Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____	4/20/2023
Sabrina S. Chow	Date

The Role of Metabolic Perturbations in Mediating the Effects of Ambient Air Pollution on Lung Cancer in the Cancer Prevention Studies

By

Sabrina S. Chow
Master of Public Health

Environmental Health-Epidemiology

Donghai Liang, PhD, MPH
Committee Chair

Jeremy A. Sarnat, ScD, MSc
Committee Chair

The Role of Metabolic Perturbations in Mediating the Effects of Ambient Air Pollution on Lung Cancer in the Cancer Prevention Studies

By

Sabrina S. Chow

B.S.
Brandeis University
2021

Thesis Committee Chairs:
Donghai Liang, Ph.D., MPH
Jeremy A. Sarnat, ScD, MSc

An abstract of
a thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Public Health
in Environmental Health-Epidemiology
2023

Abstract

The Role of Metabolic Perturbations in Mediating the Effects of Ambient Air Pollution on Lung Cancer in the Cancer Prevention Studies

By Sabrina S. Chow

Background. Exposure to ambient air pollution is an established risk factor for lung cancer. Despite this, the identification of underlying biological mechanisms of air pollution carcinogenicity remains uncertain. To address these knowledge gaps, we applied high-resolution metabolomics to detect metabolic signatures of exogeneous air pollution exposures and endogenous processes involved in lung carcinogenesis in a well-established U.S. cancer cohort. **Methods.** A total of 1,360 participants (680 matched lung cancer and control pairs) within the established Cancer Prevention Study cohorts completed comprehensive questionnaires during enrollment and follow-up to assess changes in personal and lifestyle factors and medical conditions. Participant's plasma metabolome from non-fasting blood samples was profiled with ultrahigh-performance liquid chromatography-tandem mass spectrometry. Assessment of exposure to six ambient air pollutants, including carbon monoxide (CO), nitrogen dioxide (NO_2), particulate matter (PM_{10}), fine particulate matter ($PM_{2.5}$), sulfur dioxide (SO_2), and ozone (O_3), was conducted using spatiotemporally-resolved models based on residential address at blood draw. We conducted a metabolome-wide association study using multivariate linear regression models to assess associations of air pollution and lung cancer with a meet-in-the-middle approach. Metabolites significant at the FDR < 0.2 level in the air pollution model were analyzed in the lung cancer model. High-dimensional mediation analysis was used as a secondary analysis to compare results from the meet-in-the-middle analysis. **Results.** Among 1,232 metabolic features extracted, seven were significantly associated with air pollution exposure and lung cancer incidence at the FDR < 0.2 level in the meet-in-the-middle analysis. Six features were significant via high-dimensional mediation analysis. All confirmed metabolites are enriched within peptide, lipid, and amino acid pathways. The metabolites gamma-glutamylglutamine and gamma-glutamylmethionine were each significantly associated with CO, NO_2, and PM_{10} exposure and lung cancer incidence in both analyses. **Conclusion.** This is the largest prospective metabolomics study examining biological perturbations associated with air pollution exposure and lung cancer outcomes. The findings provide an indication of association between air pollution and lung cancer mediated via peptide metabolism. Findings from this study support future studies to further clarify the specific role of these identified metabolites and pathways as mediators of air pollution carcinogenicity.

The Role of Metabolic Perturbations in Mediating the Effects of Ambient Air Pollution on Lung Cancer in the Cancer Prevention Studies

By

Sabrina S. Chow

B.S.
Brandeis University
2021

Thesis Committee Chairs:
Donghai Liang, Ph.D., MPH
Jeremy A. Sarnat, ScD, MSc

A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Public Health
in Environmental Health-Epidemiology
2023

Acknowledgements

First, and foremost, I want to thank my advisors, Drs. Donghai Liang and Jeremy A. Sarnat, for all their support and encouragement during not only the process of completing this thesis, but during my entire MPH at Rollins.

Donghai, thank you for believing in me (and hiring me) before I even moved to Atlanta two years ago. How time as flown. Had I not applied for that research position, I don't know where I would even be on my public health journey and development as a researcher. Under your support and mentorship, I have grown so much as a student and researcher. You have always been so supportive of my explorations in environmental health, celebrated my successes as your own, but most importantly, pushed me to do the best work that I could do, even if I didn't think I could.

Jeremy, you were the one that introduced me to exposure science and helped me find a space in environmental health where I can see myself for the rest of my career. I am eternally grateful for the guidance and mentorship that you have given me, engaging me in critical conversations, and giving me a multitude of opportunities to explore budding interests and receive training I never thought imaginable during a master's program. I am excited to continue on my Ph.D. journey at Emory, learning and training alongside both of you. Looks like you're stuck with me! ☺

I would also like to thank Dr. Ying Wang at the American Cancer Society (ACS) for the opportunity to work with the Cancer Prevention Studies and giving me the opportunity to see how my volunteer efforts through ACS are used on the research side. I am also grateful for the rest of the team at ACS, including Emily, Michelle, and Ryan for their assistance and guidance throughout this whole thesis process with data cleaning, management, and collection.

To my friends and family, thank you for supporting me at my worst and showing me that there is a light at the end of the tunnel. And it's here, right in front of me!

Table of Contents

Introduction

Long-term exposure to ambient air pollution is a leading environmental risk factor for respiratory diseases, including asthma, chronic obstructive pulmonary disease (COPD), and lung cancer[1-4]. Despite declining trends in lung cancer incidence and mortality, it is estimated that almost 240,000 individuals in the U.S. will be diagnosed with lung cancer in 2023, with almost 130,000 individuals dying from lung cancer[5]. Epidemiological evidence demonstrating a potential causal association between ambient air pollution exposure and lung cancer mortality has been reported in numerous studies in large prospective cohorts around the world[6-9]. However, due to the complex composition of air pollution, the biological mechanisms of air pollution carcinogenicity and the processes that may lead to individual susceptibility are still understudied. This area of research is further limited due to a lack of sensitive biomarkers to measure internal exposure and corresponding health responses.

