A cross-study analysis of prenatal exposures to environmental contaminants and the epigenome: support for stress-responsive transcription factor occupancy as a mediator of gene-specific CpG methylation patterning

Elizabeth M. Martin† and Rebecca C. Fry*

Department of Environmental Sciences and Engineering, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, NC 27599, USA

*Correspondence address. Department of Environmental Sciences and Engineering, Gillings School of Global Public Health, 135 Dauer Drive, CB 7431, University of North Carolina, Chapel Hill, NC 27599, USA. Tel: (919) 843-6864; Fax: (919) 966-7911; E-mail: rfry@unc.edu, http://orcid.org/0000-0003-0899-9018
†Elizabeth M. Martin, http://orcid.org/0000-0001-8428-7034

Abstract

A biological mechanism by which exposure to environmental contaminants results in gene-specific CpG methylation patterning is currently unknown. We hypothesize that gene-specific CpG methylation is related to environmentally perturbed transcription factor occupancy. To test this hypothesis, a database of 396 genes with altered CpG methylation either in cord blood leukocytes or placental tissue was compiled from 14 studies representing assessments of six environmental contaminants. Subsequently, an in silico approach was used to identify transcription factor binding sites enriched among the genes with altered CpG methylation in relationship to the suite of environmental contaminants. For each study, the sequences of the promoter regions (representing −1000 to +500 bp from the transcription start site) of all genes with altered CpG methylation were analyzed for enrichment of transcription factor binding sites. Binding sites for a total of 56 unique transcription factors were identified to be enriched within the promoter regions of the genes. Binding sites for the Kidney-Enriched Krupple-like Factor 15, a known responder to endogenous stress, were enriched (\( P < 0.001–0.041 \)) among the genes with altered CpG methylation associated for five of the six environmental contaminants. These data support the transcription factor occupancy theory as a potential mechanism underlying environmentally-induced gene-specific CpG methylation.

Key words: transcription factor occupancy theory; prenatal environmental exposure; CpG DNA methylation; KLF15; gene-specific methylation patterning
Introduction

Prenatal exposure to environmental contaminants has been associated with detrimental health outcomes early in life, as well as later in adulthood. Depending upon the contaminant, detrimental health effects in infants and children may include low birth weight, preterm birth, and increased incidence of infections in early childhood [1, 2]. In addition to health effects apparent at birth and in early childhood, there is the potential for prenatal environmental exposures to influence adult health. For example, later life diseases associated with early life exposures include cancers in different tissues, diabetes, and suppressed immune function [3, 4]. Still, specific genes and biological pathways that are causally related to these later life health effects of early life exposure remain largely unknown.

One potential mechanism by which prenatal exposures may drive later life health outcomes is via epigenetic modification, specifically altered CpG methylation patterning. CpG methylation has the potential to alter transcription and subsequent translational regulation, ultimately impacting disease susceptibility [5]. During mammalian fetal development, there are two major events, one during preimplantation during which DNA is passively demethylated and one during midgestation during which DNA is actively demethylated, leaving the genome completely unmethylated and vulnerable to insult [6–8]. Additionally, passive demethylation takes place throughout pregnancy resulting in multiple timeframes during which the fetal epigenome is susceptible to epigenetic modification [3, 4, 8]. These CpG modifications can be influenced via maternal exposures during the prenatal period with the potential to remain stable over the course of a lifetime [9]. This stability is important as it could provide an explanation for the relationship between prenatal exposures with later life health outcomes.

There have been several biological mechanisms proposed to underlie environmental-contaminant-driven changes in CpG methylation. For example, some environmental contaminants disrupt the process of DNA methylation by affecting DNA methyl transferase (DNMT) activity and/or modify the availability of the substrate used for methylation, namely S-adenosine methionine [10]. Environmental contaminants can adversely impact the ability of DNMT to methylate DNA either by affecting the efficiency of the enzyme and/or decreased expression levels of DNMT [11]. In other cases, contaminants may be detoxified by processes that utilize S-adenosine methionine, thus depleting the methyl donor molecule. Depletion of the methyl donor and decreased functionality of DNMT have been associated with decreased levels of global methylation [12, 13]. Importantly, while these data provide general information on contaminant-DNA methylation relationships, they do not explain how exposure to environmental contaminants results in gene-specific modifications to CpG methylation.

In prior work, we proposed that transcription factors, induced or repressed by environmental contaminants, may mediate the gene-specific patterns of DNA methylation related to in utero exposures. Specifically, we hypothesized that transcription factors respond to environmental contaminants and bind to specific target regions (binding elements) of DNA influencing access of the DNA methylation machinery to DNA [14–16]. More specifically, the “transcription factor occupancy theory” suggests that the binding of transcription factors to target regions of the genome may influence gene-specific patterns of CpG methylation by influencing the access of DNA methyltransferase (DNMT) to specific sites.

