



Integrative methylation profiling uncovers *IL10RB* hypomethylation as a mediator between environmental heavy metal exposure and lung cancer risk[☆]

Hyeongyu Shin^a, Chul-Hong Kim^b, Jeeyoung Kim^c, JungYeon Kwon^d, Min Heo^a, Nam-Eun Kim^e, Sujin Seo^f, Yong Min Cho^g, Young-Seoub Hong^d, Jung-Woong Kim^b, Woo Jin Kim^{c,*}, Sungho Won^{a,e,f,**}

^a Interdisciplinary Program of Bioinformatics, College of Natural Sciences, Seoul National University, Seoul, South Korea

^b Department of Life Science, Chung-Ang University, Seoul, South Korea

^c Department of Internal Medicine and Environmental Health Center, Kangwon National University, Chuncheon, South Korea

^d Department of Preventive Medicine, College of Medicine, Dong-A University, Busan, South Korea

^e Institute of Health and Environment, Seoul National University, Seoul, South Korea

^f Department of Public Health Sciences, Seoul National University, Seoul, South Korea

^g Department of Nano, Chemical and Biological Engineering, SeoKyeong University, Seoul, South Korea

ARTICLE INFO

Keywords:

Heavy metal

Methylation

Meet-in-the-middle analysis

Mendelian randomization

IL10RB

RARA

ABSTRACT

Residents living in areas exposed to environmental hazards, such as abandoned mines and refineries, are hypothesized to exhibit distinct omics profiles compared with those in unaffected regions. In this article, we aimed to identify the differentially methylated regions (DMRs) affected by these hazards using Forensic Research via Omics Markers in Environmental Health Vulnerable Areas (FROM) cohort. We then evaluated the potential effects of these DMRs on 23 human diseases by using data from the Korean Genome and Epidemiology Study (KoGES) cohort, which included 79,294 participants. We conducted a DMR analysis using reduced representation bisulfite sequencing on peripheral blood mononuclear cells samples, and employed the Meet-in-the-Middle analysis to identify biomarkers that enable more precise and biologically meaningful interpretation with the forensic research via FROM cohort, and significant findings were further validated with one-sample Mendelian Randomization (MR) method. To explore relationships between these biomarkers and 23 diseases, the two-sample MR approach within the KoGES cohort was conducted. Our findings indicate that long-term exposure to heavy metals in environmentally vulnerable areas alters methylation patterns, potentially affecting disease risk. Notably, a DMR in the *IL10RB* gene region was identified, where hypomethylation was associated with the development of lung cancer ($P = 1.92 \times 10^{-3}$, $\beta = -1.7830$). This study highlights the potential epigenetic impact of environmental exposures and underscores the importance of uncovering biomarkers linking exposure to disease outcome.

1. Introduction

Environmental hazards – particularly those arising from industrial activities such as mining and refining – have garnered significant attention due to their potential adverse effects on human health (Bakulski et al., 2020; Briffa et al., 2020; Cosselman et al., 2015). Heavy

metals released from these sites can contaminate soil, water, and air, ultimately entering the human body through various exposure pathways and accumulating over time. Studies have linked chronic exposure to heavy metals with increased risks of cancer, metabolic disorders, and other chronic diseases (Goyer, 1997; Tchounwou et al., 2012). Despite a growing body of epidemiological evidence, the underlying biological

[☆] This paper has been recommended for acceptance by Wen Chen.

^{*} Corresponding author. Department of Internal Medicine and Environmental Health Center, School of Medicine, Kangwon National University, Chuncheon, 24341, South Korea.

^{**} Corresponding author. Department of Public Health Science, Seoul National University, Seoul, 151-742, South Korea.

E-mail addresses: pulmo2@kangwon.ac.kr (W.J. Kim), won1@snu.ac.kr (S. Won).

<https://doi.org/10.1016/j.envpol.2025.127537>

Received 20 September 2025; Received in revised form 21 November 2025; Accepted 13 December 2025

Available online 13 December 2025

0269-7491/© 2025 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

mechanisms that link environmental hazards to disease risk remain only partially understood.

Epigenetics, and DNA methylation in particular, has emerged as a critical area of research for elucidating how environmental factors can modulate gene expression and ultimately influence disease outcomes (Martinez-Zamudio & Ha, 2011). Differentially methylated regions (DMRs) serve as potential biomarkers for environmental exposure, bridging the gap between external pollutants and the molecular pathway implicated in disease pathogenesis (Vineis et al., 2017). With the advent of high-throughput sequencing technologies, such as reduced representation bisulfite sequencing (RRBS), it is now feasible to conduct comprehensive investigations of these epigenetic marks in large, well-characterized cohorts (Meissner et al., 2005). Nonetheless, many studies simplify the relationship by linking methylation patterns to specific heavy metals through regression analysis of a single metal concentration and methylation level (Feng et al., 2020; Ha et al., 2021; Zhang et al., 2019). Identifying biologically meaningful signals in large-scale omics data necessitates the integration of exposure data and clinical outcomes through robust analytical frameworks.

In this context, the Forensic Research via Omics Markers in environmental health vulnerable areas (FROM) cohort, along with the Korean Genome and Epidemiology Study (KoGES) cohort comprising 79,294 participants provides a valuable resource for examining the relationship between environmental exposure, DNA methylation and disease risk. Meet-in-the-middle analysis (MIMA) enhances biomarker discovery by connecting exposure, DNA methylation and clinical variables, while Mendelian randomization (MR) provides critical insights into potential causal relationships between DNA methylation and diseases (Davey Smith & Ebrahim, 2003). By applying these complementary approaches, researchers can identify epigenetic changes that correlate with environmental exposures and mediate disease risk.

In this study, we aimed to elucidate the epigenetic foundations of disease risk in populations exposed to environmental hazards, focusing on DMRs influenced by heavy metal exposure. We investigated the associations between these DMRs and 27 clinical variables as well as 23 major human diseases using a comprehensive analytic framework, thereby paving the way for improved preventive and therapeutic strategies in environmentally vulnerable populations.

2. Materials and methods

2.1. Study population

Participants included in this study were drawn from the FROM cohort, which comprises residents of both exposure and control areas. Specifically, participants were recruited from three exposure areas: Goseong, South Gyeongsang Province (Mine), an abandoned metal mine area; Janghang, South Chungcheong Province (Refinery-1), a non-ferrous metal refinery area; and Seokpo, North Gyeongsang Province (Refinery-2), a metal refinery area. These areas were selected based on national environmental health data, as they are designated environmentally vulnerable regions due to long-term exposure to industrial pollutants, such as heavy metals (Kwon et al., 2024). A total of 322 participants who had lived in these areas for at least 15 years were included: 108 from Mine, 106 from Refinery-1, and 108 from Refinery-2. Participants from Mine and Refinery-1 were recruited in 2021, while those from Refinery-2 were recruited in 2022. Within the FROM cohort, the control group consisted of participants recruited from Gimhae, South Gyeongsang Province. This area is located at least 10 km away from any major industrial facilities and is not classified as environmentally vulnerable. Control areas were selected to match the exposure areas in terms of demographic characteristics such as age and sex distribution (Kwon et al., 2024). A total of 53 individuals were recruited from Gimhae: 24 in 2021 (Control-1) and 29 in 2022 (Control-2).

An independent subset of the FROM cohort ($n = 666$) with available EPIC array-based DNA methylation data, urinary cadmium

measurements, and clinical parameters was used solely for validation of the key findings. These participants were not included in the main analyses described above.

In addition to the FROM cohort, two large-scale Korean cohorts were used to generate summary statistics for the two-sample MR analysis: the KoGES and the Gene-Environmental interaction and phenotype (GENIE) cohort (Kim et al., 2017; Lee et al., 2018). KoGES is a nationwide cohort consortium launched by the Korean government to investigate the genetic and environmental causes of major chronic diseases. The GENIE cohort, established at the Seoul National University Hospital Gangnam Center, was designed to investigate gene-environment interactions in health and disease. KoGES defined disease status through standardized self-reported physician diagnoses. GENIE determined disease status using objective clinical, imaging, and laboratory assessments obtained during routine health examinations. As both cohorts capture disease status at a single time point, all disease outcomes were analyzed as prevalent cases. These two cohorts were combined for the MR analysis, and after genotype quality control, a total of 79,294 participants were included in the final integrated dataset.

2.2. Heavy metal and clinical variable measurement

For study participants, four heavy metals (Cd, Pb, Hg, and As) were measured across six heavy metal exposure variables—urinary Cd ($\mu\text{g/L}$), creatinine-adjusted urinary Cd ($\mu\text{g/g creatinine}$), blood Cd ($\mu\text{g/L}$), blood Hg ($\mu\text{g/L}$), blood Pb ($\mu\text{g/dL}$), and urinary arsenic species (AsIII, AsV, MMA)—along with 27 clinical variables, after obtaining informed consent. For the heavy metals, blood and urine samples were taken to measure heavy metal concentrations. Inductively coupled plasma mass spectrometry (ICP-MS) was used to measure cadmium (Cd) and lead (Pb) levels in both blood and urine samples. Specifically, blood Cd and Pb were analyzed using two systems: the NexION2000B (PerkinElmer, USA) and the 7700x (Agilent Technologies, USA). Urinary Cd levels were assessed solely with the 7700x. Blood Hg concentrations were determined with an automatic mercury analyzer (MA-3000, NIC, Japan). For the analysis of various arsenic species in urine, high-performance liquid chromatography coupled with ICP-MS (HPLC-ICP-MS) was performed using the NexSAR/NexION2000 system (PerkinElmer, USA).