High-resolution metabolomics (HRM), combining high-resolution mass spectrometry with chromatographic separation, has emerged over the past few decades as a key analytical platform in the identification of exogenous and endogenous metabolic features associated with air pollution-related exposures. While interest in metabolomics analyses is continuing to grow, concerns about the consistency and generalizability of results exist due to the variation across study cohorts, analytical platforms, and reported results[10]. Specifically, a majority of published studies using metabolomics has been in small-scale, panel-based studies using targeted metabolomics workflows[10]. More studies have recently used meet-in-the-middle (MITM) and high-dimensional mediation (HDMA) approaches for identifying potential intermediate biomarkers in prospective cohorts to help inform the causal pathway from external exposure to corresponding health response[11]. Molecular biomarkers associated with both exposure and response help validate causal hypotheses via the pathway perturbation paradigm[12]. Metabolic perturbations associated with air pollution exposure and adverse health outcomes such as asthma, coronary heart disease, as well as other respiratory and cardiovascular diseases have been proposed using MITM approaches[12,13]. However, no study has used either the MITM or HMDA approach to better understand potential biological pathways and intermediate markers for the air pollution and lung cancer relationship.

To address these critical knowledge gaps, we conducted a metabolome-wide association study (MWAS) using untargeted HRM with participants in the Cancer Prevention Study-II (CPS-II) Nutrition and CPS-3 Cohorts, an established prospective cancer cohort managed by the American Cancer Society (ACS), to identify metabolic perturbations associated with ambient air pollution exposure and lung cancer incidence. Our results serve to create a foundation for further studies to understand biological mechanisms associated with long-term ambient air pollution and corresponding lung cancer risk.

Methods

Study Population and Design

Participants included in this study were a part of the Cancer Prevention Study (CPS)-II Nutrition and CPS-3 Cohorts. CPS-II was a prospective cohort established by ACS in 1982 that enrolled almost 1.2 million participants across all 50 states, the District of Columbia, and Puerto Rico. The CPS-II Nutrition Cohort, a subset of the original CPS-II Cohort, was established in 1992 to investigate the relationship between lifestyle factors, exposure, and cancer. This cohort includes nearly 200,000 men and women aged 50-74 years from 21 U.S. states. CPS-3 was launched in December 2013 as the next iteration of CPS and recruited over 300,000 participants aged 30-65 years. Participants in both cohorts completed an initial questionnaire at time of enrollment and follow-up surveys every 2-3 years until study completion. Follow-up for the CPS-II Nutrition Cohort ended in 2015, and follow-up for CPS-3 will continue through 2043. Almost all participants in CPS-3 provided a non-fasting blood sample at enrollment between 2006 and 2013, while 39,200 participants in the CPS-II Nutrition Cohort provided non-fasting blood samples between 1998 and 2001. All participants were cancer-free at enrollment. The cohort and sample collection processes for CPS-II and CPS-3 are described in detail elsewhere[14,15]. Study protocols for both cohorts were approved by the Emory University (Atlanta, GA) Institutional Review Board.

In the current study, 1,162 individuals from the CPS-II Nutrition Cohort and 326 individuals from CPS-3 were included in the analysis. All participants were cancer-free at time of blood draw. During follow-up, 744 individuals between both cohorts were diagnosed and verified as lung cancer cases through medical records and state cancer registries. Cases with in-situ cases (n = 2) and missing air pollution assessment data (n = 61) were excluded from the final analysis. As a result, 521 and 159 lung cancer cases from the CPS-II Nutrition and CPS-3 Cohorts were included in the final analysis, respectively. Healthy controls were matched 1:1 to cases on sex, race/ethnicity, age at blood draw (± 6 months), and date of blood draw (± 30 days). A total of 1,360 individuals were included in the current analysis (Supplement 3).

Air Pollution Assessment

All participants provided a residential address at time of blood draw, which was used to retrospectively assign individual-level exposures to six ambient air pollutants including carbon monoxide (CO), nitrogen dioxide (NO_2), respirable particulate matter (PM_{10}), fine particulate matter ($PM_{2.5}$), ozone (O_3), and sulfur dioxide (SO_2). Spatiotemporally resolved models were used to generate participant-specific residential exposures to each of the air pollutants. Specifically, concentrations were assigned to geocoded census blocks based on participant residence at time of blood draw. The year of blood draw was used to calculate a one-year mean exposure window to estimate current level of exposure. Concentrations for all air pollutants were obtained from the Center for Air, Climate, and Energy Solutions (CACES) based on land use regression models.

Metabolomic Profiling

Metabolomic profiling on plasma samples was conducted at Metabolon, Inc. (Durham, NC) using ultrahigh-performance liquid chromatography-tandem mass spectrometry (UPLC-MS/MS). Additional analysis protocols are described elsewhere[16]. As reported by Wang et al.[17], serum samples were treated with methanol to precipitate proteins. Four sample fractions were

dried and reconstituted in different solvents for measurement under four different platforms. Two fractions were analyzed by separate reverse-phase UPLC-MS/MS methods with positive-ion-mode electrospray ionization (ESI). One fraction was analyzed by reverse-phase UPLC-MS/MS method with negative-ion-mode ESI. The final analysis was by hydrophilic interaction chromatography UPLCMS/MS with negative-ion-mode ESI. Individual metabolites were identified by comparison with a chemical library consisting of >3,300 commercially available purified standard compounds.

A total of 1,401 metabolites were detected. Duplicates of 24 samples and triplicates of 10 samples from 34 CPS-II participants, as well as samples from Metabolon, were used as quality control observations to assess the reproducibility of the platform, using intraclass correlation coefficient (ICC) calculations. The median ICC was 0.8486, indicating high analytical reproducibility.