Our theory is supported by in vitro studies demonstrating that genes in lowly methylated regions of the genome are often occupied by DNA binding proteins [17–20]. Specifically, these studies have established that transcription factor binding correlates with lowly methylated regions of the genome, suggesting that DNA-binding factors can influence patterns of DNA methylation. In addition to data derived from in vitro studies, in vivo studies have demonstrated that transcription factor binding could alter methylation patterns of key genes. Specifically, decreased maternal care in rats leads to DNA binding of the transcription factor NGFI-A, resulting in decreased DNA methylation observed at the glucocorticoid receptor (GR) gene promoter region [21]. Taken together, these data support that transcription factor binding to target genes can influence CpG methylation patterns.

In this work, an in silico approach was used to determine whether evidence exists for the transcription factor occupancy theory related to prenatal environmental exposures. Gene sets derived from studies that examined the effects of prenatal environmental exposures on genome-wide gene-specific DNA methylation were examined including those that assessed arsenic [15, 22–24], cadmium [16, 25], lead [26], manganese [27], mercury [28], and tobacco smoke [29–33]. The data from the 14 studies were integrated, resulting in a cross-contaminant database used to identify common transcription factor binding sequences in differentially methylated genes associated with environmental contaminants.

Results

Generation of a Cross-Study Database of Gene-Specific CpG Methylation

Of the 14 studies identified for analysis, the majority (n = 11, 78.6%) examined CpG methylation in cord blood leukocytes, while the remaining three studies analyzed DNA methylation in placental tissue (Table 1). A number of the published studies examined the relationship between CpG methylation and prenatal exposure to tobacco smoke (n = 5) or arsenic (n = 4). The other studies focused on cadmium (n = 2), lead (n = 1), manganese (n = 1), or mercury (n = 1). The majority (n = 12, 85.7%) of studies utilized the Illumina 450 k array for CpG methylation assessment, while one study used the Illumina 27 k array and one study used the Affymetrix MIRA methylation assay.

CpG Methylation Trends across Environmental Contaminants

We first set out to determine whether there was a common set of genes with altered CpG methylation associated with prenatal exposure to environmental contaminants. Across the 14 studies, there were 396 unique genes identified to have altered CpG methylation in relationship to exposure to six different environmental contaminants (Fig. 1). There was no overlap between the 396 genes within the compiled database and the 500 probes represented within the Houseman et al. [45]. Further detail regarding whether these genes were classified as hyper- or hypomethylated in relationship to environmental exposure is presented in Supplementary Table S1. At the gene level, there was little overlap with only 14 (3.53%) genes being shared between two or more studies, regardless of environmental contaminant.

Specifically, there was little evidence of gene-specific overlap between studies examining the same contaminant. For
| Study (PMID)       | Environmental contaminant | Subjects (n) | Number of associated genes | Measure of exposure | Range of exposures | Tissue/biological matrix | Technology                        |
|-------------------|----------------------------|--------------|----------------------------|---------------------|--------------------|--------------------------|-----------------------------------|
| Broberg et al. [22] (24965135) | Arsenic (inorganic)       | 127          | 29                         | Maternal urinary total arsenic | 20–457 μg/l | Cord blood | Illumina 450 k |
| Kile et al. [23] (24525453)     | Arsenic (inorganic)       | 44           | 6                          | Drinking water arsenic   | < LOD*–510 μg/l | Cord blood | Illumina 450 k |
| Koestler et al. [24] (23757598)  | Arsenic (inorganic)       | 134          | 25                         | Maternal urinary arsenic (quartiles) | 1.8–6.6 μg/l | Cord blood | Illumina 450 k |
| Rojas et al. [15] (25304211)     | Arsenic (inorganic)       | 38           | 16                         | Maternal urinary total arsenic | < LOD*–236 μg/l | Cord blood | Illumina 450 k |
| Kipper et al. [26] (23644563)    | Cadmium                   | 127          | 25                         | Maternal urinary cadmium | 0.21–2.4 μg/l | Cord blood | Illumina 450 k |
| Sanders et al. [16] (24169490)   | Cadmium                   | 17           | 91                         | Maternal urinary cadmium | 0–1.05 μg/l   | Cord blood | Illumina 450 k |
| Sen et al. [26] (26046694)       | Lead                      | 48           | 16                         | Maternal blood lead levels | 0.59–721 μg/dl | Cord blood | Illumina 450 k |
| Maccani et al. [27] (25982381)   | Manganese                 | 61           | 5                          | Maternal toenail manganese levels | 0.131 to 5.666 μg/g | Placental tissue | Illumina 450 k |
| Maccani et al. [28] (25748564)   | Mercury                   | 41           | 104                        | Maternal toenail mercury levels (tertiles) | < LOD*–0.38 μg/g | Placental tissue | Illumina 450 k |
| Ivorra et al. [29] (25623364)    | Tobacco smoke             | 20           | 25                         | Self-reported smoking status confirmed by cotinine levels | NA | Cord blood | Illumina 450 k |
| Joubert et al. [30] (22851337)   | Tobacco smoke             | 1062         | 10                         | Cotinine-based categorization of smoking status | <0- ->388 nmol/l | Cord blood | Illumina 450 k |
| Lee et al. [31] (25325234)       | Tobacco smoke             | 132          | 12                         | Self-reported smoking status | Smoking one cigarette per day during the second trimester | Cord blood | Illumina 450 k |
| Richmond et al. [32] (25552657)  | Tobacco smoke             | 800          | 15                         | Self-reported smoking status | Smoking in at least two trimesters during pregnancy | Cord blood | Illumina 450 k |
| Suter et al. [33] (21937876)     | Tobacco Smoke             | 36           | 33                         | Self-reported smoking status | NA | Placental tissue | Illumina 27 k |

