Published in final edited form as:

Environ Int. 2018 October ; 119: 250–263. doi:10.1016/j.envint.2018.06.024.

Multi-generational Impacts of Arsenic Exposure on Genome-wide DNA Methylation and the Implications for Arsenic-induced Skin Lesions

Xiaojuan Guo^{a,b,*}, Xushen Chen^{c,*}, Jie Wang^{d,*}, Zhiyue Liu^a, Daniel Gaile^e, Hongmei Wu^b, Guan Yu^e, Guangyun Mao^b, Zuopeng Yang^f, Zhen Di^g, Xiuqing Guo^a, Li Cao^a, Peiye Chang^a, Binxian Kang^f, Jinyu Chen^f, Wen Gao^f, and Xuefeng Ren^{a,c,#}

^aSchool of Public Health, Inner Mongolia Medical University, Hohhot, Inner Mongolia, China

^bSchool of Public Health & Management, Wenzhou Medical University, Wenzhou, Zhejiang, China

^cDepartment of Epidemiology and Environmental Health, School of Public Health and Health Professions, University at Buffalo, Buffalo, New York, USA

^dDepartment of Biochemistry, Jacobs School of Medicine and Biomedical Sciences, University at Buffalo, Buffalo, New York, USA

^eDepartment of Biostatistics, School of Public Health and Health Professions, University at Buffalo, Buffalo, New York, USA

^fCenter for Disease Control and Prevention, Wuyuan County, Inner Mongolia, China

^gCenter for Disease Control and Prevention, Hangjinhouqi County, Inner Mongolia, China

Abstract

As a nonmutagenic human carcinogen, arsenic (As)'s carcinogenic activity is likely the result of epigenetic changes, particularly alterations in DNA methylation. While increasing studies indicate a potentially important role for timing of As exposure on DNA methylation patterns and the subsequent differential risks for As toxicity and carcinogenesis, there is a lack of research that tackles these critical questions, particularly in human based populations. Here we reported a family-based study including three generations, in which each generation living in the same household had a distinctive timing of As exposure: in adulthood, *in utero* and during early childhood, and in germlines exposure for grandparents, parents, and grandchildren, respectively. We generated genome-wide DNA methylation data for 18 As-exposed families, nine control

Address correspondence to X. Ren, Department of Epidemiology and Environmental Health, School of Public Health and Health Professions, University at Buffalo, 3435 Main Street, Buffalo, New York 14214. Telephone: 716-829-5384. xuefengr@buffalo.edu.

*These authors contributed equally.

Data and Materials Availability:

Genomic methylation data are available on the Gene Expression Omnibus database (GSE Accession numbers GSE109914).

Declarations of Interest:

None declared.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

families, as well as 18 arsenical skin lesion patients. Our analysis showed that As exposure may leave detectable DNA methylation changes even though exposure occurred decades ago, and the most significant changes of global DNA methylation were observed among patients afflicted with arsenical skin lesions. As exposure across generations shared common differentially methylated DNA loci and regions (744 DML and 15 DMRs) despite the distinctive exposure timing in each generation. Importantly, based on these DML, clustering analysis grouped skin lesion patients together with grandparents in exposed families in the same cluster, separated from grandparents in control families. Further analysis identified a number of DML and several molecular pathways that were significantly distinguished between controls, exposed populations, as well as skin lesion patients. Finally, our exploratory analysis suggested that some of these DML altered by As exposure, may have the potential to be inherited affecting not only those directly exposed but also later generations. Together, our results suggest that common DML and/or DMRs associated with an increased risk for disease development could be identified regardless of when exposure to As occurred during their life span, and thus may be able to serve as biomarkers for identifying individuals at risk for As-induced skin lesions and possible cancers.

Keywords

Arsenic exposure; Arsenic-induced skin lesions; Epigenetic inheritance; Global and genome-wide DNA methylation; Multi-generational epigenetic effect

1. Introduction

Exposure to arsenic (As) through drinking water is a significant world-wide public health concern (Abernathy et al., 1999; IARC, 2004; Pershagen, 1981; Smith et al., 1992; Tapio and Grosche, 2006). As, a group I carcinogen, does not fall into the classic model of carcinogenesis, as it is not efficient at inducing point mutations or initiating and promoting the development of tumors in experimental animals (Jacobson-Kram and Montalbano, 1985; Jongen et al., 1985; Simeonova and Luster, 2000). There is mounting evidence from *in vitro*, animal, and human studies that As is a strong regulator of DNA methylation (Argos et al., 2015; Broberg et al., 2014; Kile et al., 2014; Liu et al., 2014; Rojas et al., 2015; Seow et al., 2014), likely through the induction of S-adenosyl methionine (SAM) insufficiency, changes in genetic expression and activity of DNA methyltransferases (DNMTs), as well as chromatin organization and regulation (Reichard et al., 2007; Ren et al., 2011b; Zhu et al., 2017). As-induced aberrant DNA methylation could result in deregulating gene expressions and altering key cancer relevant signaling pathways, thus playing a critical role in As-induced toxicity and carcinogenesis (Ren et al., 2011b). However, few studies, particularly involving human populations, have examined disease-relevant DNA methylation changes, which make interpretation of DNA methylation data difficult in terms of its association with As toxicity and carcinogenicity, and limits the utility of DNA methylation in developing applicable disease-relevant biomarkers for preventive and therapeutic purposes.

Further, a body of research points towards an elevated susceptibility of children to a variety of environmental toxicants. As suggested by the American Academy of Pediatrics (AAP), As may have a greater impact on children than on adults because many aspects of

organogenesis and organ maturity take place during childhood. Animal studies have demonstrated that inorganic exposure during embryogenesis, especially at an exposure dose relevant to humans, is required for As carcinogenesis in mice (Waalkes et al., 2014), and that maternal exposure during pregnancy (*in utero* As exposure) leads to phenotypic changes in infants (Fry et al., 2007; Liu et al., 2008). An increased incidence and mortality of lung and bladder cancers were reported in adults exposed to high levels of As in their early life (Smith et al., 1998, 2006; Steinmaus et al., 2014). Although the underlying mechanism(s) that link early-life exposure to later disease development have not been identified, the disruption of DNA methylation is a favorable hypothesis. Moreover, recent studies suggest that some aberrant DNA methylations, may persist after exposure and impact health status later on and even across generations (Anway et al., 2005; Chang et al., 2006; Chong and Whitelaw, 2004; Kaati et al., 2007; Morgan et al., 1999). However, the data relating the impact of the timing of As exposure to DNA methylation changes and their relevance to health outcomes is scarce.

Hetao Plain, Inner Mongolia, China, is one of the most severely As-affected areas in the world (Fujino et al., 2004; Fujino et al., 2006; Guo et al., 2003; Guo et al., 2001; Guo et al., 2006a; Guo et al., 2006b; Mao et al., 2010). More than half a century ago, the inhabitants of this area mainly used shallow well water (5–8 m in depth), which held surface water and water from the nearby river. Since then, villagers had begun to use underground water from private deeper tube wells (15–30 m in depth). Unfortunately, however, more than 60% of deep tube wells in this area were found to contain high levels of As, with concentrations higher than 50 µg/L and up to 500 µg/L (Guo et al., 2001). The consumption of drinking water containing high levels of As continued until the early 1990s. During that time, water sources were changed and water treatment plants were installed, thus the level of As in drinking water was significantly reduced (< 10 µg/L, Drinking Water Standard for As by World Health Organization). These changes created a quite unique As-exposure scenario among families living in this region, where the timing of As exposure for each generation living in the same household is completely different. Grandparents, parents, and grandchildren were exposed to As in their adulthood, *in utero*/early childhood, and in the germline, respectively.

This situation of As exposure in the Hetao Plain provides an opportunity to examine the impact of the timing of As exposure, i.e., in the germline, *in utero*/childhood, or in adulthood, on DNA methylation changes and the relevance of these changes to As-induced adverse effects. Taking advantage of this unique setting of As exposure across three generations, in late 2012, we conducted a family (including the three generations) based epidemiological study in the Hetao Plain, and collected data and bio-specimens from 18 families that were historically exposed to As and nine control families not exposed to As (Fig. 1A). To assess the link between aberrant DNA methylation and As-induced adverse effects, we also recruited 18 patients diagnosed with arsenical skin lesions who lived in the same villages as the As exposure families. We included skin lesion patients because skin lesion is an As-specific adverse effect that has been repeatedly and consistently shown to precede and predict the risk of future As-induced chronic diseases, including diabetes, cardiovascular disease and several cancers (Hsu et al., 2013; Xia et al., 2009). In this study, we reported the analytical results obtained from examining and comparing genome-wide

DNA methylation profiles across three generations between As-exposed families and control families as well as their relevance to recruited skin lesion patients.