Data Analysis

Associations between single air pollutants and metabolic features were conducted using multivariable linear regression models. The relative concentration of each metabolite was log-transformed, followed by auto-scaling to approximate a normal distribution, allowing for comparison on an equal scale. Based on existing literature and *a priori* criteria[18,19], models were adjusted for age at blood draw (continuous), sex (categorical), body mass index (BMI; continuous), race (categorical: white, non-white, unknown), smoking status (categorical: never, former, current, and unknown), highest level of education (categorical: less than high school, high school graduate, some college/technical school/2-year degree, college graduate, graduate school, unknown), multivitamin use (recently, somewhat recently, rarely/never, unknown), average weekly servings of fruits and vegetables (continuous), alcohol use (categorical: none, <1 per day, 1+ per day, unknown), and passive smoke exposure (yes, no, unknown). Benjamini-Hochberg false discovery rate (FDR) procedures were used to control for multiple comparisons[20]. All associations observed at FDR < 0.2 were considered statistically significant.

Using significant metabolic features from the air pollution MWAS, we then used a modified meet-in-the-middle (MITM) modeling to identify significant metabolic features associated with both ambient air pollution exposure and lung cancer incidence. Traditional MITM approaches use the same set of metabolites in the exposure and outcome MWAS, and then compare results to identify overlapping metabolites[11,12]. We used a modified MITM model, where in the outcome MWAS, instead of putting in the original set of metabolites, we only analyzed metabolites that were significant in exposure MWAS. Multivariable linear regression models were again used to examine the association between lung cancer status (case or control) and metabolic features. Models in this analysis were identical to the air pollution models.

As a secondary analysis we conducted a high dimensional mediation analysis (HDMA) to compare significant metabolic features from the MITM analysis. We utilized the R package, HIMA, a novel package used for estimating and testing high-dimensional mediation effects in -omics-related studies. Details on the development and statistical methods of the package are detailed elsewhere[21]. While the MITM approach is an exploratory analysis that looks at significant metabolites in exposure and outcome models separately, HDMA is a true mediation

analysis, which allows us to formally assess the steps in the causal association between air pollution exposure and lung cancer incidence. We also conducted several sensitivity analyses controlling for covariates such as occupational exposure, lung cancer stage and etiology, and found the results remained consistent.

Results

A total of 1,360 participants were included in the analysis, 1,042 from the CPS-II Nutrition Cohort and 318 from CPS-3. The average age of participants was 66.4 $\pm$ 8.83 years (Table 1). A majority of participants were former smokers (51.4%), followed by never smokers (33.4%), with most (68.7%) living with a smoker during their childhood (age 0-18 years). Just over 50% of the total study population completed some college, technical school, or a two-year degree (30%) or was a college graduate (22.6%). Levels to O_3 had the largest variance among study participants (48.2 $\pm$ 6.5 ppb), followed by NO_2 (12.1 $\pm$ 5.9 ppb) and PM_{10} (20.9 $\pm$ 5.9 $\mu g/m^3$).

We analyzed 1,401 metabolites that were detected from metabolomic profiling. Of these, 1,163 were annotated with level 1 evidence by matching to authentic chemical references. After filtering for metabolites that were undetectable in more than 90% of samples (n = 170), 1,232 metabolites were included in the analysis. Among these 1,232 metabolites, 257 metabolites were significantly associated with at least one air pollutant. Among these metabolites, 207 metabolites have been previously annotated with level 1 evidence by Metabolon, Inc., while 50 could not be annotated. Of metabolites with known identities, a majority are central in amino acid (28%), lipid (32%), and xenobiotic pathways (21%). The greatest number of significant metabolites were associated with $PM_{2.5}$ (124), followed by NO_2 (90) (Table 2). In the MITM analysis, three unique, annotated metabolites were significantly associated with at least one air pollutant and lung cancer risk (Table 3). Three metabolites were significantly associated with NO_2 exposure and lung cancer risk, two of which occur along the gamma-glutamyl amino acid pathway. Five metabolites (three confirmed metabolites with known identities and two unknown) were significant in the HDMA approach. A majority of the known metabolites in the MITM and HDMA analyses were along peptide metabolism-related super pathways (33%), most notably the gamma-glutamyl amino acid sub pathway, and unidentified (33%) (Table 3). The remaining confirmed metabolites are along the amino acid, phenylalanine metabolism pathway (17%) and lipid, diacylglycerol pathway (17%).

A single metabolite, gamma-glutamylglutamine, was positively associated with CO exposure in both the MITM and HDMA analyses (Table 3). This metabolite was also positively associated with NO_2 and PM_{10} in the MITM analysis. Gamma-glutamylmethionine was significantly associated with NO_2 and PM_{10} exposure in HDMA. Palmitoleoyl-linoleoyl-glycerol (16:1/18:2) [1]* was the only metabolite in the MITM analysis that was negatively associated with an individual air pollutant (NO_2). N-acetylphenylalanine and X-23654 were both negatively associated with NO_2 and $PM_{2.5}$ exposure in the HDMA analysis, respectively.

We conducted a number of sensitivity analyses to examine the robustness of our model while adjusting for additional potential confounders. After including family history of lung cancer and lung cancer stage in our model, we observed no differences in the number of significant

metabolites. We also conducted a sensitivity analysis using a traditional MITM approach. With this analytical approach, we found no significant metabolites among any air pollutant.

Discussion

With 1,360 participants, we characterized and identified several metabolic perturbations associated with both air pollution exposure and lung cancer incidence. Notably, we identified several pathways, including glutathione, glutamine, methionine, diacylglycerol, and phenylalanine, that have been linked to both air pollution exposure and lung cancer incidence. These pathways have implications in air pollution and cancer-related oxidative stress, nucleic acid damage, and cancer cell proliferation. Based on our knowledge, this is the largest metabolomics study on air pollution and lung cancer to date. Given the hypothesis-generating nature of untargeted metabolomic investigations, however, future hypothesis testing studies are warranted to validate our findings and explore the potential development of significant metabolites as sensitive biomarkers for assessing internal exposures to air pollution and lung cancer incidence. These specific perturbations have been associated with multiple different cancer types, including osteosarcoma, pancreatic, and kidney[22,23].