* Limit of detection (LOD)
status were the same across studies with increasing levels of the environmental contaminants associated with increased CpG methylation.

The final four genes were shared between studies of metals where cadmium and mercury exposure were associated with altered CpG methylation of Acid Trehalase-Like 1 (ATHL1). Arsenic and mercury exposure were associated with altered CpG methylation of Cyclic AMP-Responsive Element-Binding Protein 5 (CREB5) and Ribosomal Protein S6 Kinase, 90 kDa, Polypeptide 2 (RPS6K2). Arsenic and cadmium exposure were associated with altered CpG methylation of Histone Deacetylase 4 (HDAC4). In contrast to the prior findings, the observed associations between CpG methylation and the environmental contaminants were in opposite directions. ATHL1 showed cadmium-associated hypermethylation and mercury-associated hypomethylation. CREB5 displayed mercury-associated hypermethylation and arsenic-associated hypomethylation. HDAC3 showed arsenic-associated hypermethylation and cadmium-associated hypomethylation. Finally, RPS6KAS2 displayed mercury-associated hypermethylation and arsenic-associated hypomethylation (Supplementary Table S1).

Most of the summarized studies have previously characterized the associated biological functions of the genes with altered CpG methylation. In this study, we aimed to establish major functional categories of genes with altered CpG methylation to determine commonalities within or across contaminant-specific gene sets. Nearly all the contaminants (n = 4) including arsenic, cadmium, lead, and tobacco smoke were associated with CpG methylation in target genes with known roles in transcription or transcription-related processes (n = 54, P value range: 0.000749–0.043) (Supplementary Table S3). Among these genes were HDAC4, a histone deacetylase with altered CpG methylation associated with arsenic and cadmium, as well as polymerase genes DNA-Directed RNA Polymerase II Subunit E, POLR2E, DNA-Directed RNA Polymerase II Subunit J2, POLR2J2 with altered CpG methylation in relationship to cadmium, and known transcription factor genes AHRR, Translocation-Associated Notch Protein TAN-1, NOTCH1 with altered CpG methylation in relationship to tobacco smoke exposure (Supplementary Table S2). Additionally, genes that are involved with ATP binding had altered CpG methylation in relationship to exposure to arsenic (n = 11, P value = 0.019). Genes involved in DNA binding had altered CpG methylation in relationship to arsenic (n = 7, P value = 0.029), cadmium (n = 17, P value = 0.040), and tobacco smoke (n = 16, P value = 0.006). Metal binding-related genes had altered CpG methylation in relationship to prenatal exposure to arsenic (n = 22, P value = 0.001), mercury (n = 28, P value = 0.002), and tobacco smoke (n = 6, P value = 0.038) (Supplementary Table S2).

**Common Transcriptional Regulators with Enriched Binding Sites among Altered CpG Gene Sets Associated with Environmental Contaminants**

The gene sets representing altered CpG methylation associated with prenatal environmental exposures were analyzed for transcription factor binding site enrichment within the promoter regions of the genes using an in silico analysis. The gene sets were separated into hypermethylated and hypomethylated subsets. By analyzing hypermethylated genes separately from hypomethylated genes, it was possible to make in silico predictions of the relationship between transcription factors and gene-specific DNA methylation patterns. Transcription factor binding within promoter regions, upstream of the transcription start site (TSS),
may inhibit DNMT from accessing a particular location resulting in hypomethylation of the gene (Fig. 2A). In contrast, a lack of transcription factor binding within promoter regions of genes may allow DNMT access to a particular genomic location resulting in gene-specific hypermethylation.

Based on the nature of the enrichment analysis, gene sets containing fewer than seven sequences were not suitable for transcription factor binding site enrichment analysis. This resulted in the exclusion of two studies [23, 27]. Additionally, there were insufficient sequences to conduct enrichment analysis associated with hypomethylated genes in three studies [16, 30, 31] and hypermethylated genes in two studies [29, 32]. As a result of these exclusions, 11 of 14 studies, representing 341 of the 396 genes and five of the six contaminants remained for transcription factor binding-site enrichment analysis including the following contaminants: arsenic, cadmium, lead, mercury, and tobacco smoke.

