Gaseous air pollutants and DNA methylation in a methylome-wide association study of an ethnically and environmentally diverse population of U.S. adults

Holliday, Katelyn M; Gondalia, Rahul; Baldassari, Antoine; Justice, Anne E; Stewart, James D; Liao, Duanping; Yanosky, Jeff D; Jordahl, Kristina M; Bhatti, Parveen; Assimes, Themistocles L

Published in:
Environmental Research

DOI:
10.1016/j.envres.2022.113360

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2022

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license. More information can be found on the University of Groningen website: https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment.

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.
Gaseous air pollutants and DNA methylation in a methylome-wide association study of an ethnically and environmentally diverse population of U.S. adults

Katelyn M. Holliday a, b,*, Rahul Gondalia b, Antoine Baldassari b, Anne E. Justice c, James D. Stewart b, Duanping Liao d, Jeff D. Yanosky d, Kristina M. Jordahl e, Parveen Bhatti f, Themistocles L. Assimes b, James S. Pankow h, Weihua Guan j, Myriam Fornage b, k, Jan Bressler k, Kari E. North b, Karen N. Conneely b, Yun Li m, n, o, Lifang Hou p, q, Pantel S. Vokonas s, Cavin K. Ward-Caviness t, u, Rory Wilson t, u, Kathrin Wolf u, Melanie Waldenberger t, u, Josef Cyrys u, Annette Peters u, v, H. Marike Boezen w, x, Judith M. Vonk w, x, Sergi Sayols-Baixeras y, z, aa, Mikyeeong Lee bb, Andrea A. Baccarelli cc, Eric A. Whitsel b, dd

a Department of Family Medicine and Community Health, School of Medicine, Duke University, Durham, NC, USA
b Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, NC, USA
c Gesinger Health System, Danville, PA, USA
d Division of Epidemiology, Department of Public Health Sciences, Pennsylvania State University College of Medicine, Hershey, PA, USA
e Department of Epidemiology, School of Public Health, University of Washington, Seattle, WA, USA
f Cancer Control Research, BC Cancer, Vancouver, BC, Canada
g Department of Medicine, Stanford University School of Medicine, Stanford, CA, USA
h Division of Epidemiology and Community Health, University of Minnesota, Minneapolis, MN, USA
i Division of Biostatistics, University of Minnesota, Minneapolis, MN, USA
j Brown Foundation Institute of Molecular Medicine, McGovern Medical School, University of Texas Health Science Center at Houston, Houston, TX, USA
k Human Genetics Center, School of Public Health, University of Texas Health Science Center at Houston, Houston, TX, USA
l Department of Human Genetics, Emory University School of Medicine, Atlanta, GA, USA
m Department of Genetics, University of North Carolina, Chapel Hill, NC, USA
n Department of Biostatistics, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, NC, USA
o Department of Computer Science, University of North Carolina, Chapel Hill, NC, USA
p Department of Preventive Medicine, Feinberg School of Medicine, Northwestern University Chicago, Evanston, IL, USA
q Center for Population Epigenetics, Robert H. Lurie Comprehensive Cancer Center, Feinberg School of Medicine, Northwestern University Chicago, Evanston, IL, USA
r VA Normative Aging Study, VA Boston Healthcare System, Schools of Medicine and Public Health, Boston University, Boston, MA, USA
s Environmental Public Health Division, National Health and Environmental Effects Research Laboratory, 104 Mason Farm Rd, Chapel Hill, NC, 27514, USA
t Research Unit of Molecular Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, 85764, Neuherberg, Germany
u Institute of Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, 85764, Neuherberg, Germany
v Institute for Medical Information Processing, Biometry and Epidemiology, Medical Faculty, Ludwig Maximilians University, Munich, Germany
w University of Groningen, University Medical Center Groningen, Department of Epidemiology, the Netherlands
x University of Groningen, University Medical Center Groningen, GIROC Research Institute, the Netherlands
y Cardiovascular Epidemiology and Genetics Research Group, Hospital Del Mar Medical Research Institute (IMIM), Campus Del Mar, Universitat Pompeu Fabra, Barcelona, Spain
z Consorcio CIBER, M.P. Enfermedades Cardiovasculares (CIBERCV), Instituto de Salud Carlos III, Madrid, Spain
aa Molecular Epidemiology and Science for Life Laboratory, Department of Medical Sciences, Uppsala University, Uppsala, Sweden
b Epidemiology Branch, National Institute of Environmental Health Sciences, National Institutes of Health, Department of Health and Human Services, Research Triangle Park, NC, USA
c Laboratory of Environmental Epigenetics, Departments of Environmental Health Sciences and Epidemiology, Columbia University Mailman School of Public Health, New York, NY, USA
d Department of Medicine, School of Medicine, University of North Carolina, Chapel Hill, NC, USA

* Corresponding author. 2200 W. Main Street, Erwin Square Suite 600, Durham, NC, 27705.
E-mail address: katelyn.holliday@duke.edu (K.M. Holliday).