In addition, 27 clinical variables were assessed using blood and urine samples. Of these, 24 variables were directly measured via blood and urine tests: platelet count, hemoglobin, hematocrit, white blood cell count (WBC), red blood cell count (RBC), mean corpuscular volume (MCV), mean corpuscular hemoglobin (MCH), mean corpuscular hemoglobin concentration (MCHC), total protein, serum albumin, total calcium, inorganic phosphate, aspartate transaminase (AST), alanine transaminase (ALT), gamma-glutamyl transferase (GGT), alkaline phosphatase (ALP), lactic acid dehydrogenase (LDH), blood urea nitrogen (BUN), total cholesterol, triglyceride, high-density lipoprotein cholesterol (HDL), low-density lipoprotein cholesterol (LDL), urinary creatinine, and estimated glomerular filtration rate (eGFR). Three additional variables – N-acetyl- β -D-glucosaminidase (NAG), beta-2 microglobulin (β 2-MG), and serum creatinine – were measured using dedicated test kits. Further details on sample collection, clinical assessment, and heavy metal quantification are available in a previous publication (Kwon et al., 2024). Descriptive statistics were produced using Rex software (Lee et al., 2023). This study was conducted at Kangwon National University Hospital, and approved by its Institutional Review Board (KNUH-B-2021-06-001) in accordance with the Declaration of Helsinki. The research was conducted in strict adherence to all relevant guidelines and regulations.

2.3. Reduced representation bisulfite sequencing (RRBS)

Genomic DNA, isolated from PBMCs, was fragmented using the *MspI* restriction enzyme and then ligated to two distinct adaptors, adapter-A

and adaptor-B, using the Blunt/TA Ligase Master Mix (New England Biolabs). Each adaptor was designed with a CpG overhang at one end. Adaptor-A carries a molecular barcode and an Illumina R1 sequencing primer binding site, while Adaptor-B contained the binding site for Illumina R2 sequencing primer. To prevent bisulfite conversion of cytosines, all cytosines within the Illumina primer binding sites on the adaptors were protected. Following ligation, the DNA fragments underwent an end-fill reaction to remove the unligated adaptor strand, substituting dCTP with methyl-dCTP to preserve methylation information. Bisulfite conversion of unmethylated cytosines to uracil was then carried out using the EpiTect® Fast DNA Bisulfite Kit (Qiagen). The converted DNA fragments were subsequently amplified by two rounds of PCR to generate sequencing-ready libraries. The first amplification was performed with primers P1 and P2 under the following conditions: 10 cycles of 95 °C for 30 s, 60 °C for 30 s, and 72 °C for 30 s. A second PCR, incorporating index primers idxP1 and idxP2, followed the same thermal profile. Sequencing was performed on the Illumina HiSeq X platform, using a 151 bp paired-end protocol.

2.4. Methylation data preprocessing and quality control

Raw FASTQ files underwent adapter removal and quality trimming using Trim Galore (v0.6.6) with options optimized for RRBS data (`-rrbs, -paired`) (Krueger, 2015). We also specified adapter sequences (`-a, -a2`) and clipped the first 5 bp of read 2 (`-clip_R2 5`) to minimize sequencing bias. Subsequently, reads lacking a specific sequence were filtered out using a regular expression (`/^w{6,18}..A..A..T.A ... /`), removing those not containing this sequence. Following sequence preprocessing, paired-end alignment was performed using Bismark (v0.23.0) (Krueger & Andrews, 2011) with the GRCh38 reference genome, using default parameters. The aligned BAM files were further filtered and refined using SAMtools (v1.9) (Li et al., 2009). Specifically, we retained properly paired reads with a mapping quality of ≥ 20 , and the resulting files were coordinate-sorted for downstream analysis (Fig. S1).

Methylation count data were derived from BAM files using the DMRfinder Python package (Gaspar & Hart, 2017), with CpG extraction performed via `extract_CpG_data.py` using default parameters. Aggregated methylation tables were generated using `combine_CpG_sites.py` with the `-m 0` option to include all CpG sites regardless of total counts. All other parameters were set to default values. Regions with insufficient methylation counts were filtered out using the edgeR package (Robinson et al., 2010), and the remaining methylation count data were converted to M-values (Chen et al., 2017). Given the presence of batch effects within the data, batch correction was performed using the SVA package (Leek et al., 2012), which is suitable for adjusting batch effects in count data. The effectiveness of this correction was validated through PCA plots before and after batch correction (Fig. S2).

In this DNA methylation analysis, direct cell type measurements were not feasible; thus, we relied on estimates derived from the sorted BAM files. We employed the UXM tool (v0.1.0) within the wgbstools framework (v0.2.0) (Loyfer et al., 2023) to infer the proportions of possible cell types present in PBMC samples, including B-cells, T-cells, NK-cells, and Monocytes/Macrophages, using pat-level methylation patterns generated by the `bam2pat` function and the RRBS-specific cell-type marker set provided in the original publication.

2.5. SNP genotyping and quality control

Blood samples from individuals in the FROM cohort were collected and genotyped using the KoreanChip version 1.1 (Moon et al., 2019). To minimize batch effects, the K-medoid clustering technique was employed for genotype calling (Seo et al., 2019). Quality assurance was ensured using PLINK and ONETOOL (Song et al., 2018). Samples exhibiting sex discrepancies or call rates below 95 % were excluded. Similarly, single nucleotide polymorphisms (SNPs) with a call rate below 95 % or failing to meet the Hardy-Weinberg equilibrium test P

threshold of 1×10^{-6} were also discarded. For genotype imputation, we utilized the KRG reference panel in the Korean imputation server (Hwang et al., 2022). The Eagle (v2.4) facilitated pre-phasing, while Minimac4 was employed for the imputation process. Imputed SNPs were excluded if they had R-squared values below 0.8, were duplicates, or had more than 5 % missing genotypes, did not meet the HWE P threshold of the 1×10^{-6} , or had minor allele frequencies (MAFs) below 5 %. Primary sample QC was based on an identity-by-state threshold of 0.99. Using pruned SNPs, outliers were removed via multidimensional scaling plot (Fig. S3). SNP coordinates were harmonized to the human genome build hg38 using the rtracklayer R package (v1.46.0) (Lawrence et al., 2009). Ultimately, a set of 5,199,738 SNPs was selected for subsequent analyses.

We also utilized genotyped samples for two-sample MR from the KoGES (Kim et al., 2017) and the Gene-Environmental Interaction and Phenotype (GENIE) Study (Lee et al., 2018). Genotype quality control for these datasets followed the procedures described in a previously published study (Kim et al., 2024), but was applied to disease-related traits rather than clinical variables.

2.6. Permutational analysis of variance (PERMANOVA) analysis

To determine which variables have high explanatory power for methylation profiles, a permutational analysis of variance (PERMANOVA) analysis was conducted (Anderson, 2014). The methylation data was transformed into a distance matrix based on Euclidean distance to calculate the explanatory power of Area and each heavy metal exposure variable. Sex, age, body mass index (BMI), and smoking status were used as covariates, and P were empirically calculated through 10,000 permutations.

2.7. Meet in the middle analyses and Mendelian randomization

2.7.1. Detection of differentially methylated regions (DMRs)

We conducted linear regression using the Limma package (Ritchie et al., 2015; Smyth, 2005) to identify differentially methylated regions (DMRs) associated with the geographical areas. Because our RRBS data included three biological replicates, we employed a linear mixed model incorporating replicate information as random effects. We further adjusted for fixed effects, including age, sex, BMI, smoking status, and cell type composition percentages for B cells, T cells, NK cells and monocytes/macrophages to account for potential confounding. Formally, for each individual i and replicate j , the model was:

Methylation (beta) $_{ij}$ = area $_i$ + age $_i$ + sex $_i$ + BMI $_i$ + smoking status $_i$ + B cell(%) $_{ij}$ + T cell(%) $_{ij}$ + Natural killer cell(%) $_{ij}$ + monocytes/macrophages (%) $_{ij}$ + individual $_i$ + ϵ_{ij} , individual $_i \sim N(0, \sigma_i^2)$, $\epsilon_{ij} \sim N(0, \sigma^2)$

σ_i^2 denotes the variance of the individual-level random intercept, which accounts for within-person correlation. In the model, 'area' was considered as the main explanatory variable. We used this model to conduct three comparisons: (1) Mine vs. Control-1, (2) Refinery-1 vs. Control-1, and (3) Refinery-2 vs. Control-2.