Through the transcription factor binding site enrichment analysis, we identified a total of 56 transcription factors represented within the 341 genes (Fig. 3, Supplementary Tables S1 and S3). In contrast to the gene-level analysis, the transcription factor enrichment results displayed much more overlap within and across contaminants. A total of 32 (57.1%) transcription factors were shared by at least two studies. The remaining 24 (42.9%) transcription factors were unique and found to be enriched in only one study. In addition to a larger proportion of overlapping transcription factors as compared to differentially methylated genes (57.1% vs. 3.03%), there were more commonalities within studies examining the same environmental contaminant. For example, of the four studies that focused on Cpg methylation in relationship to arsenic, no common genes were identified. However, when all arsenic-associated genes were analyzed together, there were four major transcription factors identified, namely Kidney-enriched Kruppel-like factor 15 (KLF15), Kruppel-like zinc finger protein 219 (ZNF219), Myeloid zinc finger protein (MZF1), and Zinc Finger and BTB domain containing 7 A, pokemon (ZBTB7) (Supplementary Table S3). Additionally, when arsenic-associated CpG methylation was analyzed by study, there were eight common transcription factors identified. These were KLF15, ZNF219, Myc-associated zinc finger protein (MAZ), MYC-associated zinc finger protein-related transcription factor (MAZR), MZF1, Pleomorphic adenoma gene 1 (PLAG1), ZBTB7, and Zinc Finger Protein 263 (ZNF263). Similarly, between the two cadmium-associated CpG methylation studies, there were no genes that were common, but they shared binding sites for a common transcription factor, MZF1. The five studies that focused on tobacco smoke-associated CpG methylation studies, there were no genes that were common, but they shared binding sites for a common transcription factor, MZF1. The five studies that focused on tobacco smoke-associated CpG methylation identified four genes that were common to at least two studies, while the genes with altered CpG methylation were enriched for 17 transcription factors. These included Krueppel-type Zinc Fingers (KLF6), E2F transcription Factor 1 (E2F1), Ets-family member ELF2 (NERF1a), MAZ, MAZR, MZF1, PLAG1, Ras-response element binding protein 1 (RREB1), Stimulating protein 1 (SP1), Wilms Tumor Suppressor (WT1), ZBTB7, ZNF263, Zinc finger protein insulinoma-associated 1 (IA-1) functions as a transcriptional repressor (INZM1), Zinc finger transcription factor ZBP-89 (ZBP89), and Zinc finger, BED-type containing 4 (ZBED4) (Supplementary Table S3).
Interestingly, there were some transcription factors that were represented among the majority of studies. Binding sites for kidney-enriched Kruppel-like factor, KLF15, were enriched within the gene sets associated with all five contaminants analyzed (arsenic, cadmium, lead, mercury, and tobacco smoke). KLF15 was enriched (P value range: <0.0001–0.0411) in gene sets from 8 of 11 studies analyzed (Supplementary Table S3). In addition, transcription factor binding sites for Pleomorphic adenoma gene 1, PLAG1, were also enriched (P value range: <0.0001–0.425) in gene sets associated with all five contaminants and eight studies. Only one other transcription factor, myeloid zinc finger (MZF1), had binding sites that were enriched (P value range: <0.0001–0.0283) within gene sets of eight studies and was associated with four contaminants (arsenic, cadmium, lead and tobacco smoke) (Supplementary Table S3). Interestingly, 38 (65.5%) of the genes encoding for transcription factors identified in this analysis, including MZF1 and PLAG1, have glucocorticoid-responsive-related sequence elements in their promoter regions (Supplementary Table S3).

Taken together, these results provide in silico epigenome-level evidence that these environmental contaminants may trigger glucocorticoid-receptor (GR)-related signaling, resulting in the alteration of specific transcription factors such as KLF15. In addition to the transcription factors potentially controlled by the GR, others are likely GR-independent. The subsequent binding of the transcription factors to response elements in target genes is supported by an enrichment of these sites among the tested gene sets (Supplementary Table S3, Fig. 4). The differential binding/occupancy of the transcription factors between environmentally exposed individuals versus those with no/lower exposure may result in specific patterns of CpG methylation at target gene sites. Thus, the differential patterns of CpG methylation associated with environmental contaminants may in fact be an indicator of an “epigenetic environmental stress response” or “epigenetic ESR.”

