform assesses the methylation levels of a total of 486 428 individual probes each measuring the methylation levels at a single CpG site. BeadChip microarray data were collected at Expression Analysis, Inc (Durham, NC; www.expressionanalysis.com). Methylation levels were calculated and expressed as β values (β = intensity of the methylated allele (M) / (intensity of the unmethylated allele (U) + intensity of the methylated allele (M) + 100) as in Joubert *et al.* (2012). Methylation data were normalized using a quantile-based methodology (Bolstad *et al.*, 2003). A separate beta-mixture quantile normalization (BMIQ) methodology was also used for validation purposes (Teschendorff *et al.*, 2013). For data filtration, probes with high detection P-values ($P > .05$) were marked as unreliable and removed from analysis ($n = 1761$), as per the manufacturer's recommendation. Probes that represent known single nucleotide polymorphisms (SNPs) were removed (Pidsley *et al.*, 2013; $n = 59\,732$), leaving a total of 424 935 probes for further analyses. Median gene methylation was defined as the median methylation β value across subjects summarized for all probes corresponding to a particular gene.

Sites of U-tAs-associated differential DNA methylation were identified using a multivariable regression model where the dependent variable was DNA methylation and the independent variable was U-tAs. The covariates were selected based on their

association with both exposure and outcome using a bivariate analysis ($P < .05$) or based on their *a priori* status as known confounders and included the following variables: newborn gender (binary variable), and birthweight/gestational age (continuous variable). Batch effect was not a significant source of variation as evaluated using principal component analysis (PCA). Significant probes were identified based on a false discovery corrected q -value ≤ 0.05 .

Comparison of differentially methylated genes to published studies. Kile et al. identified 500 probes in the Infinium HumanMethylation450 BeadChip for which methylation changes can be used as surrogate measurements of changes in the underlying cell population mixture (Kile et al., 2014). The probes identified in the present study were compared against the cell population-related list identified by Kile et al. (2014) in order to test whether the iAs-associated changes were related to potential shifts in cell population. Additionally, the probes/genes identified in the present study were also compared with probes/genes previously identified in other human studies as having DNA methylation changes associated with iAs exposure (Chanda et al., 2006; Gribble et al., 2014; Koestler et al., 2013; Marsit et al., 2006; Smeester et al., 2011).

Assessment of DNA methylation across six different genomic regions. Region-specific DNA methylation analysis was carried out using probes annotated to one of six gene-specific regions: (1) 3' untranslated region (3'UTR), (2) gene body (Body), (3) first exon (1st Exon), (4) 5' untranslated region (5'UTR), (5) 200 bp upstream of the transcriptional start site (TSS200), and (6) 200 to 1500 bp upstream of the transcriptional start site (TSS1500). A chi-squared test was used to compare the distribution of differentially methylated probes to the overall region distribution of probes in the platform. A one-sample proportion test was run to identify each deviation from the expected proportion in each region.

Comparison of DNA methylation data to mRNA expression data. The mRNA expression data were obtained from our prior study in which RNA isolated from newborn cord blood samples were quality assessed and hybridized to the AffymetrixGeneChip® Human Gene 2.0 ST Array (Rager et al., 2014a). The detailed analytical methods used to identify U-tAs-associated gene expression is previously described (Rager et al., 2014a). Note that multiple test corrected q -values were estimated and reported previously, as well as confirmed using additional permutation-based statistical analyses (Rager et al., 2014a). Complete gene expression data from cord blood leukocytes were obtained from the same 38 subjects as used in the present DNA methylation analysis (Rager et al., 2014a).

As a first assessment in the analysis, probe methylation levels were compared with gene expression levels at baseline (ie, arsenic-independent analysis) focusing on genes with the highest expression levels ($n = 5000$) and genes with the lowest expression levels ($n = 5000$). Biological functions enriched among the highest and lowest expressed genes were identified using Ingenuity Pathway Analysis (Ingenuity Systems®, Redwood City, CA). For direct comparisons between DNA methylation and mRNA expression, fold changes in mRNA level were compared with β differences. Specifically, subjects within the highest exposure quartile (HEQ) were compared relative to subjects within the lowest exposure quartile (LEQ) as used previously to calculate iAs-associated gene expression fold changes (Rager et al., 2014a). Differences in DNA methylation were

calculated for each probeset where β difference was calculated as: (average β value HEQ) – (average β value LEQ). Matches between the DNA methylation and gene expression platform were based on Human Genome Organization (HUGO) annotations. Genes overlapping between the differentially expressed gene (DEG) list and the differentially methylated gene (DMG) list were also tested for linear correlations between expression levels and DNA methylation levels.

Genes with CpG methylation levels significantly associated with U-tAs and gene expression were further analyzed for correlations with 7 recorded birth outcomes from the BEAR subjects including gestational age, birth weight, birth weight/gestational age, newborn length, 5-minute APGAR (appearance, pulse, grimace, activity, respiration) score, placental weight, and head circumference.

Enrichment analysis of transcription factor binding sites within the promoter regions of DMGs. In order to identify potential transcriptional regulators that may be related to arsenic-associated changes in DNA methylation, enrichment analysis of upstream sequences for regulatory transcription factors (TFs) was performed among the DMGs that showed correlation with gene expression levels. Genomatix software suite (Genomatix Software GmbH, Munich, Germany) was used to retrieve the promoter regions defined as 1000 bp upstream of the transcription start site (TSS) to 50 bp downstream from the TSS. TFs with significant statistical enrichment were defined as those with P -value $< .05$ (Ho Sui et al., 2005). Additionally, TF enrichment analysis was also performed on genes represented by the top 100 probe sets found to be differentially methylated in cord blood in response to prenatal arsenic exposure (Koestler et al., 2013).