https://doi.org/10.1016/j.envres.2022.113360
Received 28 January 2022; Received in revised form 18 April 2022; Accepted 20 April 2022
Available online 30 April 2022
0013-9351/© 2022 Elsevier Inc. All rights reserved.
Epigenetic mechanisms may underlie air pollution-health outcome associations. We estimated gaseous air pollutant-DNA methylation (DNAm) associations using twelve subpopulations within Women’s Health Initiative (WHI) and Atherosclerosis Risk in Communities (ARIC) cohorts (n = 8397; mean age 61.3 years; 83% female; 46% African-American, 46% European-American, 8% Hispanic/Latino). We used geocoded participant address-specific mean ambient carbon monoxide (CO), nitrogen oxides (NO2, NOx), ozone (O3), and sulfur dioxide (SO2) concentrations estimated over the 2-, 7-, 28-, and 365-day periods before collection of blood samples used to generate Illumina 450 k array leukocyte DNAm measurements. We estimated methylene-wide, subpopulation- and race/ethnicity-stratified pollutant-DNAm associations in multi-level, linear mixed-effects models adjusted for sociodemographic, behavioral, meteorological, and technical covariates. We combined stratum-specific estimates in inverse variance-weighted meta-analyses and characterized significant associations (false discovery rate; FDR<0.05) at Cytosine-phosphate-Guanine (CpG) sites without among-strata heterogeneity (P( Cochran’s q > 0.05). We attempted replication in the Cooperative Health Research in Region of Augsburg (KORA) study and Normative Aging Study (NAS). We observed a −0.3 (95% CI: −0.4, −0.2) unit decrease in percent DNAm per interquartile range (IQR, 7.3 ppb) increase in 28-day mean NO2 concentration at cg01885635 (chromosome 3; regulatory region 290 bp upstream from ZNF621; FDR = 0.03). At intragenic sites cg21849932 (chromosome 20; LIME1; intron 3) and cg05353869 (chromosome 11; KLHL35; exon 2), we observed a −0.3 (95% CI: −0.4, −0.2) unit decrease (FDR = 0.04) and a 1.2 (95% CI: 0.7, 1.7) unit increase (FDR = 0.04), respectively, in percent DNAm per IQR (17.6 ppb) increase in 7-day mean ozone concentration. Results were not fully replicated in KORA and NAS. We identified three CpG sites potentially susceptible to gaseous air pollution-induced DNAm changes near genes relevant for cardiovascular and lung disease. Further harmonized investigations with a range of gaseous pollutants and averaging durations are needed to determine the effect of gaseous air pollutants on DNA methylation and ultimately gene expression.

1. Introduction

DNA methylation (DNAm) at Cytosine-phosphate-Guanine (CpG) sites is a physiological process that can impact gene expression. Although DNAm is stable over time and can be heritable, it can also be affected by exposure to environmental pollutants. Many environmentally induced changes in DNAm may initially be small, but they can accumulate over time (Baccarelli and Bollati, 2009). Further, experimental studies in humans (Bellavia et al., 2013) and repeated-measures occupational studies (Tarantini et al., 2009; Fan et al., 2014) have demonstrated changes in DNAm patterns after short-term exposures to air pollutants. In particular, exposure to air pollution has been associated with atypical global methylation of DNA in blood samples (De Prins et al., 2013; Madrigno et al., 2011) and atypical DNAm near specific candidate genes relevant for cardiovascular and respiratory health (Bind et al., 2014; Sofer et al., 2013; Chi et al., 2016).

Since DNAm has recently been proposed as an epigenetic mechanism by which air pollution influences health (Baccarelli and Ghosh, 2012), identifying epigenetic changes related to gaseous air pollutant exposures may inform our understanding of relevant mechanistic pathways. For example, exposure to US criteria gaseous air pollutants, including...
carbon monoxide \((\text{CO})\), oxides of nitrogen \((\text{NO}_2, \text{NO}_x)\), ozone \((\text{O}_3)\), and sulfur dioxide \((\text{SO}_2)\) may be pathophysiologically linked to cardiovascular disease (CVD) through inflammatory, oxidative, and autonomic mechanisms (Franklin et al., 2015) susceptible to DNAm-induced changes. DNAm in blood cells has also been implicated as an effect modifier of air pollution-health associations (Lepeule et al., 2014; Fu et al., 2012; Bind et al., 2012). Thus, identifying CpGs differentially methylated in response to gaseous air pollutant exposure may provide insight into the mechanisms linking gaseous air pollutants to CVD and other health outcomes.

The literature examining associations between air pollution and methylome-wide DNAm has largely focused on long-term exposure to particulate matter (PM) and NO\(_2\) air pollution (Panni et al., 2016; de F. C. Lichtenfels et al., 2018; Sayols-Baixeras et al., 2019; Plusquin et al., 2017; Lee et al., 2019; Gondalia et al., 2019); however, other pollutants and varying durations of exposure may affect methylation patterns relevant for health outcomes like CVD (Franklin et al., 2015). Existing DNAm-air pollution research is also geographically and sociodemographically limited, with much of it completed in Boston, Massachusetts and in Europe (Bind et al., 2014; Bind et al., 2012; Panni et al., 2016; de F. C. Lichtenfels et al., 2018; Sayols-Baixeras et al., 2019; Plusquin et al., 2017) among individuals of European descent. To address these gaps in research, we leveraged data from two multi-ethnic and geographically diverse populations, the Women’s Health Initiative (WHI) and the Atherosclerosis Risk in Communities Study (ARIC), to examine methylome-wide associations with short- and long-term CO, NO\(_2\), NO\(_x\), O\(_3\), and SO\(_2\) exposure.