Discussion

Across studies examining gene-specific CpG methylation related to prenatal exposure to environmental contaminants, it is currently unknown whether there are commonalities at the gene level. Furthermore, a biological mechanism underlying the gene-specific CpG methylation is unknown. To address this, in this study, first we set out to generate a database of gene-specific CpG methylation compiled from studies examining the epigenetic effects of prenatal environmental exposures assessed in cord blood leukocytes or placental tissue. The contaminants included in this study were arsenic, cadmium, lead, manganese, mercury, and tobacco smoke. This database enabled the in silico examination of promoter sequences to identify enriched transcription factor binding sites among genes with altered CpG methylation to support or refute the hypothesis of “transcription factor occupancy” as a mediator of gene-specific methylation [14–16]. At the gene level, we observed minimal overlap for studies examining the epigenetic effects even for the same contaminant, as well as those examining diverse environmental contaminants. In support of the “transcription factor occupancy” theory, there was much more commonality at the level of transcription factors where gene sets showed enrichment of binding sites for common transcription factors. Interestingly, the majority of genes with altered CpG methylation contained binding sites for KLF15, a known responder to GR signaling. Taken together, these data provide in silico support for gene-specific methylation patterns potentially reflective of an epigenetic ESR to different environmental contaminants.

There was minimal overlap of genes with altered CpG methylation across studies examining the same contaminant or across contaminants. Specifically, of the 396 genes in the database, only 14 (or 3.53%) were shared between two or more studies. The greatest overlap was observed for studies of tobacco smoke where four genes were independently identified to have altered CpG methylation across three studies analyzing tobacco smoke-associated methylation. While there was little overlap at the gene level, there are known associations between the genes with altered CpG methylation and negative health outcomes associated with the environmental contaminants. For example, the genes with altered CpG methylation in relationship to tobacco smoke exposure, CYP1A1 and GFI-1, are associated with differences in susceptibility to lung cancer and acute myeloid leukemia, respectively [31, 34]. DDAH2 has altered CpG methylation in relationship to prenatal smoke and lead and is an epigenetic marker for neural stem cell differentiation [35]. Both prenatal tobacco smoke exposure and lead have been previously associated with neurodevelopmental outcomes [36, 37]. These data highlight prenatal exposure to environmental toxins is associated with altered CpG methylation of target genes known to play a role in contaminant-related diseases.

In support of a biological mechanism for gene-specific patterns of CpG methylation, the data presented here suggest that transcription factor occupancy may play a role. In prior studies, we examined transcription factor binding site enrichment among genes with altered CpG methylation in relationship to prenatal cadmium [16], arsenic [15], and preeclampsia [14]. In the present research, we expanded upon this to examine five contaminants across 13 studies for sequence-based enrichment of transcription factor binding sites and found that 56 transcription factors have enriched binding sites within the promoter regions of 341 compiled genes. The majority (n = 33, 58.9%) of these transcription factors were represented in gene sets identified from at least two studies, with three transcription factors...
(5.36%) represented in the majority of studies, namely KLF15, MZF1, and PLAG1. MZF1 is a tumor suppressor protein and controls cell proliferation and tumorigenesis [38]. This is interesting as arsenic, cadmium, and tobacco smoke are known carcinogens and were associated with CpG methylation of target genes of MZF1. The other major transcription factor with binding sites identified to be enriched in this analysis was PLAG1, representing an imprinted and developmentally-regulated gene [39]. Decreased expression of PLAG1 has been associated with fetal growth restriction, which is a common birth outcome for fetuses exposed to environmental contaminants in utero [40]. It is also an oncogene and associated with numerous cancers [39].

Thus, transcription factor binding sites identified using sequence-based enrichment analysis may provide clues regarding proteins that regulate environmental contaminant-induced DNA methylation patterns that could, in part, underlie later life disease.

The most significant of the identified transcription factors was KLF15, with binding sites enriched within genes with altered CpG methylation upon exposure to five contaminants. KLF15 expression is upregulated in response to stress-responsive GR activation [41, 42]. GR is a transcription factor responsive to a suite of environmental contaminants including those analyzed here namely, arsenic, cadmium, lead, mercury, and tobacco smoke [43] and has previously been shown to be a potential regulator of genes and proteins altered by arsenic and cadmium [44]. While not observed in the target genes themselves, the majority of identified transcription factors include binding sites for the GR supporting that differential patterns of CpG methylation associated with environmental contaminants may be an indicator of an epigenetic ESR. This ESR may be mediated through GR signaling, with subsequent CpG methylation marks on downstream target genes.

In summary, the data from this study provide support for the transcription factor occupancy theory as a driver of environmentally altered CpG methylation patterning. While there was little overlap at the gene level, genes with altered CpG methylation in relationship to prenatal environmental contaminants share common binding sites for a set of transcription factors. These identified transcription factors have known relationships both to the contaminants of interest and associated diseases. These findings are important as they provide support for a hypothesis for a specific biological mechanism by which prenatal exposure to environmental contaminants may impart gene-specific footprints on the epigenome in both fetal blood and placental tissue. As this research utilizes specific footprints on the epigenome in both fetal blood and placental tissue, the identified target genes may be key regulators of the fetal epigenome.

