Linking Air Pollution Exposure to Blood-Based Metabolic Features in a Community-Based Aging Cohort with and without Dementia

Vrinda Kalia[a,*], Erin R. Kulick[b], Badri Vardarajan[c,d,e], Yian Gu[c,d,e,f], Jennifer J. Manly[c,e], Mitchell S.V. Elkind[d,e,f], Joel D. Kaufman[g], Dean P. Jones[h], Andrea A. Baccarelli[a], Richard Mayeux[c,d,e,f,i], Marianthi-Anna Kioumourtzoglou[a] and Gary W. Miller[a,*]

[a]Department of Environmental Health Sciences, Mailman School of Public Health, Columbia University, New York, NY, USA
[b]Department of Epidemiology and Biostatistics, Temple University College of Public Health, Philadelphia, PA, USA
[c]Taub Institute for Research on Alzheimer's Disease and the Aging Brain, Vagelos College of Physicians and Surgeons, Columbia University, New York, NY, USA
[d]The Gertrude H. Sergievsky Center, Vagelos College of Physicians and Surgeons, Columbia University, New York, NY, USA
[e]Department of Neurology, Vagelos College of Physicians and Surgeons, Columbia University and the New York Presbyterian Hospital, New York, NY, USA
[f]Department of Epidemiology, Mailman School of Public Health, Columbia University, New York, NY, USA
[g]Departments of Environmental and Occupational Health Sciences, Medicine, and Epidemiology, University of Washington, Seattle, WA, USA
[h]Department of Medicine, Clinical Biomarkers Laboratory, Emory University, Atlanta, GA, USA
[i]Department of Psychiatry, Vagelos College of Physicians and Surgeons, Columbia University, New York, NY, USA

Abstract.
Background: Long-term exposure to air pollution has been associated with changes in levels of metabolites measured in the peripheral blood. However, most research has been conducted in ethnically homogenous, young or middle-aged populations.
Objective: To study the relationship between the plasma metabolome and long-term exposure to three air pollutants: particulate matter (PM) less than 2.5 μm in aerodynamic diameter ($PM_{2.5}$), PM less than 10 μm in aerodynamic diameter (PM_{10}), and nitrogen dioxide (NO_2) in an ethnically diverse, older population.
Methods: Plasma metabolomic profiles of 107 participants of the Washington Heights and Inwood Community Aging Project in New York City, collected from 1995–2015, including non-Hispanic white, Caribbean Hispanic, and non-Hispanic Black older adults were used. We estimated the association between each metabolic feature and predicted annual mean exposure to the air pollutants using three approaches: 1) A metabolome wide association study framework; 2) Feature selection using elastic net regression; and 3) A multivariate approach using partial-least squares discriminant analysis.

*Correspondence to: Gary W. Miller, PhD or Vrinda Kalia, PhD, 722W 168th Street, New York, NY, 10032, USA. E-mails: gm2815@cumc.columbia.edu (Gary W. Miller); vk2316@cumc.columbia.edu (Vrinda Kalia).

Results: 79 features associated with exposure to $PM_{2.5}$ but none associated with PM_{10} or NO_2. $PM_{2.5}$ exposure was associated with altered amino acid metabolism, energy production, and oxidative stress response, pathways also associated with Alzheimer's disease. Three metabolites were associated with $PM_{2.5}$ exposure through all three approaches: cysteinylglycine disulfide, a diglyceride, and a dicarboxylic acid. The relationship between several features and $PM_{2.5}$ exposure was modified by diet and metabolic diseases.
Conclusions: These relationships uncover the mechanisms through which $PM_{2.5}$ exposure can lead to altered metabolic outcomes in an older population.

Keywords: Aging population, air pollution, Alzheimer's disease, dementia, metabolomics, nitrogen dioxide, particulate matter

INTRODUCTION

Air pollutants have been associated with mortality [1] and diseases [2] involving the pulmonary [3], cardiovascular [4], and nervous systems [5]. These associations have been attributed to system-wide changes induced by air pollution exposure including oxidative stress [6], inflammation [7], and changes in circulating metabolites [8]. Long term exposure to fine particulate matter with aerodynamic diameter $\leq 2.5\,\mu m$ ($PM_{2.5}$), particulate matter with aerodynamic diameter $\leq 10\,\mu m$ (PM_{10}), and nitrogen dioxide (NO_2) has been associated with decline in cognitive function [9, 10] and increased risk of Alzheimer's disease (AD) [11], exposure to $PM_{2.5}$ and NO_2 have been associated with higher incidence of dementia and AD [12], and exposure to $PM_{2.5}$ has been associated with clinical aggravation of age-related neurodegenerative diseases [13].

The metabolome comprises all metabolites and other small molecules in a biological matrix that may derive from both endogenous biochemical processes, exogenous exposures that are absorbed and metabolized by the body, as well as biochemical changes that result from environmental exposures [14]. The application of high-resolution mass spectrometry-based untargeted metabolomics allows us to capture a comprehensive profile of circulating small molecules, which include both known and as yet unknown circulating small molecules [15]. Thus, by applying untargeted metabolomics, we can determine circulating small molecules associated with long-term exposure to air pollutants in an agnostic manner. A metabolic signal in the periphery as a result of exposure to air pollutants is plausible since inhaled particles have been reported to pass into circulation from the lungs [16]. Besides, particles that reach the lower airways and alveoli can induce inflammation and oxidative stress, activating local and systemic metabolic changes [17, 18]. Indeed, studies have identified a circulat-

ing metabolic signature associated with exposure to traffic-related air pollutants [19]. However, few studies have investigated the relationship between long-term exposure to air pollution and circulating metabolic signals in an ethnically diverse aging population using an untargeted global metabolomic approach.

Individuals with chronic diseases may be more sensitive to effects of air pollution [20], possibly due to deficiencies in antioxidants and elevated levels of inflammation [21]. However, few studies have considered whether associations between exposure to air pollutants and circulating metabolites are modified by underlying metabolic diseases. Identifying factors that can modify this relationship would allow us to determine whether people with age-related diseases are particularly vulnerable to the effects of air pollution exposure given that people are living longer lives and bear a large burden of age-related co-morbidities [22]. Besides, the older adults are particularly susceptible to the effects of chronic environmental exposures in part due to age-related deterioration in physiological function as well as the presence of co-morbidities that exert a strain on the body's ability to respond to environmental insults [23]. One recently published study found that long-term exposure to air pollution is associated with metabolic perturbations in older adults; however, their study population only included non-Hispanic white men [24]. Disparities in exposure to air pollutants [25–27] and mortality due to exposure to $PM_{2.5}$ [1, 28] among racial/ethnic groups has been well documented. Thus, studying the effects of air pollution exposure in an ethnically diverse aging community is necessary.

We performed analyses to identify circulating metabolites associated with long-term exposure to $PM_{2.5}$, PM_{10}, and NO_2 and investigated whether these relationships are modified by prevalent cardiometabolic and neurodegenerative diseases. We used data from the Washington Heights and Inwood

Community Aging Project (WHICAP) in New York City, a multi-ethnic community-based cohort of older adults created to study risk factors of dementia and AD. We used three different regression methods to find reliable metabolic features associated with exposure. If any features were significantly associated (false discovery rate at 5%) with an exposure through a metabolome-wide association study framework, we additionally performed a penalized regression for feature selection, using elastic net regression, and multivariate supervised dimensionality reduction, using partial least squares differential analysis, to find metabolic features important in predicting exposure. We then report concordance of results from all three approaches. Such a multipronged approach has been used as an alternative to traditional hypothesis testing for selection of metabolic features of importance [29]. Additionally, we tested for the presence of effect modification by factors that are known to affect metabolism: sex [30], racial/ethnic group [31], dementia [32], a history of diabetes, hypertension or heart disease [33, 34], and diet [35]. We also tested for modification by the presence of an *APOE ε*4 allele since it has been associated with metabolism [36] and was found to modify the relationship between exposure to air pollutants and cognitive function in the cohort [37]. In doing so, we aimed to identify factors and morbidities that may modify the relationship between exposure and circulating metabolites.

MATERIALS AND METHODS

Study population

We leveraged previously existing metabolomics data in 119 older adults from the WHICAP cohort [32]. WHICAP is a representative community-based cohort in the northern Manhattan region of New York City comprising of individuals aged 65 years and older enrolled through a collaborative effort with the Centers for Medicare and Medicaid and through marketing rolls. All participants were enrolled after obtaining informed consent and an interview was conducted in either English or Spanish. A detailed clinical assessment with standardized medical and neurological history was collected by a trained physician. At each visit, blood was drawn and sent to the laboratory within 2 h for DNA extraction and storage of plasma and serum. At the first visit, residential information was also obtained. Results from all interviews, clinical visits, and tests were reviewed through

a consensus conference comprised of clinicians with expert knowledge in the diagnosis of AD and related dementias, who made an AD diagnosis based on accepted criteria [38]. The study comprised participants equally divided among non-Hispanic white, non-Hispanic Black and Caribbean Hispanic ancestry based on self-report, classified based on the 1990 U.S. Census guidelines. Metabolomics data was available for 59 participants who were diagnosed with clinical AD and 60 who were heathy at the time of assessment.