Results

Characteristics of the BEAR cohort. The BEAR pregnancy cohort comprises 200 women and their newborns located in Gómez Palacio, Mexico. Using a subset of 38 newborn cord blood samples from this larger cohort, the present study analyzed cord blood leukocytes as the cell type of interest and integrated DNA methylation levels with mRNA expression levels as a functional readout. The samples analyzed were selected to include subjects exposed to varying levels of arsenic as determined by both DW-iAs and U-tAs.

Specifically, cord blood samples were analyzed from newborns ($n = 38$) whose mothers were exposed to DW-iAs at levels ranging between the limit of detection (LOD) of 0.456 and 236 $\mu\text{g/l}$. The mean concentration of U-tAs in the cohort analyzed for gene expression and DNA methylation analysis was 73.87 $\mu\text{g/l}$ (median = 32.57 $\mu\text{g/l}$) and mean concentration of DW-iAs was 54.1 $\mu\text{g/l}$ (median = 24.2 $\mu\text{g/l}$; Table 1). Of the drinking water samples collected from the cohort, approximately half ($n = 21$, 55.2%) had DW-iAs levels that exceeded the WHO standard (10 $\mu\text{g/l}$). The levels of DW-iAs and U-tAs were significantly correlated in both the current subcohort ($r = 0.74$, $P < .001$, $n = 38$) and the larger BEAR cohort ($r = 0.51$, $P < .001$, $n = 200$). Additional demographic characteristics of the larger cohort participants are previously described (Rager et al., 2014a).

Identification of genes with U-tAs-associated 5-methyl cytosine levels in fetal cord leukocytes. U-tAs-associated changes in 5-methyl cytosine levels were assessed for more than 450 000 probes and, after filtering for probe quality and quantile normalization, median β values were calculated across the 38 subjects.

TABLE 1. U-tAs and DW-iAs in the BEAR Cohort and Subcohort

Arsenic Measure	Larger BEAR Cohort in $\mu\text{g/l}$ (N = 200)	Subcohort in $\mu\text{g/l}$ (N = 38)
DW-iAs		
Mean	24.6	54.1
Median	13.0	24.2
Range	<0.456 to 236.0	<0.456 to 236.0
U-tAs (SG-normalized)		
Mean	37.5	73.87
Median	23.3	32.57
Range	4.3–319.7	6.2–319.7

Limit of detection (LOD) for iAs = $0.456 \mu\text{g/l}$. Urinary arsenic was normalized using the specific gravity (SG) to adjust for changes in urine volume.

FIG. 1. Histogram of DNA methylation levels (median) for all probes analyzed ($n = 424\,935$).

The distribution of the median β values for the analyzed probes ($n = 424\,935$) exhibited a bimodal pattern where probes displayed either very low or very high median methylation levels (Fig. 1).

An adjusted multivariable regression model was used to identify individual CpG sites with arsenic-associated differences in cord blood leukocyte DNA methylation where the exposure was defined as maternal U-tAs. A total of 4771 probes, corresponding to 2919 DMGs, displayed differential U-tAs associated methylation ($q < 0.05$; Supplementary Table S1). Of those probes, 34% ($n = 1621$) displayed hypo-methylation as U-tAs levels increased and 66% ($n = 3150$) displayed hyper-methylation as U-tAs levels increased.

Comparison of differentially methylated probes with previously published studies. The U-tAs-associated DMGs were compared with a list of 500 probes that are most informative of changes in leukocyte population as defined by Kile et al. (2014) and no overlap was found. Koestler et al. (2013) published a list of top 100 probes with arsenic-associated DNA methylation changes in cord blood leukocytes from prenatally-exposed newborns. A comparison to the previously mentioned dataset revealed overlap in a probe within the gene histone cluster 1, H2ac (HIST1H2AC). Similarly, 19 probes representing 18 genes were identified as DMGs in the present study as well as in another study that evaluated DNA methylation in circulating blood leukocytes from an adult population in Mexico exposed to iAs (Smeester et al., 2011; Supplementary Table S2). In addition, genes analyzed in previous studies including Ras association domain family

FIG. 2. Enrichment of U-tAs-associated differentially methylated (DM) probes across the genomic regions. Proportions for all region-annotated probes are in gray ($n = 321\,417$) and differentially methylated (DM) probes are in black ($n = 319$). Regions with significant deviation ($P < .05$) from the platform distribution are indicated (*).

member 1 (RASSF1), estrogen receptor 1 (ESR1), cyclin-dependent kinase inhibitor 2A (CDKN2A), and tumor protein p53 (p53) demonstrated at least one probe with arsenic-related methylation changes in the present dataset ($P < .05$) (Supplementary Table S2; Chanda et al., 2006; Gribble et al., 2014; Marsit et al., 2006). Thus, numerous genes identified in the present study as DMGs were also identified in prior studies of arsenic-associated DMGs.