A key finding was the identification of several metabolic perturbations associated with peptide metabolism, specifically gamma-glutamyl amino acid metabolism, in both the MITM and HDMA analyses. In the MITM approach, exposure to CO, NO_2, and PM_{10} led to significant changes in metabolic intensities of gamma-glutamylglutamine and gamma-glutamyl methionine. Similarly, these perturbations were also significant in participants who later developed lung cancer. We saw consistent findings with HDMA, where gamma-glutamylglutamine and gamma-glutamyl methionine mediated the association between CO, NO_2, and PM_{10} exposure and lung cancer incidence. Gamma-glutamyl amino acids are formed when an amino acid and enzyme are catalyzed through gamma-glutamyl transpeptidase, with a gamma-glutamyl enzyme bound intermediate[24], and are important mediators in glutathione (GSH) metabolism pathways[25]. GSH also has antioxidant scavenging properties and supports cellular regulation, such as gene expression, DNA and protein synthesis, cytokine production, and immune response[25,26]. This may indicate chronic exposures to air pollution may lead to changes in antioxidation and cellular processes. Cancerous tumors from individuals with non-small cell lung cancer (NSCLC), which make up more than 80% of lung cancer cases, have shown high levels of gamma-glutamyl transpeptidase compared to non-cancerous tissue, increasing the uptake of GSH in cells. Previous studies have hypothesized that elevated levels of GSH and GSH detoxifying enzymes assist with tumor chemoresistance[27-29]. Mitochondrial dysfunction, a common mechanism to cause cell death, has also been found to kill malignant NSCLC carcinomas, which are resistant to conventional chemotherapy[30,31]. In our analysis, we found that the intensity of statistically significant gamma-glutamyl amino acids were higher in lung cancer cases compared to controls, which could lead to downstream increased levels of glutamine and methionine. Glutamine is essential for not only nucleotide and amino acid synthesis, but also cancer cell growth and proliferation. One study with participants who had early-stage lung cancer (Ia and Ib) had increases in glutamine levels compared to controls, while another study with all lung cancer stages (I-IV) saw overall decreases in glutamine levels compared to controls[32,33]. GSH metabolism participates in a number of different biological processes, and our observed findings are consistent with previous lung carcinogenicity-related literature, demonstrating a potential

role of GSH metabolism in mediating the association between ambient air pollution and lung cancer incidence. The current study did not control for lung cancer stage, but future analyses looking at differences in gamma-glutamylglutamine intensity based on stage could provide additional insights on changes in perturbations in participants that are cancer-free.

We also found that NO_2 exposure and lung cancer incidence were both associated with increased perturbations of N-acetylphenylalanine, a downstream metabolite in phenylalanine metabolism, created when acetyl coenzyme A (acetyl-CoA) reacts with L-phenylalanine. Acetyl-CoA plays an important role in cancer cell growth, helping to provide necessary adenosine triphosphate (ATP) to promote cancer cell growth, while inhibiting normal cell growth[34,35]. Phenylalanine, a glycogenic amino acid, can produce glucose through the citric acid cycle, which cancer cells can utilize as an energy source during rapid proliferation. One study found that L-phenylalanine and phenylalanylphenyalanine metabolite levels were increase in lung cancer patients compared to controls and verified phenylalanylphenyalanine as a diagnostic biomarker for differentiating lung cancer and tuberculosis patients[35]. Perturbations in phenylalanine pathways have previously been associated with short-term and long-term traffic-related air pollution (TRAP), specifically NO_2[36,37]. Phenylalanine levels tend to be higher in advanced lung cancer cases. As an upstream amino acid for tyrosine and neurotransmitter production, high levels of phenylalanine are caused by downregulation of metabolism-related genes in NSCLC tumors[38]. This suggests that late-stage lung cancer cells may not be able to metabolize phenylalanine[33].

Finally, we also found a perturbation associated with lipid metabolism, specifically diacylglycerol metabolism. Exposure to NO_2 was associated with a significant decrease in palmitoleoyl-linoleoyl-glycerol (16:1/18:2) metabolic intensity, which is an intermediate metabolite in triacylglycerol synthesis. Previous studies consistently show how ambient air pollution exposure leads to oxidative stress, inflammation, and nucleic acid damage[39-41]. Diacylglycerols are secondary lipid messengers that transduce signals downstream of hematopoietic cell receptors and have implications in activation, proliferation, and migration of adaptive and innate immune cells[42]. Perturbations in diacylglycerols may influence cell signaling, leading to downstream affects that affect all systems in the body, including the respiratory and immune system.

Our analysis included several notable strengths and limitations. This study, to our knowledge, is the first study that has examined metabolic perturbations associated with both air pollution exposure and lung cancer incidence, which also has the largest sample size among existing air pollution metabolomics applications. The CPS cohorts have been central in providing evidence of the association between long-term air pollution exposures and mortality[6,8,43-46]. Previous studies also mostly focus on lung cancer mortality[43-45], while our study looked at incidence. Our study also utilized both a MITM and HDMA analytical approach to examine the association between ambient air pollution exposure and lung cancer incidence, discovering consistent results between both approaches. With multiple significant metabolites between the analytical approaches, this helps to not only validate the potential role that these perturbations play in the causal relationship between ambient air pollution and lung cancer, but also show the benefit of using both analytical methods in tandem when doing untargeted metabolomics work.

However, there are a number of limitations in our study, including those inherently relating to omics-based analysis techniques and retrospective air pollution assessment. Also, even though we used spatio-temporal air pollution models to estimate individual exposures based on residential address at time of blood draw, it is possible that residual exposure measurement error may still be present and serve to obscure true associations. Additionally, blood samples used for metabolic profiling and analysis were collected while all participants were cancer-free. In some cases, lung cancer diagnoses happened over 30 years after sample collection, which combined with an imperfect proxy of air pollution exposure, may introduce bias into our true risk estimation.