2. Material and methods

2.1. Study design

We conducted subpopulation- and race/ethnicity-stratified, methylome-wide discovery analyses of gaseous pollutant-DNAm associations within WHI and ARIC subpopulations \((N = 8397)\) and completed replication analyses within the Cooperative Health Research in the Region of Augsburg study (KORA; \(N = 2141\)) and the Normative Aging Study (NAS; \(N = 773\)).

2.2. Study populations

The WHI is a large, prospective study of postmenopausal women enrolled between 1993 and 1998 at forty clinical centers in the US (The Women's Health Initiative Study Group, 1998). Women were enrolled in either the clinical trials (CT; \(N = 68,132\)) or observational study (OS; \(N = 93,676\)) cohorts. In the present analyses, we included three ancillary study subpopulations of WHI participants with available genome-wide DNAm assessed in peripheral blood leukocytes (Fig. 1):

1) Epigenetic Mechanisms of Particulate Matter-Mediated Cardiovascular Disease Risk (EMPC) was based on an exam site- and race/ethnicity-stratified, minority oversample of WHI CT participants randomly selected from the screening, third annual, and sixth annual follow-up visits with contemporaneous core analyte data and an address in the contiguous 48 US states \((N = 2200)\). A small proportion of EMPC

Fig. 1. Diagram of discovery study populations and data analysis flow (AA = African American, ARIC = Atherosclerosis Risk in Communities, AS311 = Ancillary Study 311, BAA23 = Broad Agency Award 23, CHD = Coronary Heart Disease, CpG = Cytosine-phosphate-Guanine site, CT = Clinical Trials, DNAm = DNA Methylation, EA = European American, EMPC = Epigenetic Mechanisms of Particulate Matter-Mediated Cardiovascular Disease Risk, MN = Minnesota, MS = Mississippi, N = Number, NC = North Carolina, OS = Observational Studies, WHI = Women’s Health Initiative).
participants had DNAm re-measured at the 3rd or 6th annual visits (N = 200) and during the Long Life Study 14–19 years after the screening visit (N = 43).

2) Broady Agency Award 23 (BAA23) is a case-control study of incident coronary heart disease among WHI CT (N = 1546) and OS (N = 442) participants. DNAm was assessed in peripheral blood leukocytes at the screening visit, i.e. before disease diagnosis.

3) Ancillary Study 311 (AS311) is a matched case-control study of incident bladder cancer among WHI CT (N = 405) and OS (N = 455) participants (Jordahl et al., 2018). Bladder cancer cases were matched to controls by enrollment year, number of follow-up days, age at diagnosis (=±2 years), and DNA extraction method. DNAm was assessed in peripheral blood leukocytes at the screening visit, i.e. before disease diagnosis.

In all WHI subpopulations, analyses were restricted to racial/ethnic groups including ≥100 individuals.

ARIC is a community-based, prospective cohort study of atherosclerosis in four US communities that began in 1987–1989 (The ARIC Investigators, 1989). Two sub-studies generated DNAm data for ARIC participants (Fig. 1).

4) All African American participants from Forsyth County, NC and Jackson, MS at the second or third visit (N = 2751).

5) European Americans from Forsyth County, NC or Minneapolis, MN suburbs who were part of a randomly sampled ancillary study of brain magnetic resonance imaging at the third visit (N = 1139). Peripheral blood leukocyte DNAm was assayed at the second (1990–1992) or third (1993–1995) follow-up visits.

Formal replication efforts in the KORA population-based cohort from the region of Augsburg, Southern Germany included data from S3 and S4 participants of European ancestry at follow-ups F3 (N = 459; years = 2004–2005) and F4 (N = 1682; years 2006–2008) (Holle et al., 2005; Wichmann et al., 2005). Similar replication efforts in the NAS cohort of community-dwelling elderly male veterans of European ancestry from Boston, MA involved participants at up to four follow-up exams (N = 773 participants at up to 1522 visits; years = 1999–2009; mean age 73 [range: 55–92]) (Bell et al., 1996). These participants were initially recruited by the US Veterans Administration in 1963 as healthy, residentially-stable participants aged 25–75 to study the healthy aging process.

This analysis was approved by the IRB at the University of North Carolina at Chapel Hill and all participants provided written informed consent at their local WHI or ARIC clinic with de-identified data provided through data sharing agreements.

