PRENATAL ARSENIC EXPOSURE AND THE EPIGENOME: IDENTIFYING SITES OF 5-METHYL CYTOSINE ALTERATIONS THAT PREDICT FUNCTIONAL CHANGES IN GENE EXPRESSION IN NEWBORN CORD BLOOD AND SUBSEQUENT BIRTH OUTCOMES

Daniel Rojas

A thesis submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Masters of Science in Toxicology in the Curriculum in Toxicology in the School of Medicine.

Chapel Hill
2014

Approved by:

Rebecca C. Fry

Ilona Jaspers

Bernard Weissman

ABSTRACT

Daniel Rojas: Prenatal arsenic exposure and the epigenome: Identifying sites of 5-methyl cytosine alterations that predict functional changes in gene expression in newborn cord blood and subsequent birth outcomes
(Under the direction of Rebecca C. Fry)

Prenatal exposure to inorganic arsenic (iAs) is detrimental to the health of newborns and increases the risk of disease later in life. Here we examined newborn cord blood leukocyte samples from the Biomarkers of Exposure to ARsenic (BEAR) pregnancy cohort in Mexico. Changes in iAs-associated DNA methylation were compared to corresponding gene expression levels and birth outcomes. 2,705 genes were identified with iAs-associated differences in DNA methylation. In contrast to minimal association genome-wide, site-specific analyses identified DNA methylation changes that were most predictive of gene expression levels. 16 genes were identified with correlated iAs-associated changes in DNA methylation and mRNA expression, and with enrichment for binding sites of several transcription factors. Furthermore, DNA methylation levels were associated with birth outcomes. These data highlight the complex interplay between DNA methylation and functional changes in gene expression and health outcomes and underscore the need for functional analyses coupled to epigenetic assessments.

To my family, who have always given me their love and support.

To my mentor, who provided me with invaluable advice and who gave me all her support during this journey.

ACKNOWLEDGEMENTS

This work would not have been possible without the contribution of others. I want to thank Dr. Julia E. Rager, Dr. Kathryn A. Bailey and Lisa Smeester, for their assistance throughout this research. Thanks to the Fry Lab at large, for providing a great environment to work and to discuss scientific ideas. I want to thank Drs. Zuzana Drobná, Marisela Rubio-Andrade, Miroslav Stýblo, Gonzalo García-Vargas, for their assistance with the study cohort and manuscript. Finally, Dr. Rebecca C. Fry, who conceived, developed and directed the long-term BEAR cohort project.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

3' UTR	3' Untranslated region
5' UTR	5' Untranslated region
BEAR	Biomarkers of exposure to arsenic
BMIQ	Beta-mixture quantile normalization
CDKN2A	Cyclin-dependent kinase inhibitor 2A
CTCF	CCCTC-binding factor (zinc finger protein)-like
DEG	Differentially expressed gene
DMA	Dimethylated arsenic
DMG	Differentially methylated gene
DNA	Deoxyribonucleic acid
DW-iAs	Drinking water inorganic arsenic
EGR	Eary growth response
ESR1	Estrogen receptor 1
GAK	Cyclin G associated kinase
HDBP	GLUT4 enhancer factor
HEQ	Highest exposure quartile
HES1	Hairy enhancer of split 1
HIST1H2AC	Histone cluster 1, H2ac
HUGO	Human genome organization
iAs	Inorganic arsenic
KCNQ1	Potassium voltage-gated channel KQT-like subfamily, member 1
LEQ	Lowest exposure quartile
LOD	Limit of detection
MMA	Monomethylated arsenic

mRNA	Messenger RNA
NRF2	Nuclear factor-like 2
p53	Tumor protein p53
PCA	Principal component analysis
PGAP2	Post-GPI attachment to proteins 2
PTC7	PTC7 protein phosphatase homolog
PTPRE	Protein tyrosine phosphatase, receptor type E
RASSF1	Ras association domain family member 1
RNA	Ribonucleic acid
RNF213	Ring finger protein 213
SG	Specific gravity
SNP	Single nucleotide polymorphism
TF	Transcription factor
TSS1500	200 to 1500 base pairs upstream of the transcriptional start site
TSS200	200 base pairs upstream of the transcriptional start site
UJED	Universidad Juárez del Estado de Durango
U-tAs	Total maternal urinary arsenic
WDR55	WD repeat domain 55
WHO	World health organization

CHAPTER 1: INTRODUCTION

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 non-cancer endpoints later in life (reviewed in (Bailey and Fry 2014)).

The development of iAs-associated diseases likely result from several mechanisms of toxicity including the alteration protein function via direct binding to sulfhydryl groups as well as the generation of oxidative stress. Oxidative stress, in turn, can damage cellular macromolecules such as proteins, lipids and DNA (Jomova et al. 2011). IAs exposure 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; Liu et al. 2006; Rager et al. 2014). 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).

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, resulting in life-long effects (Tobi et al. 2009). While it is generally accepted that CpG-based methylation can lead to decreased 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 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 as well as prenatal exposure to iAs is associated with DNA methylation changes in humans (Kile et al. 2012; Kile et al. 2014; Pilsner et al. 2009; Smeester et al. 2011). However, to our knowledge, 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 the relationships between prenatal iAs exposure, DNA methylation, gene expression, and birth outcomes 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. 2014). 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 (Rager et al. 2014). 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. 2014). In the present study, we expand upon this research to examine the relationship between DNA methylation levels and transcriptional changes in a gene-specific manner. At baseline (i.e. independent of iAs), the predicted negative correlation between DNA methylation and gene expression was 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. Genome-position-specific analysis identified the sites that were most predictive of functional change. A subset of genes with altered DNA methylation levels were identified that were both associated with gene expression as well as birth outcomes.

CHAPTER 2: METHODS

Study subjects

This study was approved by the University of North Carolina at Chapel Hill's Institutional Review Board (#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 described elsewhere (Rager et al. 2014).

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 as determined by iAs levels in drinking water (DW-iAs; µg/L) and the levels of total maternal urinary arsenic (U-tAs; µg/L). U-tAs is defined as the sum of the levels of iAs and its monomethylated (MMAs) and dimethylated metabolites (DMAs) (Rager et al. 2014).

Cord blood genomic and epigenomic assessment

To assess DNA methylation levels, isolated DNA was first bisulfite-converted using the EZ DNA methylation kit (Zymo Research, Irvine, CA). The converted DNA was then hybridized onto the Infinium HumanMethylation450 BeadChip (Illumina, 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. 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) as well as beta-mixture quantile normalization (BMIQ) (Teschendorff et al. 2013). For data filtration, probes with high detection p-values (p>0.05) were marked as unreliable and removed from analysis (n=1,761), as per manufacturer 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 multi-variable 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<0.05) or based on their *a priori* status as known confounders and included the following variables: newborn gender (binary variable) and birth weight/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 $\leq$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 to 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; Marsit et al. 2006; Pilsner et al. 2009; 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: (i) 3' untranslated region (3'UTR), (ii) gene body (Body), (iii) first exon (1st Exon), (iv) 5' untranslated region (5'UTR), (v) 200 base pairs upstream of the transcriptional start site (TSS200) and (vi) 200 to 1500 base pairs 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 the gene 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 Affymetrix GeneChip® Human Gene 2.0 ST Array (Rager et al. 2014). The detailed analytical methods used to identify U-tAs-associated gene expression is previously described (Rager et al. 2014). Complete gene expression data from cord blood leukoctyes were obtained from the same 38 subjects as used in the present DNA methylation analysis (Rager et al. 2014).