In this analysis, we decided to conduct single air pollutant modelling, assuming that each pollutant individually affects lung cancer incidence. Even though single pollutant modeling may not capture the true effects of air pollutant mixtures, we were able to control for co-pollutant confounding, which stems from covariance among multiple air pollutants. Future studies should use advanced modelling approaches to model multiple pollutant exposures to better assess the association between air pollution mixtures and lung cancer incidence. Due to the high volume of metabolic features in metabolomics studies, this increases the likelihood of Type I errors. To minimize the chance of a high false discovery, we utilized a Benjamini-Hochberg False Positive correction to counteract the risk of false positives from multiple comparisons and Type 1 errors. Also, since the study population was predominantly white and older, our study may not be generalizable to the general U.S. population and additional studies should be done to test to reproducibility of identified perturbations and the generalizability and success of the conducted analytical methods in more diverse populations.

Conclusions

Using advanced metabolomics profiling MITM and HDMA analytical approaches, we were able to identify multiple metabolic perturbations that may mediate the association between ambient air pollution exposure and lung cancer incidence, with 1,360 participants in the CPS-II Nutrition and CPS-3 cohorts. These perturbations show an increased risk in oxidative stress, cell signaling and proliferation, and inflammation. Collectively, these results further showcase the use of metabolic markers and metabolomics as a novel analytical method for assessing how air pollution toxicity, at the molecular level, affects lung cancer outcomes. Further identification and understanding of these metabolic perturbations may help lead to the development of sensitive biomarkers and targeted interventions to mitigate long-term adverse health effects from an early stage.

References

1. Mukharesh L, Phipatanakul W, Gaffin JM. Air pollution and childhood asthma. Curr Opin Allergy Clin Immunol 2023;23(2):100-110. DOI: 10.1097/ACI.0000000000000881.
2. Neophytou AM, Lutzker L, Good KM, et al. Associations between prenatal and early-life air pollution exposure and lung function in young children: Exploring influential windows of exposure on lung development. Environ Res 2023;222:115415. DOI: 10.1016/j.envres.2023.115415.
3. Wang Y, Du Z, Zhang Y, et al. Long-term exposure to particulate matter and COPD mortality: Insights from causal inference methods based on a large population cohort in southern China. Sci Total Environ 2023;863:160808. DOI: 10.1016/j.scitotenv.2022.160808.
4. Wang Q, Liu S. The Effects and Pathogenesis of PM2.5 and Its Components on Chronic Obstructive Pulmonary Disease. Int J Chron Obstruct Pulmon Dis 2023;18:493-506. DOI: 10.2147/COPD.S402122.
5. Siegel RL, Miller KD, Wagle NS, Jemal A. Cancer statistics, 2023. CA Cancer J Clin 2023;73(1):17-48. DOI: 10.3322/caac.21763.
6. Pope CA, 3rd, Thun MJ, Namboodiri MM, et al. Particulate air pollution as a predictor of mortality in a prospective study of U.S. adults. Am J Respir Crit Care Med 1995;151(3 Pt 1):669-74. (In eng). DOI: 10.1164/ajrccm/151.3_Pt_1.669.
7. Turner MC, Andersen ZJ, Baccarelli A, et al. Outdoor air pollution and cancer: An overview of the current evidence and public health recommendations. CA Cancer J Clin 2020 (In eng). DOI: 10.3322/caac.21632.
8. Krewski D, Jerrett M, Burnett RT, et al. Extended follow-up and spatial analysis of the American Cancer Society study linking particulate air pollution and mortality. Res Rep Health Eff Inst 2009(140):5-114; discussion 115-36. (In eng).
9. Fu J, Jiang D, Lin G, Liu K, Wang Q. An ecological analysis of PM2.5 concentrations and lung cancer mortality rates in China. BMJ Open 2015;5(11):e009452. (In eng). DOI: 10.1136/bmjopen-2015-009452.
10. Liang D, Li, Z., Vlaanderen, J., Tang, Z., Vermeulen, R., Sarnat, J.A. Systematic Review on High Resolution Metabolomics Application in Air Pollution Health Research: Current Progress, Analytical Challenges, and Recommendations for Future Direction. EHP 2023.
11. Chadeau-Hyam M, Athersuch TJ, Keun HC, et al. Meeting-in-the-middle using metabolic profiling - a strategy for the identification of intermediate biomarkers in cohort studies. Biomarkers 2011;16(1):83-8. (In eng). DOI: 10.3109/1354750x.2010.533285.
12. Vineis P, Demetriou CA, Probst-Hensch N. Long-term effects of air pollution: an exposome meet-in-the-middle approach. Int J Public Health 2020;65(2):125-127. (In eng). DOI: 10.1007/s00038-019-01329-7.
13. Jeong A, Fiorito G, Keski-Rahkonen P, et al. Perturbation of metabolic pathways mediates the association of air pollutants with asthma and cardiovascular diseases. Environ Int 2018;119:334-345. (In eng). DOI: 10.1016/j.envint.2018.06.025.
14. Calle EE, Rodriguez C, Jacobs EJ, et al. The American Cancer Society Cancer Prevention Study II Nutrition Cohort: rationale, study design, and baseline characteristics. Cancer 2002;94(2):500-11. (In eng). DOI: 10.1002/cncr.10197.