**Methods**

**Compilation of Studies Assessing Gene-Specific Changes in CpG Methylation Related to Prenatal Exposure to Environmental Contaminants**

To identify studies for inclusion in the analysis, a literature search was conducted using PubMed (US National Library of Medicine and National Institutes of Health). Studies were included if they were carried out using biological samples collected from human subjects and if maternal biomarker measures of environmental contaminants were provided. Gene-specific CpG methylation was required to be assessed using the Illumina 450k array or similar technology (Methylated CpG Island Recovery Assay (MIRA), Illumina 27 K, Illumina Golden Gate, etc.), in a relevant tissue (cord blood or placenta) for comparability between studies.

To obtain relevant studies, combinations of applicable terms were queried. The first set of terms was “prenatal,” “in utero,” and “transplacental.” The second set of terms included both the general term “environmental contaminant” and a second set of specific contaminants of interest: “air pollution,” “arsenic,” “bisphenol A,” “cadmium,” “dioxin,” “endocrine disrupters,” “flame retardants,” “lead,” “manganese,” “mercury,” “metals,” “pesticides,” “phthalates,” “polyaromatic hydrocarbons,” and “tobacco smoke.” The final set of terms pertained to CpG methylation: “methylation,” “5-methylcytosine,” and “CpG.” These yielded a total of 408 studies of interest with a total of 14 studies suitable for inclusion. Studies were excluded from analysis if they focused on non-environmental toxicant exposures (medication, alcohol), were case-control designs, or were reviews. They were additionally excluded if they did not assess gene-specific CpG methylation in cord blood or placental samples.

**Identifying Trends in Gene-Specific CpG Methylation across Studies of Diverse Environmental Contaminants**

To examine trends in gene-specific CpG methylation associated with prenatal exposure to environmental contaminants, a database of 396 genes was compiled. These genes were identified from a total of 14 total studies analyzing six contaminants: arsenic, cadmium, lead, manganese, mercury, and tobacco smoke. The genes compiled represent those determined to be statistically significant by the study authors. Most studies employed regression analysis requiring a P value < 0.05 and a false discovery rate correction [16, 23, 24, 29, 30, 32]. Others used cluster-based analysis coupled with regression models with P value selection and false discovery rate correction [26–28, 33]. Rojas et al. [15] applied a multiple regression model (P value < 0.05, q < 0.1) followed by integration with functional changes in gene expression assessed using correlation analysis [15]. Broberg et al. [22] used regression analysis presenting the most significant regression P values.

In addition to collecting all statistically significant sites for the analysis, the compiled genes were compared to the 500 probes identified by Houseman et al. [45]. These probes contained within the Infinium HumanMethylation450 BeadChip are representative of CpG methylation changes that can be used as surrogate measurements of changes in the underlying white blood cell population mixture [45]. The probes compiled in the present study were compared against the cell population-related probes to test whether the contaminant-associated changes were related to potential shifts in white blood cell population. Finally, any probes found in intergenic regions were excluded from the analysis.

After compilation of the genes from each study, each of the genes was classified as either hypermethylated or hypomethylated. Different studies used different metrics of change, with some utilizing a β-difference (Δβ) [15, 24, 28, 29, 33, 46]. Others were expressed as correlation coefficients [22, 25] or partial correlation (β) coefficients [16, 23, 26, 30, 31]. A gene’s methylation status (i.e. hyper- or hypomethylation) in relationship to the environmental exposure was assigned based upon the majority of statistically significant probes. Thus, for each of the 14 studies,
genes were classified as hyper- or hypomethylated in relationship to the environmental exposure resulting in the generation of 28 gene sets.

Identifying Biological Pathways Enriched among Genes with Altered CpG Methylation Associated with Prenatal Environmental Contaminants

Once the database of 396 genes was compiled, Database for Annotation, Visualization and Integrated Discovery (DAVID) functional annotation software (david.ncifcrf.gov) was used to conduct biological category-based enrichment analyses. Most of the summarized studies have published characterization of biological functions of the genes with altered CpG methylation. For this reason, the functional annotation tool using keyword gene ontology was used to determine enrichment of major biological functions for each study \( n = 14 \). The \( P \) value determined by DAVID represents an absolute enrichment determined by a modified Fisher’s exact test [47].

Identifying Transcription Factor Binding Site Enrichment among Genes with Altered CpG Methylation Associated with Prenatal Environmental Contaminants

The enrichment of transcription factor binding sites within the gene sets was carried out using Genomatix software (http://www.genomatix.de) [48]. In addition to the individual study-based analysis, genes were also grouped and analyzed by contaminant, resulting in six additional gene sets for analysis. Using the Gene2Promoter module, the promoter region sequences, designated as \(-1000 \text{bp} \searrow 500 \text{bp} \) of TSS, all genes within each of the gene sets were collected. Gene sets containing seven or fewer sequences were excluded as meaningful enrichment analysis could not be conducted on these gene sets. This resulted in 24 gene sets that were analyzed using the Common Transcription Factor Analysis Module. Core similarities of 95% and the presence of the consensus sequence in more than 50% of the analyzed genes were required. Conserved “core” sequences were queried across the promoter regions of all genes. These data were compared to a series of weight matrices, which are a summary of DNA sequence patterns in transcription factor binding sites. \( P \) values, representing the likelihood of obtaining an equal or greater number of sequences with a factor binding sites.