As a first assessment in the analysis, probe methylation levels were compared to gene expression at baseline (i.e. arsenic-independent analysis) focusing on genes with highest expression levels (n=5,000) and genes with the lowest expression levels (n=5,000). Biological functions enriched amongst 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 to 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. 2014). Differences in DNA methylation were calculated for each probe set 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 seven recorded birth outcomes from the BEAR subjects including gestational age, birth weight, birth weight/gestational age, newborn length, five 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 differentially methylated genes

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 was performed amongst the differentially methylated genes 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 base pairs (bp) upstream of the transcription start site (TSS) to 50 bp downstream from the TSS. Transcription factors with significant statistical enrichment were defined as those with *p*-value<0.05 (Ho Sui et al. 2005).

CHAPTER 3: 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 utilized cord blood leukocytes as the cell type of interest and integrated DNA methylation levels with mRNA expression levels as a functional read-out. The samples analyzed were selected to include subjects exposed to varying levels of arsenic as determined by both DW-iAs and U-tAs.

In the present study, 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 µg/L. The mean concentration of U-tAs in the cohort analyzed for gene expression and DNA methylation analysis was 73.87 µg/L (median=32.57 µg/L) and mean concentration of DW-iAs was 54.1 µg/L (median=24.2 µg/L) (**Table 1**).

Table 1. U-tAs stratified by DW-iAs in the BEAR cohort and subcohort. Limit of detection (LOD) for iAs=0.456 µg/L. Urinary arsenic was normalized using the Specific Gravity (SG) to adjust for changes in urine volume.

Arsenic Measure	Urinary arsenic levels in Larger BEAR cohort (N=200)	Urinary arsenic levels in Subcohort (N=38)
	N=count, Mean, median (range in µg/L)	N=count, Mean, median (range in µg/L)
DW-iAs: All	N=200 24.6, 13.0, (<0.456-236.0)	N=38 54.1, 24.2, (<0.456-236.0)
DW-iAs: < 10µg/L	N=93 1.7, <0.456, (<0.456-9.7)	N=17 1.4, <0.456, (<0.456-7.0)
DW-iAs: > 10µg/L	N=107 44.5, 25.4, (10.3-236.0)	N=21 96.8, 65.6, (16.4-236.0)
U-tAs (SG-normalized)	N=200 37.5, 23.3 (4.3-319.7)	N=38 73,87, 32.57 (6.2-319.7)

Of the drinking water samples collected from the parent cohort, approximately half (n=21, 55.2%) had DW-iAs levels that exceeded the WHO standard (10 µg/L) (**Table 1**). The levels of DW-iAs and U-tAs were significantly correlated in both the current subcohort (r = 0.74, p-value < 0.001, n=38) and the larger BEAR cohort (r = 0.51, p-value < 0.001, n=200). Additional demographic characteristics of the larger cohort participants are previously described (Rager et al. 2014).

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 beta values were calculated across the 38 subjects. The distribution of the median beta values for the analyzed probes (n=424,935) exhibited a bimodal pattern where probes displayed either very low or very high median methylation levels (Figure 1).

An adjusted multi-variable 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.

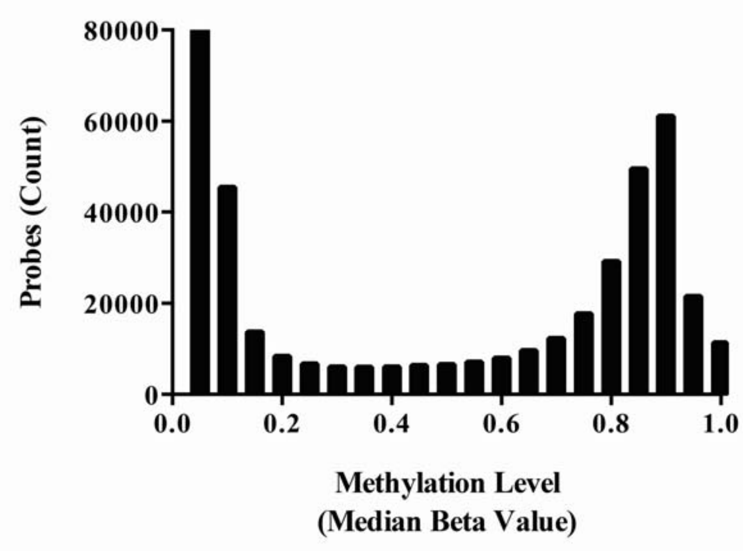

Figure 1. Median probe methylation value density for all probes analyzed (n=424,935).

A total of 4,771 probes, corresponding to 2,919 differentially methylated genes (DMGs), displayed differential U-tAs associated methylation (q<0.05) (**Table S1**). Of those probes, 34% (n=1,621) displayed hypo-methylation as U-tAs levels increased and 66% (n=3,150) displayed hyper-methylation as U-tAs levels increased.

Comparison of differentially methylated probes with previously published studies

The U-tAs-associated DMGs were compared to a list of 500 probes that are most informative of changes in leukocyte population as defined by Kile *et al.* (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) (**Table S2**). In addition, genes analyzed in previous studies including Ras association domain family 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<0.05) (**Table S2**) (Chanda et al. 2006; Gribble et al. 2014; Marsit et al. 2006; Pilsner et al. 2009). 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%) (Figure 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 five of six regions. Specifically, the 3'UTR and gene body were enriched within the probes that showed differential U-tAs associated DNA methylation, while the 1st exon, TSS200 and the 5' UTR were under-represented (Chi-Squared= 130.2, df = 5, p-value < 2.2×10^{-16}) (Figure 2).

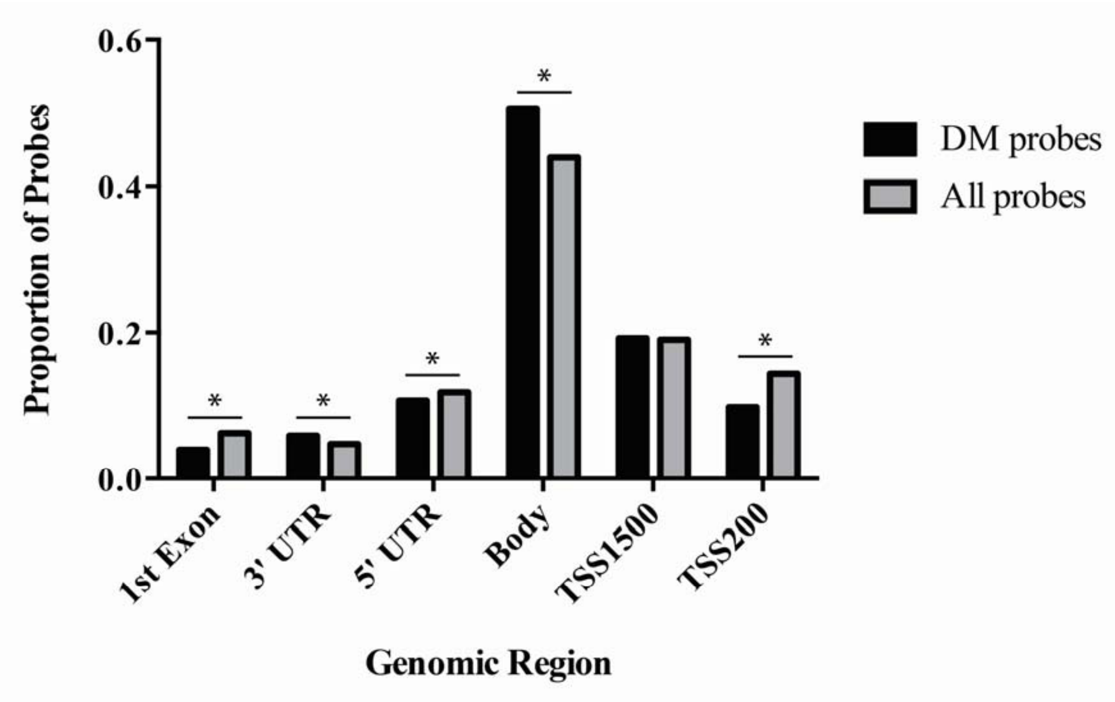

Figure 2. Enrichment of U-tAs associated differentially methylated (DM) probes across the gene associated regions. Proportion for all region-annotated probes in gray (n=321,417), annotated differentially methylated probes in black (n=319). Regions with significant deviation (p<0.05) from the platform distribution are indicated (*).