2. Methods

2.1. Information of study locations and participants

The populations studied were from six villages in Wuyuan and Hangjinhouqi counties, in the Hetao Plain, Inner Mongolia, China. Methods and protocols were carried out in accordance with guidelines approved by the Internal Review Board (IRB) both from Inner Mongolia Medical University, China, and from the University at Buffalo. Informed consent was obtained from all subjects and, for children, informed consent was approved and signed by their parents. Historical arsenic levels in drinking water from tube wells in these villages were obtained from a historical monitoring program ran by the Health Department in Inner Mongolia, China. The villages' main source of income comes from agricultural employment, and there are no other industries or mining activities around the villages. No other known notable environmental exposures were reported in this area, although some studies indicated elevated fluoride in the ground water (Zhang et al., 2014).

Two villages from each county (Wuyuan and Hangjinhouqi) were selected, with drinking water known to be highly polluted with arsenic in the past and reduced to a level below 10 µg/L around the time between 1990 and 1997. We selected one village in each county as control, with drinking water reported as having no or low arsenic contamination (undetectable to < 10 µg/L) based on information obtained from the historical arsenic monitoring program. For this study, the main selection criteria were families in which three generations lived in the same household in the selected villages and did not change their living location in the past 50 years. Another important eligibility criteria for arsenic-exposed families were that there was a clear record of using arsenic-contaminated tube-well water, and the change to arsenic free public water sources occurred no later than 1997. Other notable eligibility criteria included: 1) all family members, both in the exposed and control groups, were healthy and free from arsenic-associated diseases; and 2) the grandchildren in these families were born after 2000 and were not younger than 6 years old (school age children). The studies were conducted in late 2012 and early 2013. We first collected historical arsenic monitoring data in these villages from the Health Department and then we got consent and conducted a baseline survey among villagers, collecting basic characteristic information from the participants, including age, gender, education, smoking and alcohol drinking history, household location, health information, etc. Most adult males smoked and drank alcohol (Table 1), therefore these variables were not taken into consideration during data analysis. Based on the information collected and our selection criteria, 144 As-exposed and 87 control families were eligible for this study. In our first phase study, we randomly selected 18 exposed families matched with nine control families, and their consent to participate was obtained. Additionally, we recruited 18 villagers (of grandparent generation) diagnosed with arsenical skin disease. Villagers who had obvious skin abnormalities, including hyperkeratosis of the palms and soles and hyperpigmentation or leukomelanosis of the torso and limbs, were diagnosed at the local hospital as having arsenical skin lesions by Chinese physicians using their established clinical criteria (China, 1994). We collected 10

mL blood from all participants, and buffy coat was separated for further genomic DNA extraction (Chen et al., 2017).

2.2. Extraction and quantification of DNA

DNA was isolated from blood buffy coat using FlexiGene DNA Kit (QIAGEN, Hilden, Germany) following the manufacturer's protocol. DNA yields were first determined from the concentration of DNA, measured by absorbance at 260 nm on NanoDrop (Thermo Fisher Scientific, Waltham, MA). Then, the integrity of genomic DNA was determined by pulsed-field gel electrophoresis (PFGE) through an agarose gel. Picogreen quantitation (Quant-iT™ PicoGreen® dsDNA Reagent and Kits, Life Technologies, OR) methods were used to quantitate genomic DNA with standard spectrofluorometer using fluorescein excitation and emission wavelengths.

2.3. Global DNA methylation quantification

Global DNA methylation of all extracted DNA samples was determined by measuring 5-methylcytosine (5-mC) using a 5-mC DNA enzyme-linked immunosorbent assay (ELISA) kit (Zymo Research, CA). The kit features a unique anti-5-mC monoclonal antibody that is both sensitive to and specific for 5-mC. The protocol for measurement of 5-mC level is described in the manufacturer's instructions. Briefly, 100 ng of genomic DNA and standard controls provided by the kit were denatured and used to coat the plate wells with 5-mC coating buffer. After incubation at 37 °C for 1 hour, the wells were washed with 5-mC ELISA buffer and then an antibody mix consisting of anti-5-mC and a secondary antibody was added to each well. The plate was covered with foil and incubated at 37 °C for 1 hour. After the antibody mix was washed out from the wells with 5-mC ELISA buffer, a horseradish peroxidase (HRP) developer was added to each well and incubated at room temperature for 1 hour. The absorbance at 405 nm was measured using an ELISA plate reader (Synergy HT, BioTek, Winooski, VT). The percentage of 5-mC was calculated using the second-order regression equation of the standard curve that was constructed with negative and positive controls in the same experiment.

2.4. Infinium HumanMethylation450 BeadChip (HM450K) and image scan

500 ng (50 ng/μL) of high quality genomic DNA measured *via* Picogreen quantitation was used for a HM450K bead chip array. DNA samples were processed for array analysis using the Infinium methylation assay, as per manufacturer's protocol (Illumina, CA). Following bisulfite treatment of 500 ng genomic DNA using the EZ DNA Methylation kit (Zymo Research, CA), the bisulfite-converted DNA samples were chemically denatured neutralized using 0.1N NaOH, and amplified at 37 °C for 20–24 hours. The amplified products were then enzymatically fragmented at 37 °C for 1 hour. The fragmented DNA was precipitated using 100% 2-propanol at 4 °C for 30 minutes and centrifuged ($3,000 \times g$) at 4 °C for 20 minutes to collect a tight pellet. The supernatant was decanted by quick inversion and pellet was allowed to dry at room-temperature (22 °C) for 1 hour. The pellet was resuspended in Illumina's custom hybridization buffer and incubated at 48 °C for 1 hour followed by 95 °C for 20 minutes. The samples were loaded onto the HumanMethylation450 BeadChips and hybridized at 48 °C for 16–24 hours. The chips were then washed in Illumina's wash solutions and placed into Tecan Te-Flow chambers in preparation for staining. Unhybridized

and non-specific hybridized DNA were washed off the BeadChips. Labeled nucleotides were added to extend the primers hybridized to the DNA. The BeadChips were stained with Cy-3 and Cy-5 fluorescent dye, followed by washing, drying, and coating to protect against ozone and other elements that may be detrimental to the dyes. BeadChips were scanned using the Illumina iScan Reader. The image data was then transferred to the Illumina Genome Studio for data processing, validation of assay controls, and report generation using the methylation module. The level of methylation for the interrogated locus was determined by calculating the ratio of the fluorescent signals in the methylated vs. unmethylated sites. The results were represented in the form of β values, specifically, the average β value (AVG_Beta), representative of the average methylation level of the CpG dinucleotide. The raw data was retrieved from the Genome Studio methylation module version 1.9TM.

2.5. Genome-wide DNA methylation data analysis

Methylation Array data was analyzed using the *R* environment for statistical computing and graphics (R Development Core Team, 2009). The data was loaded, subjected to quality control inspection, filtered, and normalized in a manner that is consistent with the recommendations provided (Cazaly et al., 2016; Maksimovic et al., 2016) (see Data Load, Quality Control and Normalize Worksheet in Supplemental Material). As per the recommendations (Maksimovic et al., 2016), the methylation probe assay “beta” values were utilized primarily for data visualization purposes (notable exceptions: subset quantile normalization and cell count estimation procedures operated on beta values) while the “*M*” values (i.e., the logit transform of the beta values) were used for data modeling purposes.

2.5.1. Data quality control (QC) and filtering—Array data was loaded and inspected using the R Bioconductor package *minfi* (version 1.20.2) (Aryee et al., 2014). Relevant output for the quality control and filtering steps is provided in the Supplemental Material. Briefly, the following analyses were performed: confirmation of correct sex assignment of samples was obtained using the *getSex* function. Detection *p*-values were interrogated on a sample-wise basis with samples containing the largest mean detection probabilities flagged for inspection. A QC report was generated using the *qcReport* function. Median methylated and un-methylated intensity values were also compared using the *plotQC* function. Relative count compositions were compared across samples using the *estimateCellCounts* function. Sample data was visualized *via* multidimensional scaling (MDS), calculating the Euclidean distances between samples. The MDS plots were evaluated with respect to batch, sex, and sample pedigree information (see Data Load, Quality Control and Normalize Worksheet in Supplemental Material).

Filtering of the array features (i.e., the methylation assay probes) was performed as per the guidance provided (Maksimovic et al., 2016). Non-autosomal probes were removed from the analyses, as well as those probes that have been identified as cross-reactive (Chen et al., 2013), in close proximity to a single-nucleotide polymorphism (SNP) (using the *dropLociWithSnps* function), or with detection *p*-value greater than 0.01 for any one sample.