Supplementary data

Supplementary data are available at EnuEpiJ online.

Acknowledgements

This research was supported by grants from the National Institute of Environmental Health Sciences (T32-ES007018, R01-ES001935, P42-ES005948).

Conflict of interest: None declared.

References

1. Wigle DT, Arbuckle TE, Turner MC et al. Epidemiologic evidence of relationships between reproductive and child health outcomes and environmental chemical contaminants. J Toxicol Environ Health B Crit Rev 2008;11:373–517.

2. Windham G and Fenster L. Environmental contaminants and pregnancy outcomes. Fertil Steril 2008;89:e111–6; discussion e117.

3. Boekelheide K, Blumberg B, Chapin RE et al. Predicting later-life outcomes of early-life exposures. Environ Health Perspect 2012;120:1353–61.

4. Nicoletto SF and Rinaldi A. In the womb’s shadow. The theory of prenatal programming as the fetal origin of various adult diseases is increasingly supported by a wealth of evidence. EMBO Rep 2011;12:30–4.

5. Schubeler D. Function and information content of DNA methylation. Nature 2015;517:321–6.

6. Leitch HG, Tang WW and Surani MA. Primordial germ-cell development and epigenetic reprogramming in mammals. Curr Top Dev Biol 2013;104:149–87.

7. Morgan HD, Santos F, Green K et al. Epigenetic reprogramming in mammals. Hum Mol Genet 2005;14:R47–58.

8. Perera F and Herbstman J. Prenatal environmental exposures, epigenetics, and disease. Reprod Toxicol 2011;31:363–73.

9. Heijmans BT, Tobi EW, Stein AD et al. Persistent epigenetic differences associated with prenatal exposure to famine in humans. Proc Natl Acad Sci USA 2008;105:17046–9.

10. Ruiz-Hernandez A, Kuo CC, Rentero-Garrido P et al. Environmental chemicals and DNA methylation in adults: a systematic review of the epidemiologic evidence. Clin Epigenet 2015;7:55.

11. Subach OM, Maltseva DY, Shastry A et al. The stereochemistry of benz[a]pyrene-2’-deoxyguanosine adducts affects DNA methylation by SsII and Hhal DNA methyltransferases. FEBS J 2007;274:2121–34.

12. Anderson GS, Sant KE and Dolenoy DC. Nutrition and epigenetics: an interplay of dietary methyl donors, one-carbon metabolism and DNA methylation. J Nutr Biochem 2012;23:853–9.

13. Reichard JF, Schnekenburger M and Puga A. Long term low-dose arsenic exposure induces loss of DNA methylation. Biochem Biophys Res Commun 2007;352:188–92.

14. Martin E, Ray PD, Smeester L et al. Epigenetics and pre-eclampsia: defining functional epimutations in the pre-eclamptic placenta related to the TGF-beta pathway. PloS One 2015;10:e0141294.

15. Rojas D, Rager JE, Smeester L et al. 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.

16. Sanders AP, Smeester L, Rojas D et al. Cadmium exposure and the epigenome: Exposure-associated patterns of DNA methylation in leukocytes from mother-baby pairs. Epigenetics 2014;9:212–21.

17. Cuadrado M, Sacristán M and Antequera F. Species-specific organization of CpG island promoters at mammalian homologous genes. EMBO Rep 2001;2:586–92.

18. Feldmann A, Ivanek R, Murr R et al. Transcription factor occupancy can mediate active turnover of DNA methylation at regulatory regions. PloS Genet 2013;9:e1003994.

19. Lienert F, Wirbelauer C, Som I et al. Identification of genetic elements that autonomously determine DNA methylation states. Nat Genet 2011;43:1091–7.

20. Stadler MB, Murr R, Burger L et al. DNA-binding factors shape the mouse methylome at distal regulatory regions. Nature 2011;480:490–5.

21. Graff J, Kim D, Dobbin MM et al. Epigenetic regulation of gene expression in physiological and pathological brain processes. Physiol Rev 2011;91:603–49.
22. Broberg K, Ahmed S, Engstrom K et al. Arsenic exposure in early pregnancy alters genome-wide DNA methylation in cord blood, particularly in boys. J Dev Orig Health Dis 2014;5:288–98.

23. Kile ML, Houseman EA, Bacarelli AA et al. Effect of prenatal arsenic exposure on DNA methylation and leukocyte subpopulations in cord blood. Epigenetics 2014;9:774–82.

24. Koestler DC, Avissar-Whiting M, Houseman EA et al. Differential DNA methylation in umbilical cord blood of infants exposed to low levels of arsenic in utero. Environ Health Perspect 2013;121:971–7.