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=5, 000) versus genes with low expression (n=5, 000), the average methylation levels of the highest expressed genes (mean=0.37, SD=0.20) were significantly lower than the average methylation of the weakly expressed and/or non-expressed genes (p<0.0001; mean=0.56, SD=0.24) (Figure 3). Functional enrichment analysis comparing genes with the highest gene expression and lowest DNA methylation where β<0.3 (n=2,208 genes) versus those with the lowest expression and highest DNA methylation where β>0.7 (n=1,926 genes) was performed. High expression/low methylation genes were enriched for genes controlling core house-keeping 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 (Figure 3).

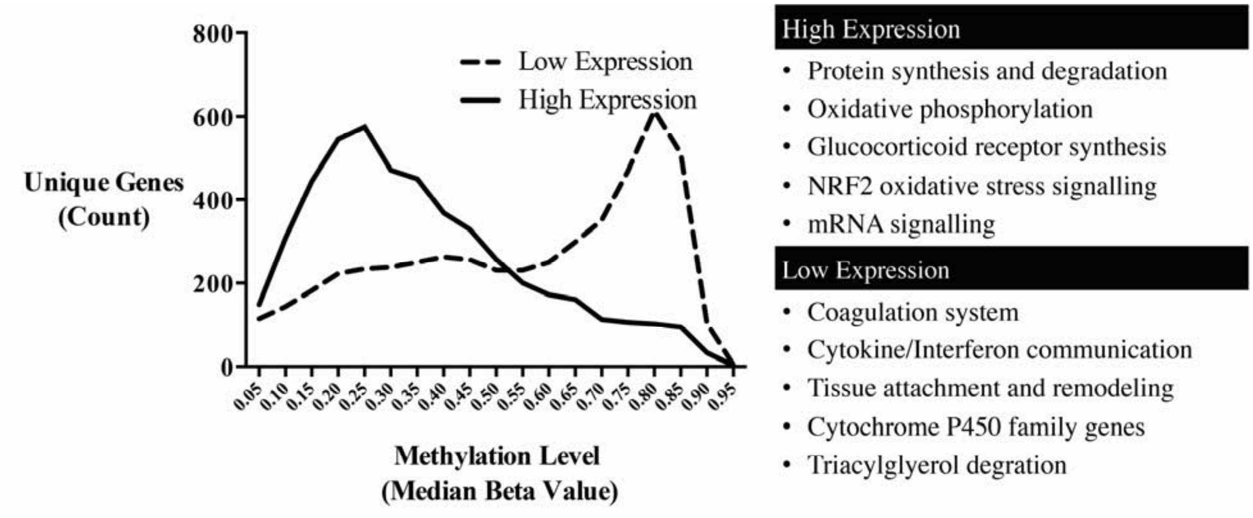

Figure 3. Density of average gene methylation values according to expression levels. The distribution in methylation levels for the genes with highest expression (n=5000) was compared to the genes with lowest expression (n=5000). Pathway analysis reveals enrichment of specific cellular functions in either high or low levels expression groups.

A total of 334 U-tAs-associated mRNAs were previously identified in cord blood leukocytes of BEAR newborns (Rager et al. 2014). It was shown that the majority of the transcripts (n=224) decreased in mRNA abundance as U-tAs increased, while 110 transcripts increased in mRNA abundance as U-tAs increased (Rager et al. 2014). In order to determine the relationship between U-tAs associated changes in gene expression (e.g. mRNA expression), and changes in DNA methylation, DNA methylation levels from these same study subjects were quantified and compared for all genes represented on both platforms (n=18, 761 genes). Of the 334 genes that showed differential gene expression, 269 were represented in the platform used to assess DNA methylation and could be matched to their corresponding methylation β-value (**Table S3**). A total of 20% (n=54 genes) of the differentially expressed genes showed at least one site of differential methylation. When analyzing the set of DMGs (n=2,705 genes) only 2% (n=54 genes) were also present in the list of differentially expressed genes (DEGs) (**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).

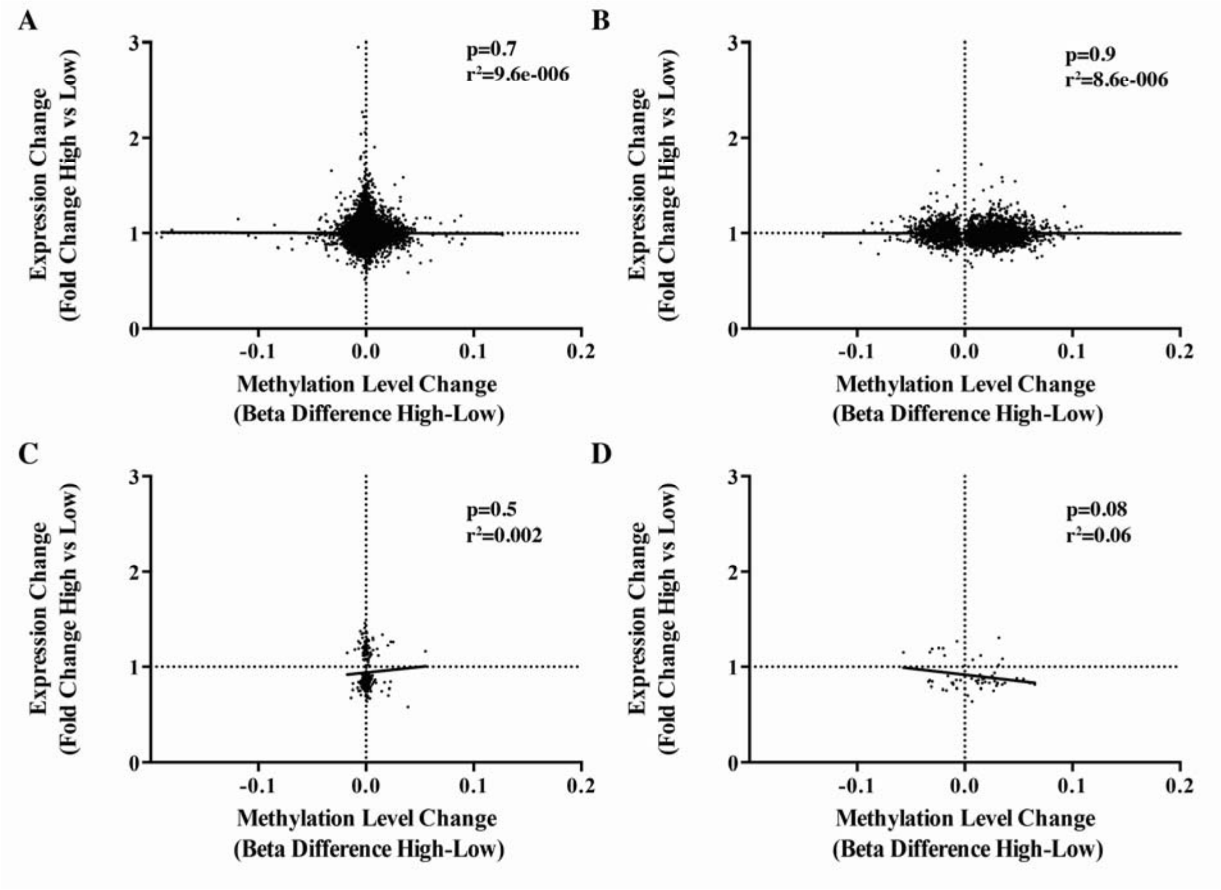

Figure 4. Comparison between U-tAs-associated DNA methylation and gene expression. Comparisons were made for: (A) all comparable genes (n=18761), (B) DMGs (n=2705), (C) DEGs (n=267), and (D) overlap of DMG and DEGs (n=54).