15. Patel AV, Jacobs EJ, Dudas DM, et al. The American Cancer Society's Cancer Prevention Study 3 (CPS-3): Recruitment, study design, and baseline characteristics. Cancer 2017;123(11):2014-2024. (In eng). DOI: 10.1002/cncr.30561.
16. Evans AM, DeHaven CD, Barrett T, Mitchell M, Milgram E. Integrated, nontargeted ultrahigh performance liquid chromatography/electrospray ionization tandem mass spectrometry platform for the identification and relative quantification of the small-molecule complement of biological systems. Anal Chem 2009;81(16):6656-67. (In eng). DOI: 10.1021/ac901536h.
17. Wang Y, Gapstur SM, Carter BD, et al. Untargeted Metabolomics Identifies Novel Potential Biomarkers of Habitual Food Intake in a Cross-Sectional Study of Postmenopausal Women. J Nutr 2018;148(6):932-943. (In eng). DOI: 10.1093/jn/nxy027.
18. Clark-Reyna SE, Grineski SE, Collins TW. Residential exposure to air toxics is linked to lower grade point averages among school children in El Paso, Texas, USA. Popul Environ 2016;37(3):319-340. DOI: 10.1007/s11111-015-0241-8.
19. Castillo MD, Kinney PL, Southerland V, et al. Estimating Intra-Urban Inequities in PM(2.5)-Attributable Health Impacts: A Case Study for Washington, DC. Geohealth 2021;5(11):e2021GH000431. DOI: 10.1029/2021GH000431.
20. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society: Series B (Methodological) 1995;57(1):289-300. DOI: https://doi.org/10.1111/j.2517-6161.1995.tb02031.x.
21. Zhang H, Zheng Y, Zhang Z, et al. Estimating and testing high-dimensional mediation effects in epigenetic studies. Bioinformatics 2016;32(20):3150-3154. (In eng). DOI: 10.1093/bioinformatics/btw351.
22. Lee GB, Lee JC, Moon MH. Plasma lipid profile comparison of five different cancers by nanoflow ultrahigh performance liquid chromatography-tandem mass spectrometry. Anal Chim Acta 2019;1063:117-126. (In eng). DOI: 10.1016/j.aca.2019.02.021.
23. Feng Z, Ou Y, Hao L. The roles of glycolysis in osteosarcoma. Front Pharmacol 2022;13:950886. (In eng). DOI: 10.3389/fphar.2022.950886.
24. Dass PD. Glutamine as a gamma-glutamyl donor for gamma-glutamyl transpeptidase: gamma-glutamyl peptide formation. Life Sci 1983;33(18):1757-62. (In eng). DOI: 10.1016/0024-3205(83)90682-3.
25. Wu G, Fang YZ, Yang S, Lupton JR, Turner ND. Glutathione metabolism and its implications for health. J Nutr 2004;134(3):489-92. (In eng). DOI: 10.1093/jn/134.3.489.
26. Chaudière J, Ferrari-Iliou R. Intracellular antioxidants: from chemical to biochemical mechanisms. Food Chem Toxicol 1999;37(9-10):949-62. (In eng). DOI: 10.1016/s0278-6915(99)00090-3.
27. Oberli-Schrämmli AE, Joncourt F, Stadler M, et al. Parallel assessment of glutathione-based detoxifying enzymes, O6-alkylguanine-DNA alkyltransferase and P-glycoprotein as indicators of drug resistance in tumor and normal lung of patients with lung cancer. Int J Cancer 1994;59(5):629-36. (In eng). DOI: 10.1002/ijc.2910590509.
28. Hida T, Kuwabara M, Ariyoshi Y, et al. Serum glutathione S-transferase-pi level as a tumor marker for non-small cell lung cancer. Potential predictive value in chemotherapeutic response. Cancer 1994;73(5):1377-82. (In eng). DOI: 10.1002/1097-0142(19940301)73:5<1377::aid-cncr2820730511>3.0.co;2-n.

29. Blair SL, Heerdt P, Sachar S, et al. Glutathione metabolism in patients with non-small cell lung cancers. Cancer Res 1997;57(1):152-5. (In eng).
30. Traverso N, Ricciarelli R, Nitti M, et al. Role of glutathione in cancer progression and chemoresistance. Oxid Med Cell Longev 2013;2013:972913. (In eng). DOI: 10.1155/2013/972913.
31. Joseph B, Marchetti P, Formstecher P, Kroemer G, Lewensohn R, Zhivotovsky B. Mitochondrial dysfunction is an essential step for killing of non-small cell lung carcinomas resistant to conventional treatment. Oncogene 2017;36(33):4818. (In eng). DOI: 10.1038/onc.2017.78.
32. Zhang X, Zhu X, Wang C, Zhang H, Cai Z. Non-targeted and targeted metabolomics approaches to diagnosing lung cancer and predicting patient prognosis. Oncotarget 2016;7(39):63437-63448. (In eng). DOI: 10.18632/oncotarget.11521.
33. Puchades-Carrasco L, Jantus-Lewintre E, Pérez-Rambla C, et al. Serum metabolomic profiling facilitates the non-invasive identification of metabolic biomarkers associated with the onset and progression of non-small cell lung cancer. Oncotarget 2016;7(11):12904-16. (In eng). DOI: 10.18632/oncotarget.7354.
34. Hao Y, Yi Q, XiaoWu X, WeiBo C, GuangChen Z, XueMin C. Acetyl-CoA: An interplay between metabolism and epigenetics in cancer. Frontiers in Molecular Medicine 2022;2 (Review) (In English). DOI: 10.3389/fmmed.2022.1044585.
35. Chen S, Li, C., Qin, Z. et al. Phenylalanylphenylalanine as a Diagnostic Biomarker for Lung Cancer and Tuberculosis. Research Square 2021 (Journal Article). DOI: https://doi.org/10.21203/rs.3.rs-775519/v1.
36. van Veldhoven K, Kiss A, Keski-Rahkonen P, et al. Impact of short-term traffic-related air pollution on the metabolome - Results from two metabolome-wide experimental studies. Environ Int 2019;123:124-131. (In eng). DOI: 10.1016/j.envint.2018.11.034.
37. Pradhan SN, Das A, Meena R, Nanda RK, Rajamani P. Biofluid metabotyping of occupationally exposed subjects to air pollution demonstrates high oxidative stress and deregulated amino acid metabolism. Sci Rep 2016;6:35972. (In eng). DOI: 10.1038/srep35972.
38. Long F, Su JH, Liang B, Su LL, Jiang SJ. Identification of Gene Biomarkers for Distinguishing Small-Cell Lung Cancer from Non-Small-Cell Lung Cancer Using a Network-Based Approach. Biomed Res Int 2015;2015:685303. (In eng). DOI: 10.1155/2015/685303.
39. Lane KJ, Levy JI, Patton AP, Durant JL, Zamore W, Brugge D. Relationship between traffic-related air pollution and inflammation biomarkers using structural equation modeling. Sci Total Environ 2023;870:161874. DOI: 10.1016/j.scitotenv.2023.161874.
40. Zhang L, Fang B, Wang H, et al. The role of systemic inflammation and oxidative stress in the association of particulate air pollution metal content and early cardiovascular damage: A panel study in healthy college students. Environ Pollut 2023;323:121345. DOI: 10.1016/j.envpol.2023.121345.
41. Liao J, Goodrich J, Walker DI, et al. Metabolic pathways altered by air pollutant exposure in association with lipid profiles in young adults. Environ Pollut 2023;327:121522. DOI: 10.1016/j.envpol.2023.121522.
42. Singh BK, Kambayashi T. The Immunomodulatory Functions of Diacylglycerol Kinase ζ. Front Cell Dev Biol 2016;4:96. (In eng). DOI: 10.3389/fcell.2016.00096.