2.5.2. Data normalization and quality determinants—The data utilized for the primary and secondary analyses was normalized using stratified quantile normalization

(Touleimat and Tost, 2012) in the initial data processing steps and then batch effects was adjusted using *ComBat* (Bioconductor sva package) (Johnson et al., 2007) following the data QC and filtering steps. The data was adjusted for batch effects attributed to the slides that the samples were run and was made with respect to model that accounted for sample pedigree information (most notably gender and generation information). MDS plots of the normalized data were evaluated with respect to batch and sample pedigree information. DNA methylation has been shown to vary systematically as a function of age (Hannum et al., 2013; Horvath et al., 2012; Petkovich et al., 2017). Participants in this study had an age range from 6 to 72 years, which provided an opportunity to assess the accuracy of DNA methylation data based on their correlation with biological ages. We thus ran an algorithm developed by Horvath (Horvath et al., 2012) and tested whether the DNA methylation pattern in a selected group of probes was correlated with chronologic age. Our analysis showed a strong association between the chronologic age of participants and the age that was predicted based on DNA methylation signature with a coefficient factor (r^2) of 0.92 (Fig. 1B), suggesting that the DNA methylation data quality was satisfactory.

2.5.3. Primary analyses to identify overlapping differential DNA methylation loci, genes and regions across generations—To identify differentially methylated DNA associated with arsenic exposure, we first compared methylation data between two groups in each generation separately. Our preliminary analysis showed that significantly more genomic loci were differentially methylated in the parent generation (Fig. S1), who were exposed to arsenic during early life (*in utero*/early childhood), indicating that this life-stage may be more susceptible to arsenic-induced DNA methylation alterations (Examples were given in Fig. S2). In addition, our data suggested that some changes in DNA methylation only occurred in one generation but not in others, indicating that DNA methylation was affected by the timing of arsenic exposure. Considering that the sample size in each individual generation is small thereby affecting the statistical power, we focused our analysis on determining and identifying genomic loci that overlap across the three generations by two statistical approaches: p -values were calculated using linear model F-test (probe-wise) and non-parametric permutation test (gene-wise), and were adjusted using the Benjamini & Hochberg method (Benjamini and Hochberg, 1995).

2.5.3.1. Identification of differentially methylated loci (DML): The first approach was probe-wise using all available samples and an assumption of normal distribution of the methylation data. In this probe-wise approach, an F-test (Bioconductor sva package) was conducted to compare the M values of all probes between the arsenic exposure and control groups with the adjustment of pedigree information for each subject; however, generational differences were not considered in the analysis. Adjusted p -values were calculated by the Benjamini Hochberg method for false discovery rate (FDR) control. Significant differentially methylated loci were defined as those matching the following two requirements: (1) FDR-corrected p -value (the q -value in F test) < 0.2 ; (2) effect size (β value) > 0.02 .

2.5.3.2. Identification of differentially methylated genomic regions (DMR): A moderated t-statistic was calculated for the differential methylation analysis on individual CpGs using

limma differential analysis pipeline, then the DMRcate function was used to combine the analyses to identify the differentially methylated regions. The distances of two significant consecutive probes in the DMR regions were less than 1000 base pairs. All of these steps were implemented by the Bioconductor package DMRcate (Peters et al., 2015). Previously, studies have suggested that environmental exposure may lead to similar alterations among adjacent CpG sites (Sofer et al., 2013).

2.5.3.3. Identification of differentially methylated genes: In this second approach, the analysis was non-parametric, rank-based, and focused on probes linked with a specific gene. Moreover, we only included families with a fully available dataset (comprising of three generations including two grandparents, one parent and one grandchild), hence, a set of 22 “complete” families (e.g., methylation array data is available for all four members) consisting of 14 arsenic exposed families and eight control families were included in the analysis. Given that the age effects on DNA methylation were not observable between the grandparental (F1) and parental (F2) generations, they were calculated together as one group in the following.

Nomenclature in this model: Let $X_{i(g), j(h)}^{F12}$ denote the sum of the PedID adjusted M values for the i^{th} probe of the g^{th} gene across the grandparental (F1) and parental (F2) generations for the h^{th} family. Let $R_{i(g), j(h)}^{F12}$ denote the corresponding rank. PedID adjustment to the M values consists of a Hodges-Lehmann estimator re-centering of the i^{th} probe assay data within each unique pedigree group (i.e., grandfather, grandmother, father, mother, son, and daughter). Let $X_{i(g), j(h)}^{F3}$ denote the PedID adjusted M value for the grandchildren (F3 generation) from the h^{th} family, and let $R_{i(g), j(h)}^{F3}$ denote the corresponding rank score.

Permutation test: The following rank-based test statistic was utilized to identify genes containing probes that are differentially methylated with respect to the exposed and control sample groups: $T_{i(g)} = \sum_{h \in Control} (R_{i(g), j(h)}^{F12} + R_{i(g), j(h)}^{F3})$. The test statistic for each probe

within each gene was then evaluated against a permutation null distribution. A permutation scheme was employed which assumed that the labels of “Exposed” and “Control” were interchangeable with respect to families, i.e., the identity of the eight control families was permuted under the null hypothesis. A gene-wise p -value was obtained by way of a minP permutation adjustment similar to the arm-wise adjustment (Gaile et al., 2007). The permutation adjusted gene-wise p -values remove the dependence between the minimum p -values across gene-wise probe sets and the size of the probe sets (i.e., unadjusted minimum p -values are negatively correlated to the size of the probe sets). In total, 19,868 genes were analyzed using the same set of 50,000 family-wise permutations for each minP adjustment. The 574 genes with an adjusted gene-wise p -value less than or equal to 0.04 and with less than 500 probe sets, were re-analyzed using all $\binom{22}{8} = 319,770$ possible permutations, the purpose being to better order the most significant genes. The permutation adjusted gene-wise p -values were then further adjusted using the method of Benjamini and Hochberg (Benjamini and Hochberg, 1995) such that the false discovery rate was controlled at a nominal level of 0.10, and at the same time with an effect size > 0.02 .

2.5.4. Pathway analysis by Ingenuity Pathway Analysis (IPA) package—All genes associated with identified DML were input into the IPA analysis tools to assess canonical signaling pathway (Knight et al., 2014). A right-tailed Fisher's exact test was performed for the enrichment of these genes in the IPA hand-curated canonical pathway database. Here, the p -value calculated for a pathway measures the probability of being randomly selected from all of the curated pathways. In our study, we used a cut-off of the corrected p -value less than 0.05 to define significant pathways.

2.5.5. Clustering analysis—The hierarchical cluster analysis was performed on all the samples, including the patient samples, based on the M -values of the 744 significant probes identified in the previous section. The Euclidean distances were measured for all sample pairs wisely.

2.5.6. Exploratory analysis to identify the potential inheritable aberrant DNA methylation caused by arsenic exposure—In this section, based on the M -values of the DNA methylation data, we used a probe-wise linear mixed effect model and the intersection union test (Berger and Hsu, 1996) to identify probes (and therefore the corresponding genes) with one of the following two patterns: Pattern 1: for each generation, the mean M -value of the arsenic-exposed group is significantly less than the mean M -value of the non-exposed group; and Pattern 2: for each generation, the mean M -value of the arsenic-exposed group is significantly greater than the mean M -value of the non-exposed group.

To be specific, for each probe, we used Y_{ijk} to denote the M -value of the k^{th} (k can be 1 or 2) individual from the j^{th} (j can be 1, 2, or 3) generation in the i^{th} family, and used a binary variable E_i to denote the group indicator of the i -th family, where $E_i=1$ for the exposure group and $E_i=0$ for the control group. Our statistical analysis used the following linear mixed effect model:

$$Y_{ijk} = \mu + \delta_1 I(j=2) + \delta_2 I(j=3) + \alpha_1 I(E_i=1) * I(j=1) + \alpha_2 I(E_i=1) * I(j=2) + \alpha_3 I(E_i=1) * I(j=3) + f_i + \varepsilon_{ijk},$$

where $I(\cdot)$ is an indicator function, the parameters μ , δ_1 , δ_2 , α_1 , α_2 , α_3 are fixed effects, f_i is the family-specific random effect, and ε_{ijk} is the random error.

To identify probes with Pattern 1, we used the test:

$$H_0: \bigcup_{i=1}^3 \{\alpha_i \geq 0\} \quad \text{vs.} \quad H_1: \alpha_1 < 0, \alpha_2 < 0, \alpha_3 < 0$$

To identify probes with Pattern 2, we used the test:

$$H_0: \bigcup_{i=1}^3 \{\alpha_i \leq 0\} \quad \text{vs.} \quad H_1: \alpha_1 > 0, \alpha_2 > 0, \alpha_3 > 0$$

We used the intersection union test (Berger, 1997) for each of the above two hypothesis testing problems. The Bonferroni correction method was used to control the type I error to be 0.05. Our analysis found 21 probes with Pattern 1 and seven probes with Pattern 2. The approximated *p*-values of these selected probes are shown in Supplemental Material (Table S6). In the boxplots, we use G1, G2, and G3 to denote the three generations respectively.