High resolution mass spectrometry-based metabolomics

The details on the acquisition of the untargeted metabolomics has been previously described [32, 39]. The data were generated in the Clinical Biomarkers Laboratory at Emory University. Briefly, the metabolites were extracted from plasma using acetonitrile and the extracts were injected in triplicate on two chromatographic columns: a hydrophilic interaction column (HILIC) under positive ionization (HILIC+) and a C18 column under negative ionization (C18-), to obtain three technical replicates per sample per column. After separation and ionization, the ions produced were analyzed in full scan mode for molecules within 85–1250 kDa on a Thermo Orbitrap HF Q-Exactive mass spectrometer. The untargeted metabolomic data were processed through a computational pipeline that leverages open source feature detection and peak alignment software, *apLCMS* [40] and *xMSanalyzer* [41]. The feature tables were generated containing information on the mass-to-charge (*m/z*) ratio, retention time, and the abundance/intensity of each ion for each sample. Correction for batch effects was performed using ComBat, which uses an empirical Bayesian framework to adjust for known batches in which the samples were run [42]. Each of these ions are referred to as metabolic features. For the analysis, metabolic features detected in at least 70% of the samples were retained, leaving 6,375 features from the HILIC+ column and 3,759 features from the C18- column for further analysis. Zero-intensity values were considered below the detection limit of the instrument and were imputed with half the minimum intensity observed for each metabolic feature. The intensity of each metabolic feature was log10 transformed, quantile normalized, and auto-scaled for normalization and standardization.

Exposure assessment

The methods used for exposure assessment have been described in detail previously [9]. In brief, participants' residential addresses obtained during the first neuropsychological examination were geocoded and long-term air pollution exposures in the calendar year prior to a clinical visit were estimated for NO_2 (ppb), PM_{10} ($\mu g/m^3$), and $PM_{2.5}$, ($\mu g/m^3$), using regionalized universal kriging models [43, 44]. Measurements of the air pollutants were obtained from the US EPA Air Quality System database. For geocodes without monitoring data, the annual averages were used in the kriging models to predict concentrations. To improve predictions, information on land use, distance to roadways, and other geographic variables were included using partial least square methods. The cross-validated MSE-based R^2 was 0.88 and 0.85 for $PM_{2.5}$ and NO_2, respectively [43, 44]. The models did not apply a grid but made predictions at the actual residential location based on calculated geographic covariates at that location. The exposure data were analyzed as continuous variables in linear and penalized regressions and dichotomized for the partial least square discriminant analysis. In this study we considered 107 participants for whom both metabolomics data and air pollution exposure estimates were available. These participants had characteristics similar to all those with metabolomics and exposure data available.

Statistical analysis

We used a metabolome-wide association study (MWAS) framework with correction for multiple comparisons by controlling the false discovery rate (FDR) at 5%. In case significant associations between metabolic features and an air pollutant were observed, we then performed two additional tests, elastic net regression (e-net) and partial least squares discriminant analysis (PLS-DA). All models were adjusted for the year of blood draw in order to adjust for temporal variation in the data. We also adjusted all models for total number of education years completed as a proxy for socioeconomic status. The overlap of metabolic features in the results from all three analyses was determined using the *VennDetail* package (version 1.14.0) which extracts details of multi-set interactions that can be used for visualization. All analyses were conducted in R (version 3.6.3).

Metabolome-wide association with exposure to air pollutants

We estimated the relationship between estimated long-term exposure to $PM_{2.5}$, PM_{10}, and NO_2 and each metabolic feature using multiple single-pollutant models, adjusted for potential confounders and predictors of the outcome including age (years), sex (women/men), racial/ethnic group (non-Hispanic white, non-Hispanic Black, Caribbean Hispanic), clinical diagnosis of AD (case/control (no dementia)), year of blood draw, and years of education. We used single pollutant models since the mechanisms and pathways of toxicity through which these pollutants have been related to health tend to be distinct. Besides, there was a strong positive correlation between all three pollutants and that raised issues of collinearity in our models (Supplementary Table 2). The analyses were conducted separately for data from each column. We corrected for multiple comparisons using an FDR of 5% and q-values were estimated using the Benjamini-Hochberg (BH) method.

Penalized regression for selection of metabolic features that predict exposure to PM2.5

The MWAS only detected significant associations with $PM_{2.5}$; we thus, performed elastic net regression to determine which features from both columns were predictors of $PM_{2.5}$ exposure. The covariates included were the same as the MWAS framework. We used five-fold cross-validation, repeated five times, using the *caret* package (version 6.0–90) with a custom grid search where alpha ranged from 0.1 to 1 with 0.1 increments, and lambda ranged from 0.0001 to 1 with 0.053 increments, to determine the optimal penalization and mixing parameters [45]. Based on the root mean square error, the parameters chosen were a lambda of 0.158 and alpha of 1 (i.e., lasso was the optimal fit).

Partial least squares discriminant analysis

Finally, we conducted a partial least squares discriminant analysis to determine which metabolic features were important in discriminating between low and high exposure to $PM_{2.5}$ using a multivariate approach. We created a binary categorization of the $PM_{2.5}$ exposure by first calculating the residuals from regressing $PM_{2.5}$ exposure on year of blood draw, since there was a strong negative correlation between $PM_{2.5}$ exposure and year of blood draw. Second, we categorized exposure as high or low at the median of the residuals. Then, we regressed the intensity of each metabolic feature from both columns on the same

covariates as above in "Metabolome-wide association with exposure to air pollutants" and stored the residuals. Finally, we used the de-trended metabolic feature intensities in a PLS-DA model using the *mixOmics* package (version 6.11.1), which contains methods for multivariate methods for omics data [46]. We used a variable importance (VIP) score cut-off of 3.0 for component 1 and 2.5 for component 2, to deem a metabolic feature important in discriminating between high and low exposure. These cut-offs were chosen to find the most important features based on the distribution of VIP scores.

Pathway analysis

To determine the biological relevance of the metabolic features associated with $PM_{2.5}$, we conducted pathway analysis using the "functional analysis" module in *MetaboAnalyst* (version 5.0) [47], a web-based interface for comprehensive metabolomic data analysis. We used the MWAS results from both columns and applied a nominal *p*-value cut-off of 0.01 to determine metabolic pathway enrichment using the mummichog algorithm and the human MFN reference database [48]. We present results for pathways with a Fisher's exact test *p*-value < 0.1.

Metabolite annotation

Metabolite annotations were made using an internal library and by matching to the Human Metabolome Database (HMDB) [49], the Kyoto Encyclopedia of Genes and Genomes (KEGG) database, and LIPIDMAPS using the R package *xMSannotator* (version 1.3.2) [50]. This uses a multistage clustering algorithm method to determine metabolic pathway associations, intensity profiles, retention time, mass defect, and isotope/adduct patterns to assign putative annotations to metabolic features. In cases where a feature had multiple matches, we used the following rules to assign an annotation: first, we screened features based on the confidence score assigned by xMSannotator, and the annotation with the highest score was used. Second, if all annotations had the same score, we chose the annotation with the lowest difference in expected and observed mass (delta parts per million (ppm)). Finally, if all features had the same score and delta ppm, we indicated the identity as "multiple matches" since we couldn't decipher a unique putative annotation (10.5%). If a feature did not match any database entries, it was denoted as "unknown" (35.2%). The confidence in annotation was based on criteria defined by Schymanski et al. [51], where level 1 corresponds to a confirmed structure identified through MS/MS and/or comparison to an authentic standard; level 2 to a probable structure identified through spectral matches to a database; level 3 to a putative identification with a speculative structure; level 4 to an unequivocal molecular formula but with insufficient evidence to propose a structure; and level 5 to an exact mass but not enough information to assign a formula.

Effect modification and sensitivity analysis

We tested for the presence of effect modification only in the metabolic features that were significantly associated with long-term exposure to $PM_{2.5}$ after correction for multiple comparisons through the MWAS approach. This 2-step approach that utilizes a screening step has been shown to improve power to detect gene-by-environment interactions [52]. We compared models with and without an interaction term between $PM_{2.5}$ and the effect modifier using likelihood ratio tests and controlled the FDR at 20% to look for noteworthy interactions using the Benjamini-Hochberg method.

We tested whether racial/ethnic group, sex, history of heart disease, hypertension or diabetes, an AD diagnosis, or an *APOE* ε4 allele were effect modifiers of the observed associations. We ran separate linear models for each modifier, adjusted for the same covariates as above in "Metabolome-wide association with exposure to air pollutants" and included an interaction term for each modifier and $PM_{2.5}$. This model was compared to a lower-order model with the modifier included only as a covariate, without an interaction term.

To test for effect modification by diet we used data from a 61-item version of the Willet semiquantitative food frequency questionnaire (SFFQ). This questionnaire has been validated for use in older adults [53] and in the WHICAP cohort [54–56]. The validity of the SFFQ data has been previously determined in a subsample of the WHICAP cohort ($n = 78$) and a moderate intraclass correlation was observed for energy-adjusted nutrients (0.30 for total calories, 0.28 for carbohydrates, 0.41 for fats, and 0.33 for protein) [55]. The questionnaire was administered in English or Spanish to participants of the WHICAP cohort to determine average food consumption in the year prior to baseline assessment. The 61 food items were categorized into 30 groups and intake of each group was calculated by summing intakes of member food items. The daily intake of nutrients was determined by multiplying the frequency of consumption of each

portion of every food by the nutrient content of each portion [57, 58]. In our study, we had the SFFQ data available for 77 of the observations (72%). We used data on estimated intake of four macronutrients: total carbohydrate, total protein, animal fat, and vegetable fat. The estimated intake of the macronutrients was adjusted for total caloric intake using the residual method. We used principal component analysis to create linear combinations of the four macronutrients to discern patterns in macronutrient intake using the *prcomp()* function in R. Each of the principal components (PC) was used to test for effect modification by including an interaction term between each PC and $PM_{2.5}$ in a linear model like that above and additionally, adjusting for total caloric intake.