2.3. Epigenome-wide DNA methylation

Participants provided fasting blood from which peripheral blood leukocytes were isolated and DNA was extracted. DNAm at up to 485,577 CpG sites was measured using the Illumina 450 K Infinium Methylation BeadChip (Illumina Inc.; San Diego, CA) and quantitatively represented by the methylation β value, the proportion of DNA methylated at each CpG site (=methylated signal divided by the sum of methylated and unmethylated signals). DNAm was Beta-Mixture Quantile (BMQ)-normalized to correct for differences otherwise attributable to the design of Type I and II probes (Wu et al., 2014) and batch-corrected using random intercepts for plate and chip and a fixed effect for row in all subpopulations except for WHI EMPC, which used empirical Bayes methods (Johnson et al., 2007) to adjust for plate.

2.4. Air pollution concentrations

Primary exposures were gaseous air pollutants (in ppb) regulated by the U.S. Environmental Protection Agency (EPA) under the Clean Air Act according to National Ambient Air Quality Standards (NAAQS): CO, NOx, NO2, O3, and SO2. Measurements encompassed a range of short- and long-term concentration averages to provide insight into potential DNAm-mediated acute and chronic effects of gaseous pollutants on various measures of CVD. National-scale, log-normal ordinary kriging (Liao et al., 2006) was used to estimate geocoded participant address-specific mean pollutant concentrations averaged over the 2-, 7-, 28-, and 365-day periods before the blood draw used for DNAm quantification. The models were based on daily (or for O3, maximum 8-h) mean concentrations from the US Environmental Protection Agency Air Quality System. Cross-validation statistics for these models are available in Supplemental Table 1.

2.5. Covariates

Covariates included:

1) Study design factors: WHI clinical trial arm, case-control status, matching factors for case-control studies, study center

2) Sociodemographic characteristics: age at blood draw, sex, race (analysis stratification variable), education (<high school vs. ≥ high school), neighborhood socioeconomic status (Diez Roux et al., 2001)

3) Behavioral and health characteristics: smoking status and alcohol use (never, former, current), physical activity (total energy expenditure in MET-hours/week), body mass index (BMI; weight in kilograms/height in meters)2

4) Meteorological covariates: dew point in °Celsius, temperature in °Celsius, barometric pressure in kPa (all expressed as means estimated over the air pollution exposure averaging duration from National Climatic Data Center stations within 50 km of the participant’s home)

5) Season of blood draw: spring, summer, fall, winter

6) Estimated leukocyte proportions: CD8+ T cell, CD4+ T cell, B cell natural killer cell, monocyte, granulocyte (imputed using the Houseman method and reference sample data (Houseman et al., 2012) for most populations). Among ARIC African American participants, 175 had complete data on measured leukocyte proportions (lymphocytes, monocytes, neutrophils, eosinophils and basophils). These 175 measures were used as a reference sample to impute cell type proportions in the remainder using the Houseman method (Demerath et al., 2015).

7) Ancestry principal components: PCs (Price et al., 2006) as available (unavailable for AS311 participants)

2.6. Missing data

Eight individuals did not have DNAm data and were excluded. To impute missing covariate and exposure data for the remaining individuals (Supplemental Tables 2 and 3), we generated ten imputed datasets for each study subpopulation with non-missing covariates and exposures using a fully conditional specification multiple imputation method implemented in SAS 9.4 (Cary, NC) (Fig. 1). We applied logistic and discriminant functions to impute binary and categorical variables and predictive means matching using the k-nearest neighbor method (k = 5) to impute continuous variables (see Supplemental Table 4 for a complete list of variables included in the imputation).

2.7. Statistical analyses

To estimate gaseous air pollutant-DNAm associations, we used study subpopulation- and race/ethnicity-stratified linear mixed models (LMM) with a harmonized covariate adjustment strategy and pooled estimates across all ten imputed datasets (Fig. 1) (van Buuren and Groothuis-Oudshoorn, 2011). The specification of random mixed model components varied by study subpopulation, depending on study design and repeated measure availability. Specifically, analyses of WHI EMPC’s
repeated DNA methylation measures involved a three-level LMM while those of WHI BAA23 and WHI AS311 involved a two-level, cross-sectional LMM, and those of ARIC involved a one-level, cross-sectional LMM. In WHI EMPC, BAA23, and AS311, the models included a center-level random intercept and slope to account for within-site correlations whereas models in ARIC controlled for two centers as a fixed effect. See Supplemental Table 4 for the complete stratum-specific adjustment strategy. We fit the stratified models with maximum likelihood estimation supplemented in the Julia v0.6 MixedModels package.

We then used Cochran’s Q test statistics to examine homogeneity of the pooled study- and race/ethnicity-specific gaseous air pollutant-DNA methylation associations. After excluding heterogeneous associations (Cochran’s Q < 0.05), we combined stratum-specific estimates in fixed-effects, inverse variance-weighted meta-analyses and ranked results by FDR adjusted p-values (Fig. 1). We used FDR < 0.05 to identify methylome-wide significant associations and reported a ranked list of associations. We also focused on associations for CpG sites ranked in the top 5 of this list with associations for ≥2 pollutants or averaging durations. We reported estimates of the unit increase or decrease in percent DNA methylation (95% confidence interval (CI)) per interquartile range (IQR) increase in gaseous air pollution concentration.