Results from these three comparisons were aggregated using the Brown method (Brown, 1975), as implemented in the `poolR` R package (Cinar & Viechtbauer, 2022). Multiple testing correction was then applied to the combined P using Benjamini-Hochberg (BH) method (Benjamini & Hochberg, 1995). Methylation regions that achieved a 0.05 BH-adjusted significance level were selected as primary DMR candidates.

2.7.2. Meet-in-the-middle analysis (MIMA)

Primary DMR candidates identified through the Brown method were further investigated with MIMA to explore environmental exposure relevant to residents in environmentally vulnerable areas and their effects on clinical variables as described in previous studies that aimed at

identifying biomarkers (Liang et al., 2021; Vineis & Perera, 2007). Environmental exposure in vulnerable areas was assessed based on six heavy metal exposure variables, and effect of DMR was evaluated for 27 clinical variables. These included associations between heavy metal-DMR, DMR-clinical variable, DMR-area, clinical variable-area, and clinical variable-heavy metal. In the association tests involving DMRs – heavy metal-DMR, DMR-clinical variable, and DMR-area – both significance ($P < 0.05$) and direction (indicated by the sign of the beta value) were used to identify prominent DMRs. However, for the remaining two associations, clinical variable-area and clinical variable-heavy metal, only the direction was taken into account. For area-related associations, only those showing consistent directions across all three areas were retained, and where statistical significance was required (e.g., DMR-area), associations had to be significant ($P < 0.05$) in all three comparisons.

To ensure consistent and accurate interpretation of heavy metal influences in environmentally vulnerable areas, the analysis adhered to two key conditions: directionality and its pattern combinations. First, given that the heavy metal concentrations were significantly higher in the environmentally vulnerable areas than in the controls, the directionality of the DMR-area association had to match that of the DMR-heavy metal association; similarly, the directionality of clinical variable-area association had to align with the clinical variable-heavy metal association. Second, when considering three directional steps: (A) area (heavy metal) to DMR, (B) DMR to clinical variable, and (C) clinical variable to area (heavy metal), there were four acceptable combinations: (i) A, B, and C all positive; (ii) A positive, B and C negative; (iii) A negative, B positive, and C negative; and (iv) A and B negative, but C positive. Deviations from these combinations indicate inconsistent directionality, complicating biological interpretation. Cases that did not meet these two conditions were excluded from the results (Fig. S4).

All analyses were conducted separately for three groups: Male-only, Female-only, and Total group including both males and females. To facilitate better understanding, a flowchart illustrating the analysis workflow and DMR selection process is presented alongside (Fig. 1).

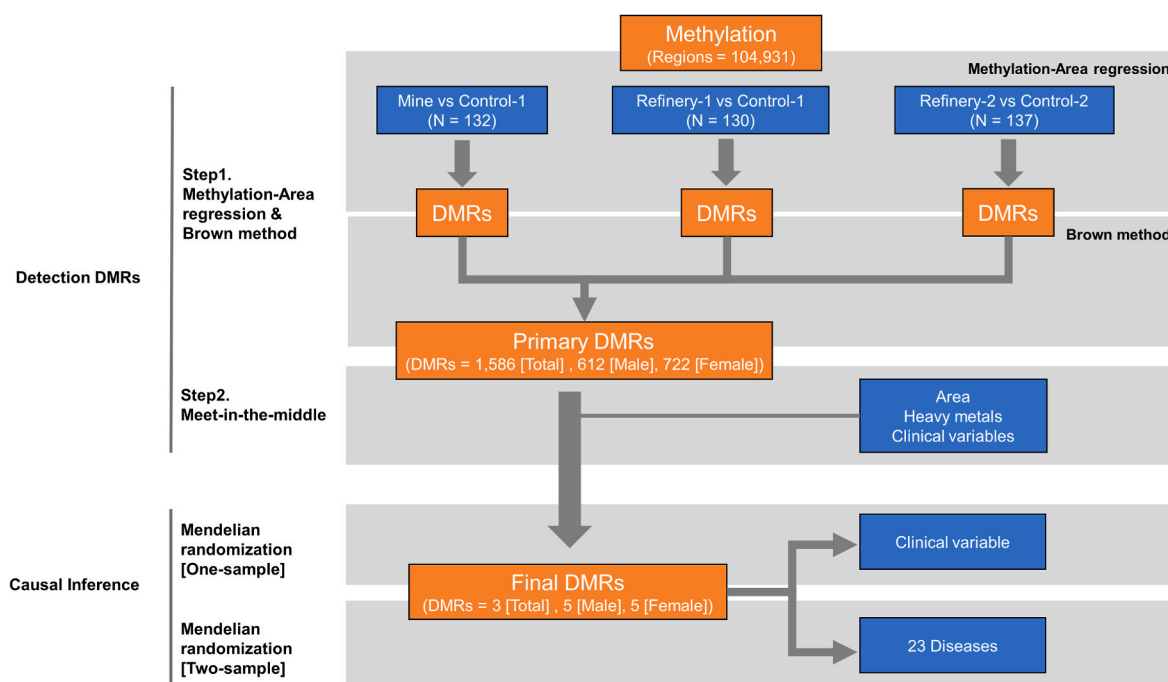
2.7.3. Sensitivity analysis with one sample Mendelian Randomization

We conducted a one-sample Mendelian Randomization (MR) analysis using DMRs identified through both the Brown method and MIMA to evaluate potential causal relationships with clinical variables. We set the DMRs as the exposure and the clinical variables as the outcome, including smoking status, age, sex and BMI as covariates. The selection of SNPs as instrumental variables (IVs) was based on the following criteria: (i) P of less than 5×10^{-5} in the methylation quantitative trait loci (meQTL) analysis, (ii) P greater than 5×10^{-5} from regressions of other covariates or outcome on the SNPs, and (iii) clumping with an r^2 threshold < 0.001 to ensure the independence of the selected SNPs (Glymour et al., 2012). All SNPs meeting these criteria were utilized as instrumental variables. Together with evaluating instrument strength using F-statistics, these selection criteria satisfy the core assumptions of MR: the relevance assumption through strong meQTL associations and LD-independent instruments; the independence assumption by excluding SNPs associated with covariates or outcomes; and the exclusion restriction assumption was assessed using the Sargan over-identification test (Emdin et al., 2017). Subsequently, a two-stage least squares regression (TSLS) analysis was performed using the ivreg package in R (version 0.6.1). Results that were statistically significant but challenging to interpret biologically were excluded.

2.7.4. Assessment of DMR effect on diseases using two-sample Mendelian Randomization

We conducted two-sample MR using the same set of DMRs used in the one-sample MR, all of which had been identified through both the Brown method and MIMA. This analysis aimed to evaluate the effect of DMRs on 23 disease outcomes, using DMR-related SNP data from this study and disease-related SNP data from two separate sources: a combined dataset from KoGES and GENIE.

The process for selecting instrumental variables was identical to that used in the one-sample MR. The two-sample MR analyses were conducted using the TwoSampleMR R package (version 0.5.6). MR-PRESSO was applied to detect and correct for horizontal pleiotropy and outlier effects that could bias causal estimates (Verbanck et al., 2018). In addition, Cochran's Q-statistics were used to evaluate heterogeneity across instruments. Diseases significant at the BH-adjusted P threshold



of 0.05 and biologically interpretable were selected.

2.8. Gene set analysis

DMRs identified through MIMA, which were significant in at least one area, were annotated using the Ensembl DB (Kinsella et al., 2011). For DMRs located outside of gene regions, annotation was performed using the name of the nearest gene. To investigate the broader biological programs associated with these genes, we performed gene set enrichment analysis using the enrichR package (Kuleshov et al., 2016) and specifically utilized the GO Biological Process 2023 database (Consortium, 2019), which provides comprehensive functional annotations of biological processes. For the GO analysis, separate analyses were conducted for Male-only, Female-only, and Total group. In each group, GO analysis was performed separately for hypermethylated regions and hypomethylated regions.

2.9. Independent validation analysis

To validate the key associations identified in the main analysis, we performed an independent validation using EPIC array-based DNA methylation data from the FROM cohort, which were not included in the primary analyses. An epigenome-wide association study (EWAS) was conducted using the Limma package in R. Because the validation dataset did not include repeated samples, random effects were not applied. Instead, a linear model was fitted for each CpG site, using the same covariate structure applied in the DMR discovery analysis—adjusting for age, sex, smoking status, body mass index (BMI), and estimated cell-type proportions. As independent variables, we included the heavy metal and clinical variable that showed significant associations with DMRs in the main analysis to assess whether these CpG–variable relationships could be replicated in the validation dataset.

3. Results

3.1. Descriptive statistics

We compared variations in heavy metal concentrations among participants from different areas (Table S1). All six heavy metal exposure variables showed higher average concentrations in the exposure areas compared to the control group. Notably, urine cadmium (uCd) and blood mercury (bHg) levels were significantly higher in Mine, while blood cadmium (bCd) showed the highest concentration in Refinery-1, and blood lead (bPb) levels were the highest in Refinery-2. For arsenic species (As^{III} , As^V , and MMA), both Control-2 and Refinery-2 showed significantly higher concentrations compared to the other areas. The Kruskal–Wallis test confirmed that the differences in heavy metal concentrations across all areas were statistically significant. For descriptive comparisons of covariates, continuous variables were assessed using the Kruskal–Wallis test, whereas categorical variables (sex and smoking status) were evaluated using the chi-square test.