Sensitivity analysis was conducted to determine whether smoking history confounded the relationship observed between long-term $PM_{2.5}$ exposure and the metabolic features by additionally adjusting the model above in "Metabolome-wide association with exposure to air pollutants" for smoking history. The participants were asked the question "Have you ever smoked?" and their responses were recorded as "Yes/No", providing a binary variable for their smoking history. The results were compared by plotting the coefficients for $PM_{2.5}$ from the two models to determine whether adjusting for smoking history produced coefficients different from the main analysis.

RESULTS

The mean age of the participants was 89.2 years (standard deviation (SD) = 8.44), women made up 82% of the population, 34.6% were non-Hispanic Black, 33.6% were non-Hispanic white, and 31.8% were Caribbean Hispanic, and the mean number of years of education was 9.42 (SD = 4.66). Roughly half of the participants were diagnosed with AD (53.3%), 27.1% of the participants carried at least one APOE ε4 allele, 20.6% of the participants had a history of diabetes, 37.4% had a history of heart disease, and 71% had a history of hypertension. The metabolomic data consisted of 6375 metabolic features from the HILIC+ column and 3759 features from the C18- column after data filtering and cleaning. The average long-term exposures in these participants was 31.7 (SD = 7.03) ppb for NO_2, 21.0 (SD = 7.73) μg/m^3 for PM_{10} and 12.9 (SD = 2.41) μg/m^3 for $PM_{2.5}$ (Table 1) and correlation between pollutants ranged from 0.73–0.91 (Supplementary Table 2).

Using an MWAS framework, we found 79 metabolic features significantly associated with $PM_{2.5}$ (61 from the HILIC+ column and 18 from the C18- column, Fig. 1, Supplementary Table 3) after correcting for multiple comparisons. No significant associations were found between the metabolic features and PM_{10} or NO_2 (Fig. 1). Through the penalized elastic net regression, we found 32 metabolic features that had non-zero coefficients (18 from HILIC+ column and 14 from C18-, Supplementary Table 4). The results from the PLS-DA showed separation along component 1 and 2 of people exposed to $PM_{2.5}$ below or above the median (Fig. 2A). Thirty-six metabolic features had a variable importance score > 3 on component 1 or a VIP score > 2.5 on component 2 (32 from HILIC+ column and 4 from C18- column, Supplementary Table 5).

When comparing the results from the three approaches, we found 6 features associated with $PM_{2.5}$ through all approaches, i.e., the MWAS, elastic net and PLS-DA (Fig. 2B). These features were putatively annotated as cysteinylglycine disulphide (cys-gly), a diglyceride, alphachloralose, a dicarboxylic acid, and one feature had multiple matches in HMDB (Table 2). The metabolic features associated with $PM_{2.5}$ enriched several metabolic pathways including: alanine and aspartate metabolism, the TCA cycle, glutamate metabolism, glycolysis and gluconeogenesis, butanoate metabolism, pyruvate metabolism, methionine and cysteine metabolism, tyrosine metabolism, fatty acid oxidation, vitamin A metabolism, glycerophospholipid metabolism, and aminosugars metabolism. These represented pathways related to energy production, redox homeostasis, and amino acid metabolism (Fig. 2C).

None of the relationships identified as significant through MWAS were modified by sex, racial/ethnic group, history of heart disease, or history of hypertension. The relationship between a lysophosphatidylethanolamine and $PM_{2.5}$ was modified by both AD diagnosis and *APOE* ε4 allele status, such that people with AD or carriers of at least one ε4 allele had a negative association between exposure and levels of lysophosphatidylethanolamine while those without AD or an ε4 allele had a positive association (Table 3, Supplementary Figure 1). The relationship between two other features annotated as lysophosphatidylethanolamine was also modified by APOE ε4 allele status in a similar way, i.e., those with at least one ε4 allele had a negative association between exposure and lysophosphatidylethanolamine (Table 3, Supplemen-

Table 1
Characteristics of the study population in the Washington Heights and Inwood Community Aging Project. After regressing out the effect of year of blood draw, the PM2.5 exposure levels were dichotomized at the median, to create low and high exposure groups. This was also used to determine features associated with exposure through a partial least square discriminant analysis

	Low PM$_{2.5}$ ($N=54$)	High PM$_{2.5}$ ($N=53$)	Overall ($N=107$)
Age (y)			
Mean (SD)	90.7 (7.31)	87.6 (9.25)	89.2 (8.44)
Sex			
Women	44 (81.5%)	44 (83.0%)	88 (82.2%)
Men	10 (18.5%)	9 (17.0%)	19 (17.8%)
Racial/Ethnic group			
Non-Hispanic Black	22 (40.7%)	15 (28.3%)	37 (34.6%)
Non-Hispanic white	16 (29.6%)	20 (37.7%)	36 (33.6%)
Caribbean Hispanic	16 (29.6%)	18 (34.0%)	34 (31.8%)
Education (y)			
Mean (SD)	9.74 (4.81)	9.09 (4.53)	9.42 (4.66)
Alzheimer's disease			
Case	28 (51.9%)	29 (54.7%)	57 (53.3%)
Control	26 (48.1%)	24 (45.3%)	50 (46.7%)
APOE ε4 carrier			
No ε4 allele	39 (72.2%)	39 (73.6%)	78 (72.9%)
At least one ε4 allele	15 (27.8%)	14 (26.4%)	29 (27.1%)
History of diabetes			
No	45 (83.3%)	40 (75.5%)	85 (79.4%)
Yes	9 (16.7%)	13 (24.5%)	22 (20.6%)
History of heart disease			
No	36 (66.7%)	31 (58.5%)	67 (62.6%)
Yes	18 (33.3%)	22 (41.5%)	40 (37.4%)
History of hypertension			
No	21 (38.9%)	10 (18.9%)	31 (29.0%)
Yes	33 (61.1%)	43 (81.1%)	76 (71.0%)
PM$_{2.5}$ (μg/m3)			
Mean (SD)	11.8 (2.65)	14.0 (1.48)	12.9 (2.41)
PM$_{10}$ (μg/m3)			
Mean (SD)	20.7 (8.82)	21.2 (6.51)	21.0 (7.73)
NO$_2$ (ppb)			
Mean (SD)	29.8 (7.96)	33.7 (5.29)	31.7 (7.03)

tary Figure 2). The relationship between three features and PM$_{2.5}$ was modified by a history of diabetes, such that people with a history of diabetes had a greater positive association between the features and exposure to PM$_{2.5}$ (Supplementary Table 6, Supplementary Figure 3), however; the features did not match any database entries.

The principal component analysis of macronutrients from the dietary data generated four PCs (Supplementary Figure 4). PC1 (explained 57.1% of the total variance in the data) had positive loadings from carbohydrates and vegetable fat and negative loadings from total protein and animal fat and modified the relationship between one metabolic feature and exposure to PM$_{2.5}$ such that a unit increase in the score on PC1 reduced the slope of the relationship between the metabolic feature and PM$_{2.5}$. PC2 (28% of variance explained) and PC3 (13.9% of variance explained) did not modify any relationships discovered but the relationship between 39 of the

metabolic features, about half of the features identified in MWAS, and PM$_{2.5}$ was modified by PC4 (explained 0.01% of the variance in the data), which had positive loadings from all four macronutrients (Table 3, Supplementary Table 6, Supplementary Figure 4).

In sensitivity analysis, the associations between exposure and the metabolic features were not affected by the inclusion of smoking history as a covariate in the model (Supplementary Figure 5).