3. Results

3.1. General characteristics of studying participants

Descriptive characteristics of the study participants are presented in Table 1. Due to the one-child policy in China, most families have only one grandchild. For the parents' generation, those who later came into the family due to marriage were not included in the analysis because their arsenic exposure histories were unclear. A total of 67 participants from 18 exposure families, 35 individuals from nine control families, and 18 arsenical skin lesion patients were included in this study. The average age in each generation of the arsenic exposure group is slightly younger than their respective control group but no statistical significance was found. When looking at the grandparents' generation, the patients' group has more male participants than both arsenic-exposed and control groups. The average arsenic concentrations in drinking water from family tube-wells were collected from historical records and were 266.35 ± 182.40 $\mu\text{g/L}$ and 258.78 ± 212.22 $\mu\text{g/L}$ in arsenic-exposed participants and in patients, respectively (Table 1).

We quantified global DNA 5-mC levels. Global DNA hypomethylation as a function of age was observed in both arsenic-exposed and control families, in which the global DNA methylation level was the highest in grandchildren and the lowest in the grandparent generation. Moreover, global DNA methylation levels were slightly reduced when comparing arsenic-exposed families with unexposed families for each generation but were not statistically significant. However, global DNA methylation was significantly reduced in patients' with arsenical skin lesions ($6.10\% \pm 1.07\%$) when compared to the age corresponding grandparent generation in both control ($7.04\% \pm 1.49\%$) and arsenic-exposed groups ($6.95\% \pm 1.10\%$) with and without a controlling gender factor (Fig. 1B&C).

3.2. Differentially methylated DNA loci and regions and genes across generations between exposure and control families

Under our primary analysis approach (Analysis 2.5.3.1.), we applied the cut off value: FDR-corrected *p*-value < 0.2 and effect size > 0.02, with the main consideration of minimizing the false negative rate given the fact that the sample size of exposure and control groups is small thus affecting the statistical power. There were a total of 744 CpG sites identified as significant that were differentially methylated in all three generations between the arsenic-exposed and control groups (Fig. 2). When FDR was 0.05 or 0.1, respectively, 85 and 217 probes remained significant. These DML were distributed among all chromosomes, with chromosomes one and six each containing more than 70 DML (Fig. 2A). Among the 744 DML, 417 were hyper-methylated and 330 were hypo-methylated in the arsenic-exposed population compared to the control (Fig. 2B). We examined the DML locations in terms of their relationship with functional genes. A large number of identified DML were indeed

found in the gene promoter regions, with 17% and 6% of DML located within TSS1500 and TSS200, respectively (Fig. 2C).

The top 150 significant DML were listed in Supplemental Material (Table S1). Examples of the methylation β value plot for several identified top DML are shown (Fig. 3A), in which the changes in methylation patterns at these genomic loci (obtained by comparing each generation in the exposed to its respective control) for each generation are similar to each other. We then performed an analysis (Analysis 2.5.3.2.) to identify the genomic regions with clustered CpG sites that had differential methylation levels between the arsenic-exposed and control groups. Under the condition of FDR p -value < 0.05 , we found 15 CpG clusters (Table S2). Examples of three clusters located on chromosomes 6, 1, and 12 are shown in Fig. 3B. To understand the underlying mechanism and to explain the differential patterns of DNA methylation observed between arsenic exposed and control populations, we inputted all 744 DML into the IPA package. The IPA identified 25 canonical pathways commonly affected by arsenic exposure across generations with a p -value < 0.05 (Table S3). Three pathways, including Neuropathic Pain-Signaling in Dorsal Horn Neurons, Arsenate detoxification I pathways and Chondroitin and Dermatan Biosynthesis, were highly statistically significant with a p -value < 0.01 (Fig. 3C).

In gene-wise analysis (Analysis 2.5.3.3.), we selected the best probe of each gene based on adjusted minimum p -values that best represented the methylation status of the gene. We then performed non-parametric permutation analysis, and a total of 1,804 genes were identified if we kept the cut-off value at FDR-corrected p -value < 0.2 and effect size > 0.02 . When the FDR-corrected p -value was 0.1 or 0.05, respectively, there were 134 and six genes remaining on the list (Table S4). The six genes at FDR-corrected p -value < 0.05 were *MLKL* (mixed lineage kinase domain like pseudokinase), *VIPR2* (vasoactive intestinal peptide receptor 2), *SYNGR3* (synaptogyrin 3), *REEP1* (receptor accessory protein 1), *ACY3* (aminoacylase 3), and *SLC16A3* (solute carrier family 16 member 3) (Fig. 4A). A large number of genes identified by gene-wise analysis were also found in the probe-wise analysis (Fig. S2). Among the six genes with FDR < 0.05 , *MLKL*, *VIPR2*, *ACY3* and *SLC16A3* were also included in the top 150 DML list generated in the probe-wise analysis (Table S1 and S4). IPA analysis was also performed using the genes identified from gene-wise analysis. Twenty-four pathways with p -value < 0.05 and eight pathways with p -value < 0.01 were identified (Table S5). Among these, the arsenate detoxification I and chondroitin and dermatan biosynthesis pathways were once again identified as some of the top pathways with a p -value < 0.01 (Fig. 4B).

3.3. Identified DML in arsenical skin lesion patients

Few studies have evaluated the relevance of DNA methylation changes to arsenic-induced adverse effects. Utilizing obtained biospecimens from arsenical skin lesion patients, we examined whether the identified DML and DMRs were not only altered by exposure to arsenic but also present in patients with arsenic-associated diseases. We performed a clustering analysis of samples, including all patients and grandparents in the arsenic-exposed and control groups, based on the identified 744 DML. All patient samples were grouped into one big cluster together with grandparents in the arsenic-exposed group (red dash line), but

separated from grandparents in the control group, indicating that grouping was mainly based on historical arsenic-exposure status (Fig. 5). At the same time, the DNA methylation patterns of patients were also distinguishable from the grandparents in the arsenic exposed group that did not develop diseases, as the majority of patients were grouped together in a small cluster within the big cluster (marked by the blue dash line). Further analysis showed that, among the 744 DML identified, 21.8% (162 DML) of the DML in patients had a comparable DNA methylation level with grandparents in the arsenic-exposed group which were significantly distinguished from grandparents in the control group ($p < 0.01$). In comparison, only 8.6% of the DML in patients had a similar level as grandparents in the control group (Fig. 6A). Moreover, 41 of the DML with a similar methylation level in patients and grandparents in the exposed group were part of the top 100 DML identified in the probe-wise analysis (Fig. 6A). We redrew Figs. 3A & B including only grandparents and patient samples. As shown, among the top DML and DMRs, DNA methylation levels in patients were more similar to grandparents in the exposed group than to grandparents in the control group (Figs. 6B & C).

3.4. Inheritable potential of DNA methylation alterations caused by arsenic exposure

While limited, emerging evidence from *in vitro* and animal studies suggests that some DNA methylations altered by environmental factors exhibit meiotic stability, can be maintained, and can even be transmitted from one generation to the next (Chong and Whitelaw, 2004), in a phenomenon called epigenetic transgenerational inheritance. In general, it requires observing at least four generations in order to determine this effect; our inclusion of three generations makes us unable to directly test this question. However, our study with three generations remains valuable and can provide much-needed insight on the manifestation of this phenomenon in humans. We thus conducted a simple test with the idea that if arsenic-induced DNA methylation changes in the grandparents' generation are stable and can be transmitted to future generations, then the aberrant DNA methylation will be maintained in the parents' and grandchildren's generations and remain significantly distinguished from the corresponding generations of the control population. In other words, if the inheritance does occur, the change in DNA methylation at specific locations compared to control in all generations will be similar; the level and direction of change in each generation will be comparable to each other. By applying a linear mixed effect model and the intersection union test, our analysis, using a controlled Bonferroni corrected type I error of 0.05, found 21 hypo-methylated probes and seven hyper-methylated probes that matched this hypothesis (Table S6). Two examples of these CpG loci are shown in Fig. 7A. The penetrant pattern of these two loci in each family of both the exposed and control groups is drawn and shown in Fig. 7B.

4. Discussion

Altered epigenetic modifications (e.g., DNA methylation) have been proposed as the underlying mechanism for arsenic-induced carcinogenesis (Ren et al., 2011b). This family-based study attempts to examine the impact of the timing of arsenic exposure on DNA methylation in a multi-generational setting and identify commonly aberrant DML and/or DMRs associated with an increased risk for disease development regardless of the distinct

life-stages in which the exposure to arsenic occurred. We show that historical arsenic exposure may leave detectable DNA methylation changes, with the most significant changes observable in people who developed arsenic-associated diseases, suggesting that arsenic exposure induced DNA hypo-methylation has a long-lasting effect and is associated with increased risk for arsenic toxicity. A recent study suggests that arsenic carcinogenic effects manifest 40 years after exposure reduction (Smith et al., 2017). The long-lasting DNA methylation changes could be the underlying mechanism behind the long latencies of arsenic-induced adverse effects. More than 700 DML are identified across three generations of people who were exposed to arsenic in adulthood, early childhood, and germ cells. Based on these DML, the DNA methylation patterns in arsenical skin lesion patients more closely resembled those of people exposed to arsenic, and are distinguished from those of people without arsenic exposure history. Finally, our exploratory analysis indicates that some arsenic-induced DNA methylation changes have the potential to be inherited, and thus arsenic could exert effects affecting multiple generations through its impact on DNA methylation.