25. Kippler M, Engstrom K, Mlakar SJ et al. Sex-specific effects of arsenic exposure. Neurotoxicology 2014;44:1–7.

26. Sen A, Cingolani P, Senut MC et al. Lead exposure induces DNA methylation changes in 5-hydroxymethylcytosine clusters in CpG islands in human embryonic stem cells and umbilical cord blood. Epigenetics 2015;10:607–21.

27. Maccani JZ, Koestler DC, Houseman EA et al. DNA methylation changes in the placenta are associated with fetal manganese exposure. Reprod Toxicol 2015;57:43–9.

28. Maccani JZ, Koestler DC, Lester B et al. Placental DNA methylation related to both infant toenail mercury and adverse neurobehavioral outcomes. Environ Health Perspect 2015;123:723–9.

29. Ivorra C, Fraga MF, Bayon GF et al. DNA methylation patterns in newborns exposed to tobacco in utero. J Transl Med 2015;13:25.

30. Joubert BR, Haberg SE, Nilsen RM et al. 450 K epigenome-wide scan identifies differential DNA methylation in newborns related to maternal smoking during pregnancy. Environ Health Perspect 2012;120:1425–31.

31. Lee KW, Richmond R, Hu P et al. Prenatal exposure to maternal cigarette smoking and DNA methylation: epigenome-wide association in a discovery sample of adolescents and replication in an independent cohort at birth through 17 years of age. Environ Health Perspect 2015;123:193–9.

32. Richmond RC, Simpkin AJ, Woodward G et al. Prenatal exposure to maternal smoking and offspring DNA methylation across the life course: findings from the Avon Longitudinal Study of Parents and Children (ALSPAC). Hum Mol Genet 2015;24:2201–17.

33. Suter M, Ma J, Harris A et al. (2011). Maternal tobacco use modestly alters correlated epigenome-wide placental DNA methylation and gene expression. Epigenetics 2015;6:1284–94.

34. Khandanpour C, Thiede C, Valk PJ et al. A variant allele of Growth Factor Independence 1 (GFI1) is associated with acute myeloid leukemia. Blood 2010;115:2462–72.

35. Backdahl L, Herberth M, Wilson G. Gene body methylation of the dimethylarginine dimethylamino-hydrolase 2 (Ddah2) gene is an epigenetic biomarker for neural stem cell differentiation. Epigenetics 2009;4:248–54.

36. Liu J, Gao D, Chen Y et al. Lead exposure at each stage of pregnancy and neurobehavioral development of neonates. Neurotoxicology 2014;44:1–7.

37. Polanska K, Jurewicz J and Hanke W. Smoking and alcohol drinking during pregnancy as the risk factors for poor child neurodevelopment—a review of epidemiological studies. Int J Occup Med Environ Health 2015;28:419–43.

38. Peterson FC, Hayes PL, Waltner JK et al. Structure of the SCAN domain from the tumor suppressor protein MZF1. J Mol Biol 2006;363:137–47.

39. Van Dyck F, Declercq J, Braem CV et al. PLAG1, the prototype of the PLAG gene family: versatility in tumour development (review). Int J Oncol 2007;30:765–74.

40. Tang Q, Wu W, Xu X et al. miR-141 contributes to fetal growth restriction by regulating PLAG1 expression. PloS One 2013;8:e58737.

41. Asada M, Rauch A, Shimizu H et al. DNA binding-dependent glucocorticoid receptor activity promotes adipogenesis via Kruppel-like factor 15 gene expression. Lab Invest 2011;91:203–15.

42. Masuno K, Haldar SM, Jeyaraj D et al. Expression profiling identifies Klf15 as a glucocorticoid target that regulates airway hyperresponsiveness. Am J Respir Cell Mol Biol 2011;45:642–9.

43. Marketon JI and Sternberg EM. The glucocorticoid receptor: a revisited target for toxins. Toxins 2010;2:1357–80.

44. Rager JE, Yosim A and Fry RC. Prenatal exposure to arsenic and cadmium impacts infectious disease-related genes within the glucocorticoid receptor signal transduction pathway. Int J Mol Sci 2014;15:22374–91.

45. Houseman EA, Accomando WP, Koestler DC et al. DNA methylation arrays as surrogate measures of cell mixture distribution. BMC Bioinformatics 2012;13:86.

46. Akbal A, Yilmaz H and Tütkun E. Arsenic exposure associated with decreased bone mineralization in male. Aging Male 2014;17:256–8.

47. Huang da W, Sherman BT and Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 2009;4:44–57.

48. Cartharius K, Frech K, Grote K et al. MatInspector and beyond: promoter analysis based on transcription factor binding sites. Bioinformatics 2005;21:2933–42.

49. Ho Sui SJ, Mortimer JR, Arenillas DJ et al. oPOSSUM: identification of over-represented transcription factor binding sites in co-expressed genes. Nucleic Acids Res 2005;33:3154–64.