43. Pope CA, 3rd, Burnett RT, Turner MC, et al. Lung cancer and cardiovascular disease mortality associated with ambient air pollution and cigarette smoke: shape of the exposure-response relationships. Environ Health Perspect 2011;119(11):1616-21. DOI: 10.1289/ehp.1103639.
44. Pope CA, 3rd, Krewski D, Gapstur SM, Turner MC, Jerrett M, Burnett RT. Fine Particulate Air Pollution and Mortality: Response to Enstrom's Reanalysis of the American Cancer Society Cancer Prevention Study II Cohort. Dose Response 2017;15(4):1559325817746303. DOI: 10.1177/1559325817746303.
45. Turner MC, Krewski D, Diver WR, et al. Ambient Air Pollution and Cancer Mortality in the Cancer Prevention Study II. Environ Health Perspect 2017;125(8):087013. DOI: 10.1289/EHP1249.
46. Turner MC, Krewski D, Pope CA, 3rd, Chen Y, Gapstur SM, Thun MJ. Long-term ambient fine particulate matter air pollution and lung cancer in a large cohort of never-smokers. Am J Respir Crit Care Med 2011;184(12):1374-81. DOI: 10.1164/rccm.201106-1011OC.

Tables and Figures

Table 1 Demographic Characteristics for Participants in CPS-II Nutrition and CPS-3 Cohorts (n = 1,360)

	n = 1360
Age at Blood Draw	
Mean (SD)	66.4 (8.83)
Median [Min, Max]	68.0 [31.0, 85.0]
Sex	
Female	679 (49.9%)
Male	681 (50.1%)
Race	
White	1311 (96.4%)
Non-White	48 (3.5%)
Missing	1 (0.1%)
BMI	
Mean (SD)	26.5 (4.82)
Median [Min, Max]	25.8 [15.0, 50.9]
Missing	14 (1.0%)
Passive Smoke Exposure	
Yes	934 (68.7%)
No	350 (25.7%)
Missing	76 (5.6%)
Average Weekly Consumption of Fruits and Vegetables	
Mean (SD)	27.9 (15.9)
Median [Min, Max]	25.9 [0, 133]
Missing	46 (3.4%)
Smoking Status	
Never	454 (33.4%)
Current	154 (11.3%)
Former	699 (51.4%)
Missing	53 (3.9%)
Highest Educational Level	
Less than HS	53 (3.9%)
HS Graduate	263 (19.3%)
Some College/Tech School/2-year Degree	408 (30.0%)
College Graduate	307 (22.6%)
Graduate School	268 (19.7%)
Missing	61 (4.5%)
Multivitamin Use	
Recently	538 (39.6%)
Somewhat Recently	43 (3.2%)
Rarely/Never	340 (25.0%)
Missing	439 (32.3%)

Table 1 Demographic Characteristics for Participants in CPS-II Nutrition and CPS-3 Cohorts (n = 1,360) *(Continued)*

	n = 1360
Family History of Lung Cancer	
Yes	118 (8.7%)
No	1184 (87.1%)
Missing	58 (4.3%)
Passive Smoke Exposure	
Yes	934 (68.7%)
No	350 (25.7%)
Missing	76 (5.6%)
Alcohol Use	
None	426 (31.3%)
<1 per day	553 (40.7%)
1+ per day	281 (20.7%)
Missing	100 (7.4%)
Marital Status	
Single	26 (1.9%)
Married	1051 (77.3%)
Other	283 (20.8%)
Age at Lung Cancer Diagnosis	
Mean (SD)	72.2 (10.5)
Median [Min, Max]	75.0 [33.0, 95.0]
Missing	677 (49.8%)
Lung Cancer Stage	
Localized	157 (11.5%)
Regional	171 (12.6%)
Distant	320 (23.5%)
Unknown	35 (2.6%)
Missing	677 (49.8%)

Table 2 Metabolic features significantly associated with air pollutant exposure models* (n = 1,360)

Air pollutant	**FDR q < 0.2**		
	Air Pollution Model	**MITM[i]**	**HDMA**
CO Exposure	62	1	1
NO_2 Exposure	90	3	2
O_3 Exposure	44	0	1
PM_{10} Exposure	72	0	1
$PM_{2.5}$ Exposure	124	0	1
SO_2 Exposure	16	0	0

*Both Benjamini-Hochberg false discovery rate (FDR) procedure (q value) and raw p values were used to identify a reasonable number of significant metabolic features.