Identification of genomic regions enriched within the U-tAs-associated probes. To note, the probe distribution within the Illumina Infinium HumanMethylation450 BeadChip platform does not uniformly assess for DNA methylation patterning across the genome. Specifically, of the gene-annotated probes, the BeadChip contains more probes ($n = 161\,677$ or 44%) that assess CpG methylation in the gene body versus probes that assess methylation in the 3'UTR ($n = 17\,494$, 5%; Fig. 2). To determine if there was an enrichment of probes identified with U-tAs-associated differential methylation for a given genomic region, analysis was performed relative to the probe distribution for the Illumina platform. The distribution of differentially methylated probes was significantly different from the platform distribution for 5 of 6 regions. Specifically, the 3'UTR and gene body were enriched within the probes that showed differential U-tAs-associated DNA methylation, whereas the 1st Exon, TSS200 and the 5'UTR were under-represented (chi-squared = 130.2, $df = 5$, $P\text{-value} < 2.2 \times 10^{-16}$; Fig. 2).

Comparison of U-tAs-associated DNA methylation differences and mRNA expression changes. When comparing arsenic-independent levels of DNA methylation between genes with high expression ($n = 5000$) versus genes with low expression ($n = 5000$), the average DNA methylation levels of the highest expressed genes (mean = 0.37, $SD = 0.20$) were significantly lower than the average DNA methylation of the weakly expressed and/or nonexpressed genes ($P < .0001$; mean = 0.56, $SD = 0.24$; Fig. 3). Functional enrichment analysis comparing genes with the highest gene expression and lowest DNA methylation where $\beta < 0.3$ ($n = 2208$ genes) versus those with the lowest expression and highest DNA methylation where $\beta > 0.7$ ($n = 1926$ genes) was performed. High expression/low methylation genes were

FIG. 3. Histogram of gene methylation levels (median) according to expression levels. The distribution of DNA methylation levels for the genes with highest mRNA/gene expression ($n = 5000$) was compared with the genes with lowest mRNA/gene expression ($n = 5000$). Cellular functions identified as enriched through pathway analysis are listed for high and low expression level groups.

enriched for genes controlling core housekeeping functions such as mRNA synthesis and oxidative phosphorylation. Low expression/high methylation genes were enriched in functions not typically recognized as being performed in leukocytes such as cytochrome p450 synthesis and triacylglycerol degradation (Fig. 3).

A total of 334 U-tAs-associated mRNAs were previously identified to have differential expression in cord blood leukocytes of BEAR newborns (Rager et al., 2014a). It was shown that the majority of the transcripts ($n = 224$) decreased in mRNA abundance as U-tAs increased, whereas 110 transcripts increased in mRNA abundance as U-tAs increased (Rager et al., 2014a). In order to determine the relationship between U-tAs-associated changes in gene expression (eg, mRNA expression), and changes in DNA methylation assessed here, DNA methylation levels from these same study subjects were quantified and compared for all genes represented on both platforms ($n = 18761$ genes). Of the 334 genes that showed differential U-tAs-associated gene expression, 269 were represented in the platform used to assess DNA methylation and could be matched to their corresponding methylation β value (Supplementary Table S3). A total of 20% ($n = 54$ genes) of the differentially expressed genes (DEGs) showed at least one site of differential methylation. Strikingly, when analyzing the set of expression matched DMGs ($n = 2705$ genes) only 2% ($n = 54$ genes) were also present in the list of DEGs (Supplementary Table S3).

Regions of DNA methylation identified as predictors of gene expression.

For those genes that were represented on both gene expression and DNA methylation platforms, correlations were calculated between the U-tAs-associated mRNA fold change and U-tAs-associated differential methylation (β difference). The range of U-tAs-associated β difference was -0.19 to 0.13 , representing up to a 19% difference in methylation levels between the HEQ and the LEQ groups. Correlations between the expression fold change and methylation β difference were assessed for: (1) all genes ($n = 18761$; Fig. 4A); (2) the DMGs ($n = 2705$; Fig. 4B); (3) DEGs ($n = 267$; Fig. 4C); or (4) both DMGs and DEGs ($n = 54$; Fig. 4D). None of these correlations reached statistical significance ($P > .05$); however, a marginal correlation was observed between those genes that were common between the DMGs and the DEGs ($n = 54$; $r^2 = 0.06$, $P = .08$; Fig. 4D). Similar correlation results were observed when data were normalized using

the alternate BMIQ method (data not shown). Furthermore, to establish whether this lack of overlap was dependent upon the expression level of the genes being analyzed, the correlation was also run for genes belonging to two different expression categories. Specifically, the relationships were assessed between fold change in expression level and DNA methylation β differences for genes with low expression ($n = 5000$) and genes with high expression ($n = 5000$). The correlations were not significant within either of these groups ($P > .05$). Further analysis focused on the DEGs by focusing the analysis by the following gene-associated regions: 3' untranslated region (3'UTR), gene body (Body), first exon (1st Exon), 5' untranslated region (5'UTR), 200 bp upstream of the TSS (TSS200), and from 200 to 1500 bp upstream of the TSS (TSS1500; Table 2). Of the genomic regions analyzed, differences in DNA methylation that occurred within the first exon were significantly correlated with changes in gene expression ($P = .04$). Furthermore, when the analysis focused only on probes located within CpG islands, a highly significant association ($P < .05$) was observed between U-tAs-associated changes in gene expression and methylation in all regions ($P = .02$). A significant positive association was also observed for the correlation between DNA methylation of probes located within CpG islands in the first exon ($P = .02$). Furthermore, a significant negative correlation identified for those probes within CpG islands in the 5'UTR ($P = .014$) and TSS200 ($P = .026$; Table 2).