DISCUSSION

We found several plasma metabolic features associated with predicted outdoor residential annual concentrations of PM$_{2.5}$ in a racially and ethnically diverse urban population of older adults. Pathway analysis revealed perturbations in the metabolism of several amino acids, the citrate cycle, glycolysis and gluconeogenesis, butanoate, pyruvate, fatty acid,
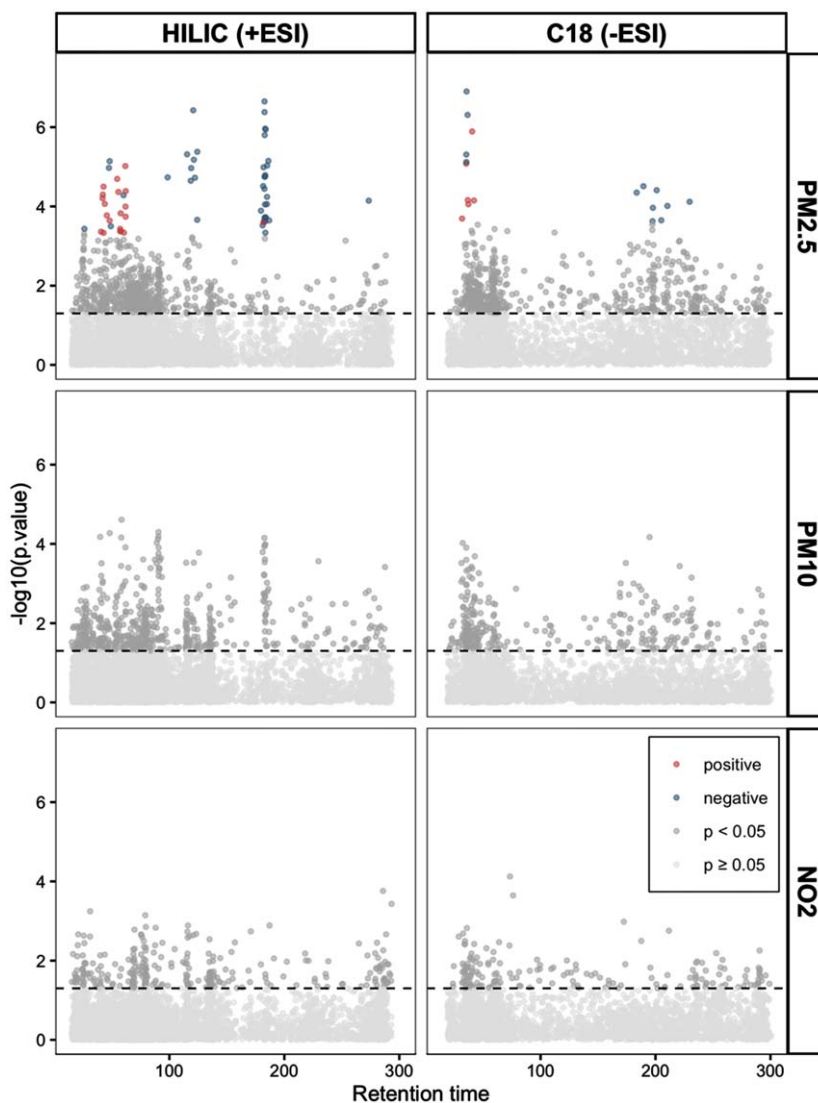
Fig. 1. Manhattan plots show the metabolic features associated with long-term exposure to air pollutants determined through a metabolome wide association study framework. In the grid, the two columns represent data from the HILIC (on the left) and C18 chromatographic column under positive and negative ionization respectively. The rows show data for the respective air pollutants. All models were adjusted for age (years), sex (men/women), racial/ethnic group (non-Hispanic white, non-Hispanic black, Hispanic Caribbean), Alzheimer's disease status (case/control), year of blood draw and years of education. The horizontal line represents nominal p-value < 0.05. The colored dots represent features significantly associated with the pollutant after correction for multiple comparisons (FDR < 0.05) with blue dots representing a negative association between that metabolic feature and air pollutant and the red dots a positive association.

vitamins and co-factors, glycerophospholipid, and aminosugars. These metabolic pathways have been previously associated with $PM_{2.5}$ and include indicators of oxidative stress. These signals could help us understand the mechanisms through which $PM_{2.5}$ exposure can lead to altered health outcomes.

This is the first study of metabolic features associated with long-term exposure to $PM_{2.5}$ in a racially and ethnically diverse aging population of both men

and women. Additionally, we used three different statistical approaches to determine features associated with exposure to $PM_{2.5}$. By looking for concordance among results from diverse regression methods, we found a reliable set of features that are associated with, and can predict exposure to, $PM_{2.5}$. While others have used such multipronged approaches for metabolomic data analysis [29], this is the first study to consider such an approach when studying the effect
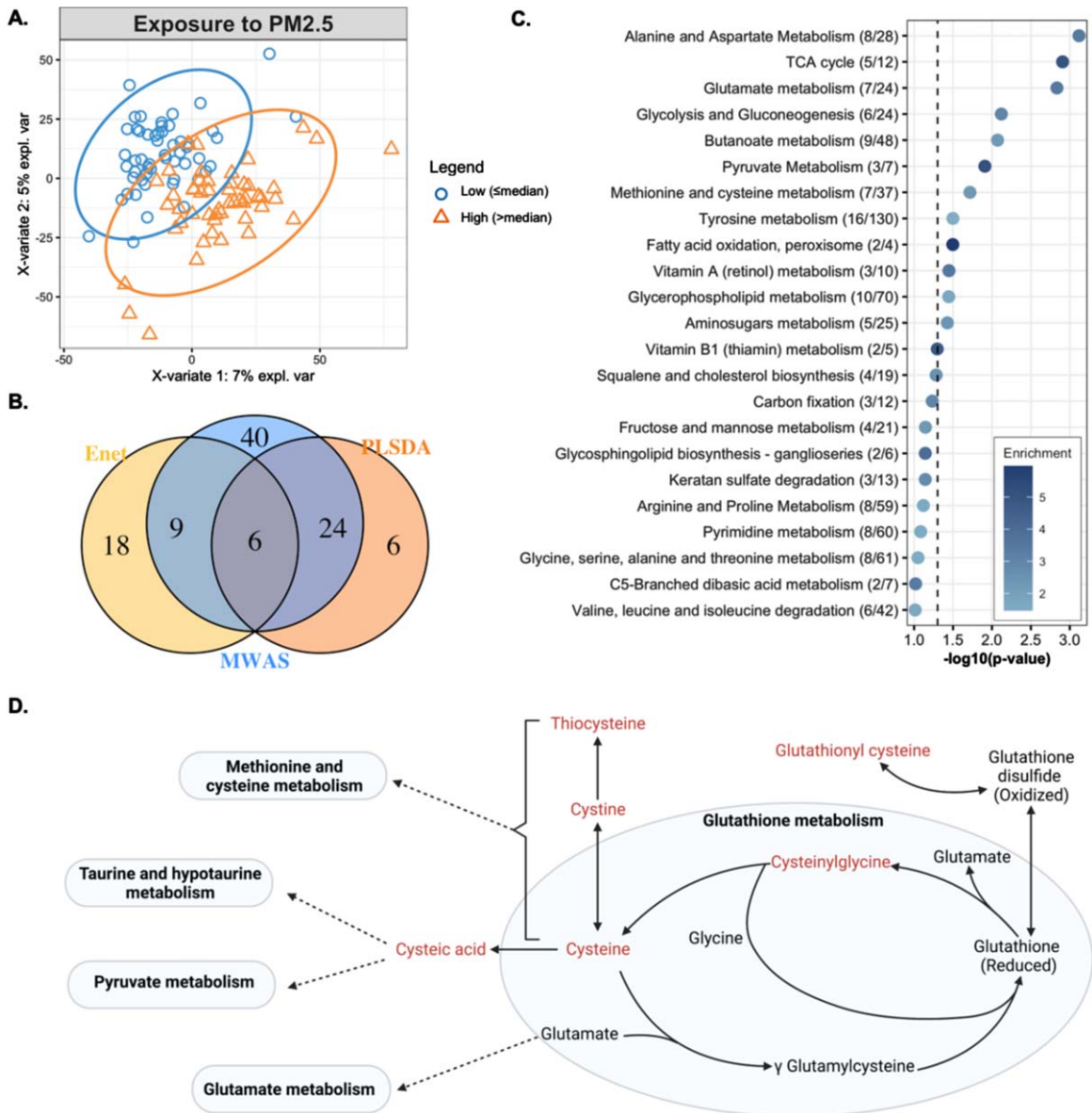
Fig. 2. In A, the biplot shows clustering of observations with high (orange triangles) and low (blue circles) long-term exposure to $PM_{2.5}$ after adjusting for age, sex, ethnic group, Alzheimer's disease status, year of blood draw and years of education. In B, the overlap in metabolic features associated with long-term $PM_{2.5}$, estimated through three different approaches, a metabolome association study (MWAS), elastic net regression (enet) and partial least squares discriminant analysis (PLSDA). In C, the metabolic pathways enriched by the metabolic features associated with long-term $PM_{2.5}$ exposure, determined through the MWAS approach. Enrichment represents that ratio between observed number of significant hits/expected number of hits. In D, a representation of glutathione metabolism and the biochemical relationship between members of glutathione metabolism and other pathways. Metabolites in red text were negatively associated with $PM_{2.5}$ through either MWAS, enet, or PLS-DA.

of long-term exposure to $PM_{2.5}$ on the circulating metabolome. We identified six features associated with $PM_{2.5}$ through this approach, two of which were annotated as cys-gly (an $M + 2H$ adduct and an isotope of the $M + H$ adduct), one as dicarboxylic acid, and another as diglyceride with level 3 confidence,

and two were annotated with level 5 confidence, one as alphachloralose and another with multiple database matches.

Both features annotated as cys-gly were negatively associated with exposure to $PM_{2.5}$. A previous study in the Normative Aging Study also reported

Table 2
The putative annotations of the six features associated with long-term PM2.5 through all three approaches (Fig. 2B)

m/z	Time (s)	Annotation	Formula	Delta ppm	Adduct	ID score*	ESI	β	p	Pathway/source
300.0478	120.7	Cysteinylglycine disulfide, multiple other matches	C8H15N3O5S2	0	M+H [+2]	3	+	–0.48	3.7E-07	Glutathione metabolism; Redox homeostasis
149.53	124.3	Cysteinylglycine disulfide	C8H15N3O5S2	0.47	M+2H	3	+	–0.45	4.2E-06	Glutathione metabolism; Redox homeostasis
355.0733	115.4	Multiple matches	–	–	–	5	+	–0.44	4.8E-06	–
154.9871	183.3	Alphachloralose	C8H11Cl3O6	8	M+2H	5	+	–0.45	1.1E-06	Rodenticide
775.6017	48	A diglyceride	C45H72O5	4.32	M+2ACN+H	3	+	–0.43	7.2E-06	Lipid metabolism
129.0194	36.5	A dicarboxylic acid	C5H6O4	0.54	M-H	3	–	–0.49	4.9E-07	Lipid metabolism

m/z, mass-to-charge ratio; Time, Retention time; Delta ppm, mass difference in parts per million; ID score, confidence in annotation based on Schymanski scale (1 being the highest and 5 the lowest); ESI, electrospray ionization.

a significant negative relationship between cys-gly and long-term exposure to $PM_{2.5}$ [24]. Cys-gly is formed by the degradation of reduced glutathione metabolism catalyzed by glutathione specific γ-glutamylcyclotransferase [59, 60]. This cleavage of reduced glutathione is important for the turnover of glutathione (Fig), which is required for maintaining redox homeostasis [60], and in cellular stress response [61]. A reduction in glutathione levels has also been observed in the liver of mice exposed to $PM_{2.5}$ [62].