2.8. Replication association analyses

KORA and NAS collaborators carried out harmonized gaseous pollutant-, averaging duration-, and CpG site-specific associations identified as significant in the discovery analyses. For replication analyses, we used a Bonferroni significance threshold [corrected for the number of CpG sites (N) carried from discovery into replication (P < 0.05 ÷ N)] and considered sites meeting the significance threshold with directionally consistent estimates between discovery and replication analyses as replicated.

2.9. Quality control

We followed established protocols for excluding low quality participant samples and CpG sites (Supplemental Table 5). To conform with LMM distribution assumptions and filter out variation due to single participant samples and CpG sites (Supplemental Table 5), we used FDR < 0.05, we combined stratum-specific estimates in fixed-effects, inverse variance-weighted meta-analyses and ranked results by FDR adjusted p-values (Fig. 1). We used FDR < 0.05 to identify methylome-wide significant associations and reported a ranked list of associations. We also focused on associations for CpG sites ranked in the top 5 of this list with associations for ≥2 pollutants or averaging durations. We reported estimates of the unit increase or decrease in percent DNA methylation (95% confidence interval (CI)) per interquartile range (IQR) increase in gaseous air pollution concentration.

2.10. Functional annotation

We functionally annotated significant CpG sites using the UCSC Genome Browser (Feb. 2009 GRC37/hg19) (Kuhn et al., 2013) with data from the Encyclopedia of DNA elements (ENCODE) (Rosenbloom et al., 2012) and Roadmap Epigenomics Project (Bernstein et al., 2010).

3. Results

The study included data from twelve subpopulations in WHI and ARIC (Table 1). Forty-six percent of the 8397 participants were African American, 46% were European American, and 8% were Hispanic/Latino (from WHI only). Seventeen percent of participants were male (from ARIC only). The overall mean (SD) age was 61.3 (7.4) years with subpopulation-specific mean ages ranging from 56.6 to 67.8 years. Air pollution exposures varied among subpopulations (Table 1, Supplemental Table 3). For example, the overall mean (SD) 28-day NO2 concentration was 18.0 (7.3) ppb but ranged from 14.9 (3.6) ppb in ARIC African Americans to 21.5 (9.4) ppb in WHI BAA23 CT African Americans (Table 1).

Discovery analyses showed minimal evidence of inflation across pollutants and averaging durations [median (range) λ = 0.98 (0.90–1.25); Fig. 2], and identified three CpG sites significant at FDR < 0.05 (Table 2, Fig. 3). These sites demonstrated consistency in the direction and magnitude of the change in %DNA methylation across subpopulations (Fig. 4) and across additional exposure averaging durations for the same pollutant (see Supplemental Table 6 for ranked list).

At the intergenic site cg01885635 (chromosome 3; 290 bp upstream from ZNF621), we observed a −0.3 (95% CI: −0.4, −0.2) unit decrease in percent DNA methylation per IQR (7.3 ppb) increase in 28-day mean NO2 concentrations (FDR = 0.03; Table 2, Fig. 4A). At intragenic sites cg21849932 (chromosome 20; LIME1; intron 3) and cg05353869 (chromosome 11; KLHL35; exon 2), we also observed a −0.3 (95% CI: −0.4, −0.2) unit decrease (FDR = 0.04, Table 2, Fig. 4B) and a 1.2 (95% CI: 0.7, 1.7) unit increase (FDR = 0.04, Table 2, Fig. 4C), respectively, in percent DNA methylation per IQR (17.6 ppb) increase in 7-day mean O3.