Our data analysis was centered on the question, “Does arsenic exposure induce common DNA methylation changes among people exposed to arsenic at different life-stages?” Under two very different statistical approaches, we identified many common DML across three generations. For example, four of the top six loci identified using the non-parameter permutation test were also included in the top 150 list identified using the F-test. In addition, two canonical pathways, the arsenate detoxification I and chondroitin and dermatan biosynthesis pathways, were included on the top list of canonical pathways via IPA analysis and reached a statistical significance of p -value < 0.01 for both analyses. The efficiency of arsenic metabolism in the body is a well-established factor that is associated with differential risk for the toxic and carcinogenic activities of arsenic (Ahsan et al., 2007; Chen et al., 2017; Chung et al., 2010; De Chaudhuri et al., 2008; Lin et al., 2006; Ren et al., 2011a; Sampayo-Reyes et al., 2010; Valenzuela et al., 2009; Zhang et al., 2015). The methylation levels of *As3MT* (arsenite methyltransferase) and *GLRX2* (Glutaredoxin 2) genes, two critical enzymes involved in arsenate detoxification I pathways, were increased across the three generations in the arsenic exposed population. *GLRX2* reductase is responsible for converting arsenate to arsenite, which is then methylated via an *As3MT* mediated reaction. The hyper-methylation of these two genes could result in lower gene expressions, thus lowering the efficiency of arsenic metabolism in the body, and ultimately increasing arsenic toxicity (Jansen et al., 2016). Chondroitin sulfate, a glycosaminoglycan, is involved in many biological processes, including cell proliferation and morphogenesis (Sirko et al., 2007; Winship et al., 2017). Aberrant DNA methylation in genes involved in the chondroitin biosynthesis pathway was reported previously in infants who were exposed to arsenic *in utero* (Kaushal et al., 2017), however, the functional role of chondroitin sulfate in arsenic toxicity and carcinogenicity has yet to be determined. Additionally, we identified 15 genomic regions with differentially methylated clustered CpG sites in the arsenic-exposed group compared to the control group. These data suggest that there are common DNA methylation changes that can be identified in people exposed to arsenic at different life stages.

Many studies, in particular human-population based, have been conducted to investigate DNA methylation alterations due to arsenic exposure (Argos et al., 2015; Broberg et al., 2014; Kile et al., 2014; Liu et al., 2014; Rojas et al., 2015; Seow et al., 2014). Many of these studies have observed novel regions of the genome associated with arsenic exposure, but few have evaluated the biological and functional relevance of these DNA methylation changes. In this study, we included 18 biospecimens from arsenical skin lesion patients who were from the same villages as the arsenic-exposed group. We showed that people with arsenic-induced skin lesions had a statistically significant lower level of global DNA methylation in comparison to both age-matched arsenic-exposed and unexposed participants. At the gene-specific level, a cluster analysis was able to group samples based on their exposure status. More importantly, results from the cluster analysis also indicated that the DNA methylation patterns of patients' samples were also distinguishable from people exposed to arsenic but did not develop any diseases. Further, we found 162 DML that were not only related to arsenic exposure but were also associated with arsenical skin disease, indicating that signature disease-relevant DNA methylation changes could be identified and established.

It is known that global DNA methylation gradually decreases during a lifetime (Alisch et al., 2012; Horvath et al., 2012; Weidner et al., 2014; Weidner and Wagner, 2014). This was also observed in our study population across three generations. Although not statistically significant, the level of global DNA methylation was slightly lower for each generation in the arsenic-exposed population compared to the corresponding control population. This DNA hypo-methylation has been reported in many studies both *in vitro* and *in vivo* (Ren et al., 2011b). However, this is the first time that the global DNA methylation level in grandchildren, who were exposed to arsenic only via the embryo forming cells, is lower than that of children of the same age without arsenic exposure. Until recently, it was assumed that epigenetic alterations induced by environmental factors conferred adaptive capabilities to individuals during their lifetime but were not passed on to future generations. However, increasing evidence suggests that while the majority of DNA methylation in the germline is erased in early embryogenesis, some epigenetic modifications altered by environmental factors exhibit meiotic stability and can be maintained and further transmitted from one generation to the next (Chong and Whitelaw, 2004), in which the expression of genes in offspring is affected by the life experience of earlier generations (Anway et al., 2005; Chang et al., 2006; Kaati et al., 2007; Morgan et al., 1999). Evidence in humans for these multi-generational epigenetic impacts induced by environmental factors, including arsenic, is scant. Our exploratory analysis identified 28 DML that showed an inheritable potential of which 21 DML were hypo-methylated DML. Among the 744 DML identified by comparing the exposed and control groups, there were slightly more hyper-methylated DML. Given the characterized global DNA hypo-methylation of arsenic exposure, it will be interesting to determine whether the hypo-methylated loci are more likely to escape the demethylation that occurs during early embryogenesis and destined to be inherited. Our study is not capable of determining whether the DNA methylation level changes in grandchildren is the result of their grandparents' exposure or the exposure of germ cells. However, even if the DNA methylation changes in grandchildren is due to germ cell exposure, it still suggests that DNA methylation changes caused by arsenic exposure could escape the erasure in early embryogenesis, suggesting potential transgenerational effects of arsenic-induced DNA

methylation changes. This may serve as important evidence for the idea of Lamarckism, which poses a big challenge to the central dogma of genetic inheritance.

Our study has limitations, particularly, the relatively small sample size. This led to a limited statistical power in our data analysis. Therefore, we set the FDR adjusted p -value at 0.2 in order to avoid false negatives. However, at the same time, this increases the possibility of false positive discoveries. We thus also applied two very different statistical approaches, with the aim of examining the results and identifying the loci that were significant in both statistical models. This strategy apparently works and many of the top loci identified are shared between these two approaches. Further studies (i.e. larger study) are warranted to test if these observations are true. We are in the position of advancing this study and validating the results on a larger scale, as more families have been recruited and more samples are available. The arsenic exposure information was based on historical monitoring data, and there is risk of exposure misclassifications. However, while there are slight variations in the data measured in different years, there is no evidence that suggests the arsenic concentrations in the wells were altered significantly during the drinking periods. In addition, these tube wells were household based and used solely for each individual family. Overall, the risk of exposure misclassification is small. The regulation of DNA methylation is a complex process and many factors including genetics, environment, lifestyle, and nutritional status are capable of modifying DNA methylation. Arsenic exposed and control populations come from the same relatively isolated communities, thus the variance due to genetics, diet and culture is limited, and the potential noise associated with analyzing a mixed human population is greatly reduced. However, there is still a slight possibility that other factors may not be completely balanced between the exposed and control population, and that the observed correlation between arsenic exposure and altered DNA methylation may not accurately reflect the true association. We expect that our planned study with larger sample size and detailed information of covariates will enable us to eliminate other confounding factors.

5. Summary

Collectively, our data suggest that aberrant DNA methylations may remain stable and observable decades after exposure remediation. Arsenic exposure induces common DML/DMRs across generations (with exposure occurring at different life-stages in each generation), which not only reflects arsenic exposure but may be also relevant to arsenic-induced diseases. Further, our study is first that we know of indicating for the possible transgenerational epigenetic impact of arsenic exposure. The observation of arsenic exposure affecting multiple generations and the potential transgenerational inheritance of aberrant DNA methylation highlights the importance of public health efforts to increase public awareness of arsenic adverse health effects, eliminate exposure, and provide long-term monitoring and intervention.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Dr. Allan Smith from University of California at Berkeley for critically reviewing the manuscript. These studies were supported by the National Natural Science Foundation of China grant (81260413) (to Z.L. and X.G.), the National Center for Advancing Translational Sciences of the National Institutes of Health (NIH) under award number UL1TR001412 to the University at Buffalo, and NIH grants (ES022629) to X.R.