Dicarboxylic acids are produced by breakdown of amino acids and fatty acids and can feed into the TCA cycle for energy production [63, 64]. The observed association may suggest changes in the energy production from amino acids and fatty acids because of long-term exposure to $PM_{2.5}$. In fact, pathway analysis did reveal alterations in pathways related to energy production, namely the TCA cycle and glycolysis and gluconeogenesis. *In vivo* exposure to $PM_{2.5}$ in mice also showed altered TCA cycle, glucose, and lipid metabolism in the liver [62].

Diglycerides (DG) consist of a glyceride with two fatty acid chains. DGs serve as substrates to the diacylglycerol acyltransferase (*Dgat*) enzyme, which converts DGs to triglycerides, metabolites that are key regulators of lipid transport and deposition in adipose tissue [65, 66]. Mice exposed to $PM_{2.5}$ had higher expression of *Dgat* in white adipose tissue [67], suggesting an increase in the synthesis of triglycerides from DGs. This response to exposure to $PM_{2.5}$ would also be in line with the negative association between the circulating DG and $PM_{2.5}$ observed in our population.

Finally, two features had low confidence in their annotation: one was putatively annotated as alphachloralose while the other had multiple database matches. Given the low confidence in annotation, we do not discuss their potential role in association with $PM_{2.5}$ but plan to confirm the identity of these features in the future.

Several of our findings confirm results from previous analyses. Long-term exposure to $PM_{2.5}$ over 2000–2016 in an aging population of white men living in the Boston area, with mean age of 75 years, was also associated with alterations in glycerophospholipid, alanine, and glutathione metabolism [24]. Similar pathways were also reported to be altered by $PM_{2.5}$ exposure in a subset of the UKTwins study, including pyruvate metabolism, glycolysis, and gluconeogenesis [68]. In a study of younger individuals, long-term exposure to $PM_{2.5}$ was associated with circulating phospholipids, similar to the enrichment in glycerophospholipid metabolism in our study [69]. Nearly all the metabolomic pathways affected by long-term $PM_{2.5}$ in our study were also reported in a study of exposure to ultrafine particulate matter [70]. Additionally, animal and *in vitro* studies have also reported similar changes in metabolic pathways. Changes in the TCA cycle, amino acid biosynthesis, and glutathione metabolism were reported in lung epithelial cells exposed to fine particulate matter [71]. In mice chronically exposed to ambient $PM_{2.5}$, serum metabolomic analysis revealed changes in glycerophospholipid, sphingolipids, glycerolipids and lysophospholipids, and in pathways related to protein digestion and absorption, glycine, serine, threonine metabolism, alanine metabolism, carbon

Table 3

Effect modification. The relationship between several metabolic features and PM2.5 was modified by Alzheimer's disease diagnosis, *APOE ε*4 allele, history of diabetes, and diet

Modifier	Number of features[#]	m/z	Time (s)	FDR from LR test	Putative annotation	Adduct	Delta ppm	ID score[*]	ESI
Sex	0								
Racial/Ethnic group	0								
History of heart disease	0								
History of hypertension	0								
Clinical diagnosis of Alzheimer's disease	1	475.2654	189.5	0.102	A lysophosphatidylethanolamine	M-H [-1]	–	3	–
*APOE ε*4 allele status	4	475.2654	189.5	0.08	A lysophosphatidylethanolamine	M-H [-1]	–	3	–
		500.2779	197.7	0.101	A lysophosphatidylethanolamine	M-H	0.72	3	–
		503.292	201	0.08	A lysophosphatidylethanolamine	M-H [-1]	–	3	–
History of diabetes	3				*Unknown (see supplemental table 6)*				
Diet: PC1 ↑ carbohydrate, vegetable fat ↓ protein, animal fat	1				*Unknown (see supplemental table 6)*				
Diet: PC2	0								
Diet: PC3	0								
Diet: PC4 ↑ carbohydrates, animal fat, protein, and vegetable fat	39	241.0308	181.6	0.063	Cystine	M+H	1.33	3	+
		774.5957	47.4	0.0012	A phosphatidylcholine or phosphatidylethanolamine	M+H	6.49	3	+
		170.0128	179.7	0.138	Cysteic acid	M+H	6.06	3	+
		122.027	183.9	0.087	Cysteine	M+H	0.25	3	+
		141.5325	181.9	0.063	Cystine	M+ACN+2H	0.14	3	+
		243.0263	182.7	0.063	Cystine	M+H [+2]	–	3	+
		263.0126	183.2	0.063	Cystine	M+Na	1.79	3	+
		149.53	124.3	0.158	Cysteinylglycine disulfide	M+2H	0.47	3	+
		298.0517	119	0.109	Cysteinylglycine disulfide	M+H	2.99	3	+
		300.0478	120.7	0.158	Cysteinylglycine disulfide, multiple other matches	M+H [+2]	–	3	+
		324.3257	41.9	0.078	N-(14-Methylhexadecanoyl)pyrrolidine	M+H	1.2	3	+
		378.7789	61.9	0.115	A diglyceride	M+ACN+H	7.42	5	+
		796.5254	57.2	0.076	A phosphatidylethanolamine	M+H-H2O	3.45	5	+
		822.5468	54.6	0.063	A phosphatidylethanolamine	M+H-H2O	3.65	5	+
		824.5603	55.7	0.058	A phosphatidylethanolamine	M+H-H2O	1.02	5	+
		98.5181	183.4	0.11	Diethyl trisulfide, multiple other matches	M+ACN+2H	3.25	5	+
		242.0344	181.1	0.081	Glucosamine 6-sulfate/ Azorhodine 2G	M+H-H2O/M+H+NH4	3.76/4.88	5	+
		152.9913	182.8	0.063	Multiple matches			5	+
		142.0342	182.5	0.071	Polyvidone	M+H-2H2O	2.53	5	+
		132.9872	60.1	0.103	Pyruvic acid, multiple other matches	M+2Na-H	–	5	+
		98.0164	186.8	0.063	Thiocysteine	M+ACN+2H	0.61	5	+
		195.0259	184.9	0.094	Thiocysteine, other multiple matches	M+ACN+H	1.28	5	+
		586.0316	182.9	0.109	Zoledronate	2M+ACN+H	8.75	5	+

[#]A noteworthy interaction was defined at an FDR < 0.2. LR, Likelihood ratio test; m/z, mass-to-charge ratio; Time, retention time; Delta ppm, mass difference in parts per million; ID score, confidence in annotation based on Schymanski scale (1 being the highest and 5 the lowest); ESI, electrospray ionization. See supplemental table 6 for information on features with no known database match.

metabolism, saccharides, fatty acids, sterols, and stress hormones [72]. Interestingly, several metabolic pathways associated with $PM_{2.5}$ exposure have also been implicated in AD. These include cysteine and glutathione metabolism [73], methionine, glycolysis and gluconeogenesis, and TCA cycle metabolism [74]. In a prior analysis in the cohort, we reported a relationship between AD and alanine and aspartate metabolism, and glycerophospholipid metabolism [32]. These pathways may provide some insight into possible mechanisms through which exposure to $PM_{2.5}$ may be related to incidence and risk of AD.

Several studies of short-term $PM_{2.5}$ exposure have reported changes in arginine, histidine, linoleate, and leukotriene metabolism [19, 70, 75, 76]. We did not find significant changes in levels of these metabolites associated with long-term exposure to $PM_{2.5}$ in our study, suggesting that there might be exposure-window specific changes in circulating metabolites.

Through our analysis, we also wanted to discover whether the associations identified between circulating metabolites and exposure to long-term $PM_{2.5}$ were modified by sex, racial/ethnic group, underlying metabolic diseases (history of diabetes, dementia diagnosis, history of heart disease, history of hypertension), or diet. The relationship between two of the six features associated with $PM_{2.5}$, identified through all three statistical approaches, both annotated as cysgly, was modified by diet. None of the other possible modifiers affected the associations between the six features and exposure.

APOE ε4 is a known risk factor for AD, the gene may underlie the observed modification by the dementia diagnosis. The *APOE ε4* gene plays a role in transporting cholesterol and other fats in the bloodstream, especially in the brain [77]. Lysophosphatidylethanolamine (LysoPE) is a derivative of phosphatidylethanolamine, a phospholipid typically found in the cell membrane, especially of the central nervous system [78]. Our findings suggest that people with at least one *APOE ε4* allele may have a different response to long-term exposure to $PM_{2.5}$ possibly due to altered lipid transport. Studies in *APOE* null mice have reported increased lipid deposition when exposed to concentrated $PM_{2.5}$ [79] and altered atherosclerotic plaque formation upon diesel exhaust exposure [80] compared to mice that were unexposed.