Table 1

<table>
<thead>
<tr>
<th>Study</th>
<th>Race/Ethnicity</th>
<th>N</th>
<th>% Female</th>
<th>Age (years) Mean (SD)</th>
<th>N CpGs Maximum</th>
<th>28-day NO2 (ppb) Mean (IQR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARIC</td>
<td>AA</td>
<td>2664</td>
<td>63</td>
<td>56.6 (5.9)</td>
<td>463,431</td>
<td>14.9 (3.6)</td>
</tr>
<tr>
<td></td>
<td>EA</td>
<td>1100</td>
<td>58</td>
<td>59.9 (5.4)</td>
<td>462,543</td>
<td>18.6 (3.5)</td>
</tr>
<tr>
<td>WHI</td>
<td>AS311</td>
<td>CT</td>
<td>351</td>
<td>64.7 (7.1)</td>
<td>461,136</td>
<td>20.0 (7.7)</td>
</tr>
<tr>
<td></td>
<td>EA</td>
<td>395</td>
<td>100</td>
<td>66.2 (6.9)</td>
<td>461,136</td>
<td>19.9 (8.5)</td>
</tr>
<tr>
<td></td>
<td>OA</td>
<td>100</td>
<td>371</td>
<td>61.8 (6.3)</td>
<td>461,014</td>
<td>21.5 (9.4)</td>
</tr>
<tr>
<td></td>
<td>OA</td>
<td>100</td>
<td>926</td>
<td>67.8 (6.2)</td>
<td>461,014</td>
<td>19.4 (8.3)</td>
</tr>
<tr>
<td></td>
<td>HL</td>
<td>100</td>
<td>220</td>
<td>60.7 (6.4)</td>
<td>461,014</td>
<td>18.8 (11.8)</td>
</tr>
<tr>
<td></td>
<td>OS</td>
<td>100</td>
<td>529</td>
<td>62.8 (6.8)</td>
<td>461,014</td>
<td>21.4 (9.4)</td>
</tr>
<tr>
<td></td>
<td>HL</td>
<td>100</td>
<td>174</td>
<td>62.8 (7.3)</td>
<td>461,014</td>
<td>19.8 (10.7)</td>
</tr>
<tr>
<td></td>
<td>OA</td>
<td>100</td>
<td>1072</td>
<td>64.6 (7.1)</td>
<td>461,014</td>
<td>18.6 (8.0)</td>
</tr>
<tr>
<td></td>
<td>OA</td>
<td>100</td>
<td>312</td>
<td>61.5 (6.1)</td>
<td>461,014</td>
<td>18.1 (11.6)</td>
</tr>
<tr>
<td>Overall</td>
<td>AA: 46%</td>
<td>8397</td>
<td>83</td>
<td>61.3 (7.4)</td>
<td>463,916</td>
<td>18.0 (7.3)</td>
</tr>
<tr>
<td></td>
<td>EA: 46%</td>
<td></td>
<td></td>
<td></td>
<td>463,054</td>
<td></td>
</tr>
<tr>
<td></td>
<td>HL: 8%</td>
<td></td>
<td></td>
<td></td>
<td>461,346</td>
<td></td>
</tr>
</tbody>
</table>

Abbreviations: AA, African American; ARIC, Atherosclerosis Risk in Communities Study; AS311, Ancillary Study 311; BAA23, Broad Agency Award 23; CpG, cytosine-phosphate-guanine site; CT, clinical trial; EA, European American; EMPC, Epigenetic Mechanisms of Particulate Matter Mediated Cardiovascular Disease Risk; HL, Hispanic Latino; IQR, interquartile range; N, number; NO2, nitrogen dioxide; OS, observational study; ppb, parts per billion; SD, standard deviation; WHI, Women’s Health Initiative.

* Median imputed mean (IQR) from 10 imputed datasets.

* At the 1st visit. DNA methylation data were also available among 185 EMPC participants in visit years 3 or 6 and 43 EMPC participants in study years 14–19.
Fig. 2. Quantile-quantile (QQ) plot of observed vs. expected -log_{10} p-value for each CpG site from multi-ethnic, fixed-effects meta-analyses of 2-, 7-, 28-, and 365-day mean gaseous air pollutant concentrations.

Table 2
Gaseous air pollutant-DNAm associations significant at FDR < 0.05 among ARIC and WHI Subpopulations.

<table>
<thead>
<tr>
<th>Top Rank</th>
<th>Chr</th>
<th>CpG Site</th>
<th>Position</th>
<th>Gene</th>
<th>Pollutant</th>
<th>Averaging Duration (days)</th>
<th>Change in %DNAm (95% CI)</th>
<th>FDR</th>
<th>N</th>
<th>P_{Cochran}'s Q</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3</td>
<td>cg01885635</td>
<td>40,566,085; intergenic</td>
<td>ZNF621</td>
<td>NO_{2}</td>
<td>28</td>
<td>-0.3 (-0.4, -0.2)</td>
<td>0.03^b</td>
<td>8613</td>
<td>0.9</td>
</tr>
<tr>
<td>2</td>
<td>20</td>
<td>cg21849932</td>
<td>62,369,462; intron 3</td>
<td>LIME1</td>
<td>O_{3}</td>
<td>7</td>
<td>-0.3 (-0.4, -0.2)</td>
<td>0.04^b</td>
<td>8623</td>
<td>0.5</td>
</tr>
<tr>
<td>3</td>
<td>11</td>
<td>cg05353869</td>
<td>75,139,544; exon 2</td>
<td>KLIHL35</td>
<td>O_{3}</td>
<td>7</td>
<td>1.2 (0.7, 1.7)</td>
<td>0.04^b</td>
<td>8621</td>
<td>0.8</td>
</tr>
</tbody>
</table>

Abbreviations: ARIC, Atherosclerosis Risk in Communities; CpG, Cytosine-phosphate-Guanine; Chr, chromosome; DNAm, DNA methylation; FDR, false discovery rate; N, number; NO_{2}, nitrogen dioxide; O_{3}, ozone; WHI, Women’s Health Initiative.

^a Per IQR (ppb) increase in gaseous air pollutant concentration (NO_{2} 28-day = 7.3; O_{3} 7-day = 17.6).

^b FDR<0.05.