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 LEQ groups. Correlations between the expression fold change and methylation β difference were assessed for: (1) all genes (n=18,761) (Figure 4, A), (2) the DMGs (n=2,705) (Figure 4, B), (3) DEGs (n=267) (Figure 4, C) or (4) both DMGs and DEGs (n=54) (Figure 4, D). None of these correlations reached statistical significance (p>0.05), however a marginal correlation was observed between those genes that were common between the DMGs and DEGs (n=54) (r^2=0.06, p=0.08) (Figure 4, D). Similar correlation results were observed when data were normalized using the alternate BMIQ method (data not shown). 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 (1stExon), 5' untranslated region (5'UTR), 200 base pairs upstream of the transcriptional start site (TSS200), and from 200 to 1500 base pairs upstream of the transcriptional start site (TSS1500) (**Table 2**).

Table 2. Summary of correlations between iAs-associated changes in DNA methylation (median beta difference) and gene expression changes (FC-fold change) including all differentially expressed genes (DEGs). 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 1500bp-200bp upstream of the transcription start site (TSS1500), promoter region 200bp-0bp upstream of the transcription start site (TSS200). For each of these focus regions, correlations were calculated including either all CpG probes or only those in CpG islands.

	All Probes				**Only CpG Island Probes**			
Region	**Slope**	**RSq**	**p-value**	**n**	**Slope**	**RSq**	**p-value**	**n**
All	1.20E+00	1.82E-03	4.87E-01	269	-3.77E+00	2.10E-02	2.85E-02	228
1st Exon	3.50E+00	3.15E-02	4.02E-02	134	5.68E+00	5.03E-02	2.15E-02	105
3'UTR	1.54E-01	1.43E-04	8.79E-01	165	7.06E+00	1.61E-01	5.74E-02	23
5'UTR	-7.93E-01	1.94E-03	6.33E-01	120	-4.62E+00	7.27E-02	1.43E-02	82
Body	-2.19E-01	1.00E-04	8.76E-01	246	-4.68E+00	2.56E-02	5.36E-02	146
TSS1500	-4.00E-02	6.09E-06	9.69E-01	245	4.17E+00	3.51E-02	6.32E-02	99
TSS200	-7.54E-01	1.21E-03	6.13E-01	214	-4.90E+00	3.48E-02	2.62E-02	142

Of the genomic regions analyzed, differences in DNA methylation that occurred within the first exon were significantly correlated with changes in gene expression (p=0.04). Furthermore, when the analysis focused only on probes located within CpG islands, a highly significant association (p<0.05) was observed between U-tAs associated changes in gene expression and methylation in all regions (p=0.02). Significance was also observed for the correlation between probes located within CpG islands in the first exon (p=0.02). Furthermore, a significant negative correlation identified for those probes within CpG islands in the 5' UTR and TSS200 (p=0.03) (**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**).

Table 3. iAs-associated, differentially methylated CpG sites with significant association to gene expression levels. Associations were calculated using Pearson correlation.

CpG probe ID	Refgene Symbol	Gene Name	Expression Fold Change (HEQ /LEQ)	p-value (U-tAs)	q-value (U-tAs)	p -value (Expression)
cg16360836	*HDAC4*	Histone deacetylase 4	9.36E-01	7.02E-05	2.03E-02	2.00E-04
cg24671666	*PLEKHG3*	Pleckstrin homology domain containing, family G (with RhoGef domain) member 3	7.64E-01	3.30E-04	3.62E-02	1.90E-03
cg10632215	*CERK*	Ceramide kinase	8.61E-01	5.06E-04	4.30E-02	4.00E-03
cg15601244	*SH2B2*	SH2B adaptor protein 2	8.49E-01	7.31E-05	2.05E-02	4.10E-03
cg26489875	*MALAT1*	Metastasis associated lung adenocarcinoma transcript 1 (non-protein coding)	1.12E+00	4.82E-04	4.20E-02	4.40E-03
cg15883181	*RNF213*	Ring finger protein 213	8.28E-01	7.63E-05	2.07E-02	4.90E-03
cg18054302	*GAK*	Cyclin G associated kinase	8.86E-01	3.06E-04	3.53E-02	5.30E-03
cg22798758	*PTPRE*	Protein tyrosine phosphatase, receptor type, E	8.61E-01	2.95E-04	3.48E-02	6.00E-03
cg09112262	*WDR55*	WD repeat domain 55	8.28E-01	4.63E-04	4.13E-02	6.10E-03
cg20851790	*RNF213*	Ring finger protein 213	8.28E-01	4.25E-04	3.98E-02	6.10E-03
cg07199894	*ULK1*	Unc-51 like autophagy activating kinase 1	1.15E+00	6.51E-04	4.69E-02	7.80E-03
cg00627621	*TRIB1*	Tribbles homolog 1 (Drosophila)	8.41E-01	3.35E-04	3.65E-02	1.22E-02
cg15574972	*MALAT1*	Metastasis associated lung adenocarcinoma transcript 1 (non-protein coding)	1.12E+00	1.82E-04	2.89E-02	1.30E-02
cg18312113	*SSU72*	SSU72 RNA polymerase II CTD phosphatase homolog (S. cerevisiae)	8.76E-01	4.22E-04	3.97E-02	1.72E-02
cg20689294	*PTPRE*	Protein tyrosine phosphatase, receptor type, E	8.61E-01	5.48E-04	4.40E-02	1.77E-02
cg06719391	*KCNQ1*	Potassium voltage-gated channel, KQT-like subfamily, member 1	8.59E-01	2.91E-05	1.49E-02	1.80E-02
cg02582997	*GAK*	Cyclin G associated kinase	8.86E-01	5.43E-04	4.40E-02	2.07E-02

cg03041730	*GAK*	Cyclin G associated kinase	8.86E-01	2.99E-05	1.50E-02	2.30E-02
cg24725201	*KCNQ1*	Potassium voltage-gated channel, KQT-like subfamily, member 1	8.59E-01	1.08E-04	2.36E-02	2.67E-02
cg24089935	*KCNQ1*	Potassium voltage-gated channel, KQT-like subfamily, member 1	8.59E-01	7.98E-05	2.10E-02	2.77E-02
cg06107260	*HDAC4*	Histone deacetylase 4	9.36E-01	9.26E-06	1.04E-02	2.94E-02
cg03926050	*PPTC7*	PTC7 protein phosphatase homolog (S. cerevisiae)	8.13E-01	2.93E-04	3.47E-02	3.86E-02
cg01693193	*KCNQ1*	Potassium voltage-gated channel, KQT-like subfamily, member 1	8.59E-01	3.04E-05	1.51E-02	3.94E-02
cg04295928	*PGAP2*	Post-GPI attachment to proteins 2	8.69E-01	3.41E-05	1.57E-02	4.62E-02
cg08438705	*SBNO2*	Strawberry notch homolog 2 (Drosophila)	8.51E-01	4.91E-04	4.24E-02	4.79E-02

To note, when contrasted to the total number of DMGs (n=2,705), this represents <1% of the total. Representative plots for probes with significant correlation between expression and methylation are shown (Figure 5).