References

- Abernathy CO, Liu YP, Longfellow D, Aposhian HV, Beck B, Fowler B, Goyer R, Menzer R, Rossman T, Thompson C, Waalkes M. Arsenic: health effects, mechanisms of actions, and research issues. *Environ Health Perspect.* 1999; 107:593–597. [PubMed: 10379007]
- Ahsan H, Chen Y, Kibriya MG, Slavkovich V, Parvez F, Jasmine F, Gamble MV, Graziano JH. Arsenic metabolism, genetic susceptibility, and risk of premalignant skin lesions in Bangladesh. *Cancer Epidemiol Biomarkers Prev.* 2007; 16:1270–1278. [PubMed: 17548696]
- Alisch RS, Barwick BG, Chopra P, Myrick LK, Satten GA, Conneely KN, Warren ST. Age-associated DNA methylation in pediatric populations. *Genome Res.* 2012; 22:623–632. [PubMed: 22300631]
- Anway MD, Cupp AS, Uzumcu M, Skinner MK. Epigenetic transgenerational actions of endocrine disruptors and male fertility. *Science.* 2005; 308:1466–1469. [PubMed: 15933200]
- Argos M, Chen L, Jasmine F, Tong L, Pierce BL, Roy S, Paul-Brutus R, Gamble MV, Harper KN, Parvez F, Rahman M, Rakibuz-Zaman M, Slavkovich V, Baron JA, Graziano JH, Kibriya MG, Ahsan H. Gene-specific differential DNA methylation and chronic arsenic exposure in an epigenome-wide association study of adults in Bangladesh. *Environ Health Perspect.* 2015; 123:64–71. [PubMed: 25325195]
- Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, Irizarry RA. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics.* 2014; 30:1363–1369. [PubMed: 24478339]
- Benjamini Y, Hochberg Y. Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B-Methodological.* 1995; 57:289–300.
- Berger RL. Likelihood Ratio Tests and Intersection-Union Tests. In: Panchapakesan S, Balakrishnan N, editors *Advances in Statistical Decision Theory and Applications*. Statistics for Industry and Technology. Birkhäuser; Boston: 1997. 225–237.
- Berger RL, Hsu JC. Bioequivalence trials, intersection-union tests and equivalence confidence sets. *Statistical Science.* 1996; 11:283–302.
- Broberg K, Ahmed S, Engstrom K, Hossain MB, Jurkovic Mlakar S, Bottai M, Grandt M, Raqib R, Vahter M. Arsenic exposure in early pregnancy alters genome-wide DNA methylation in cord blood, particularly in boys. *J Dev Orig Health Dis.* 2014; 5:288–298. [PubMed: 24965135]
- Cazaly E, Thomson R, Marthick JR, Holloway AF, Charlesworth J, Dickinson JL. Comparison of pre-processing methodologies for Illumina 450k methylation array data in familial analyses. *Clin Epigenetics.* 2016; 8:75. [PubMed: 27429663]
- Chang HS, Anway MD, Rekow SS, Skinner MK. Transgenerational epigenetic imprinting of the male germline by endocrine disruptor exposure during gonadal sex determination. *Endocrinology.* 2006; 147:5524–5541. [PubMed: 16973722]
- Chen X, Guo X, He P, Nie J, Yan X, Zhu J, Zhang L, Mao G, Wu H, Liu Z, Aga D, Xu P, Smith M, Ren X. Interactive Influence of N6AMT1 and As3MT Genetic Variations on Arsenic Metabolism in the Population of Inner Mongolia, China. *Toxicol Sci.* 2017; 155:124–134. [PubMed: 27637898]
- Chen YA, Lemire M, Choufani S, Butcher DT, Grafodatskaya D, Zanke BW, Gallinger S, Hudson TJ, Weksberg R. Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. *Epigenetics.* 2013; 8:203–209. [PubMed: 23314698]
- China. Hygiene Ministry. Office for Control and Treatment of Endemic. *J Endem Dis Cont Treat Inner Mongolia.* 1994. Clinical diagnostic criteria for arsenism in Inner Mongolia; 19third cover page

- Chong S, Whitelaw E. Epigenetic germline inheritance. *Curr Opin Genet Dev.* 2004; 14:692–696. [PubMed: 15531166]
- Chung CJ, Pu YS, Su CT, Chen HW, Huang YK, Shiue HS, Hsueh YM. Polymorphisms in one-carbon metabolism pathway genes, urinary arsenic profile, and urothelial carcinoma. *Cancer Causes Control.* 2010; 21:1605–1613. [PubMed: 20532609]
- De Chaudhuri S, Ghosh P, Sarma N, Majumdar P, Sau TJ, Basu S, Roychoudhury S, Ray K, Giri AK. Genetic variants associated with arsenic susceptibility: study of purine nucleoside phosphorylase, arsenic (+3) methyltransferase, and glutathione S-transferase omega genes. *Environ Health Perspect.* 2008; 116:501–505. [PubMed: 18414634]
- Fry RC, Navasumrit P, Valiathan C, Svensson JP, Hogan BJ, Luo M, Bhattacharya S, Kandjanapa K, Soontararuks S, Nookabkaew S, Mahidol C, Ruchirawat M, Samson LD. Activation of inflammation/NF-kappaB signaling in infants born to arsenic-exposed mothers. *PLoS Genet.* 2007; 3:e207. [PubMed: 18039032]
- Fujino Y, Guo X, Liu J, You L, Miyatake M, Yoshimura T. Japan Inner Mongolia Arsenic Pollution Study G. Mental health burden amongst inhabitants of an arsenic-affected area in Inner Mongolia, China. *Soc Sci Med.* 2004; 59:1969–1973. [PubMed: 15312930]
- Fujino Y, Guo X, Shirane K, Liu J, Wu K, Miyatake M, Tanabe K, Kusuda T, Yoshimura T. Japan Inner Mongolia Arsenic Pollution Study G. Arsenic in drinking water and peripheral nerve conduction velocity among residents of a chronically arsenic-affected area in Inner Mongolia. *J Epidemiol.* 2006; 16:207–213. [PubMed: 16951540]
- Gaile DP, Schifano ED, Miecznikowski JC, Java JJ, Conroy JM, Nowak NJ. Estimating the arm-wise false discovery rate in array comparative genomic hybridization experiments. *Stat Appl Genet Mol Biol.* 2007; 6:Article32. [PubMed: 18052915]
- Guo X, Fujino Y, Chai J, Wu K, Xia Y, Li Y, Lv J, Sun Z, Yoshimura T. The prevalence of subjective symptoms after exposure to arsenic in drinking water in Inner Mongolia, China. *J Epidemiol.* 2003; 13:211–215. [PubMed: 12934964]
- Guo X, Fujino Y, Kaneko S, Wu K, Xia Y, Yoshimura T. Arsenic contamination of groundwater and prevalence of arsenical dermatosis in the Hetao plain area, Inner Mongolia, China. *Mol Cell Biochem.* 2001; 222:137–140. [PubMed: 11678595]
- Guo X, Fujino Y, Ye X, Liu J, Yoshimura T. Association between multi-level inorganic arsenic exposure from drinking water and skin lesions in China. *Int J Environ Res Public Health.* 2006a; 3:262–267. [PubMed: 16968972]
- Guo X, Liu Z, Huang C, You L. Levels of arsenic in drinking-water and cutaneous lesions in Inner Mongolia. *J Health Popul Nutr.* 2006b; 24:214–220. [PubMed: 17195562]
- Hannum G, Guinney J, Zhao L, Zhang L, Hughes G, Sada S, Klotzle B, Bibikova M, Fan JB, Gao Y, Deconde R, Chen M, Rajapakse I, Friend S, Ideker T, Zhang K. Genome-wide methylation profiles reveal quantitative views of human aging rates. *Mol Cell.* 2013; 49:359–367. [PubMed: 23177740]
- Horvath S, Zhang Y, Langfelder P, Kahn RS, Boks MP, van Eijk K, van den Berg LH, Ophoff RA. Aging effects on DNA methylation modules in human brain and blood tissue. *Genome Biol.* 2012; 13:R97. [PubMed: 23034122]
- Hsu LI, Chen GS, Lee CH, Yang TY, Chen YH, Wang YH, Hsueh YM, Chiou HY, Wu MM, Chen CJ. Use of arsenic-induced palmoplantar hyperkeratosis and skin cancers to predict risk of subsequent internal malignancy. *Am J Epidemiol.* 2013; 177:202–212. [PubMed: 23299695]
- IARC Working Group on the Evaluation of Carcinogenic Risks to Humans. Some drinking-water disinfectants and contaminants, including arsenic. *IARC Monogr Eval Carcinog Risks Hum.* 2004; 84:1–477. [PubMed: 15645577]
- Jacobson-Kram D, Montalbano D. The reproductive effects assessment group's report on the mutagenicity of inorganic arsenic. *Environ Mutagen.* 1985; 7:787–804. [PubMed: 3899634]
- Jansen RJ, Argos M, Tong L, Li J, Rakibuz-Zaman M, Islam MT, Slavkovich V, Ahmed A, Navas-Acien A, Parvez F, Chen Y, Gamble MV, Graziano JH, Pierce BL, Ahsan H. Determinants and Consequences of Arsenic Metabolism Efficiency among 4,794 Individuals: Demographics, Lifestyle, Genetics, and Toxicity. *Cancer Epidemiol Biomarkers Prev.* 2016; 25:381–390. [PubMed: 26677206]

- Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics*. 2007; 8:118–127. [PubMed: 16632515]
- Jongen WM, Cardinaals JM, Bos PM, Hagel P. Genotoxicity testing of arsenobetaine, the predominant form of arsenic in marine fishery products. *Food Chem Toxicol*. 1985; 23:669–673. [PubMed: 3928470]
- Kaati G, Bygren LO, Pembrey M, Sjöström M. Transgenerational response to nutrition, early life circumstances and longevity. *Eur J Hum Genet*. 2007; 15:784–790. [PubMed: 17457370]
- Kaushal A, Zhang H, Karmaus WJJ, Everson TM, Marsit CJ, Karagas MR, Tsai SF, Wen HJ, Wang SL. Genome-wide DNA methylation at birth in relation to in utero arsenic exposure and the associated health in later life. *Environ Health*. 2017; 16:50. [PubMed: 28558807]
- Kile ML, Houseman EA, Baccarelli AA, Quamruzzaman Q, Rahman M, Mostofa G, Cardenas A, Wright RO, Christiani DC. Effect of prenatal arsenic exposure on DNA methylation and leukocyte subpopulations in cord blood. *Epigenetics*. 2014; 9:774–782. [PubMed: 24525453]
- Knight JM, Davidson LA, Herman D, Martin CR, Goldsby JS, Ivanov IV, Donovan SM, Chapkin RS. Non-invasive analysis of intestinal development in preterm and term infants using RNA-Sequencing. *Sci Rep*. 2014; 4:5453. [PubMed: 24965658]
- Lin GF, Du H, Chen JG, Lu HC, Guo WC, Meng H, Zhang TB, Zhang XJ, Lu DR, Golka K, Shen JH. Arsenic-related skin lesions and glutathione S-transferase P1 A1578G (Ile105Val) polymorphism in two ethnic clans exposed to indoor combustion of high arsenic coal in one village. *Pharmacogenet Genomics*. 2006; 16:863–871. [PubMed: 17108810]
- Liu J, Yu L, Tokar EJ, Bortner C, Sifre MI, Sun Y, Waalkes MP. Arsenic-induced aberrant gene expression in fetal mouse primary liver-cell cultures. *Ann N Y Acad Sci*. 2008; 1140:368–375. [PubMed: 18991936]
- Liu X, Zheng Y, Zhang W, Zhang X, Lloyd-Jones DM, Baccarelli AA, Ning H, Fornage M, He K, Liu K, Hou L. Blood methylomics in response to arsenic exposure in a low-exposed US population. *J Expo Sci Environ Epidemiol*. 2014; 24:145–149. [PubMed: 24368509]
- Maksimovic J, Phipson B, Oshlack A. A cross-package Bioconductor workflow for analysing methylation array data. *F1000Res*. 2016; 5:1281. [PubMed: 27347385]
- Mao G, Guo X, Kang R, Ren C, Yang Z, Sun Y, Zhang C, Zhang X, Zhang H, Yang W. Prevalence of disability in an arsenic exposure area in Inner Mongolia, China. *Chemosphere*. 2010; 80:978–981. [PubMed: 20646734]
- Morgan HD, Sutherland HG, Martin DI, Whitelaw E. Epigenetic inheritance at the agouti locus in the mouse. *Nat Genet*. 1999; 23:314–318. [PubMed: 10545949]
- Pershagen G. The carcinogenicity of arsenic. *Environ Health Perspect*. 1981; 40:93–100. [PubMed: 7023936]
- Peters TJ, Buckley MJ, Statham AL, Pidsley R, Samaras K, RVL, Clark SJ, Molloy PL. De novo identification of differentially methylated regions in the human genome. *Epigenetics Chromatin*. 2015; 8:6. [PubMed: 25972926]
- Petkovich DA, Podolskiy DI, Lobanov AV, Lee SG, Miller RA, Gladyshev VN. Using DNA Methylation Profiling to Evaluate Biological Age and Longevity Interventions. *Cell Metab*. 2017; 25:954–960e956. [PubMed: 28380383]
- R Development Core Team. R: A language and environment for statistical computing. R foundation for statistical computing; Vienna, Austria: 2009.
- Reichard JF, Schnekenburger M, Puga A. Long term low-dose arsenic exposure induces loss of DNA methylation. *Biochem Biophys Res Commun*. 2007; 352:188–192. [PubMed: 17107663]
- Ren X, Aleshin M, Jo WJ, Dills R, Kalman DA, Vulpe CD, Smith MT, Zhang L. Involvement of N-6 adenine-specific DNA methyltransferase 1 (N6AMT1) in arsenic biomethylation and its role in arsenic-induced toxicity. *Environ Health Perspect*. 2011a; 119:771–777. [PubMed: 21193388]
- Ren X, McHale CM, Skibola CF, Smith AH, Smith MT, Zhang L. An emerging role for epigenetic dysregulation in arsenic toxicity and carcinogenesis. *Environ Health Perspect*. 2011b; 119:11–19. [PubMed: 20682481]
- Rojas D, Rager JE, Smeester L, Bailey KA, Drobna Z, Rubio-Andrade M, Styblo M, Garcia-Vargas G, Fry RC. Prenatal arsenic exposure and the epigenome: identifying sites of 5-methylcytosine

alterations that predict functional changes in gene expression in newborn cord blood and subsequent birth outcomes. *Toxicol Sci.* 2015; 143:97–106. [PubMed: 25304211]

- Sampayo-Reyes A, Hernandez A, El-Yamani N, Lopez-Campos C, Mayet-Machado E, Rincon-Castaneda CB, de Limones-Aguilar ML, Lopez-Campos JE, de Leon MB, Gonzalez-Hernandez S, Hinojosa-Garza D, Marcos R. Arsenic induces DNA damage in environmentally exposed Mexican children and adults. Influence of GSTO1 and AS3MT polymorphisms. *Toxicol Sci.* 2010; 117:63–71. [PubMed: 20547570]
- Seow WJ, Kile ML, Baccarelli AA, Pan WC, Byun HM, Mostofa G, Quamruzzaman Q, Rahman M, Lin X, Christiani DC. Epigenome-wide DNA methylation changes with development of arsenic-induced skin lesions in Bangladesh: a case-control follow-up study. *Environ Mol Mutagen.* 2014; 55:449–456. [PubMed: 24677489]
- Simeonova PP, Luster MI. Mechanisms of arsenic carcinogenicity: genetic or epigenetic mechanisms? *J Environ Pathol Toxicol Oncol.* 2000; 19:281–286. [PubMed: 10983894]
- Sirko S, von Holst A, Wizenmann A, Gotz M, Faissner A. Chondroitin sulfate glycosaminoglycans control proliferation, radial glia cell differentiation and neurogenesis in neural stem/progenitor cells. *Development.* 2007; 134:2727–2738. [PubMed: 17596283]
- Smith AH, Goycolea M, Haque R, Biggs ML. Marked increase in bladder and lung cancer mortality in a region of Northern Chile due to arsenic in drinking water. *Am J Epidemiol.* 1998; 147:660–669. [PubMed: 9554605]
- Smith AH, Hopenhayn-Rich C, Bates MN, Goeden HM, Hertz-Picciotto I, Duggan HM, Wood R, Kosnett MJ, Smith MT. Cancer risks from arsenic in drinking water. *Environ Health Perspect.* 1992; 97:259–267. [PubMed: 1396465]
- Smith AH, Marshall G, Roh T, Ferreccio C, Liaw J, Steinmaus C. Lung, Bladder, and Kidney Cancer Mortality 40 Years After Arsenic Exposure Reduction. *J Natl Cancer Inst.* 2017
- Smith AH, Marshall G, Yuan Y, Ferreccio C, Liaw J, von Ehrenstein O, Steinmaus C, Bates MN, Selvin S. Increased mortality from lung cancer and bronchiectasis in young adults after exposure to arsenic in utero and in early childhood. *Environ Health Perspect.* 2006; 114:1293–1296. [PubMed: 16882542]
- Sofer T, Schifano ED, Hoppin JA, Hou L, Baccarelli AA. A-clustering: a novel method for the detection of co-regulated methylation regions, and regions associated with exposure. *Bioinformatics.* 2013; 29:2884–2891. [PubMed: 23990415]
- Steinmaus C, Ferreccio C, Acevedo J, Yuan Y, Liaw J, Duran V, Cuevas S, Garcia J, Meza R, Valdes R, Valdes G, Benitez H, VanderLinde V, Villagra V, Cantor KP, Moore LE, Perez SG, Steinmaus S, Smith AH. Increased lung and bladder cancer incidence in adults after in utero and early-life arsenic exposure. *Cancer Epidemiol Biomarkers Prev.* 2014; 23:1529–1538. [PubMed: 24859871]
- Tapio S, Grosche B. Arsenic in the aetiology of cancer. *Mutat Res.* 2006; 612:215–246. [PubMed: 16574468]
- Touleimat N, Tost J. Complete pipeline for Infinium((R)) Human Methylation 450K BeadChip data processing using subset quantile normalization for accurate DNA methylation estimation. *Epigenomics.* 2012; 4:325–341. [PubMed: 22690668]
- Valenzuela OL, Drobna Z, Hernandez-Castellanos E, Sanchez-Pena LC, Garcia-Vargas GG, Borja-Aburto VH, Styblo M, Del Razo LM. Association of AS3MT polymorphisms and the risk of premalignant arsenic skin lesions. *Toxicol Appl Pharmacol.* 2009; 239:200–207. [PubMed: 19538983]
- Waalkes MP, Qu W, Tokar EJ, Kissling GE, Dixon D. Lung tumors in mice induced by “whole-life” inorganic arsenic exposure at human-relevant doses. *Arch Toxicol.* 2014; 88:1619–1629. [PubMed: 25005685]
- Weidner CI, Lin Q, Koch CM, Eisele L, Beier F, Ziegler P, Bauerschlag DO, Jockel KH, Erbel R, Muhleisen TW, Zenke M, Brummendorf TH, Wagner W. Aging of blood can be tracked by DNA methylation changes at just three CpG sites. *Genome Biol.* 2014; 15:R24. [PubMed: 24490752]
- Weidner CI, Wagner W. The epigenetic tracks of aging. *Biol Chem.* 2014; 395:1307–1314. [PubMed: 25205717]