3.2. DNA methylation analysis

Before performing downstream analyses, quality control and basic descriptive statistics of the DNA methylation data were summarized (Table S2). After filtering, a total of 104,931 regions were retained for analysis. The average bisulfite conversion rate across samples was 99.47 %, indicating high conversion efficiency. The mean β -value across all samples was 0.757.

3.2.1. Permutational analysis of variance (PERMANOVA) analysis

PERMANOVA analysis revealed that both area and sex remained significant after the BH adjustment (BH-adjusted $P < 0.05$) (Table S3). The significance of the area variable supports the initial hypothesis of this study, suggesting that methylation profiles differ according to the

area of residence. Notably, the strong significance observed for sex is particularly interesting, especially considering that the analysis was restricted to autosomal regions. This result suggests that sex may have a substantial impact on methylation profiles, highlighting the necessity of conducting sex-stratified analyses to better understand these effects. In addition, bHg was significant at the $\alpha = 0.05$ level, suggesting that heavy metal exposure may contribute to variations in methylation profiles.

3.2.2. Discovery of differentially methylated regions (DMRs) based on area differences

The DMR results from area comparisons were visualized using a volcano plot (Fig. 2) (Table S4–6). In the comparison between Mine and Control-1, the number of DMRs satisfying FDR < 0.05 was relatively similar for both hypo- and hypermethylation. However, in the remaining area comparisons, hypomethylated DMRs were markedly more prevalent than hypermethylated DMRs. This trend, where hypomethylated DMRs were markedly more prevalent than hypermethylated DMRs, was also observed in the separate analyses conducted for the sex-stratified analyses (Fig. S5). To further explore the chromosomal distribution of DMRs, a Circos plot was generated using the circlize R package (v0.4.15) (Gu et al., 2014). The Circos plot confirmed the dominance of hypomethylated DMRs, mirroring the trend observed in both the overall and sex-stratified analyses (Fig. S5–6, Tables S4–12).

Following the linear regression analysis using the Limma package, the results of the three comparisons (Mine vs. Control-1, Refinery-1 vs. Control-1, Refinery-2 vs. Control-2) were aggregated using the Brown method. Methylation regions that achieved a 0.05 BH-adjusted significance level were selected as primary DMR candidates. The results of these analyses revealed that a total of 1,586 differentially methylated regions (DMRs) were identified as significant in the overall comparison. When the data were stratified by sex, the male subgroup had 612 significant DMRs, while the female subgroup had 722 significant DMRs (Tables S13–15).

3.2.3. Meet-in-the-middle analysis (MIMA)

Using the primary DMR candidates identified through the Brown method, the subsequent Meet-in-the-middle analysis (MIMA) was performed. A total of 11 DMRs were finally selected from MIMA, all of which were significant across all area comparison analyses. The log2 fold change values and P from each area comparison for these DMRs, along with the meta P calculated based on these, are presented in Table S16. To provide a comprehensive summary of the relationship between beta values and P for associations involving DMRs–areas, DMRs–heavy metals, as well as DMRs–clinical variables, these associations are graphically illustrated in Fig. 3(B) and Fig. S7. Additionally, the beta values for associations involving clinical variables–heavy metals and clinical variables–areas are visualized in the form of a heatmap (Fig. S8). A Venn diagram further highlights the overlap among the 11 selected DMRs, indicating which were significant in the Total, Female-only, and Male-only analyses (Fig. 3(A)). Notably, the DMRs significant in the Male-only analysis did not overlap with any other group, whereas two overlapping DMRs were identified between the Female-only and Total analyses. This similarity between the Female-only and Total results is also evident in the heatmap shown in Fig. 3(B) and Fig. S7.

3.3. Mendelian randomization

3.3.1. Sensitivity assessment via one-sample Mendelian Randomization

We performed one-sample MR analysis to identify significant relationships between the DMRs and clinical variables detected through MIMA. Among the 11 DMRs identified as significant in MIMA, 2 were found to be significant at the 0.05 significance level in the sensitivity analysis using one-sample MR, with consistent directions of effect compared to those observed in the MIMA analysis (Table S17). The DMR located within the RARA gene region on chromosome 17, which

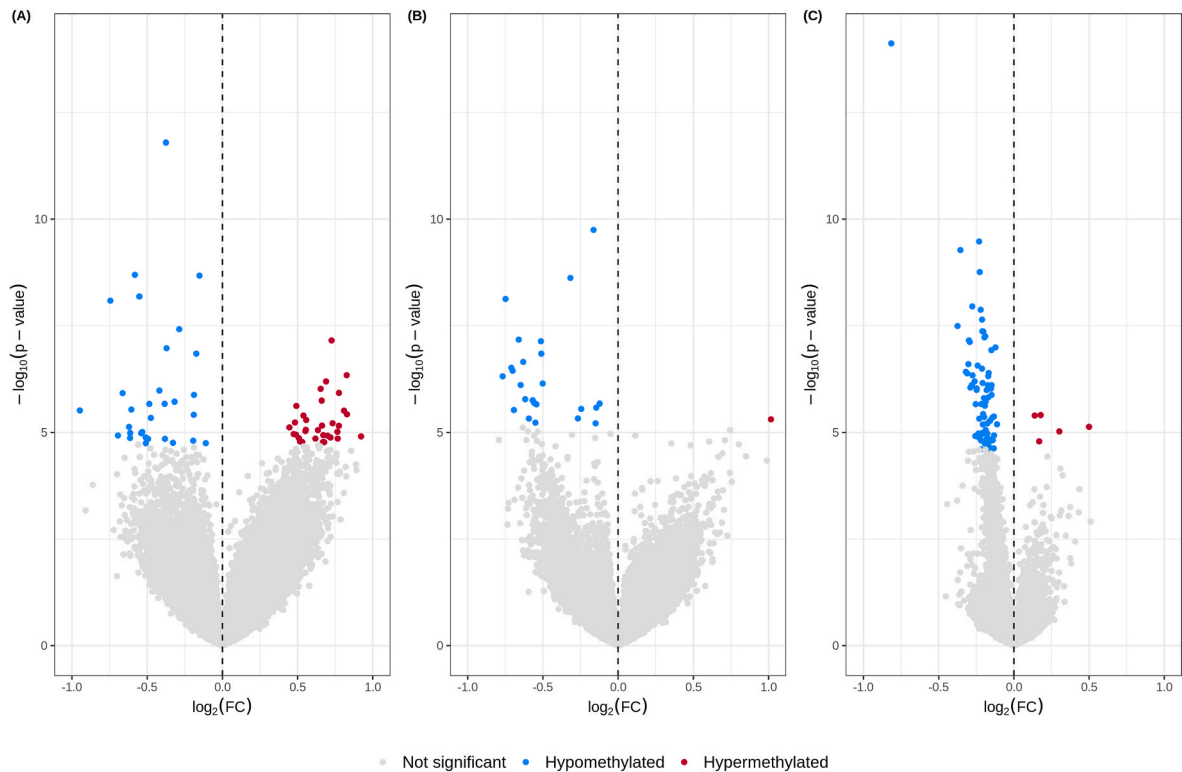


Fig. 2. Volcano plots representing the results of DMR analysis based on regions. Volcano plots (A–C) present the results of DMR analysis for the comparisons of Mine vs. Control-1, Refinery-1 vs. Control-1, and Refinery-2 vs. Control-2, based on data from the combined male and female group. DMRs meeting the BH-adjusted significance levels (0.05) are highlighted, with hypermethylated DMRs shown in red and hypomethylated DMRs in blue. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