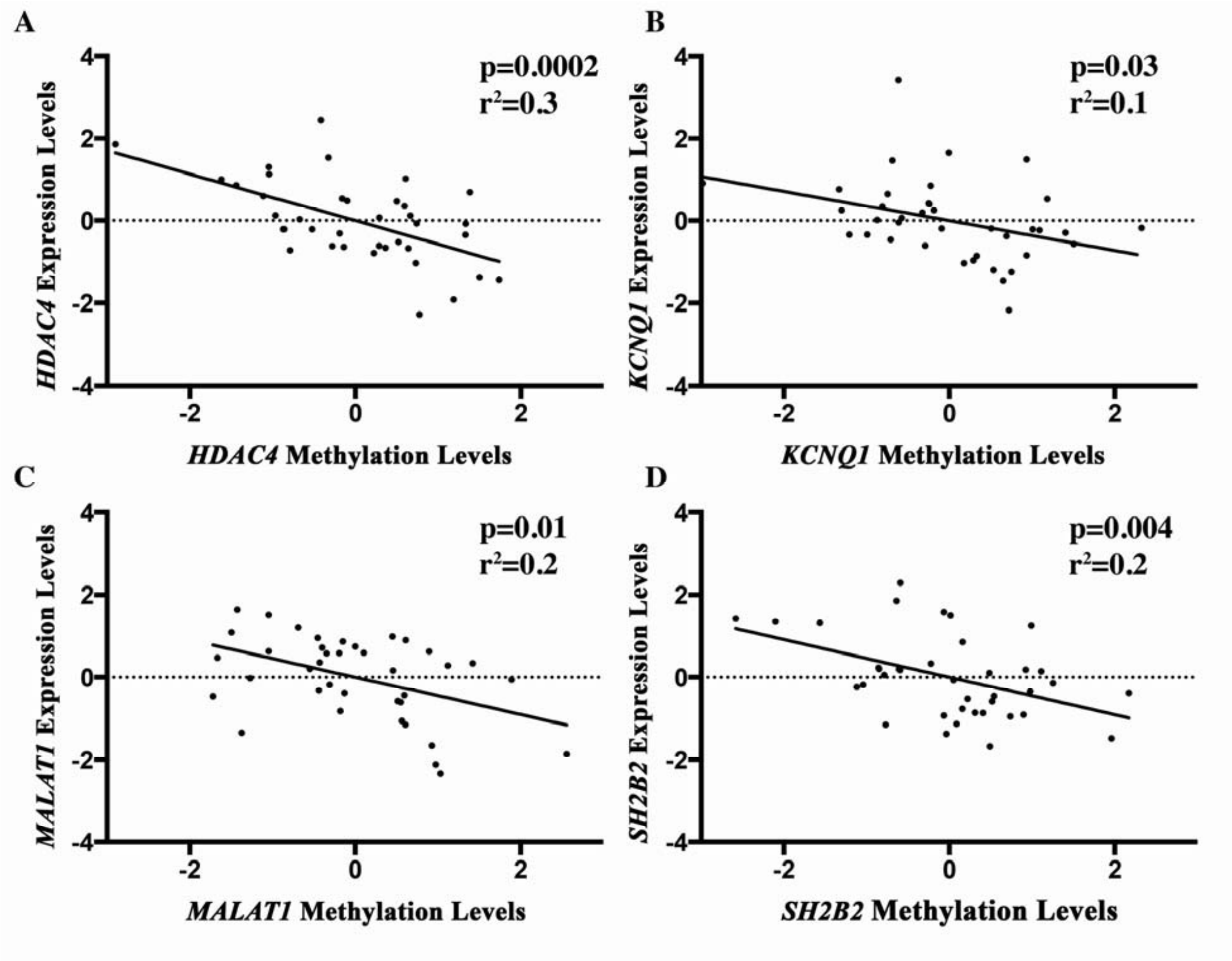

Figure 5. Representative plots of genes with significant correlation between U-tAs associated DNA methylation and U-tAs associated gene expression.

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). 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 213 (*RNF213*), and PTC7 protein phosphatase homolog (*PPTC7*) (**Table 4**).

Table 4. CpG sites associated with birth outcomes in the BEAR cohort. Associations between outcomes and CpG sites were calculated using Pearson correlation.

Birth Outcome	Refgene Symbol	Gene Name	CpG probe ID	p-value	R	n
Gestational age	*PTPRE*	Protein tyrosine phosphatase, receptor type, E	cg22798758	3.74E-02	-0.34	38
Gestational age	*WDR55*	WD repeat domain 55	cg09112262	4.08E-02	-0.33	38
Gestational age	*PTPRE*	Protein tyrosine phosphatase, receptor type, E	cg20689294	1.76E-02	-0.38	38
Gestational age	*KCNQ1*	Potassium voltage-gated channel, KQT-like subfamily, member 1	cg06719391	2.63E-02	-0.36	38
Gestational age	*GAK*	Cyclin G associated kinase	cg02582997	4.35E-02	-0.33	38
Gestational age	*PGAP2*	Post-GPI attachment to proteins 2	cg04295928	2.75E-03	-0.47	38
Placenta weight	*RNF213*	Ring finger protein 213	cg15883181	4.72E-02	0.33	36
Placenta weight	*PTPRE*	Protein tyrosine phosphatase, receptor type, E	cg20689294	3.45E-03	0.47	36
Placenta weight	*KCNQ1*	Potassium voltage-gated channel, KQT-like subfamily, member 1	cg06719391	1.61E-02	0.4	36
Placenta weight	*GAK*	Cyclin G associated kinase	cg02582997	3.53E-02	0.35	36
Placenta weight	*PGAP2*	Post-GPI attachment to proteins 2	cg04295928	1.20E-02	0.41	36
Head circumference	*PPTC7*	PTC7 protein phosphatase homolog (S. cerevisiae)	cg03926050	4.73E-02	0.33	36

Transcription factor enrichment analysis in differentially methylated genes and differentially expressed genes

In order to identify potential upstream regulators that could impact locations of genomic DNA methylation, gene expression levels and health outcomes, enrichment analysis for transcription factor 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 transcription factors (TFs) were enriched within the gene set when compared to genes with altered expression but no correlation with methylation (**Table S4**). The top five most significant TFs enriched in the gene set included CCCTC-binding factor (zinc finger protein)-like (CTCF) (p<0.00001), early growth response (EGR) (p<0.00001), zinc finger and BTB domain containing 14 (ZF5), GLUT4 enhancer factor (HDBP) (p<0.00001), hairy and enhancer of split 1 (HES1) (p=0.00006) (**Table S4**).