Having a history of diabetes modified the relationship between $PM_{2.5}$ and three features. These features did not match any entries in HMDB, KEGG, or LIPIDMAPS. However, one of these features was also modified by dietary PC1, suggesting the feature may derive from food or some other exogenous source.

Several studies have reported that the relationship between exposure to air pollutants and health outcomes is modified by diet [81]. For example, vitamin C supplementation has been shown to protect from adverse effects of ozone [68, 82, 83]. Similarly, a Mediterranean diet has been shown to modify the risk of long-term exposure to $PM_{2.5}$ and cardiovascular disease mortality [84]. Consumption of animal-based food also modified the relationship between exposure to air pollution and gestational diabetes [85], while a plant-based diet has been shown to protect against adverse effects of $PM_{2.5}$ on cognitive function [86].

We used macronutrient data to find whether broad changes in dietary intake influenced any of the relationships identified through the MWAS approach. To capture patterns in the macronutrient data, we used principal component analysis. We found that one feature was modified by PC1, which captured the most variance in the data, had positive loadings from carbohydrates and vegetable fat, and had negative loadings from protein and animal fat, suggesting a diet with low protein and animal-based products. However, this feature did not match entries in HMDB, KEGG, or LIPIDMAPS. Nearly half of the features that were significantly associated with $PM_{2.5}$ had a significant interaction with PC4, which captured < 1% of the variance in the dietary data and had positive loadings from all four macronutrient categories. This finding suggests that macronutrient intake modifies the relationship between circulating metabolites and long-term exposure to $PM_{2.5}$, albeit only a small variation in the macronutrient intake data was responsible for this. It could also suggest that differences in the intake of nutrients other than the four macronutrients may modify the relationship between these metabolic features and $PM_{2.5}$. Analysis using additional dietary data will be needed to confirm these findings.

We would like to acknowledge some limitations in the analyses presented here. First, due to our small sample size we may have been underpowered to discover associations with small effect sizes. This could explain why we found no associations between circulating metabolites and long-term exposure to PM_{10} and NO_2, which others have reported [24, 87, 88]. Second, we could only provide level 3 confidence in metabolite identification. The accurate identity of these metabolites will be determined as we confirm more metabolites on our platform using chemical standards. Third, due to the small sample size we were

likely underpowered to detect significant interactions, especially with the dietary data, for which we only have 77 observations (∼72% of the total study sample). Fourth, the estimates of air pollution exposure were predicted for each participant at baseline and maybe prone to exposure measurement error. Additionally, the estimates did not account for time spent outside of the home, thus we are unable to account for exposures experienced in geographical locations outside the home, like exposures in occupational settings. However, since most of the participants were retired at the time of the study, occupational exposure is unlikely to confound the relationships observed here. Fifth, since the samples were selected for equal representation of racial/ethnic groups and AD status, we may be prone to selection bias; however, the mean levels of air pollutant exposure were similar between AD cases and controls (Supplementary Table 1). Sixth, our study included people with and without AD but is not a case-control study which may reduce the generalizability of our study to the general population. Seventh, since we were interested in discovering whether presence of metabolic diseases, *APOE ε*4 allele and diet can modify the relationships observed, we used an FDR threshold of 20%, which may be considered lenient and not robust to false positive associations. However, we chose this threshold given the exploratory nature of our question and small sample size, and we believe the findings of effect modification warrant further exploration in larger studies. Eighth, the assumptions of normality required for the regression framework in the metabolome-wide association study framework are not always met, especially when considering the thousands of features investigated for association. Ninth, by restricting our effect modification analyses to features that were significantly associated with exposure, we missed the opportunity to investigate features that did not reach statistical significance and we may overestimate the effect of features selected given that we had previously analyzed the feature in order to decide on its inclusion for further analysis. Finally, while we did adjust our models for known confounders, we are unable to rule out residual confounding from unknown variables.

Despite these limitations, our study has several strengths. Our sample was representative of the neighborhoods in northern Manhattan. We applied high-resolution mass spectrometry-based metabolomics to capture the circulating metabolome in an agnostic manner. Additionally, we had information on history of chronic disease with metabolic consequences and dietary information that has been well validated in the cohort. This study is also a critical step as we expand our analysis to the larger WHICAP cohort.

In conclusion, using an untargeted metabolomics analysis, we applied a multipronged approach to find circulating biochemical signals associated with estimated long-term exposure to $PM_{2.5}$. We found that circulating cysteinylglycine disulfide, a diglyceride, and a dicarboxylic acid are associated with long-term exposure to $PM_{2.5}$, suggesting changes in glutathione turnover, lipid transport and storage, and energy production. We also found changes in several metabolic pathways, that have previously been associated with long-term exposure to $PM_{2.5}$. The relationship between long-term $PM_{2.5}$ and cysteinylglycine disulfide was modified by diet suggesting that dietary changes may influence this relationship. In the future, we plan to identify the features putatively annotated in this analysis. Finally, we plan to investigate the relationship between long-term exposure to air pollutants and circulating metabolites in the larger WHICAP study. This will provide more power to also consider the relationship between long-term exposure to NO_2 and PM_{10} and explore the role of diet in the relationship between long-term exposure to air pollutants and circulating metabolites.

ACKNOWLEDGMENTS

FUNDING

CONFLICT OF INTEREST

The authors have no conflict of interest to report.

DATA AVAILABILITY

Data are available upon request through the WHICAP resource request portal.

SUPPLEMENTARY MATERIAL

The supplementary material is available in the electronic version of this article: https://dx.doi.org/10.3233/JAD-230122.

REFERENCES

[1] Wang Y, Shi L, Lee M, Liu P, Di Q, Zanobetti A, Schwartz JD (2017) Long-term exposure to PM2.5 and mortality among older adults in the Southeastern US. *Epidemiology* **28**, 207-214.

[2] Manisalidis I, Stavropoulou E, Stavropoulos A, Bezirtzoglou E (2020) Environmental and health impacts of air pollution: A review. *Front Public Health* **8**, 14.

[3] Rice MB, Ljungman PL, Wilker EH, Dorans KS, Gold DR, Schwartz J, Koutrakis P, Washko GR, O'Connor GT, Mittleman MA (2015) Long-term exposure to traffic emissions and fine particulate matter and lung function decline in the Framingham heart study. *Am J Respir Crit Care Med* **191**, 656-664.

[4] Ljungman PLS, Li W, Rice MB, Wilker EH, Schwartz J, Gold DR, Koutrakis P, Benjamin EJ, Vasan RS, Mitchell GF, Hamburg NM, Mittleman MA (2018) Long- and short-term air pollution exposure and measures of arterial stiffness in the Framingham Heart Study. *Environ Int* **121**, 139-147.

[5] Béjot Y, Reis J, Giroud M, Feigin V (2018) A review of epidemiological research on stroke and dementia and exposure to air pollution. *Int J Stroke* **13**, 687-695.

[6] Nwanaji-Enwerem JC, Wang W, Nwanaji-Enwerem O, Vokonas P, Baccarelli A, Weisskopf M, Herndon LW, Wiggs JL, Park SK, Schwartz J (2019) Association of long-term ambient black carbon exposure and oxidative stress allelic variants with intraocular pressure in older men. *JAMA Ophthalmol* **137**, 129-137.

[7] Chen J-C, Schwartz J (2008) Metabolic syndrome and inflammatory responses to long-term particulate air pollutants. *Environ Health Perspect* **116**, 612-617.

[8] Gruzieva O, Jeong A, He S, Yu Z, Bont J de, Pinho MGM, Eze IC, Kress S, Wheelock CE, Peters A, Vlaanderen J, Hoogh K de, Scalbert A, Chadeau-Hyam M, Vermeulen RCH, Gehring U, Probst-Hensch N, Melén E (2022) Air pollution, metabolites and respiratory health across the life-course. *Eur Respir Rev* **31**, 220038.

[9] Kulick ER, Wellenius GA, Boehme AK, Joyce NR, Schupf N, Kaufman JD, Mayeux R, Sacco RL, Manly JJ, Elkind MSV (2020) Long-term exposure to air pollution and trajectories of cognitive decline among older adults. *Neurology* **94**, e1782-e1792.

[10] Gao X, Coull B, Lin X, Vokonas P, Spiro A, Hou L, Schwartz J, Baccarelli AA (2021) Short-term air pollution, cognitive performance and nonsteroidal anti-inflammatory drug use in the Veterans Affairs Normative Aging Study. *Nat Aging* **1**, 430-437.

[11] Fu P, Yung KKL (2020) Air pollution and Alzheimer's disease: A systematic review and meta-analysis. *J Alzheimers Dis* **77**, 701-714.

[12] Shi L, Steenland K, Li H, Liu P, Zhang Y, Lyles RH, Requia WJ, Ilango SD, Chang HH, Wingo T, Weber RJ, Schwartz J (2021) A national cohort study (2000–2018) of long-term air pollution exposure and incident dementia in older adults in the United States. *Nat Commun* **12**, 6754.

[13] Nunez Y, Boehme AK, Weisskopf MG, Re DB, Navas-Acien A, van Donkelaar A, Martin RV, Kioumourtzoglou MA (2021) Fine particle exposure and clinical aggravation in neurodegenerative diseases in New York State. *Environ Health Perspect* **129**, 27003.