Association analysis was run between mRNA expression and DNA methylation across all subjects for each of the 54 DEG and DEM probes. A total of 16 genes displayed a significant correlation (Table 3). To note, when contrasted to the total number of matched DMGs ($n = 2705$), this represents $<1\%$ of the total. Representative plots for probes with significant correlation between expression and methylation are shown (Fig. 5). Further analysis of the association between the methylation levels of these 16 genes and birth outcomes identified significant associations to gestational age ($n = 6$ probes, $n = 5$ genes), placental weight ($n = 5$ probes, $n = 5$ genes), and head circumference ($n = 1$ probe, $n = 1$ gene; Table 4). Genes with methylation at CpG sites associated with U-tAs levels and birth outcomes included: post-GPI attachment to proteins 2 (PGAP2), protein tyrosine phosphatase, receptor type, E (PTPRE), potassium voltage-gated channel, KQT-like subfamily, member 1 (KCNQ1), cyclin G associated kinase (GAK), WD repeat domain 55 (WDR55), ring finger protein

FIG. 4. Comparison between U-tAs-associated DNA methylation and gene expression. Comparisons were made for: A, All comparable genes ($n = 18761$). B, DMGs ($n = 2705$). C, DEGs ($n = 267$). D, Overlap of DMG and DEGs ($n = 54$).

TABLE 2. Summary of Correlations between U-tAsAs-associated Changes in DNA Methylation (median β difference) and mRNA/Gene Expression Changes (FC-Fold Change) Including All DEGs

Region	All Probes			CpG Island Probes		
	R	P-value	n	R	P-value	n
All	0.043	0.487	269	-0.145	0.029	228
1st Exon	0.177	0.040	134	0.224	0.022	105
3'UTR	0.012	0.879	165	0.401	0.057	23
5'UTR	-0.044	0.633	120	-0.270	0.014	82
Body	-0.010	0.876	246	-0.160	0.054	146
TSS1500	-0.002	0.969	245	0.187	0.063	99
TSS200	-0.035	0.613	214	-0.187	0.026	142

Correlations were calculated including all gene-associated regions (all), or one of the following: first exon (1st Exon), 3' untranslated region (3'UTR), 5' untranslated region (5'UTR), gene body (Body), promoter region 1500–200bp upstream of the TSS (TSS1500), promoter region 200–0bp upstream of the TSS (TSS200). For each of these focus regions, correlations were calculated including either all CpG probes or only those positioned within CpG islands.

213 (RNF213), and PTC7 protein phosphatase homolog (PTC7; Table 4).

TF enrichment analysis in DMGs and DEGs. In order to identify potential upstream regulators that could impact locations of genomic DNA methylation, gene expression levels and health outcomes, enrichment analysis for TF binding sites was performed. This analysis focused on the 16 genes that showed association between differential methylation and gene expression. Binding sites for a total of 36 TFs were enriched within the gene set when compared with genes with altered expression but no correlation with methylation (Supplementary Table S4). The top 5 most significant TFs enriched in the gene set included CCCTC-binding factor (zinc finger protein)-like (CTCF; $P < .00001$), EGR ($P < .00001$), zinc finger and BTB domain

containing 14 (ZF5), GLUT4 enhancer factor (HDBP; $P < .00001$) and hairy and enhancer of split 1 (HES1; $P = .00006$; Supplementary Table S4). As a comparative analysis, the list of 100 CpGs identified in Koestler et al. (2013) were analyzed. Of the 36 TFs identified, 3 overlapped with those identified here including CTCF, nuclear respiratory factor 1 (NRF1), and paired box 5 (PAX5; Supplementary Table S4).

DISCUSSION

In the present study, we aimed to better understand the potential functional implications of DNA methylation changes associated with prenatal arsenic exposure by coupling the data with mRNA expression levels as a functional readout along with measures on subsequent birth outcomes. Using samples obtained from the BEAR cohort in Mexico, site-specific DNA methylation patterns were analyzed in cord blood leukocytes of newborns as they relate to concentrations of maternal urinary arsenic (U-tAs). Many genes (>2000) displayed altered DNA methylation patterning associated with U-tAs, some of which have been observed in previous arsenic-exposed cohorts. However, these changes in DNA methylation were largely unrelated to active measures of gene expression. This finding is highly relevant to ongoing research that does not include functional measures of active transcription or translation. Importantly, these data support that future studies should focus on the genomics regions of CpG islands within the first exon, the 5'UTR and 200bp upstream of the TSS as DNA methylation changes within these regions were most highly predictive of arsenic-associated changes in gene expression. Furthermore, a set of 7 DMGs were identified that were predictive of functional change at the mRNA level and also associated with birth outcomes including gestational age. These genes were enriched for binding sites for TFs supporting a hypothesis of DNA methylation patterns as “environmental footprints” (Sanders et al., 2014) reflective of TF occupancy.

TABLE 3. CpG sites Displaying Association between U-tAs DNA Methylation and mRNA Expression Levels