CHAPTER 4: DISCUSSION, LIMITATIONS, FUTURE DIRECTIONS AND CONCLUSIONS

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 as a functional read-out along with 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 (>2,000) 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 gene expression level, a finding that is highly relevant to ongoing research that does not include functional readouts at a transcriptional or proteomic level. In contrast to earlier work estimating that ~20% of the genomic response to prenatal iAs is controlled by miRNAs (Rager et al. 2014), the results here suggest that the transcriptional response may be minimally (~5% or 16/267) controlled by DNA methylation changes. Importantly a set of seven of the differentially methylated genes that were predictive of functional change at the mRNA level were also associated with birth outcomes including gestational age. These genes were enriched for binding sites for transcription factors and support a hypothesis of DNA methylation patterns as "environmental footprints" (Sanders et al. 2014) of transcription factor occupancy.

When analyzed in an arsenic-independent manner, the 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, and genes with low expression displayed higher levels of methylation, a finding that 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 amongst 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). 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=18,761 genes), a result that did not depend on the data normalization method used. These results highlight that changes in mRNA expression associated with U-tAs are only weakly correlated to changes in gene methylation on a genome-wide scale. Given the focus of much literature pertaining to CpG methylation on the gene-silencing effects of DNA methylation (Jones 2012) these results may seem unexpected. It is important to note, however, that such low correlations between gene expression and DNA methylation levels have been observed previously. For example, the Meissner lab studied genome-wide DNA methylation and expression changes 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 differentially methylated genes 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 changes in 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.

We identified CpG methylated sites that best predict gene expression levels by performing region-specific analysis. Probes within CpG islands located in the first exon, 5' UTR and TSS200 were identified as the most predictive of gene expression, however, each representing differential effects on transcription. Our research supports previous findings of promoter methylation associated with transcriptional repression and conversely, first exon DNA methylation as a hallmark of transcribed genes (Jjingo et al. 2012; Jones 2012). Together our data reveal the functional differences between CpG methylation in distinct genomic regions and demonstrate that DNA methylation changes positioned within CpG islands of the first exon and TSS200 most 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 mechanism by which to 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, seven of these 16 genes also displayed an association between DNA methylation and birth outcomes, specifically gestational age, placental weight and head circumference. Interestingly, two of the five 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). Furthermore, these sixteen functionally consequential genes display an enrichment for binding sites of specific transcription factors including EGR and CTCF, both known to be altered by arsenic and their modulation may impact various cellular signaling pathways (Simeonova et al. 2000; Xu et al. 2013). The current research provides insight into the transcriptional regulation that can potentially influence DNA methylation patterning and elicit functional change. 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 transcription factors amongst these genes. This finding suggested the potential for "environmental footprints" of prior transcription factor occupancy during times of DNA methylation (Sanders et al. 2014). The present study provides *in silico* support for the hypothesis that occupancy of U-tAs-associated transcription factors may impact the methylation status of a subset of differentially expressed and functionally consequential genes.

This study is not without limitations. The sample size in our study is relatively small thus these findings should be further validated in larger cohorts. Nevertheless, the fact that many of the gene targets displaying altered methylation and particularly those with association to birth outcomes have been found in other studies, provides support for biological relevance. Future work should also investigate whether the changes in DNA methylation and/or gene expression observed here are 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.

APENDIX 1: TABLE S1

Table S1. U-tAs associated differentially methylated probes (q<0.05) First page shown.

CpG ID	RefGene Symbol2	Gene Name	Partial Correlation (U-tAs)	p-value (U-tAs)	q-value (U-tAs)
cg19815813	*A2LD1*	gamma-glutamylamine cyclotransferase	5.71E-01	2.72E-04	3.39E-02
cg20367788	*AAGAB*	alpha- and gamma-adaptin binding protein	5.73E-01	2.09E-04	3.06E-02
cg06048605	*ABCA2*	ATP-binding cassette, sub-family A (ABC1), member 2	6.33E-01	1.81E-05	1.27E-02
cg09632163	*ABCA3*	ATP-binding cassette, sub-family A (ABC1), member 3	7.14E-01	8.06E-07	5.48E-03
cg25793628	*ABCA5*	ATP-binding cassette, sub-family A (ABC1), member 5	5.48E-01	3.48E-04	3.68E-02
cg27656398	*ABCB9*	ATP-binding cassette, sub-family B (MDR/TAP), member 9	-5.32E-01	3.33E-04	3.64E-02
cg18016288	*ABCC4*	ATP-binding cassette, sub-family C (CFTR/MRP), member 4	-5.77E-01	1.51E-04	2.70E-02
cg08499057	*ABCF2*	ATP-binding cassette, sub-family F (GCN20), member 2	5.81E-01	1.20E-04	2.47E-02
cg11113753	*ABCG5*	ATP-binding cassette, sub-family G (WHITE), member 5	-5.86E-01	8.04E-05	2.11E-02
cg23527366	*ABHD14A*	abhydrolase domain containing 14A	5.36E-01	6.71E-04	4.74E-02
cg15077643	*ABI2*	abl-interactor 2	6.68E-01	7.74E-06	9.58E-03
cg25004737	*ABI2*	abl-interactor 2	6.44E-01	2.03E-05	1.31E-02
cg27169846	*ABLIM2*	actin binding LIM protein family, member 2	6.85E-01	1.85E-06	6.11E-03
cg19635501	*ABLIM2*	actin binding LIM protein family, member 2	5.84E-01	1.56E-04	2.73E-02
cg04514683	*ABP1*	drebrin-like	4.42E-01	2.53E-04	3.29E-02
cg17411020	*ABR*	active BCR-related	-5.76E-01	2.14E-04	3.09E-02
cg02110776	*ABR*	active BCR-related	-5.54E-01	3.80E-04	3.81E-02
cg07646714	*ABTB2*	ankyrin repeat and BTB (POZ) domain containing 2	-6.38E-01	2.39E-05	1.38E-02
cg01369908	*ABTB2*	ankyrin repeat and BTB (POZ) domain containing 2	-6.09E-01	5.39E-05	1.85E-02
cg01178601	*ABTB2*	ankyrin repeat and BTB (POZ) domain containing 2	-5.95E-01	7.58E-05	2.07E-02
cg07834934	*ACACA*	acetyl-CoA carboxylase alpha	6.22E-01	4.94E-05	1.78E-02

APENDIX 2: TABLE S2

Table S2. Differentially methylated probes identified in previously published arsenic research and their relationship to current findings in the BEAR cohort. First page shown.

Author	**CpG probe ID**	**RefGene Symbol**	**Gene Name**	**Direction of change in high arsenic group (Previous research)**	**BEAR Partial Correlation (U-tAs)**	**BEAR q-value (U-tAs)**	**BEAR p-value (U-tAs)**	**Differentially Methylated in BEAR (Yes/No)**
Chanda et al. (2006)	cg01620719	*TP53*	tumor protein p53	Hypermethylated promoter	4.28E-01	1.15E-01	6.62E-03	No
Chanda et al. (2006)	cg02045224	*TP53*	tumor protein p53	Hypermethylated promoter	1.12E-01	5.97E-01	4.96E-01	No
Chanda et al. (2006)	cg02087342	*TP53*	tumor protein p53	Hypermethylated promoter	-3.33E-01	2.41E-01	4.40E-02	No
Chanda et al. (2006)	cg02166782	*TP53*	tumor protein p53	Hypermethylated promoter	-5.15E-02	6.87E-01	7.64E-01	No
Chanda et al. (2006)	cg02842899	*TP53*	tumor protein p53	Hypermethylated promoter	1.28E-01	5.75E-01	4.45E-01	No
Chanda et al. (2006)	cg03079681	*p16*	cyclin-dependent kinase inhibitor 2A	Hypermethylated promoter	-7.40E-02	6.58E-01	6.66E-01	No
Chanda et al. (2006)	cg06317056	*TP53*	tumor protein p53	Hypermethylated promoter	-8.21E-02	6.47E-01	6.30E-01	No
Chanda et al. (2006)	cg06365412	*TP53*	tumor protein p53	Hypermethylated promoter	3.14E-01	2.50E-01	4.84E-02	No
Chanda et al. (2006)	cg07562918	*p16*	cyclin-dependent kinase inhibitor 2A	Hypermethylated promoter	-2.34E-01	3.77E-01	1.40E-01	No
Chanda et al. (2006)	cg07760161	*TP53*	tumor protein p53	Hypermethylated promoter	2.04E-01	4.50E-01	2.23E-01	No
Chanda et al. (2006)	cg07991600	*TP53*	tumor protein p53	Hypermethylated promoter	-2.12E-01	4.38E-01	2.07E-01	No