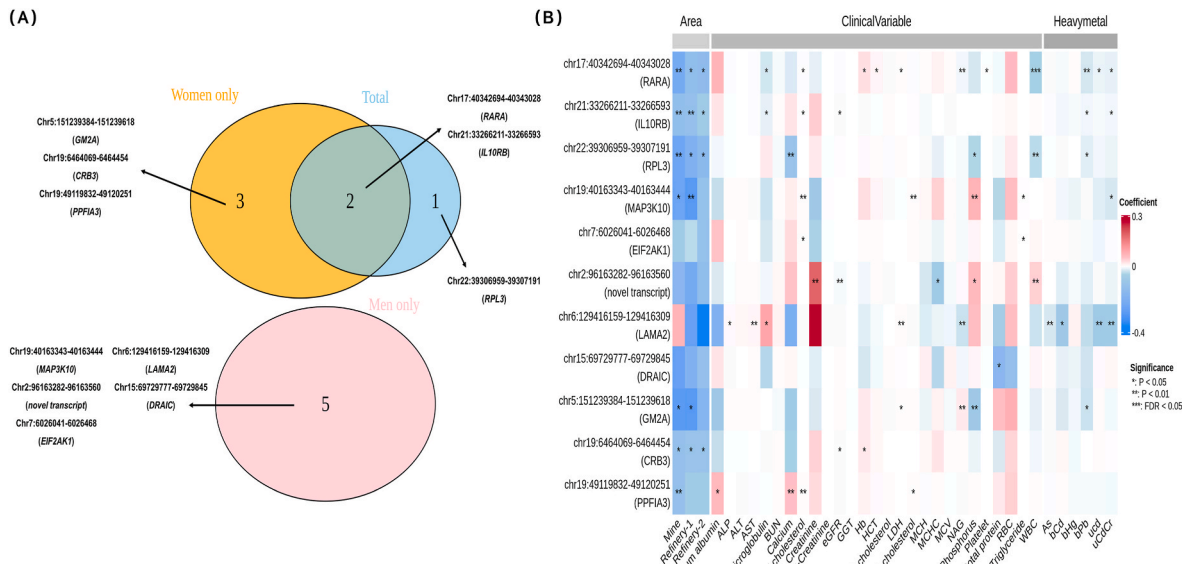


Fig. 3. Visualization of final selected DMRs identified via Meet-in-the-Middle analysis. Venn diagram (A) and heatmap (B) illustrate the final selected DMRs identified through MIMA. The Venn diagram (A) clearly illustrates which of the significant DMRs were observed in each of the Total (Male + Female), Male-only, and Female-only analyses. Heatmap (B) provides a comprehensive visualization of beta values and P for the associations between DMRs and areas, DMRs and heavy metals, as well as DMRs and clinical variables. This heatmap represents the results for the Total, allowing for a detailed comparison of methylation patterns.

exhibited a negative relationship with uCd, showed a positive relationship with hemoglobin and a negative relationship with WBC. In the Male-only dataset, the DMR located in the *EIF2AK1* gene region on chromosome 7 was found to have a negative association with β 2-MG. Detailed information on IVs used in the one-sample MR

analysis—including rsID, chromosome, position, effect sizes, standard errors, *P* from the meQTL analysis, and allele frequencies—is provided in [Tables S18–21](#), for each DMR included as an exposure.

3.3.2. Assessment of DMR effect on diseases using two-sample Mendelian Randomization

We conducted a two-sample MR analysis to evaluate the relationship between the DMRs and 23 diseases at the BH-adjusted significance level of 0.05 (Table S22). Through this analysis, the DMR located in the *IL10RB* gene region on chromosome 21 was found to be significantly associated with lung cancer at the BH-adjusted 0.05 significance level, and the complete association results for all 23 diseases—including the *IL10RB* DMR—are summarized in the updated Table S23. Notably, lung cancer was the only disease that remained statistically significant after BH correction, whereas none of the other 22 diseases reached BH-adjusted significance, indicating that the *IL10RB* DMR exhibits a lung cancer-specific association pattern. The remaining results that satisfied the threshold of $P < 0.05$ are summarized in Table S24. Detailed information on the IVs used in the two-sample MR analysis is provided in Tables S25–34. For each DMR used as the exposure, the table includes the rsID, chromosome, position, effect and non-effect alleles in both exposure and outcome datasets, effect sizes, standard errors, P from the exposure meQTL and outcome GWAS analyses, and allele frequencies.

The significant MR results, including the direction of effects, are summarized in Fig. 4(A). In addition, the key genomic and statistical findings for the *IL10RB* and *RARA* regions are presented in Table 1.

3.4. Gene set analysis

The gene set analysis identified significant GO terms (BH $P < 0.05$) in both the hypermethylated and hypomethylated regions of the Total group and the Female-only group (Fig. 4(B–C), Fig. S9). However, in the Male-only group, no GO terms met the adjusted P threshold (< 0.05), and the results have been summarized in the Supplementary section (Fig. S10).

3.5. Independent validation analysis

To validate the key associations presented in Fig. 4(A), we performed an independent replication analysis using EPIC array data from the FROM cohort, which was not included in the main analyses. The CpG site cg18506172 within the *IL10RB* DMR region showed a significant association with uCdCr, and cg05824218 within the *RARA* DMR region was significantly associated with WBC in the female subset. Both CpGs exhibited consistent effect directions with the primary results and are located precisely within the genomic coordinates of the DMRs identified in this study, providing additional support for the robustness of the main findings (Table S35).

4. Discussion

This study was driven by the hypothesis that DNA methylation patterns differ among individuals in environmentally vulnerable areas, particularly those exposed to heavy metals. Our primary objective was to identify methylation markers capable of assessing the long-term effects of such exposure. Instead of relying on single-model regressions, we employed a more comprehensive approach by integrating multiple regression outcomes through MIMA and MR. MIMA provides a strategic framework for linking environmental exposures to health outcomes by identifying intermediate biomarkers associated with both exposure levels and disease-relevant clinical traits. By incorporating both upstream (exposure) and downstream (clinical effect) biological signals, MIMA addresses a key challenge in molecular epidemiology: drawing causal inferences in complex exposure–response relationships. This approach, as emphasized by the National Institute of Environmental Health Sciences and others (Vineis & Perera, 2007), enhances causal interpretation, helps identify key environmental contributors and is

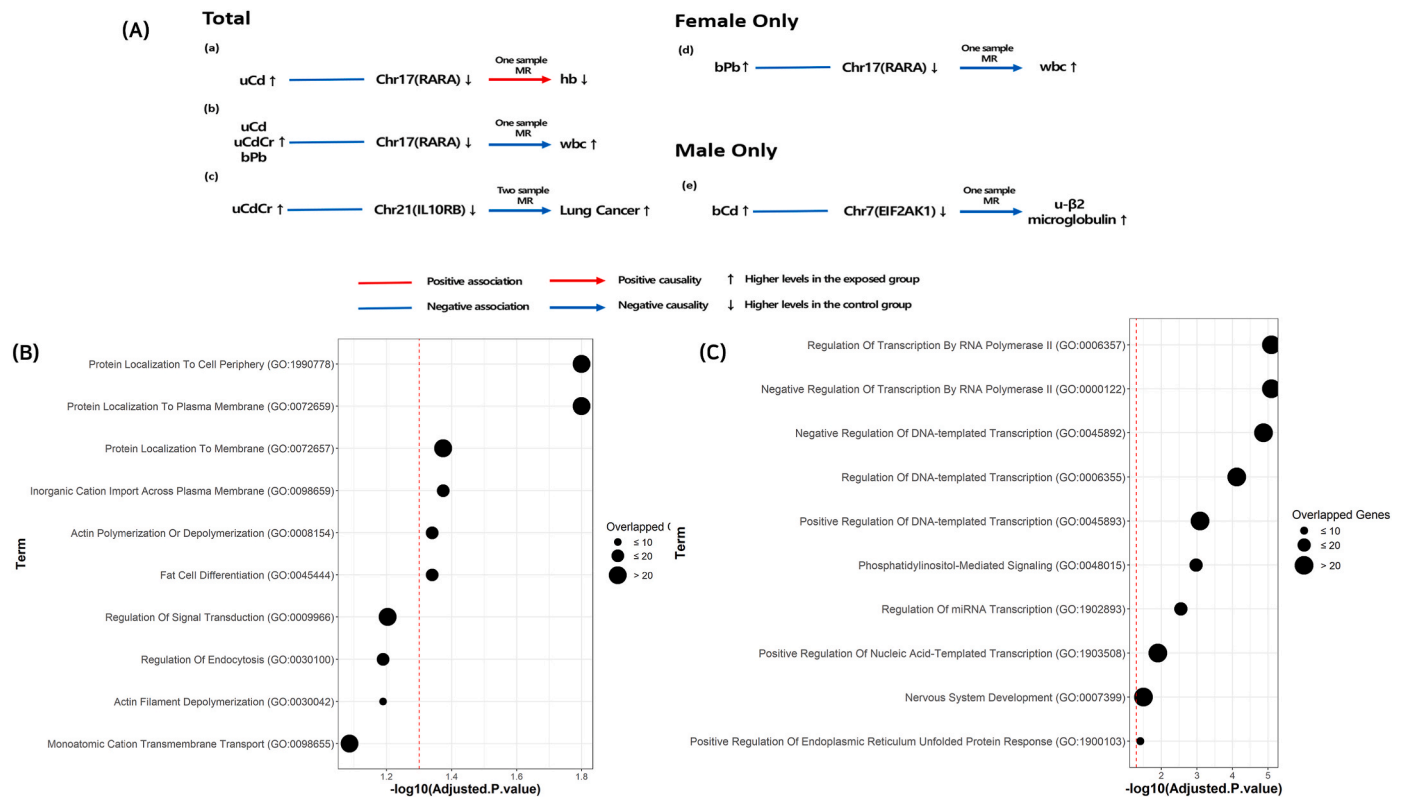


Fig. 4. Potential causal relationships between identified DMRs, clinical variables, related diseases, and significant GO terms. (A) summarizes the relationships identified through MIMA and MR analyses, showing associations between heavy metals, DMRs, clinical variables, and diseases. The figure highlights findings from the Total, Male-only, and Female-only analyses, with arrows indicating the direction and type of associations. (B) presents the gene set analysis results for hypermethylated regions in the Total group, and (C) presents the results for hypomethylated regions in the Total group. The red dashed vertical line represents the BH-adjusted P of 0.05. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Table 1

Summary of key genomic characteristics and statistical findings for the *RARA*, *IL10RB* regions. This table summarizes the genomic position and gene annotation of each DMR, as well as the direction of methylation change in exposed groups relative to controls. For each DMR, the associated heavy metal(s) and clinical outcome are listed together with the effect directions for both heavy metal–DMR and DMR–outcome relationships. The raw P-values from the MR analysis and the dataset in which each result was observed (Total, Male, or Female) are also provided.