CpG Probe ID	Refgene Symbol	Gene Name	Expression Fold Change (HEQ/LEQ)	P-value (U-tAs)	q-value (U-tAs)	R	P-value (Expression)
cg16360836	HDAC4	Histone deacetylase 4	0.94	7.02E-05	2.03E-02	−0.57	2.00E-04
cg24671666	PLEKHG3	Pleckstrin homology domain containing, family G member 3	0.76	3.30E-04	3.62E-02	0.49	1.90E-03
cg10632215	CERK	Ceramide kinase	0.86	5.06E-04	4.30E-02	0.46	4.00E-03
cg15601244	SH2B2	SH2B adaptor protein 2	0.85	7.31E-05	2.05E-02	−0.46	4.10E-03
cg26489875	MALAT1	Metastasis associated lung aadenocarcinoma transcript 1	1.12	4.82E-04	4.20E-02	−0.45	4.40E-03
cg15883181	RNF213	Ring finger protein 213	0.83	7.63E-05	2.07E-02	−0.45	4.90E-03
cg18054302	GAK	Cyclin G associated kinase	0.89	3.06E-04	3.53E-02	−0.44	5.30E-03
cg22798758	PTPRE	Protein tyrosine phosphatase, receptor type, E	0.86	2.95E-04	3.48E-02	−0.44	6.00E-03
cg09112262	WDR55	WD repeat domain 55	0.83	4.63E-04	4.13E-02	−0.44	6.10E-03
cg20851790	RNF213	Ring finger protein 213	0.83	4.25E-04	3.98E-02	−0.44	6.10E-03
cg07199894	ULK1	Unc-51 like autophagy activating kinase 1	1.15	6.51E-04	4.69E-02	−0.43	7.80E-03
cg00627621	TRIB1	Tribbles homolog 1 (Drosophila)	0.84	3.35E-04	3.65E-02	−0.40	1.22E-02
cg15574972	MALAT1	Metastasis associated lung adenocarcinoma transcript 1	1.12	1.82E-04	2.89E-02	−0.40	1.30E-02
cg18312113	SSU72	SSU72 RNA polymerase II CTD phosphatase homolog (<i>Saccharomyces cerevisiae</i>)	0.88	4.22E-04	3.97E-02	−0.39	1.72E-02
cg20689294	PTPRE	Protein tyrosine phosphatase, receptor type, E	0.86	5.48E-04	4.40E-02	−0.38	1.77E-02
cg06719391	KCNQ1	Potassium voltage-gated channel, KQT-like subfamily, member 1	0.86	2.91E-05	1.49E-02	−0.38	1.80E-02
cg02582997	GAK	Cyclin G associated kinase	0.89	5.43E-04	4.40E-02	−0.37	2.07E-02
cg03041730	GAK	Cyclin G associated kinase	0.89	2.99E-05	1.50E-02	−0.37	2.30E-02
cg24725201	KCNQ1	Potassium voltage-gated channel, KQT-like subfamily, member 1	0.86	1.08E-04	2.36E-02	−0.36	2.67E-02
cg24089935	KCNQ1	Potassium voltage-gated channel, KQT-like subfamily, member 1	0.86	7.98E-05	2.10E-02	−0.36	2.77E-02
cg06107260	HDAC4	Histone deacetylase 4	0.94	9.26E-06	1.04E-02	−0.35	2.94E-02
cg03926050	PPTC7	PTC7 protein phosphatase homolog (<i>S. cerevisiae</i>)	0.81	2.93E-04	3.47E-02	0.34	3.86E-02
cg01693193	KCNQ1	Potassium voltage-gated channel, KQT-like subfamily, member 1	0.86	3.04E-05	1.51E-02	0.34	3.94E-02
cg04295928	PGAP2	Post-GPI attachment to proteins 2	0.87	3.41E-05	1.57E-02	−0.33	4.62E-02
cg08438705	SBNO2	Strawberry notch homolog 2 (<i>Drosophila</i>)	0.85	4.91E-04	4.24E-02	−0.32	4.79E-02

Associations were calculated using Pearson correlations.

When analyzed in an arsenic-independent manner, a significant relationship between DNA methylation and functional gene expression was observed. As may have been anticipated, genes with high levels of DNA methylation had, on average, lower expression levels, whereas genes with low expression displayed higher levels of DNA methylation. This finding is consistent with previous observations comparing the methylation levels of genes according to their expression category (higher versus lower expression; [Bell et al., 2011](#)). Also consistent with prior observations, housekeeping genes were enriched among the genes with low CpG methylation levels thus potentially contributing to high housekeeping gene expression levels, as has been observed previously ([Fernandez et al., 2012](#)). These data suggest that DNA methylation in cord blood leukocytes is associated with baseline gene expression levels.

Surprisingly, when analyzed in an arsenic-dependent manner in the context of U-tAs-associated changes in DNA methylation, there was minimal association and no statistical significance between DNA methylation and gene expression on a genome-wide level ($n = 18761$ genes), a result that was not dependent upon the data normalization method used nor on the relative expression of the genes. These results highlight that genome-wide changes in DNA methylation associated with U-tAs were only weakly correlated to changes in active mRNA

expression in cord leukocytes. Given the focus of much literature pertaining to CpG methylation on the gene-silencing effects of such DNA methylation ([Jones 2012](#)) these results may seem unexpected. It is important to note, however, that such low correlations between active measures of gene expression and DNA methylation levels have been observed previously. For example, the Meissner lab studied genome-wide DNA methylation and expression profiles across cell types and observed minimal correlation ([Bock et al., 2012](#)). Similarly, a weak correlation was observed between gene expression and gene methylation in neocortex cells ([Dosunmu et al., 2012](#)). Highly relevant to the present study, a study of lung cells from arsenic-exposed mice showed that overall correlation with gene expression among the DMGs was not statistically significant ([Boellmann et al., 2010](#)). Taken together, coupling our data from the present study with data from the aforementioned studies highlights the very important finding that not all sites with altered DNA methylation are associated with concomitant changes in active gene expression. Therefore, it is an important consideration that some changes in DNA methylation may represent permissive marks in the regulation of gene expression as has been proposed ([Boellmann et al., 2010](#)), as opposed to eliciting active changes in functional gene expression. An additional consideration is that mRNA expression is both time and

FIG. 5. Representative plots of genes with significant correlation between U-tAs-associated DNA methylation and U-tAs-associated mRNA/gene expression.