Fig. 3. Manhattan plot of -log_{10} p-value vs. chromosomal position of each CpG site from multi-ethnic, fixed-effects meta-analyses of 2-, 7-, 28-, and 365-day mean gaseous air pollutant concentrations. Sites meeting the methylome-wide significance thresholds (FDR < 0.05) were circled instead of plotting all 20 horizontal reference lines identifying the significance thresholds calculated separately for each combination of 5 pollutants and 4 averaging durations.
concentration.

The highest ranked pollutant-DNAm associations also had suggestive associations at other pollutant averaging durations (Supplemental Table 6). Additionally, at a third intragenic site, cg15008743 (chromosome 19; ZNF83), we note five non-FDR significant associations with 28- and 365-day mean NO2 and 2-, 7-, and 28-day mean NO2 concentrations (Supplemental Table 6; Supplemental Figure 1).

The NAS, KORA F3, and KORA F4 populations have been described previously (Panni et al., 2016; Ward-Caviness et al., 2016) and basic sociodemographic characteristics are listed in Supplemental Table 7. The 28-day mean NO2-DNAm association at cg01885635 meta-analyzed across NAS, KORA F3, and KORA F4 was significant, corresponding to a 0.5 (95% CI: 0.3, 0.7) unit increase in percent DNAm per IQR increase. However, this estimate was directionally inconsistent with the −0.3 (95% CI: −0.4, −0.2) unit decrease in percent DNAm observed in the meta-analyzed WHI and ARIC populations. Examination of sex as a potential explanation for the observed difference in direction of association among the predominately female WHI and ARIC discovery populations and predominately male NAS and KORA replication populations did not explain the differences, although the discovery population was limited to 17% male, thereby limiting power. KORA could not supply replication analyses for the remaining sites due to unavailing O3 exposure data for this cohort. In NAS, the association between DNAm at cg21849932 and 7-day O3 exposure data for this cohort. In NAS, the association between DNAm at cg21849932 and 7-day O3 exposure data for this cohort. In NAS, the association between DNAm at cg21849932 and 7-day O3 exposure data for this cohort. In NAS, the association between DNAm at cg21849932 (P Cochran’s q = 0.9), B) per IQR (17.6 ppb) increase in 7-day O3 at cg21849932 (P Cochran’s q = 0.5), C) per IQR (17.6 ppb) increase in 7-day O3 at cg05353869 (P Cochran’s q = 0.8).

Fig. 4. Unit change in %DNAm (95%CI) A) per IQR (7.3 ppb) increase in 28-day NO2 at cg01885635 (P Cochran’s q = 0.9), B) per IQR (17.6 ppb) increase in 7-day O3 at cg21849932 (P Cochran’s q = 0.5), C) per IQR (17.6 ppb) increase in 7-day O3 at cg05353869 (P Cochran’s q = 0.8).

4. Discussion

This methylome-wide study of demographically and geographically diverse participants identified medium duration NO2- and O3-associated changes in DNAm at three CpG sites. The results of these multi-ethnic meta-analyses were significant, precisely estimated, and largely homogeneous across study subpopulation and racial/ethnic strata. Nevertheless, further investigation is needed to determine the effects of gaseous air pollutants on DNAm and resulting gene expression given that our results were not replicable in NAS and KORA, in part due to limited availability of ozone exposure data.

The most significant association in the present study was between 28-day NO2 exposure and decreased DNAm at a CpG site on chromosome 3. This CpG site is upstream from ZNF621 in a region that is enriched with regulatory elements including transcription factor binding, histone protein modification, and DNase hypersensitivity sites (Supplemental Figure 2) (Rosenbloom et al., 2012). ZNF621 encodes Zinc Finger Protein 621, one of many such proteins involved in e.g. transcription, apoptosis, and protein packaging (Laity et al., 2001), which is universally expressed in lung, coronary artery, and other tissues affected by air pollution and atherosclerosis (Rosenbloom et al., 2012). A study investigating loci associated with type 2 diabetes (T2D) in populations of European and African ancestry found that one such locus increased T2D risk via cis-regulation of ZNF621 expression in adipose tissue (Lau et al., 2017). As multiple studies have implicated air pollution as a risk factor for T2D (Rzez et al., 2015; Park and Wang, 2014; Balti et al., 2014), altered DNAm in a regulatory region of ZNF621 may be a mechanistic pathway underlying this association.

Air pollution-associated changes in DNAm at the two remaining FDR-significant sites may also have biological links to cardiorespiratory disease via LIME1 and KLHL35. LIME1 encodes the Lck-interacting transmembrane adaptor 1 protein, which propagates T and B cell receptor signals (Hur et al., 2003; Hörejš et al., 2004). It is expressed in liver, spleen, brain, and whole blood (Kuhn et al., 2013) as well as hematopoietic stem cells and lung tissue (Hur et al., 2003). In particular, LIME1 has a role in T-cell activation and resulting IL-2 expression (Hur et al., 2003). Observed changes in DNAm within the third intron of LIME1 therefore suggest a plausible link between air pollution exposures, local/systemic inflammatory responses to these pollutants, and their established cardiorespiratory effects, although they also may reflect among-participant variation in minor lymphocyte proportions not captured by the Houseman method (Houseman et al., 2012). KLHL35 encodes Kelch-like protein 35. Kelch-like proteins are involved in posttranslational transfer of ubiquitins to other proteins (Dhanoa et al., 2013), thereby targeting them for intracellular translocation and/or degradation (Mukhopadhyay and Riezman, 2007). Although KLHL35 is expressed in a wide variety of tissues, including the testes, brain, and arteries (Kuhn et al., 2013), its specific function is unknown. Chromatin interaction with the promoter of KLHL35 has been suggested as a mechanism by which the single nucleotide polymorphism rs90121, a susceptibility variant for coronary artery disease (Howson et al., 2017), functions (van der Harst and Verweij, 2018).