Chr	Position	Gene name	Direction of DMR (exposure)	Related heavy metal	Effect direction with heavy metal	Related outcome	Effect direction with outcome	P (MR)	Dataset
17	40,342,694 - 40,343,028	<i>RARA</i>	Hypomethylated	uCd	Negative	Hemoglobin	Positive	4.80×10^{-02}	Total
				uCd	Negative	WBC	Negative	2.27×10^{-03}	Total
				uCdCr bPb bPb	Negative	WBC	Negative	9.07×10^{-03}	Female
				uCdCr	Negative	Lung Cancer	Negative	1.92×10^{-03}	Total
21	33,266,211 - 33,266,593	<i>IL10RB</i>	Hypomethylated	uCdCr	Negative	Lung Cancer	Negative	1.92×10^{-03}	Total

compatible with both prospective and retrospective study designs. MR complements this by using genetic variants as instrumental variables to minimize confounding and reverse causation often encountered in observational studies (Avery et al., 2022). This integrated methodology provides a more robust framework for assessing the long-term health effects of heavy metal exposure.

A key finding of this study is the potential causal relationship between the hypomethylation of the *IL10RB* gene and an increased risk of lung cancer. We posit a mechanistic model in which heavy metal exposure induces *IL10RB* hypomethylation, which in turn acts as an epigenetic ‘switch’ to promote immune evasion within the tumor microenvironment. The proposed mechanism unfolds in several steps. First, our data indicates that heavy metal exposure, particularly cadmium, is associated with *IL10RB* hypomethylation. Second, this epigenetic change likely leads to the overexpression of its protein product, the IL-10 receptor B. The IL-10 receptor is essential for mediating the function of IL-10, a potent immunosuppressive cytokine. Third, previous studies in non-small cell lung cancer have shown that signaling through IL-10 and its receptor correlates with tumor size. Furthermore, IL-10 expression on tumor-infiltrating regulatory T-cells may contribute to immune evasion by suppressing anti-tumor immunity through the PD-1/PD-L1 pathway (Vahl et al., 2017). Notably, under tumor-like metabolic stress, *IL-10R* expression can be upregulated, thereby inhibiting apoptosis in tumor cells and inducing therapeutic resistance. These findings suggest that *IL10RB* hypomethylation may reflect increased IL-10 signaling within the tumor microenvironment, potentially promoting tumor progression.

Our two-sample MR analysis further supports a causal link between *IL10RB* hypomethylation and an increased risk of lung cancer. In summary, heavy metal exposure—via *IL10RB* hypomethylation—may sensitize the tumor microenvironment to the immunosuppressive effects of IL-10, thereby weakening the anti-tumor immune response. This hypothesis is strongly corroborated by a study showing that smoking, a major source of cadmium exposure, is associated with *IL10RB* hypomethylation. In that study, the *IL10RB* locus was identified as a significant DMR that distinguishes smokers from never-smokers and was found to be hypomethylated in lung adenocarcinoma (Dong et al., 2020). Therefore, this DMR could serve as a molecular indicator for assessing damage from environmental pollutants.

Our GO enrichment analysis provides additional mechanistic context that complements the uCdCr–*IL10RB*–lung cancer axis described above. Specifically, significant GO terms such as actin polymerization/depolymerization (GO:0008154) and phosphatidylinositol-mediated signaling (GO:0048015) appear to reflect biological processes tightly linked to cadmium-induced cellular stress and are also consistent with molecular pathways that have been implicated in lung cancer. The link between cadmium and the actin cytoskeleton is well-documented; cadmium exposure is known to disrupt actin dynamics, increase F-actin formation in lung cells, and cause oxidative damage to related proteins

(Das & Al-Naemi, 2019). Thus, the epigenetic changes we observed in actin-regulating genes are likely a direct consequence of cadmium-induced cellular stress, which could create a vicious cycle that enhances the migratory and invasive capabilities of cancer cells. For GO:0048015, many annotated genes belong to the PI3K/Akt cascade, a pathway that is frequently hyperactivated in non-small cell lung cancer and plays a central role in tumorigenesis, cancer cell survival, and resistance to EGFR-TKI therapy (Cheng et al., 2014). Importantly, PI3K/Akt signaling is also a well-established downstream axis of IL-10/IL-10R activation, with *IL10RB* serving as an essential co-receptor that facilitates PI3K-dependent anti-inflammatory and immunoregulatory responses in macrophages (Antoniv & Ivashkiv, 2011; Saraiva et al., 2019). This pathway has also been reported as a molecular target of cadmium toxicity (Kulkarni et al., 2020), suggesting that cadmium exposure may influence signaling processes that are relevant to lung cancer biology. In summary, the GO analysis suggests that heavy metal exposure contributes to lung carcinogenesis through multiple molecular axes: epigenetic dysregulation, cytoskeletal remodeling, and activation of pro-survival signaling, all of which lend biological plausibility to our MR findings.

Another notable finding was a DMR located within the *RARA* gene region (chr17:40,342,694–40,343,028). The *RARA* gene, encoding the retinoic acid receptor, plays a crucial role in cell differentiation and growth. The analysis revealed that individuals in environmentally vulnerable areas with high urine cadmium concentrations exhibited hypomethylation at this DMR compared to the control group, which was associated with lower hemoglobin levels. This finding is consistent with several previous studies, reporting a negative correlation between cadmium exposure and hemoglobin concentration, as cadmium can induce anemia via hemolysis, iron deficiency, and insufficient erythropoietin production (Horiguchi et al., 2011). The link between cadmium exposure and the *RARA* gene is supported by previous research, which identified a cadmium-related DMR within the *RARA* gene region (Domingo-Reloso et al., 2020), and showed significant *RARA* hypomethylation in smokers compared to non-smokers (Lee et al., 2020). Although direct evidence linking *RARA* to hemoglobin is limited, the association is supported by its functional interaction with *RXRA*. These two proteins form a heterodimer essential for retinoic acid signaling. A recent study found that *RXRA* hypomethylation was significantly associated with increased hemoglobin levels during high-altitude acclimatization, highlighting the role of *RXRA*–*PPARA* complex in regulating hematopoiesis (Childebayeva et al., 2019). Given that *RARA* and *RXRA* are transcriptional partners, it is biologically plausible that *RARA* hypomethylation in cadmium-exposed individuals may contribute to hematological phenotypes. Furthermore, *RARA* is well-known for its involvement in hematopoietic cell differentiation, most notably in acute promyelocytic leukemia, where the *PML*–*RARA* fusion protein blocks myeloid differentiation (Yilmaz et al., 2021). This established role in hematopoiesis reinforces the biological plausibility of the observed

association between *RARA* methylation and blood cell traits in our study.

Additionally, we observed the association between *EIF2AK1* and β 2-MG levels in the context of blood cadmium exposure. Although direct evidence linking *EIF2AK1*, β 2-MG, and cadmium exposure remains sparse, our findings suggest that cadmium-induced hypomethylation at *EIF2AK1* region may contribute to elevated β 2-MG levels. Further experimental studies are needed to substantiate this hypothesis.

In this study, we conducted both sex-combined and sex-stratified analyses to comprehensively assess DNA methylation changes associated with environmental exposure. The sex-combined model, which included sex as a covariate, revealed significant DMRs across the entire cohort. In addition, sex-stratified analyses identified several DMRs that were significant only in either males or females, including genes such as *GM2A*, *CRB3*, and *EIF2AK1*. Notably, some genes such as *GM2A* and *PPFIA3* exhibited opposite directions of methylation changes between males and females. These findings suggest that certain methylation responses to environmental exposures may manifest in a sex-specific manner and could be overlooked in pooled analyses. This observation is consistent with previous studies showing that autosomal DNA methylation patterns differ between men and women, even beyond the influence of sex chromosomes (Gatev et al., 2021; Grant et al., 2022). Moreover, although most evidence for sex-dependent epigenetic changes in response to environmental exposures such as heavy metals has come from studies in animal models or developmental cohorts, such findings raise the possibility that similar mechanisms may also operate in adults (Sen et al., 2015; Sobolewski et al., 2018). These effects may be mediated by sex-specific differences in immune function, hormonal regulation, or metabolic pathways. These findings support the inclusion of both sex-combined and sex-stratified approaches in environmental epigenetic research and underscore the importance of considering sex as a biological factor in interpreting DNA methylation patterns.