TABLE 4. Genes with CpG Methylation Associated with Birth Outcomes in the BEAR cohort

Birth Outcomes ^a	Ref gene Symbol	Gene Name	CpG probe ID	P-value	R	n
GA	PTPRE	Protein tyrosine phosphatase, receptor type, E	cg22798758	3.74E-02	-0.34	38
GA	WDR55	WD repeat domain 55	cg09112262	4.08E-02	-0.33	38
GA	PTPRE	Protein tyrosine phosphatase, receptor type, E	cg20689294	1.76E-02	-0.38	38
GA	KCNQ1	Potassium voltage-gated channel, KQT-like subfamily, member 1	cg06719391	2.63E-02	-0.36	38
GA	GAK	Cyclin G associated kinase	cg02582997	4.35E-02	-0.33	38
GA	PGAP2	Post-GPI attachment to proteins 2	cg04295928	2.75E-03	-0.47	38
PW	RNF213	Ring finger protein 213	cg15883181	4.72E-02	0.33	36
PW	PTPRE	Protein tyrosine phosphatase, receptor type, E	cg20689294	3.45E-03	0.47	36
PW	KCNQ1	Potassium voltage-gated channel, KQT-like subfamily, member 1	cg06719391	1.61E-02	0.4	36
PW	GAK	Cyclin G associated kinase	cg02582997	3.53E-02	0.35	36
PW	PGAP2	Post-GPI attachment to proteins 2	cg04295928	1.20E-02	0.41	36
HC	PPTC7	PTC7 protein phosphatase homolog (<i>S. cerevisiae</i>)	cg03926050	4.73E-02	0.33	36

Associations between outcomes and CpG sites were calculated using Pearson correlation.

^aGA, gestational age; PW, placental weight; HC, head circumference.

tissue-specific (Rager *et al.*, 2014b). The data from the present study combined with our former research highlight that other epigenetic mechanisms besides DNA methylation may be more influential in determining arsenic-associated gene expression. Specifically, although these results indicate that the transcriptional response may be minimally (~2% to 5% or 16/267) controlled by DNA methylation changes, previous work from our laboratory estimated that ~20% of the genomic response to prenatal iAs is controlled by miRNAs (Rager *et al.*, 2014a).

Comparisons across genomic regions identified CpG methylated sites that were most predictive of changes in gene expression levels. By comparing the differential effect of changes in DNA methylation by location and the mRNA expression of the associated genes, we identified that probes within CpG islands

located in the first exon, 5'UTR and TSS200 were the most predictive of gene expression levels. The data in the present study support findings by the Encyclopedia Of DNA Elements (ENCODE) project, which described the association between active gene expression and unmethylated gene promoters and methylated gene bodies (Ball *et al.*, 2009). Together our data reveal the functional differences between CpG methylation in distinct genomic regions and demonstrate that DNA methylation changes positioned within specific regions can more accurately predict functional consequences at the gene expression level. These findings of genome-position effects are highly relevant in the context of other studies of iAs-associated DNA methylation changes as they highlight a method by which investigators may prioritize analysis to better predict potentially functional CpG methylation.

A total of 16 genes, representing <1% of the total DMGs, were identified to have significant statistical correlation between changes in U-tAs-associated gene expression and DNA methylation across study subjects. Furthermore, 7 of these 16 genes also displayed an association between DNA methylation and birth outcomes, specifically gestational age, placental weight, and head circumference. In support of our data, 2 of the 5 genes associated with gestational age, namely *KCNQ1* and *GAK*, have also been identified in a separate study investigating differentially methylated regions (DMRs) associated with gestational age (Lee et al., 2012). Interestingly, these 16 functionally consequential genes display an enrichment for binding sites of specific TFs including EGR and CTCF, both known to be altered by arsenic and known to impact various cellular signaling pathways when modulated (Simeonova et al., 2000; Xu et al., 2013). The current research provides insight into the transcriptional regulation that can potentially influence iAs-associated DNA methylation patterning and elicit functional changes. These results build upon recent work where we demonstrated that prenatal exposure to cadmium is associated with changes in CpG methylation in cord blood leukocytes with an enrichment of binding sites for TFs among these genes. This finding suggested the potential for “environmental footprints” of prior TF occupancy during times of DNA methylation (Sanders et al., 2014). It was striking to observe a minimal overlap (ie, a single probe overlap) between the genes/probes with differential DNA methylation identified here and those published previously as arsenic-associated in cord blood leukocytes from prenatally-exposed newborns (Koestler et al., 2013). Interestingly, although the genes that were identified in the two studies were different, the DMGs from both studies were found to share common binding sequences for TFs such as CTCF. The identified TFs may thus be common regulators of arsenic-associated DNA methylation patterning, representing a nonrandom biological mechanism underlying the target genes. Taken together, the data from the present study provide *in silico* support for the hypothesis that occupancy of U-tAs-associated TFs may impact the DNA methylation status of a subset of differentially expressed and functionally consequential genes.

This study is not without limitations. Necessitated by the cost of deep biological assessment of DNA methylation and mRNA levels within samples and subjects, the sample size in our study is relatively small—thus these findings need to be validated in larger cohorts. Nevertheless, the fact that many of the gene targets displaying altered DNA methylation patterns and particularly those with association to birth outcomes have been identified in other studies, provides support for biological relevance. Future work should also investigate which of the changes in DNA methylation and/or gene expression observed may be stable throughout time. Together, these results increase the current understanding of the complex relationships between iAs exposure during pregnancy and epigenetic control of cellular signaling events in the infant cord blood that may impact human health.