[14] Uppal K, Walker DI, Liu K, Li S, Go Y-M, Jones DP (2016) Computational metabolomics: A framework for the million metabolome. *Chem Res Toxicol* **29**, 1956-1975.

[15] Li S, Dunlop AL, Jones DP, Corwin EJ (2016) High-resolution metabolomics. *Biol Res Nurs* **18**, 12-22.

[16] Nemmar A, Hoet PHM, Vanquickenborne B, Dinsdale D, Thomeer M, Hoylaerts MF, Vanbilloen H, Mortelmans L, Nemery B (2002) Passage of inhaled particles into the blood circulation in humans. *Circulation* **105**, 411-414.

[17] Xing Y-F, Xu Y-H, Shi M-H, Lian Y-X (2016) The impact of PM2.5 on the human respiratory system. *J Thorac Dis* **8**, E69-E74.

[18] Jia H, Liu Y, Guo D, He W, Zhao L, Xia S (2021) PM2.5-induced pulmonary inflammation via activating of the NLRP3/caspase-1 signaling pathway. *Environ Toxicol* **36**, 298-307.

[19] Liang D, Moutinho JL, Golan R, Yu T, Ladva CN, Niedzwiecki M, Walker DI, Sarnat SE, Chang HH, Greenwald R, Jones DP, Russell AG, Sarnat JA (2018) Use of high-resolution metabolomics for the identification of metabolic signals associated with traffic-related air pollution. *Environ Int* **120**, 145-154.

[20] Bateson TF, Schwartz J (2004) Who is sensitive to the effects of particulate air pollution on mortality? A case-crossover analysis of effect modifiers. *Epidemiology* **15**, 143-149.

[21] Schwartz J, Park SK, O'Neill MS, Vokonas PS, Sparrow D, Weiss S, Kelsey K (2005) Glutathione-S-transferase M1, obesity, statins, and autonomic effects of particles: Gene-by-drug-by-environment interaction. *Am J Respir Crit Care Med* **172**, 1529-1533.

[22] Wiener JM, Tilly J (2002) Population ageing in the United States of America: Implications for public programmes. *Int J Epidemiol* **31**, 776-781.

[23] Kalia V, Belsky DW, Baccarelli AA, Miller GW (2022) An exposomic framework to uncover environmental drivers of aging. *Exposome* **2**, osac002.

[24] Nassan FL, Kelly RS, Kosheleva A, Koutrakis P, Vokonas PS, Lasky-Su JA, Schwartz JD (2021) Metabolomic signatures of the long-term exposure to air pollution and temperature. *Environ Health* **20**, 3.

[25] Colmer J, Hardman I, Shimshack J, Voorheis J (2020) Disparities in PM2.5 air pollution in the United States. *Science* **369**, 575-578.

[26] Jbaily A, Zhou X, Liu J, Lee T-H, Kamareddine L, Verguet S, Dominici F (2022) Air pollution exposure disparities across US population and income groups. *Nature* **601**, 228-233.

[27] Liu J, Clark LP, Bechle MJ, Hajat A, Kim S-Y, Robinson AL, Sheppard L, Szpiro AA, Marshall JD (2021) Disparities in air pollution exposure in the United States by race/ethnicity and income, 1990–2010. *Environ Health Perspect* **129**, 127005.

[28] Kioumourtzoglou M-A, Schwartz J, James P, Dominici F, Zanobetti A (2016) PM2.5 and mortality in 207 US cities: Modification by temperature and city characteristics. *Epidemiology* **27**, 221-227.

[29] Perttula K, Schiffman C, Edmands WMB, Petrick L, Grigoryan H, Cai X, Gunter MJ, Naccarati A, Polidoro S, Dudoit S, Vineis P, Rappaport SM (2018) Untargeted lipidomic features associated with colorectal cancer in a prospective cohort. *BMC Cancer* **18**, 996.

[30] Costanzo M, Caterino M, Sotgiu G, Ruoppolo M, Franconi F, Campesi I (2022) Sex differences in the human metabolome. *Biol Sex Differ* **13**, 30.

[31] Patel MJ, Batch BC, Svetkey LP, Bain JR, Turer CB, Haynes C, Muehlbauer MJ, Stevens RD, Newgard CB, Shah SH (2013) Race and sex differences in small-molecule metabolites and metabolic hormones in overweight and obese adults. *OMICS* **17**, 627-635.

[32] Vardarajan B, Kalia V, Manly J, Brickman A, Reyes-Dumeyer D, Lantigua R, Ionita-Laza I, Jones DP, Miller GW, Mayeux R (2020) Differences in plasma metabolites related to Alzheimer's disease, APOE ε4 status, and ethnicity. *Alzheimers Dement (N Y)* **6**, e12025.

[33] Iliou A, Mikros E, Karaman I, Elliott F, Griffin JL, Tzoulaki I, Elliott P (2021) Metabolic phenotyping and cardiovascular disease: An overview of evidence from epidemiological settings. *Heart* **107**, 1123-1129.

[34] Newgard CB (2017) Metabolomics and metabolic diseases: Where do we stand? *Cell Metab* **25**, 43-56.

[35] Tebani A, Bekri S (2019) Paving the way to precision nutrition through metabolomics. *Front Nutr* **6**, 41.

[36] Peña-Bautista C, Roca M, López-Cuevas R, Baquero M, Vento M, Cháfer-Pericás C (2020) Metabolomics study to identify plasma biomarkers in alzheimer disease: ApoE genotype effect. *J Pharm Biomed Anal* **180**, 113088.

[37] Kulick ER, Elkind MSV, Boehme AK, Joyce NR, Schupf N, Kaufman JD, Mayeux R, Manly JJ, Wellenius GA (2020) Long-term exposure to ambient air pollution, APOE-ε4 status, and cognitive decline in a cohort of older adults in Northern Manhattan. *Environ Int* **136**, 105440.

[38] McKhann GM, Knopman DS, Chertkow H, Hyman BT, Jack CR, Kawas CH, Klunk WE, Koroshetz WJ, Manly JJ, Mayeux R, Mohs RC, Morris JC, Rossor MN, Scheltens P, Carrillo MC, Thies B, Weintraub S, Phelps CH (2011) The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement* **7**, 263-269.

[39] Liu KH, Walker DI, Uppal K, Tran V, Rohrbeck P, Mallon TM, Jones DP (2016) High-resolution metabolomics assessment of military personnel: Evaluating analytical strategies for chemical detection. *J Occup Environ Med* **58**(8 Suppl 1), S53-S61.

[40] Yu T, Park Y, Johnson JM, Jones DP (2009) apLCMS—adaptive processing of high-resolution LC/MS data. *Bioinformatics* **25**, 1930-1936.

[41] Uppal K, Soltow QA, Strobel FH, Pittard WS, Gernert KM, Yu T, Jones DP (2013) xMSanalyzer: Automated pipeline for improved feature detection and downstream analysis of large-scale, non-targeted metabolomics data. *BMC Bioinformatics* **14**, 15.

[42] Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD (2012) The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* **28**, 882-883.

[43] Sampson PD, Richards M, Szpiro AA, Bergen S, Sheppard L, Larson TV, Kaufman JD (2013) A regionalized national universal kriging model using Partial Least Squares regression for estimating annual PM2.5 concentrations in epidemiology. *Atmos Environ (1994)* **75**, 383-392.

[44] Young MT, Bechle MJ, Sampson PD, Szpiro AA, Marshall JD, Sheppard L, Kaufman JD (2016) Satellite-based no2 and model validation in a national prediction model based on universal kriging and land-use regression. *Environ Sci Technol* **50**, 3686-3694.

[45] Kuhn M (2008) Building predictive models in R using the caret package. *J Stat Softw* **28**, 1-26.

[46] Kim-Anh Le Cao FR (2016) mixOmics: Omics Data Integration Project. R package version 6.1.1.

[47] Pang Z, Zhou G, Ewald J, Chang L, Hacariz O, Basu N, Xia J (2022) Using MetaboAnalyst 5.0 for LC–HRMS spectra processing, multi-omics integration and covariate adjustment of global metabolomics data. *Nat Protoc* **17**, 1735-1761.

[48] Li S, Park Y, Duraisingham S, Strobel FH, Khan N, Soltow QA, Jones DP, Pulendran B (2013) Predicting network activity from high throughput metabolomics. *PLoS Comput Biol* **9**, e1003123.

[49] Wishart DS, Guo A, Oler E, Wang F, Anjum A, Peters H, Dizon R, Sayeeda Z, Tian S, Lee BL, Berjanskii M, Mah R, Yamamoto M, Jovel J, Torres-Calzada C, Hiebert-Giesbrecht M, Lui VW, Varshavi D, Varshavi D, Allen D, Arndt D, Khetarpal N, Sivakumaran A, Harford K, Sanford S, Yee K, Cao X, Budinski Z, Liigand J, Zhang L, Zheng J, Mandal R, Karu N, Dambrova M, Schiöth HB, Greiner R, Gautam V (2022) HMDB 5.0: The Human Metabolome Database for 2022. *Nucleic Acids Res* **50**, D622-D631.

[50] Uppal K, Walker DI, Jones DP (2017) xMSannotator: An R package for network-based annotation of high-resolution metabolomics data. *Anal Chem* **89**, 1063-1067.

[51] Schymanski EL, Jeon J, Gulde R, Fenner K, Ruff M, Singer HP, Hollender J (2014) Identifying small molecules via high resolution mass spectrometry: Communicating confidence. *Environ Sci Technol* **48**, 2097-2098.