Despite the significance of these results, there are several limitations. Firstly, the presence of heavy metals such as cadmium, lead, and mercury in cigarettes creates a significant correlation between smoking status and heavy metal concentrations. This correlation complicates the interpretation of elevated heavy metal levels in residents' urine and blood – whether they arise from residing in environmentally vulnerable areas or smoking habits. Our data indicated that the proportion of individuals with a smoking history was higher in the control group compared to the exposed areas (Table S1). To address this issue, we adjusted for the effect of smoking status by including it as a covariate in our regression analysis. Nevertheless, the potential for residual confounding effect remains a concern and should be considered when interpreting the findings. Secondly, while we have proposed potential links between specific DMRs and disease or clinical variables through extensive literature review, the detailed biological mechanisms underlying these associations are not yet fully elucidated. Future experimental research, grounded in various theoretical frameworks, is essential to provide more concrete evidence and deepen our understanding of these relationships. Thirdly, the cross-sectional design of this study limits the ability to infer temporality or establish causal relationships between environmental exposure, DNA methylation changes, and clinical outcomes. Since both exposure and outcomes were measured at the same time point, the possibility of reverse causation cannot be ruled out. Longitudinal studies will be necessary to establish the directionality of these relationships. Finally, although we utilized reference-based deconvolution methods to estimate cell-type proportions from bulk methylation data, we were unable to directly measure or validate cell-type compositions. This limitation may introduce bias, especially in the interpretation of findings presumed to be cell type-specific. Future studies incorporating single-cell methylation profiling or flow cytometry-based validation are warranted to address this limitation.

5. Conclusions

In conclusion, our study demonstrates that individuals residing in environmentally vulnerable areas, such as mining or refineries, show notable alterations in DNA methylation patterns, which may reflect the influence of various environmental factors including potential heavy metal exposure. These methylation changes may be associated with specific clinical or disease-related traits, as partially supported by our MIMA and MR analyses. In particular, the results of DMRs located in the *IL10RB* and *RARA* gene regions are noteworthy because they showed results consistent with previously known literature. These DMRs are also expected to serve as biomarkers for the continuous monitoring of the residents' conditions in environmentally vulnerable areas, which was the original aim of the analysis. This approach can provide invaluable insights into the long-term impact of environmental contaminants on human health, emphasizing the need for proactive measures to mitigate these effects.

CRedit authorship contribution statement

Hyeongyu Shin: Writing – review & editing, Writing – original draft, Visualization, Methodology, Formal analysis, Data curation, Conceptualization. **Chul-Hong Kim:** Resources, Investigation. **Jeeyoung Kim:** Writing – review & editing, Investigation. **JungYeon Kwon:** Investigation. **Min Heo:** Methodology, Formal analysis, Data curation. **Nam-Eun Kim:** Data curation. **Sujin Seo:** Formal analysis. **Yong Min Cho:** Investigation, Funding acquisition. **Young-Seoub Hong:** Project administration, Investigation, Funding acquisition. **Jung-Woong Kim:** Writing – review & editing, Resources, Investigation. **Woo Jin Kim:** Writing – review & editing, Project administration, Investigation, Funding acquisition. **Sungho Won:** Writing – review & editing, Supervision, Funding acquisition.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Sungho Won reports financial support was provided by Korea Ministry of Environment. Sungho Won reports financial support was provided by Korea National Institute of Health. HyeonGyu Shin has patent #10-2025-0021946 pending with RexSoft. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This study was supported by the grant from the Korea Environmental Industry and Technology Institute through the Core Technology Development Project for Environmental Disease Prevention and Management, funded by the Korea Ministry of Environment (RS-2021-KE001380), and the Korea National Institute of Health (KNIH) research project (2024ER210800). This study was conducted with bioresources from National Biobank of Korea, the Korea Disease Control and Prevention Agency, Republic of Korea (KBN-2020-101). Statistical analyses were supported by the national supercomputing center with supercomputing resources including technical support (KSC-2023-CRE-0117). Large data transfer was supported by KREONET, which is managed and operated by KISTI.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.envpol.2025.127537>.

Data availability

Data will be made available on request.