SUPPLEMENTARY DATA

Supplementary data are available online at <http://toxsci.oxfordjournals.org/>.

FUNDING

National Institute of Environmental Health Sciences (T32ES007018, R01ES019315, P30ES010126, P42ES005948).

ACKNOWLEDGMENTS

We would like to acknowledge Dr. Michael Wu, Dr. Andrew F. Olshan, Mr. Brian Barkley and Mrs. Elizabeth Martin for their assistance with the article.

REFERENCES

- Andrew, A. S., Karagas, M. R., and Hamilton, J. W. (2003). Decreased DNA repair gene expression among individuals exposed to arsenic in United States drinking water. *Int. J. Cancer* **104**, 263–268.
- ATSDR (2007). *Agency for Toxic Substances and Disease Registry. Toxicological Profile for Arsenic*. CAS#: 7440-38-2, i-500.
- Bailey, K., and Fry, R. C. (2014). Long-term health consequences of prenatal arsenic exposure: links to the genome and the epigenome. *Rev. Environ. Health* **29**, 9–12.
- Ball, M. P., Li, J. B., Gao, Y., Lee, J. H., LeProust, E. M., Park, I. H., Xie, B., Daley, G. Q., and Church, G. M. (2009). Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells. *Nat. Biotechnol.* **27**, 361–368.
- Bell, J. T., Pai, A. A., Pickrell, J. K., Gaffney, D. J., Pique-Regi, R., Degner, J. F., Gilad, Y., and Pritchard, J. K. (2011). DNA methylation patterns associate with genetic and gene expression variation in HapMap cell lines. *Genome Biol.* **12**, R10.
- Bock, C., Beerman, I., Lien, W. H., Smith, Z. D., Gu, H., Boyle, P., Gnirke, A., Fuchs, E., Rossi, D. J., and Meissner, A. (2012). DNA methylation dynamics during *in vivo* differentiation of blood and skin stem cells. *Mol. Cell* **47**, 633–647.
- Boellmann, F., Zhang, L., Clewell, H. J., Schroth, G. P., Kenyon, E. M., Andersen, M. E., and Thomas, R. S. (2010). Genome-wide analysis of DNA methylation and gene expression changes in the mouse lung following subchronic arsenate exposure. *Toxicol. Sci.* **117**, 404–417.
- Bolstad, B. M., Irizarry, R. A., Astrand, M., and Speed, T. P. (2003). A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* **19**, 185–193.
- Chanda, S., Dasgupta, U. B., Guhamazumder, D., Gupta, M., Chaudhuri, U., Lahiri, S., Das, S., Ghosh, N., and Chatterjee, D. (2006). DNA hypermethylation of promoter of gene p53 and p16 in arsenic-exposed people with and without malignancy. *Toxicol. Sci.* **89**, 431–437.
- Dosunmu, R., Alashwal, H., and Zawia, N. H. (2012). Genome-wide expression and methylation profiling in the aged rodent brain due to early-life Pb exposure and its relevance to aging. *Mech. Ageing Dev.* **133**, 435–443.
- Fernandez, A. F., Assenov, Y., Martin-Subero, J. I., Balint, B., Siebert, R., Taniguchi, H., Yamamoto, H., Hidalgo, M., Tan, A. C., Galm, O., et al. (2012). A DNA methylation fingerprint of 1628 human samples. *Genome Res.* **22**, 407–419.
- Fry, R. C., Navasumrit, P., Valiathan, C., Svensson, J. P., Hogan, B. J., Luo, M., Bhattacharya, S., Kandjanapa, K., Soontararuks, S., Nookabkaew, S., et al. (2007). Activation of inflammation/NF-kappaB signaling in infants born to arsenic-exposed mothers. *PLoS Genet.* **3**, 2180–2189.
- Gribble, M. O., Tang, W. Y., Shang, Y., Pollak, J., Umans, J. G., Francesconi, K. A., Goessler, W., Silbergeld, E. K., Guallar, E., Cole, S. A., et al. (2014). Differential methylation of the arsenic (III) methyltransferase promoter according to arsenic exposure. *Arch. Toxicol.* **88**, 275–282.
- Ho Sui, S. J., Mortimer, J. R., Arenillas, D. J., Brumm, J., Walsh, C. J., Kennedy, B. P., and Wasserman, W. W. (2005). oPOSSUM: identification of over-represented transcription factor