[52] Murcray CE, Lewinger JP, Gauderman WJ (2009) Gene-environment interaction in genome-wide association studies. *Am J Epidemiol* **169**, 219-226.

[53] Smith W, Mitchell P, Reay EM, Webb K, Harvey PW (1998) Validity and reproducibility of a self-administered food frequency questionnaire in older people. *Aust N Z J Public Health* **22**, 456-463.

[54] Luchsinger JA, Tang M-X, Siddiqui M, Shea S, Mayeux R (2004) Alcohol intake and risk of dementia. *J Am Geriatr Soc* **52**, 540-546.

[55] Luchsinger JA, Tang M-X, Shea S, Mayeux R (2002) Caloric intake and the risk of Alzheimer disease. *Arch Neurol* **59**, 1258-1263.

[56] Luchsinger JA, Tang M-X, Shea S, Mayeux R (2003) Antioxidant vitamin intake and risk of Alzheimer disease. *Arch Neurol* **60**, 203-208.

[57] Gu Y, Nieves JW, Stern Y, Luchsinger JA, Scarmeas N (2010) Food combination and Alzheimer disease risk: A protective diet. *Arch Neurol* **67**, 699-706.

[58] Willett W (1998) *Nutritional Epidemiology*, Oxford University Press, New York, NY.

[59] Liu Y, Hyde AS, Simpson MA, Barycki JJ (2014) Emerging regulatory paradigms in glutathione metabolism. *Adv Cancer Res* **122**, 69-101.

[60] Kaur A, Gautam R, Srivastava R, Chandel A, Kumar A, Karthikeyan S, Bachhawat AK (2017) ChaC2, an enzyme for slow turnover of cytosolic glutathione. *J Biol Chem* **292**, 638-651.

[61] Mungrue IN, Pagnon J, Kohannim O, Gargalovic PS, Lusis AJ (2009) CHAC1/MGC4504 is a novel proapoptotic component of the unfolded protein response, downstream of the ATF4-ATF3-CHOP cascade. *J Immunol* **182**, 466-476.

[62] Reyes-Caballero H, Rao X, Sun Q, Warmoes MO, Lin P, Sussan TE, Park B, Fan TW-M, Maiseyeu A, Rajagopalan S, Girnun GD, Biswal S (2019) Air pollution-derived particulate matter dysregulates hepatic Krebs cycle, glucose and lipid metabolism in mice. *Sci Rep* **9**, 17423.

[63] Mortensen PB (1992) Formation and degradation of dicarboxylic acids in relation to alterations in fatty acid oxidation in rats. *Biochim Biophys Acta* **1124**, 71-79.

[64] Tserng KY, Jin S-J (1991) Metabolic conversion of dicarboxylic acids to succinate in rat liver homogenates. A stable isotope tracer study. *J Biol Chem* **266**, 2924-2929.

[65] Spector AA (1984) Plasma lipid transport. *Clin Physiol Biochem* **2**, 123-134.

[66] Ahmadian M, Wang Y, Sul HS (2010) Lipolysis in adipocytes. *Int J Biochem Cell Biol* **42**, 555-559.

[67] Mendez R, Zheng Z, Fan Z, Rajagopalan S, Sun Q, Zhang K (2013) Exposure to fine airborne particulate matter induces macrophage infiltration, unfolded protein response, and lipid deposition in white adipose tissue. *Am J Transl Res* **5**, 224-234.

[68] Menni C, Metrustry SJ, Mohney RP, Beevers S, Barratt B, Spector TD, Kelly FJ, Valdes AM (2015) Circulating levels of antioxidant vitamins correlate with better lung function and reduced exposure to ambient pollution. *Am J Respir Crit Care Med* **191**, 1203-1207.

[69] Chu H, Huang F-Q, Yuan Q, Fan Y, Xin J, Du M, Wang M, Zhang Z, Ma G (2021) Metabolomics identifying biomarkers of PM2.5 exposure for vulnerable population: Based on a prospective cohort study. *Environ Sci Pollut Res Int* **28**, 14586-14596.

[70] Walker DI, Lane KJ, Liu K, Uppal K, Patton AP, Durant JL, Jones DP, Brugge D, Pennell KD (2019) Metabolomic assessment of exposure to near-highway ultrafine particles. *J Expo Sci Environ Epidemiol* **29**, 469-483.

[71] Huang Q, Zhang J, Luo L, Wang X, Wang X, Alamdar A, Peng S, Liu L, Tian M, Shen H (2015) Metabolomics reveals disturbed metabolic pathways in human lung epithelial cells exposed to airborne fine particulate matter. *Toxicol Res* **4**, 939-947.

[72] Xu Y, Wang W, Zhou J, Chen M, Huang X, Zhu Y, Xie X, Li W, Zhang Y, Kan H, Ying Z (2019) Metabolomics analysis of a mouse model for chronic exposure to ambient PM2.5. *Environ Pollut* **247**, 953-963.

[73] Paul BD, Sbodio JI, Snyder SH (2018) Cysteine metabolism in neuronal redox homeostasis. *Trends Pharmacol Sci* **39**, 513-524.

[74] Trushina E, Dutta T, Persson X-MT, Mielke MM, Petersen RC (2013) Identification of altered metabolic pathways in plasma and CSF in mild cognitive impairment and Alzheimer's disease using metabolomics. *PLoS One* **8**, e63644.

[75] Hood RB, Liang D, Tang Z, Kloog I, Schwartz J, Laden F, Jones D, Gaskins AJ (2022) Length of PM2.5 exposure and alterations in the serum metabolome among women undergoing infertility treatment. *Environ Epidemiol* **6**, e191.

[76] Wang J, Lin L, Huang J, Zhang J, Duan J, Guo X, Wu S, Sun Z (2022) Impact of PM2.5 exposure on plasma metabolome in healthy adults during air pollution waves: A randomized, crossover trial. *J Hazard Mater* **436**, 129180.

[77] Liu C-C, Kanekiyo T, Xu H, Bu G (2013) Apolipoprotein E and Alzheimer disease: Risk, mechanisms, and therapy. *Nat Rev Neurol* **9**, 106-118.

[78] Tracey TJ, Steyn FJ, Wolvetang EJ, Ngo ST (2018) Neuronal lipid metabolism: Multiple pathways driving functional outcomes in health and disease. *Front Mol Neurosci* **11**, 10.

[79] Du X, Jiang S, Zeng X, Zhang J, Pan K, Zhou J, Xie Y, Kan H, Song W, Sun Q, Zhao J (2018) Air pollution is associated with the development of atherosclerosis via the cooperation of CD36 and NLRP3 inflammasome in ApoE-/- mice. *Toxicol Lett* **290**, 123-132.

[80] Campen MJ, Lund AK, Knuckles TL, Conklin DJ, Bishop B, Young D, Seilkop S, Seagrave J, Reed MD, McDonald JD (2010) Inhaled diesel emissions alter atherosclerotic plaque composition in ApoE(-/-) mice. *Toxicol Appl Pharmacol* **242**, 310-317.

[81] Whyand T, Hurst JR, Beckles M, Caplin ME (2018) Pollution and respiratory disease: Can diet or supplements help? A review. *Respir Res* **19**, 79.

[82] Samet JM, Hatch GE, Horstman D, Steck-Scott S, Arab L, Bromberg PA, Levine M, McDonnell WF, Devlin RB (2001) Effect of antioxidant supplementation on ozone-induced lung injury in human subjects. *Am J Respir Crit Care Med* **164**, 819-825.

[83] Romieu I, Sienra-Monge JJ, Ramírez-Aguilar M, Téllez-Rojo MM, Moreno-Macías H, Reyes-Ruiz NI, del Río-Navarro BE, Ruiz-Navarro MX, Hatch G, Slade R, Hernández-Avila M (2002) Antioxidant supplementation and lung functions among children with asthma exposed to high levels of air pollutants. *Am J Respir Crit Care Med* **166**, 703-709.

[84] Lim CC, Hayes RB, Ahn J, Shao Y, Silverman DT, Jones RR, Thurston GD (2019) Mediterranean diet and the association between air pollution and cardiovascular disease mortality risk. *Circulation* **139**, 1766-1775.

[85] Hehua Z, Yang X, Qing C, Shanyan G, Yuhong Z (2021) Dietary patterns and associations between air pollution and gestational diabetes mellitus. *Environ Int* **147**, 106347.

[86] Zhu A, Chen H, Shen J, Wang X, Li Z, Zhao A, Shi X, Yan L, Zeng Y, Yuan C, Ji JS (2022) Interaction between plant-based dietary pattern and air pollution on cognitive function: A prospective cohort analysis of Chinese older adults. *Lancet Reg Health West Pac* **20**, 100372.

[87] Ward-Caviness CK, Nwanaji-Enwerem JC, Wolf K, Wahl S, Colicino E, Trevisi L, Kloog I, Just AC, Vokonas P, Cyrys J, Gieger C, Schwartz J, Baccarelli AA, Schneider A, Peters A (2016) Long-term exposure to air pollution is associated with biological aging. *Oncotarget* **7**, 74510-74525.

[88] Liao J, Gheissari R, Thomas DC, Gilliland FD, Lurmann F, Islam KT, Chen Z (2022) Transcriptomic and metabolomic associations with exposures to air pollutants among young adults with childhood asthma history. *Environ Pollut* **299**, 118903.