APENDIX 3: TABLE S3

Table S3. Differentialy expressed genes and their matched changes in gene methylation. Partial correlation denotes the coefficient of the association between the methylation at the given CpG site to U-tAs using a multivariate model. Adjusted p-values for false discovery rate are represented as q-values. First page shown.

Column ID	Refgene Symbol	Gene Name	Expression FC (HEQ /LEQ)	Methylation Partial Correlation (U-tAs)	Methylation q-value (U-tAs)	Methylation p-value (U-tAs)
cg22534545	*ABCD1*	ATP-binding cassette, sub-family D (ALD), member 1	8.83E-01	2.10E-01	3.34E-01	1.02E-01
cg27024381	*ABCD1*	ATP-binding cassette, sub-family D (ALD), member 1	8.83E-01	-3.59E-02	7.05E-01	8.34E-01
cg00091063	*ACAA1*	acetyl-CoA acyltransferase 1	8.45E-01	8.39E-02	6.40E-01	6.11E-01
cg00343022	*ACAA1*	acetyl-CoA acyltransferase 1	8.45E-01	-1.09E-01	6.01E-01	5.06E-01
cg00884680	*ACAA1*	acetyl-CoA acyltransferase 1	8.45E-01	-3.33E-02	7.08E-01	8.45E-01
cg01353464	*ACAA1*	acetyl-CoA acyltransferase 1	8.45E-01	-3.43E-01	1.93E-01	2.50E-02
cg02314846	*ACAA1*	acetyl-CoA acyltransferase 1	8.45E-01	2.62E-01	3.50E-01	1.15E-01
cg06069310	*ACAA1*	acetyl-CoA acyltransferase 1	8.45E-01	6.80E-02	6.62E-01	6.77E-01
cg06239037	*ACAA1*	acetyl-CoA acyltransferase 1	8.45E-01	1.96E-01	4.71E-01	2.52E-01
cg06808967	*ACAA1*	acetyl-CoA acyltransferase 1	8.45E-01	-2.18E-01	4.33E-01	2.01E-01
cg07895684	*ACAA1*	acetyl-CoA acyltransferase 1	8.45E-01	3.01E-01	2.82E-01	6.61E-02
cg10548708	*ACAA1*	acetyl-CoA acyltransferase 1	8.45E-01	-3.25E-01	2.58E-01	5.23E-02
cg15348640	*ACAA1*	acetyl-CoA acyltransferase 1	8.45E-01	2.09E-01	3.93E-01	1.56E-01
cg15427004	*ACAA1*	acetyl-CoA acyltransferase 1	8.45E-01	3.85E-01	1.58E-01	1.51E-02
cg17034030	*ACAA1*	acetyl-CoA acyltransferase 1	8.45E-01	-1.33E-01	5.73E-01	4.39E-01
cg17696234	*ACAA1*	acetyl-CoA acyltransferase 1	8.45E-01	2.01E-01	4.58E-01	2.34E-01
cg24368167	*ACAA1*	acetyl-CoA acyltransferase 1	8.45E-01	4.74E-02	6.92E-01	7.83E-01
cg25525163	*ACAA1*	acetyl-CoA acyltransferase 1	8.45E-01	3.76E-01	1.88E-01	2.34E-02
cg00337466	*ADAM8*	ADAM metallopeptidase domain 8	8.37E-01	1.43E-01	4.72E-01	2.54E-01
cg00805360	*ADAM8*	ADAM metallopeptidase domain 8	8.37E-01	4.07E-01	1.40E-01	1.11E-02

APENDIX 4: TABLE S4

Table S4. Enriched transcription factor binding sites in DEGs with correlated changes in gene expression and gene methylation (n=16) versus DEGs with no correlation between changes in genes expression and DNA methylation (n=318).

Transcription Factor Binding Site	**Description**	**p-score (DEG)**
CTCF	CCCTC-binding factor (zinc finger protein)	1.54E-09
EGR1	early growth response 1	1.70E-07
ZF5	zinc finger and BTB domain containing 14	1.37E-06
HDBP1	SLC2A4 regulator	1.19E-05
HES1	hairy and enhancer of split 1, (Drosophila)	6.61E-05
AP2	fatty acid binding protein 4, adipocyte	1.45E-04
MYCMAX	c-Myc/Max heterodimer	3.85E-04
XCPE1	X gene core promoter element 1	6.99E-04
GAGA	GAGA-Box	9.67E-04
FTF	nuclear receptor subfamily 5, group A, member 2	1.24E-03
SP1	Sp1 transcription factor	1.52E-03
DICE	Downstream Immunoglobulin Control Element	2.53E-03
E2F2	E2F transcription factor 2	2.98E-03
ATF	plasminogen activator, urokinase	3.39E-03
HMTE	Human motif ten element	3.85E-03
NRF1	nuclear respiratory factor 1	4.37E-03
CREB	cAMP responsive element binding protein 1	4.95E-03
NKX11	NK1 homeobox 1, SAX2	5.44E-03
USF	Upstream stimulating factor	5.78E-03
CDE	Cell cycle-dependent element, CDF-1 binding site (CDE/CHR tandem elements regulate cell cycle dependent repression)	6.14E-03
NRSE	Neural-restrictive-silencer-element	6.93E-03
CMYC	Myelocytomatosis oncogene (c-myc proto-oncogene)	7.59E-03
NMYC	v-myc avian myelocytomatosis viral oncogene neuroblastoma derived homolog	1.02E-02

E2F4	E2F transcription factor 4, p107/p130-binding	1.31E-02
WHN	forkhead box N1	1.31E-02
MAX	MYC associated factor X	1.60E-02
GCM1	glial cells missing homolog 1 (Drosophila)	1.93E-02
MYOGENIN	Myogenic bHLH protein myogenin (myf4)	1.98E-02
PAX5	paired box 5	3.00E-02
BACH2	BTB and CNC homology 1, basic leucine zipper transcription factor 2	3.08E-02
NANOG	Nanog homeobox	3.24E-02
MYBL1	v-myb avian myeloblastosis viral oncogene homolog-like 1	3.57E-02
CREB1	cAMP responsive element binding protein 1	3.85E-02
STAT3	signal transducer and activator of transcription 3 (acute-phase response factor)	3.94E-02
HELT	helt bHLH transcription factor	4.14E-02
E2F3	E2F transcription factor 3	4.34E-02

REFERENCES

Andrew AS, Karagas MR, Hamilton JW. 2003. Decreased DNA repair gene expression among individuals exposed to arsenic in united states drinking water. International journal of cancer Journal international du 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, Fry RC. 2014. Long-term health consequences of prenatal arsenic exposure: Links to the genome and the epigenome. Reviews on environmental health 29:9-12.