- Winship A, Van Sinderen M, Heffernan-Marks A, Dimitriadis E. Chondroitin sulfate proteoglycan protein is stimulated by interleukin 11 and promotes endometrial epithelial cancer cell proliferation and migration. *Int J Oncol*. 2017; 50:798–804. [PubMed: 28098860]
- Xia Y, Wade TJ, Wu K, Li Y, Ning Z, Le XC, He X, Chen B, Feng Y, Mumford JL. Well water arsenic exposure, arsenic induced skin-lesions and self-reported morbidity in Inner Mongolia. *Int J Environ Res Public Health*. 2009; 6:1010–1025. [PubMed: 19440430]
- Zhang H, Ge Y, He P, Chen X, Carina A, Qiu Y, Aga DS, Ren X. Interactive Effects of N6AMT1 and As3MT in Arsenic Biomethylation. *Toxicol Sci*. 2015; 146:354–362. [PubMed: 25997655]
- Zhang Y, Ma R, Li Z. Human health risk assessment of groundwater in Hetao Plain (Inner Mongolia Autonomous Region, China). *Environ Monit Assess*. 2014; 186:4669–4684. [PubMed: 24705813]
- Zhu J, Wang J, Chen X, Tsompana M, Gaile D, Buck M, Ren X. A time-series analysis of altered histone H3 acetylation and gene expression during the course of MMAIII-induced malignant transformation of urinary bladder cells. *Carcinogenesis*. 2017; 38:378–390. [PubMed: 28182198]

Highlights

- Historical arsenic exposure could leave detectable DNA methylation changes.
- Arsenic exposure induces common DNA methylation changes across multiple generations.
- Aberrant DNA methylations are enriched in patients with arsenic-induced skin lesions.
- Potential for the transgenerational epigenetic impact of arsenic exposure is explored.

Fig. 1. Illustration of study plan and characters of study population

A. The study was conducted in Hetao Basin, Inner Mongolia, China (Historical drinking water arsenic concentration may be higher than $500 \mu\text{g/L}$). F1: Grandparents exposed to arsenic at adulthood; F2: Parents exposed to arsenic *in utero* and early childhood; F3: Grandchildren with germ cell arsenic exposure. B. Significantly reduced global DNA methylation in patients compared to grandparents in the control and arsenic exposed groups; and C. Horvath's DNAm age of all samples based on genome-wide DNA methylation data. The chronological ages of volunteers are strongly correlated with predicted Horvath's predictor based on DNAm profiles ($r^2 = 0.92$).

Fig. 2. Summary of the identified DML

A. Summary of DML distributions in chromosomes. The cytogenetic bands of each chromosome are presented in the first circle (in Mb), with the centromere highlighted in red. The next circle shows the distribution of DML identified across chromosomes (red color: hyper-methylation, blue color: hypo-methylation). The cream-colored circle reports the names of genes found within 50 kb of each associated CpG (black letters). The central circle reports 28 identified CpGs with an inheritable potential. B. Summary of DML identified based on various parameters (red color: hyper-methylation, blue color: hypo-methylation). The Y-axis indicates the frequency of hyper- and hypo-methylation. Frequency was illustrated by the end of each bar. The X-axis indicates groups of effective size. C. Summary of DML locations and the relationship with genes.

Fig. 3. Examples of the top DML, DMRs and canonical pathways

A. β value plot of selected examples from top DML including all generations from both arsenic exposed and control groups. B. Plots of three top DMRs, the Y-axis indicates β value and the X-axis indicates genomic location and relevant genes. C. Top three canonical pathways identified by IPA analysis using all DML identified.

Fig. 4. Summary of the results from gene-wise permutation test

A. β value plot of top six genes with associated CpG differentially methylated (FDR < 0.05).

B. Canonical pathways with a p -value < 0.01.

Fig. 5. Unsupervised hierarchical clustering of DML for arsenical skin lesion patients and grandparents from arsenic exposed and control groups

The gray and red dashed rectangles indicate members of the two main clusters. The color bar at the base of the figure indicates samples belonging groups. The blue dashed rectangle indicates patients that have DNA methylation patterns distinguishable from grandparents in the arsenic-exposed group (age-matched population that did not develop any diseases).

Fig. 6. DML in arsenical skin lesion patients

A. Comparison of DML similarity between patients and grandparents in the arsenic-exposed or in control group. B & C. Redrawing of the DML and DMRs included in Figures 3A & B with patients' samples added.

Fig. 7. Examples of CpG sites with inheritable potential

A. Bean plot of M value for two selected CpG sites with inheritable potential. Upper and bottom panel indicate hypo- and hyper-methylation of CpG sites, respectively. B. Illustration of potential penetrant patterns of the two sites in both control and arsenic exposed families.

Table 1

General characteristics of study participants.

Characteristic	Non-Exposure Group				Exposure Group			Patients with skin lesions
	Generation 1 ^a	Generation 2 ^b	Generation 3 ^c		Generation 1	Generation 2	Generation 3	
<i>n</i>	17	9	9		31	18	18	
Age (years) ^d	65.24 ± 7.52	40.00 ± 6.95	13.13 ± 6.29		58.81 ± 8.01	33.59 ± 5.22	10.83 ± 3.45	60.94 ± 11.22
Gender								
Male, <i>n</i> (%)	8 (47.1%)	8 (88.8%)	4 (44.4%)		16 (51.6%)	16 (88.8%)	8 (44.4%)	12 (66.7%)
Smoking (Yes, <i>n</i> (%))	8 (100%)	7 (87.5%)	0 (0%)		15 (93.8%)	16 (100%)	0 (0%)	10 (83.3%)
Drinking (Yes, <i>n</i> (%))	6 (75.0%)	6 (75.0%)	0 (0%)		14 (87.5%)	14 (87.5%)	0 (0%)	9 (75.0%)
Female, <i>n</i> (%)	9 (52.9%)	1 (11.2%)	5 (55.6%)		15 (48.4%)	2 (11.2%)	10 (55.6%)	6 (33.3%)
Smoking (Yes, <i>n</i> (%))	2 (22.2%)	0 (0%)	0 (0%)		5 (33.3%)	0 (0%)	0 (0%)	2 (33.3%)
Drinking (Yes, <i>n</i> (%))	1 (11.1%)	0 (0%)	0 (0%)		1 (6.6%)	0 (0%)	0 (0%)	0 (0%)
Education								
No school, <i>n</i> (%)	7 (41.2%)	0 (0%)	0 (0%)		16 (51.6%)	0 (0%)	0 (0%)	8 (44.4%)
Elementary, <i>n</i> (%)	6 (35.3%)	0 (0%)	3 (33.3%)		10 (32.3%)	0 (0%)	11 (61.1%)	5 (27.8%)
Middle, <i>n</i> (%)	3 (17.6%)	1 (11.1%)	3 (33.3%)		3 (9.7%)	0 (0%)	6 (33.3%)	3 (16.7%)
High, <i>n</i> (%)	1 (5.9%)	8 (88.9%)	3 (33.3%)		2 (6.5%)	18 (100%)	1 (5.6%)	2 (11.1%)
Arsenic Concentration (ppb) in tube-wells ^d	ND ^e	ND ^e	N/A		266.35 ± 182.40	266.35 ± 182.40	N/A	258.78 ± 212.22
DNA Global Methylation (% 5-mC) ^d	7.04% ± 1.49%	7.26% ± 1.36%	7.86% ± 1.21%		6.95% ± 1.10%	6.96% ± 0.89%	7.28% ± 2.12%	6.10% ± 1.07%* ^f

^aGeneration 1: Grandparent generation.^bGeneration 2: Parent generation.^cGeneration 3: Grandchildren generation.^dValues are mean ± SD.^eND: Non-detectable or below 10 µg/L.^fStatistical significance of global DNA methylation in patients vs. non-exposure and arsenic exposure groups is presented as **p*-value < 0.05.