References

- Anderson, M.J., 2014. Permutational Multivariate Analysis of Variance (PERMANOVA). Wiley statsref: statistics reference online, pp. 1–15.
- Antoniv, T.T., Ivashkiv, L.B., 2011. Interleukin-10-induced gene expression and suppressive function are selectively modulated by the PI3K-Akt-GSK3 pathway. *Immunology* 132 (4), 567–577.
- Avery, C.L., Howard, A.G., Ballou, A.F., Buchanan, V.L., Collins, J.M., Downie, C.G., Engel, S.M., Graff, M., Highland, H.M., Lee, M.P., 2022. Strengthening causal inference in exposomics research: application of genetic data and methods. *Environ. Health Perspect.* 130 (5), 055001.
- Bakulski, K.M., Seo, Y.A., Hickman, R.C., Brandt, D., Vadari, H.S., Hu, H., Park, S.K., 2020. Heavy metals exposure and Alzheimer's disease and related dementias. *J. Alzheim. Dis.* 76 (4), 1215–1242.
- Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Roy. Stat. Soc. B* 57 (1), 289–300.
- Briffa, J., Sinagra, E., Blundell, R., 2020. Heavy metal pollution in the environment and their toxicological effects on humans. *Heliyon* 6 (9), e04691.
- Brown, M.B., 1975. 400: a method for combining non-independent, one-sided tests of significance. *Biometrics* 987–992.
- Chen, Y., Pal, B., Visvader, J.E., Smyth, G.K., 2017. Differential methylation analysis of reduced representation bisulfite sequencing experiments using edgeR. *F1000Research* 6.
- Cheng, H., Shcherba, M., Pendurti, G., Liang, Y., Piperdi, B., Perez-Soler, R., 2014. Targeting the PI3K/AKT/mTOR pathway: potential for lung cancer treatment. *Lung cancer management* 3 (1), 67–75.
- Childebayeva, A., Harman, T., Weinstein, J., Goodrich, J.M., Dolinoy, D.C., Day, T.A., Bigham, A.W., Brutsaert, T.D., 2019. DNA methylation changes are associated with an incremental ascent to high altitude. *Front. Genet.* 10, 1062.
- Cinar, O., Viechtbauer, W., 2022. The poolr package for combining independent and dependent p values. *J. Stat. Software* 101, 1–42.
- Consortium, G.O., 2019. The gene ontology resource: 20 years and still GOing strong. *Nucleic Acids Res.* 47 (D1), D330–D338.
- Cosselman, K.E., Navas-Acien, A., Kaufman, J.D., 2015. Environmental factors in cardiovascular disease. *Nat. Rev. Cardiol.* 12 (11), 627–642.
- Das, S.C., Al-Naemi, H.A., 2019. Cadmium Toxicity: Oxidative Stress, Inflammation and Tissue Injury.
- Davey Smith, G., Ebrahim, S., 2003. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int. J. Epidemiol.* 32 (1), 1–22.
- Domingo-Rellosa, A., Rifo-Campos, A.L., Haack, K., Rentero-Garrido, P., Ladd-Acosta, C., Fallin, D.M., Tang, W.Y., Herreros-Martinez, M., Gonzalez, J.R., Bozack, A.K., 2020. Cadmium, smoking, and human blood DNA methylation profiles in adults from the strong heart study. *Environ. Health Perspect.* 128 (6), 067005.
- Dong, Y.-M., Li, M., He, Q.-E., Tong, Y.-F., Gao, H.-Z., Zhang, Y.-Z., Wu, Y.-M., Hu, J., Zhang, N., Song, K., 2020. Epigenome-wide tobacco-related methylation signature identification and their multilevel regulatory network inference for lung adenocarcinoma. *BioMed Res. Int.* 2020 (1), 2471915.
- Emdin, C.A., Khera, A.V., Kathiresan, S., 2017. Mendelian randomization. *JAMA* 318 (19), 1925–1926.
- Feng, L., Guo, X., Li, T., Yao, C., Xia, H., Jiang, Z., Jia, J., Fang, Y., Shi, L., Lu, C.A., 2020. Novel DNA methylation biomarkers for hexavalent chromium exposure: an epigenome-wide analysis. *Epigenomics* 12 (3), 221–233.
- Gaspar, J.M., Hart, R.P., 2017. DMRfinder: efficiently identifying differentially methylated regions from MethylC-seq data. *BMC Bioinf.* 18, 1–8.
- Gatev, E., Inkster, A.M., Negri, G.L., Konwar, C., Lussier, A.A., Skakkebaek, A., Sokolowski, M.B., Gravholt, C.H., Dunn, E.C., Kober, M.S., 2021. Autosomal sex-associated co-methylated regions predict biological sex from DNA methylation. *Nucleic Acids Res.* 49 (16), 9097–9116.
- Glymour, M.M., Tchetgen Tchetgen, E.J., Robins, J.M., 2012. Credible Mendelian randomization studies: approaches for evaluating the instrumental variable assumptions. *Am. J. Epidemiol.* 175 (4), 332–339.
- Goyer, R.A., 1997. Toxic and essential metal interactions. *Annu. Rev. Nutr.* 17 (1), 37–50.
- Grant, O.A., Wang, Y., Kumari, M., Zabet, N.R., Schalkwyk, L., 2022. Characterising sex differences of autosomal DNA methylation in whole blood using the illumina EPIC array. *Clin. Epigenet.* 14 (1), 62.
- Gu, Z., Gu, L., Eils, R., Schlesner, M., Brors, B., 2014. "Circlize" Implements and Enhances Circular Visualization in R.
- Ha, F., Li, N., Long, C., Zheng, P., Hu, G., Jia, G., Wang, T., 2021. The effect of global DNA methylation on PDCD5 expression in the PBMC of occupational chromate exposed workers. *J. Occup. Environ. Med.* 63 (7), 600–608.
- Horiguchi, H., Oguma, E., Kayama, F., 2011. Cadmium induces anemia through interdependent progress of hemolysis, body iron accumulation, and insufficient erythropoietin production in rats. *Toxicol. Sci.* 122 (1), 198–210.
- Hwang, M.Y., Choi, N.-H., Won, H.H., Kim, Y.J., 2022. Analyzing the Korean reference genome with meta-imputation increased the imputation accuracy and spectrum of rare variants in the Korean population. *Front. Genet.* 13, 1008646.
- Kim, M.J., Heo, M., Kim, S.J., Song, H.E., Lee, H., Kim, N.-E., Shin, H., Do, A.R., Kim, J., Cho, Y.M., 2024. Associations between plasma metabolites and heavy metal exposure in residents of environmentally polluted areas. *Environ. Int.* 187, 108709.
- Kim, Y., Han, B.-G., Group, K., 2017. Cohort profile: the Korean genome and epidemiology study (KoGES) consortium. *Int. J. Epidemiol.* 46 (2), e20.
- Kinsella, R.J., Kähäri, A., Haider, S., Zamora, J., Proctor, G., Spudich, G., Almeida-King, J., Staines, D., Derwent, P., Kerhornou, A., 2011. Ensembl BioMart: a hub for data retrieval across taxonomic space. *Database* 2011.
- Krueger, F., 2015. Trim Galore!: a Wrapper Around Cutadapt and FastQC to Consistently Apply Adapter and Quality Trimming to FastQ Files, with Extra Functionality for RRBS Data. Babraham Institute.
- Krueger, F., Andrews, S.R., 2011. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* 27 (11), 1571–1572.
- Kuleshov, M.V., Jones, M.R., Rouillard, A.D., Fernandez, N.F., Duan, Q., Wang, Z., Koplev, S., Jenkins, S.L., Jagodnik, K.M., Lachmann, A., 2016. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* 44 (W1), W90–W97.
- Kulkarni, P., Dasgupta, P., Bhat, N.S., Hashimoto, Y., Saini, S., Shahryari, V., Yamamura, S., Shiina, M., Tanaka, Y., Dahiya, R., 2020. Role of the PI3K/Akt pathway in cadmium induced malignant transformation of normal prostate epithelial cells. *Toxicol. Appl. Pharmacol.* 409, 115308.
- Kwon, J.-Y., Kim, W.J., Cho, Y.M., Kim, B.-G., Lee, S., Rho, J.H., Eom, S.-Y., Han, D., Choi, K.-H., Lee, J.-H., 2024. Introduction to the forensic research via omics markers in environmental health vulnerable areas (FROM) study. *Epidemiology and health* 46, e2024062.
- Lawrence, M., Gentleman, R., Carey, V., 2009. rtracklayer: an R package for interfacing with genome browsers. *Bioinformatics* 25 (14), 1841.
- Lee, B., An, J., Lee, S., Won, S., 2023. Rex: R-linked EXcel add-in for statistical analysis of medical and bioinformatics data. *Genes & genomics* 45 (3), 295–305.
- Lee, C., Choe, E.K., Choi, J.M., Hwang, Y., Lee, Y., Park, B., Chung, S.J., Kwak, M.-S., Lee, J.-E., Kim, J.S., 2018. Health and prevention enhancement (H-PEACE): a retrospective, population-based cohort study conducted at the Seoul National University Hospital Gangnam Center, Korea. *BMJ Open* 8 (4), e019327.
- Lee, J.-E., Kim, H.-R., Lee, M.-H., Kim, N.-H., Wang, K.-M., Lee, S.-h., Park, O., Hong, E.-J., Youn, J.-W., Kim, Y.-Y., 2020. Smoking-related DNA methylation is differentially associated with cadmium concentration in blood. *Biochem. Genet.* 58 (4), 617–630.
- Leek, J.T., Johnson, W.E., Parker, H.S., Jaffe, A.E., Storey, J.D., 2012. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* 28 (6), 882–883.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., Subgroup, G.P.D.P., 2009. The sequence alignment/map format and SAMtools. *Bioinformatics* 25 (16), 2078–2079.
- Liang, Y., Zeng, T., Tian, J., Yan, J., Lan, Z., Chen, J., Xin, X., Lei, B., Cai, Z., 2021. Long-term environmental cadmium exposure induced serum metabolic changes related to renal and liver dysfunctions in a female cohort from Southwest China. *Sci. Total Environ.* 798, 149379.
- Loyfer, N., Magenheimer, J., Peretz, A., Cann, G., Bredno, J., Klochendler, A., Fox-Fisher, I., Shabi-Porat, S., Hecht, M., Pelet, T., 2023. A DNA methylation atlas of normal human cell types. *Nature* 613 (7943), 355–364.
- Martinez-Zamudio, R., Ha, H.C., 2011. Environmental epigenetics in metal exposure. *Epigenetics* 6 (7), 820–827.
- Meissner, A., Gnirke, A., Bell, G.W., Ramsahoye, B., Lander, E.S., Jaenisch, R., 2005. Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. *Nucleic Acids Res.* 33 (18), 5868–5877.
- Moon, S., Kim, Y.J., Han, S., Hwang, M.Y., Shin, D.M., Park, M.Y., Lu, Y., Yoon, K., Jang, H.-M., Kim, Y.K., 2019. The Korea biobank array: design and identification of coding variants associated with blood biochemical traits. *Sci. Rep.* 9 (1), 1382.
- Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., Smyth, G.K., 2015. Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 43 (7), e47.
- Robinson, M.D., McCarthy, D.J., Smyth, G.K., 2010. edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26 (1), 139–140.
- Saraiva, M., Vieira, P., O'garra, A., 2019. Biology and therapeutic potential of interleukin-10. *J. Exp. Med.* 217 (1), e20190418.
- Sen, A., Land, S., Dereski, M., Ruden, D., 2015. Early life lead exposure causes gender-specific changes in the DNA methylation profile of DNA extracted from dried blood spots. *Epigenomics* 7 (3), 379–393.
- Seo, S., Park, K., Lee, J.J., Choi, K.Y., Lee, K.H., Won, S., 2019. SNP genotype calling and quality control for multi-batch-based studies. *Genes & genomics* 41, 927–939.
- Smyth, G.K., 2005. Limma: linear models for microarray data. In: *Bioinformatics and Computational Biology Solutions Using R and Bioconductor*. Springer, pp. 397–420.
- Sobolewski, M., Varma, G., Adams, B., Anderson, D.W., Schneider, J.S., Cory-Slechta, D. A., 2018. Developmental lead exposure and prenatal stress result in sex-specific reprogramming of adult stress physiology and epigenetic profiles in brain. *Toxicol. Sci.* 163 (2), 478–489.
- Song, Y.E., Lee, S., Park, K., Elston, R.C., Yang, H.-J., Won, S., 2018. ONETOOL for the analysis of family-based big data. *Bioinformatics* 34 (16), 2851–2853.
- Tchounwou, P.B., Yedjou, C.G., Patlolla, A.K., Sutton, D.J., 2012. Heavy metal toxicity and the environment. Molecular, clinical and environmental toxicology: volume 3: environmental toxicology 133–164.
- Vahl, J.M., Friedrich, J., Mittler, S., Trump, S., Heim, L., Kachler, K., Balabko, L., Fuhrich, N., Geppert, C.-I., Trüfa, D.I., 2017. Interleukin-10-regulated tumour tolerance in non-small cell lung cancer. *Br. J. Cancer* 117 (11), 1644–1655.

- Verbanck, M., Chen, C.-Y., Neale, B., Do, R., 2018. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. *Nat. Genet.* 50 (5), 693–698.
- Vineis, P., Chadeau-Hyam, M., Gmuender, H., Gulliver, J., Herceg, Z., Kleinjans, J., Kogevinas, M., Kyrtopoulos, S., Nieuwenhuijsen, M., Phillips, D.H., 2017. The exposome in practice: design of the EXPOsOMICS project. *Int. J. Hyg Environ. Health* 220 (2), 142–151.
- Vineis, P., Perera, F., 2007. Molecular epidemiology and biomarkers in etiologic cancer research: the new in light of the old. *Cancer Epidemiol. Biomark. Prev.* 16 (10), 1954–1965.
- Yilmaz, M., Kantarjian, H., Ravandi, F., 2021. Acute promyelocytic leukemia current treatment algorithms. *Blood Cancer J.* 11 (6), 123.
- Zhang, X.x., He, Z., Feng, B., Shao, H., 2019. An epigenome-wide DNA methylation study of workers with an occupational exposure to lead. *J. Appl. Toxicol.* 39 (9), 1311–1319.