Other CpG sites that may be of interest for cardiometabolic and respiratory disease were not statistically significant according to a priori FDR-based significance thresholds, but appeared within our top ranked list. For example, cg15008743 appeared as part of five NO2/NOx-DNAm associations in our top ranked list. This CpG lies within ZNF83, which encodes another zinc finger protein (Laity et al., 2001). ZNF83 is expressed in tissues including the female reproductive tract, brain, thyroid, and pituitary (Kuhn et al., 2013). Other highly ranked but non-FDR significant associations included those near genes that have frequently been implicated in cardiometabolic disease (e.g. CPT1A

K.M. Holliday et al.

Environmental Research 212 (2022) 113360
We were unable to replicate the top hits for 28-day NO2 in NAS and KORA and 7-day O3 in NAS. Potentially important differences exist between the NAS and KORA replication populations and the WHI and ARIC discovery populations that may help explain the lack of replication. While NAS is a population of elderly, white men in one U.S. city and KORA of residents in one region of Germany, WHI and ARIC included a racially/ethnically diverse population of mostly women from throughout the U.S. In addition, the standard deviation of the 28-day NO2 concentration in the more geographically restricted NAS and KORA cohorts was approximately half that among WHI subgroups, despite similar means across populations. Also, the 7-day O3 concentration was lower in NAS than in WHI and ARIC (23 ppb vs 40 ppb).

Various air pollutants may have different effects on biologic processes over acute and chronic exposures (Franklin et al., 2015). Our results highlighted shorter-term (7- and 28-day average) NO2 and ozone concentrations as potentially important for DNAm. Because availability of data for some of these exposures was limited in the planned replication analyses in KORA and NAS, we also attempted to look up results of analyses in several recently published methylome-wide association studies involving gaseous air pollutants (de F. C. Lichtenfels et al., 2018; Sayols-Baixeras et al., 2019; Lee et al., 2019). However, none of the published studies included O3 or shorter-term NO2 exposures that would have enabled a relatively well-harmonized comparison of the top hits identified by the present study. Comparison of our non-significant results for longer duration (28- and 365-day) NO2 exposures with these published studies showed a similarly inverse and significant annual mean NO2-DNAm association at cg01885635 in the LifeLines cohort in the Netherlands (de F. C. Lichtenfels et al., 2018). However, the annual (Lee et al., 2019) and 10-year (Sayols-Baixeras et al., 2019) mean NO2-DNAm associations in other studies were neither significant nor inverse.

4.1. Limitations

This study may be limited by the effects of exposure measurement error, reduced power inherent in multiple testing, and inability to fully replicate results. Although twelve demographically, geographically, and environmentally diverse participant subpopulations contributed to this study, we were still able to harmonize exposure estimation, variable definitions, and quality control across them. We also used national-scale air pollution exposure estimation strategies to generate consistent estimates of ambient exposures, yet these potentially underestimate the true magnitude of personal air pollutant-outcome associations (Holliday et al., 2014). We used FDR to account for multiple testing across CpG sites, but we did not correct for multiple testing of several gaseous air pollutants and averaging durations as exposure variables. The KORA and NAS replication populations were limited to individuals of European ancestry from single geographic sites with more homogeneous gaseous air pollution exposures than those among the racially diverse WHI and ARIC participants from across the U.S., but they were the only identified studies with shorter-term air pollution exposures. Ozone exposure data were unavailable in KORA, limiting replication attempts for two CpG sites to the NAS population. We were unable to compare results with those of shorter-term NO2 or O3 exposure in other studies due to lack of data, although we observed similar results with longer-term NO2 in one study. Finally, if our results regarding effects of gaseous air pollutants on DNAm are replicable in future studies, further examination of the effects of these DNAm changes on gene expression will be necessary to understand their potential impact on disease.

5. Conclusions

Gaseous air pollutants may be associated with DNA methylation of cardiovascular disease-relevant gene regions. Further harmonized replication efforts in similarly diverse cohorts with a range of shorter-term pollutant exposures is warranted.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.envres.2022.113360.

References


Bellovia, A., Urich, B., Speck, M., et al., 2013. DNA hypomethylation, ambient particulate matter, and increased blood pressure: findings from controlled human exposure experiments. J. Am. Heart Assoc. 2 (3).


K.M. Holliday et al.

Environ. Health Perspect. 124 (6), 566