- binding sites in co-expressed genes. *Nucleic Acids Res.* **33**, 3154–3164.
- Jomova, K., Jenisova, Z., Feszterova, M., Baros, S., Liska, J., Hudecova, D., Rhodes, C. J., and Valko, M. (2011). Arsenic: toxicity, oxidative stress and human disease. *J. Appl. Toxicol.* **31**, 95–107.
- Jones, P. A. (2012). Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat. Rev. Genet.* **13**, 484–492.
- Joubert, B. R., Haberg, S. E., Nilsen, R. M., Wang, X., Vollset, S. E., Murphy, S. K., Huang, Z., Hoyo, C., Midttun, O., Cupul-Uicab, L. A., et al. (2012). 450K epigenome-wide scan identifies differential DNA methylation in newborns related to maternal smoking during pregnancy. *Environ. Health Perspect.* **120**, 1425–1431.
- Kile, M. L., Baccarelli, A., Hoffman, E., Tarantini, L., Quamruzzaman, Q., Rahman, M., Mahiuddin, G., Mostofa, G., Hsueh, Y. M., Wright, R. O., and Christiani, D. C. (2012). Prenatal arsenic exposure and DNA methylation in maternal and umbilical cord blood leukocytes. *Environ. Health Perspect.* **120**, 1061–1066.
- Kile, M. L., Houseman, E. A., Baccarelli, A., Quamruzzaman, Q., Rahman, M., Mostofa, G., Cardenas, A., Wright, R. O., and Christiani, D. C. (2014). Effect of prenatal arsenic exposure on DNA methylation and leukocyte subpopulations in cord blood. *Epigenetics* **9**, 774–782.
- Koestler, D. C., Avissar-Whiting, M., Houseman, E. A., Karagas, M. R., and Marsit, C. J. (2013). Differential DNA methylation in umbilical cord blood of infants exposed to low levels of arsenic in utero. *Environ. Health Perspect.* **121**, 971–977.
- Lee, H., Jaffe, A. E., Feinberg, J. I., Tryggvadottir, R., Brown, S., Montano, C., Aryee, M. J., Irizarry, R. A., Herbstman, J., Witter, F. R., et al. (2012). DNA methylation shows genome-wide association of NFIX, RAPGEF2 and MSRB3 with gestational age at birth. *Int. J. Epidemiol.* **41**, 188–199.
- Liu, J., Xie, Y., Ducharme, D. M., Shen, J., Diwan, B. A., Merrick, B. A., Grissom, S. F., Tucker, C. J., Paules, R. S., Tennant, R., and Waalkes, M. P. (2006). Global gene expression associated with hepatocarcinogenesis in adult male mice induced by in utero arsenic exposure. *Environ. Health Perspect.* **114**, 404–411.
- Liu, J., Xie, Y., Ward, J. M., Diwan, B. A., and Waalkes, M. P. (2004). Toxicogenomic analysis of aberrant gene expression in liver tumors and nontumorous livers of adult mice exposed in utero to inorganic arsenic. *Toxicol. Sci.* **77**, 249–257.
- Marsit, C. J., Karagas, M. R., Danaee, H., Liu, M., Andrew, A., Schned, A., Nelson, H. H., and Kelsey, K. T. (2006). Carcinogen exposure and gene promoter hypermethylation in bladder cancer. *Carcinogenesis* **27**, 112–116.
- Pidsley, R., Wong, C. C. Y., Volta, M., Lunnon, K., Mill, J., and Schalkwyk, L. C. (2013). A data-driven approach to preprocessing Illumina 450K methylation array data. *BMC Genomics* **14**, 293.
- Rager, J. E., Bailey, K. A., Smeester, L., Miller, S. K., Parker, J. S., Laine, J. E., Drobna, Z., Currier, J., Douillet, C., Olshan, A. F., et al. (2014a). Prenatal arsenic exposure and the epigenome: altered microRNAs associated with innate and adaptive immune signaling in newborn cord blood. *Environ. Mol. Mutagen.* **55**, 196–208.
- Rager, J. E., Moeller, B. C., Miller, S. K., Kracko, D., Doyle-Eisele, M., Swenberg, J. A., and Fry, R. C. (2014b). Formaldehyde-associated changes in microRNAs: tissue and temporal specificity in the rat nose, white blood cells, and bone marrow. *Toxicol. Sci.* **138**, 36–46.
- Sanders, A. P., Smeester, L., Rojas, D., Debusscher, T., Wu, M. C., Wright, F. A., Zhou, Y. H., Laine, J. E., Rager, J. E., Swamy, G. K., et al. (2014). Cadmium exposure and the epigenome: exposure-associated patterns of DNA methylation in leukocytes from mother-baby pairs. *Epigenetics* **9**, 212–221.
- Simeonova, P. P., Wang, S., Toriuma, W., Komminen, V., Matheson, J., Unimye, N., Kayama, F., Harki, D., Ding, M., Vallyathan, V., and Luster, M. I. (2000). Arsenic mediates cell proliferation and gene expression in the bladder epithelium: association with activating protein-1 transactivation. *Cancer Res.* **60**, 3445–3453.
- Smeester, L., Rager, J. E., Bailey, K. A., Guan, X., Smith, N., Garcia-Vargas, G., Del Razo, L. M., Drobna, Z., Kelkar, H., Styblo, M., and Fry, R. C. (2011). Epigenetic changes in individuals with arsenicosis. *Chem. Res. Toxicol.* **24**, 165–167.
- Smith, Z. D., Chan, M. M., Mikkelsen, T. S., Gu, H., Gnirke, A., Regev, A., and Meissner, A. (2012). A unique regulatory phase of DNA methylation in the early mammalian embryo. *Nature* **484**, 339–344.
- Teschendorff, A. E., Marabita, F., Lechner, M., Bartlett, T., Tegner, J., Gomez-Cabrero, D., and Beck, S. (2013). A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450k DNA methylation data. *Bioinformatics* **29**, 189–196.
- Tobi, E. W., Lumey, L. H., Talens, R. P., Kremer, D., Putter, H., Stein, A. D., Slagboom, P. E., and Heijmans, B. T. (2009). DNA methylation differences after exposure to prenatal famine are common and timing- and sex-specific. *Hum. Mol. Genet.* **18**, 4046–4053.
- WHO (2006). *World Health Organization. Guidelines for Drinking Water Quality. First addendum to 3rd addition, Volume 1.*
- Xu, H., Lam, S. H., Shen, Y., and Gong, Z. (2013). Genome-wide identification of molecular pathways and biomarkers in response to arsenic exposure in zebrafish liver. *PLoS One* **8**, e68737.