Bell JT, Pai AA, Pickrell JK, Gaffney DJ, Pique-Regi R, Degner JF, et al. 2011. DNA methylation patterns associate with genetic and gene expression variation in hapmap cell lines. Genome biology 12:R10.

Bock C, Beerman I, Lien WH, Smith ZD, Gu H, Boyle P, et al. 2012. DNA methylation dynamics during in vivo differentiation of blood and skin stem cells. Molecular cell 47:633-647.

Boellmann F, Zhang L, Clewell HJ, Schroth GP, Kenyon EM, Andersen ME, et al. 2010. Genome-wide analysis of DNA methylation and gene expression changes in the mouse lung following subchronic arsenate exposure. Toxicological sciences : an official journal of the Society of Toxicology 117:404-417.

Bolstad BM, Irizarry RA, Astrand M, Speed TP. 2003. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. Bioinformatics 19:185-193.

Chanda S, Dasgupta UB, Guhamazumder D, Gupta M, Chaudhuri U, Lahiri S, et al. 2006. DNA hypermethylation of promoter of gene p53 and p16 in arsenic-exposed people with and without malignancy. Toxicological sciences : an official journal of the Society of Toxicology 89:431-437.

Dosunmu R, Alashwal H, Zawia NH. 2012. Genome-wide expression and methylation profiling in the aged rodent brain due to early-life pb exposure and its relevance to aging. Mechanisms of ageing and development 133:435-443.

Fernandez AF, Assenov Y, Martin-Subero JI, Balint B, Siebert R, Taniguchi H, et al. 2012. A DNA methylation fingerprint of 1628 human samples. Genome research 22:407-419.

Fry RC, Navasumrit P, Valiathan C, Svensson JP, Hogan BJ, Luo M, et al. 2007. Activation of inflammation/nf-kappab signaling in infants born to arsenic-exposed mothers. PLoS genetics 3.

Gribble MO, Tang WY, Shang Y, Pollak J, Umans JG, Francesconi KA, et al. 2014. Differential methylation of the arsenic (iii) methyltransferase promoter according to arsenic exposure. Archives of toxicology 88:275-282.

Ho Sui SJ, Mortimer JR, Arenillas DJ, Brumm J, Walsh CJ, Kennedy BP, et al. 2005. Opossum: Identification of over-represented transcription factor binding sites in co-expressed genes. Nucleic acids research 33:3154-3164.

Jjingo D, Conley AB, Yi SV, Lunyak VV, Jordan IK. 2012. On the presence and role of human gene-body DNA methylation. Oncotarget 3:462-474.

Jomova K, Jenisova Z, Feszterova M, Baros S, Liska J, Hudecova D, et al. 2011. Arsenic: Toxicity, oxidative stress and human disease. Journal of applied toxicology : JAT 31:95-107.

Jones PA. 2012. Functions of DNA methylation: Islands, start sites, gene bodies and beyond. Nature reviews Genetics 13:484-492.

Joubert BR, Haberg SE, Nilsen RM, Wang X, Vollset SE, Murphy SK, et al. 2012. 450k epigenome-wide scan identifies differential DNA methylation in newborns related to maternal smoking during pregnancy. Environmental health perspectives 120:1425-1431.

Kile ML, Baccarelli A, Hoffman E, Tarantini L, Quamruzzaman Q, Rahman M, et al. 2012. Prenatal arsenic exposure and DNA methylation in maternal and umbilical cord blood leukocytes. Environmental health perspectives 120:1061-1066.

Kile ML, Houseman EA, Baccarelli A, Quamruzzaman Q, Rahman M, Mostofa G, et al. 2014. Effect of prenatal arsenic exposure on DNA methylation and leukocyte subpopulations in cord blood. Epigenetics : official journal of the DNA Methylation Society 9.

Lee H, Jaffe AE, Feinberg JI, Tryggvadottir R, Brown S, Montano C, et al. 2012. DNA methylation shows genome-wide association of nfix, rapgef2 and msrb3 with gestational age at birth. International journal of epidemiology 41:188-199.

Liu J, Xie Y, Ward JM, Diwan BA, Waalkes MP. 2004. Toxicogenomic analysis of aberrant gene expression in liver tumors and nontumorous livers of adult mice exposed in utero to inorganic arsenic. Toxicological sciences : an official journal of the Society of Toxicology 77:249-257.

Liu J, Xie Y, Ducharme DM, Shen J, Diwan BA, Merrick BA, et al. 2006. Global gene expression associated with hepatocarcinogenesis in adult male mice induced by in utero arsenic exposure. Environmental health perspectives 114:404-411.

Marsit CJ, Karagas MR, Danaee H, Liu M, Andrew A, Schned A, et al. 2006. Carcinogen exposure and gene promoter hypermethylation in bladder cancer. Carcinogenesis 27:112-116.

Pidsley R, CC YW, Volta M, Lunnon K, Mill J, Schalkwyk LC. 2013. A data-driven approach to preprocessing illumina 450k methylation array data. BMC genomics 14:293.

Pilsner JR, Liu X, Ahsan H, Ilievski V, Slavkovich V, Levy D, et al. 2009. Folate deficiency, hyperhomocysteinemia, low urinary creatinine, and hypomethylation of leukocyte DNA are risk factors for arsenic-induced skin lesions. Environmental health perspectives 117:254-260.

Rager JE, Bailey KA, Smeester L, Miller SK, Parker JS, Laine JE, et al. 2014. Prenatal arsenic exposure and the epigenome: Altered micrornas associated with innate and adaptive immune signaling in newborn cord blood. Environmental and molecular mutagenesis 55:196-208.

Sanders AP, Smeester L, Rojas D, Debussycher T, Wu MC, Wright FA, et al. 2014. Cadmium exposure and the epigenome: Exposure-associated patterns of DNA methylation in leukocytes from mother-baby pairs. Epigenetics : official journal of the DNA Methylation Society 9:212-221.

Simeonova PP, Wang S, Toriuma W, Kommineni V, Matheson J, Unimye N, et al. 2000. Arsenic mediates cell proliferation and gene expression in the bladder epithelium: Association with activating protein-1 transactivation. Cancer research 60:3445-3453.

Smeester L, Rager JE, Bailey KA, Guan X, Smith N, Garcia-Vargas G, et al. 2011. Epigenetic changes in individuals with arsenicosis. Chemical research in toxicology 24:165-167.

Smith ZD, Chan MM, Mikkelsen TS, Gu H, Gnirke A, Regev A, et al. 2012. A unique regulatory phase of DNA methylation in the early mammalian embryo. Nature 484:339-344.

Teschendorff AE, Marabita F, Lechner M, Bartlett T, Tegner J, Gomez-Cabrero D, et al. 2013. A beta-mixture quantile normalization method for correcting probe design bias in illumina infinium 450 k DNA methylation data. Bioinformatics 29:189-196.

Tobi EW, Lumey LH, Talens RP, Kremer D, Putter H, Stein AD, et al. 2009. DNA methylation differences after exposure to prenatal famine are common and timing- and sex-specific. Human molecular genetics 18:4046-4053.

WHO. 2006. World health organization. Guidelines for drinking water quality. (First addendum to 3rd addition, Volume 1). Geneva, Switzerland:WHO Press.

Xu H, Lam SH, Shen Y, Gong Z. 2013. Genome-wide identification of molecular pathways and biomarkers in response to arsenic exposure in zebrafish liver. PloS one 8:e68737.