[i]Significant features at q < 0.2 in the air pollution model were put in the meet-in-the-middle model

Table 3 Metabolites significantly associated with air pollutant exposure and lung cancer incidence using meet-in-the-middle (MITM) and high dimensional mediation analysis (HDMA) analytical approaches

Air pollutant	Metabolites	Platform	Super Pathway	Sub Pathway	FDR	
					MITM	HDMA
CO	gamma-glutamylglutamine	Pos Early	Peptide	Gamma-glutamyl Amino Acid	0.181	0.012
NO_2	N-acetylphenylalanine	Pos Early	Amino Acid	Phenylalanine Metabolism	-	0.085
	gamma-glutamylmethionine	Pos Early	Peptide	Gamma-glutamyl Amino Acid	0.172	0.002
	palmitoleoyl-linoleoyl-glycerol (16:1/18:2) [1]*	Pos Early	Amino Acid	Phenylalanine Metabolism	0.132	-
	gamma-glutamylglutamine	Pos Early	Peptide	Gamma-glutamyl Amino Acid	0.132	-
O_3	X-26111	Negative	Unknown	Unknown	-	0.0480
PM_{10}	gamma-glutamylmethionine	Pos Early	Peptide	Gamma-glutamyl Amino Acid	0.207	0.003
	gamma-glutamylglutamine	Pos Early	Peptide	Gamma-glutamyl Amino Acid	0.207	-
$PM_{2.5}$	X-23654	Pos Early	Unknown	Unknown	-	0.048

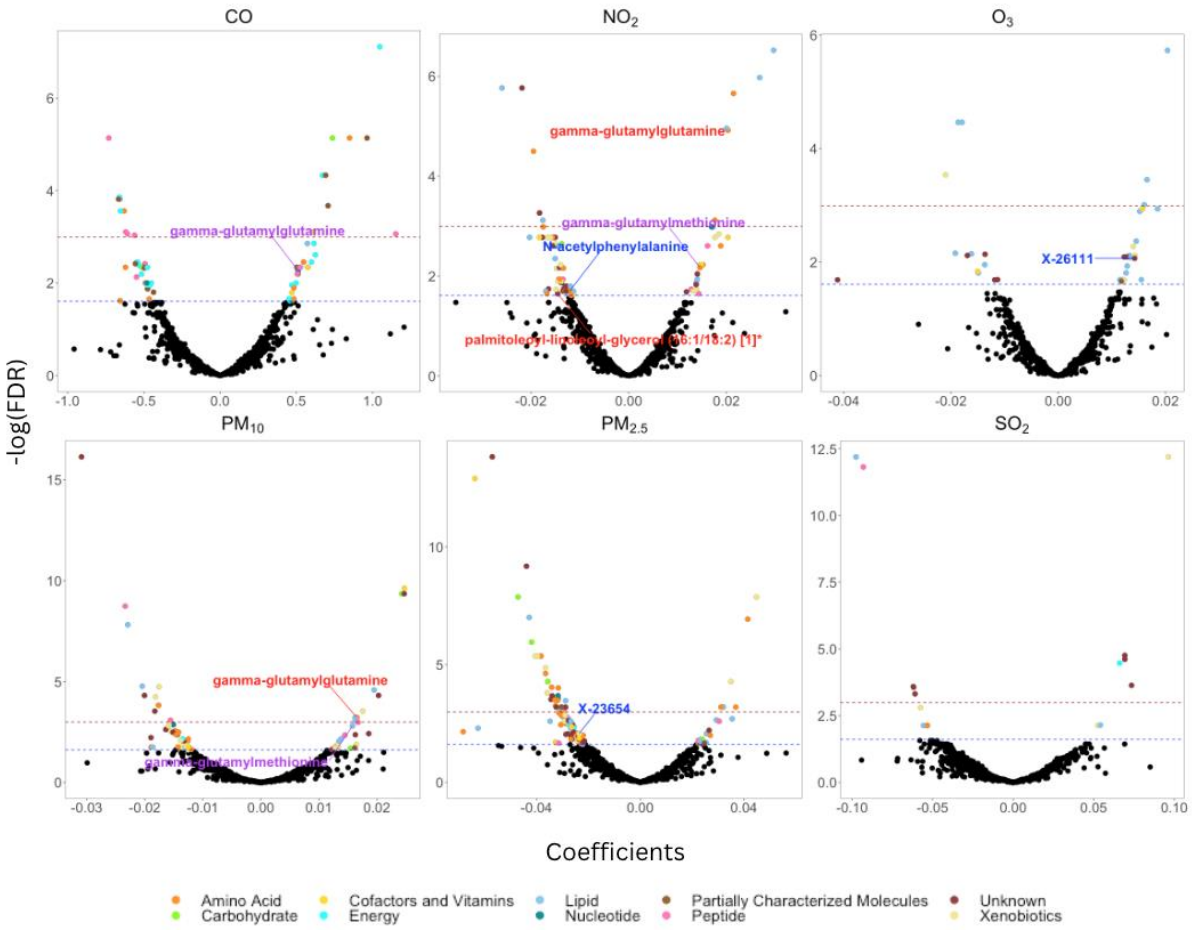


Figure 1 Volcano plots of associations between changes in metabolite intensities and air pollutants. X-axis denotes the coefficients of metabolite-pollutant associations. Y-axis denotes the negative natural log of false discovery rate (FDR) in metabolite-pollutant association. Different colors were used to represent different pathways where the metabolites are involved. Dark red dashed line represents FDR = 0.05 and blue dashed line represents FDR = 0.2. Metabolites labeled with red color were significant in the meet-in-the-middle (MITM) analysis, blue-labeled metabolites were significant in the high-dimensional mediation analysis (HDMA) analysis, and purple-labeled metabolites were significant in both the MITM and HDMA analyses.

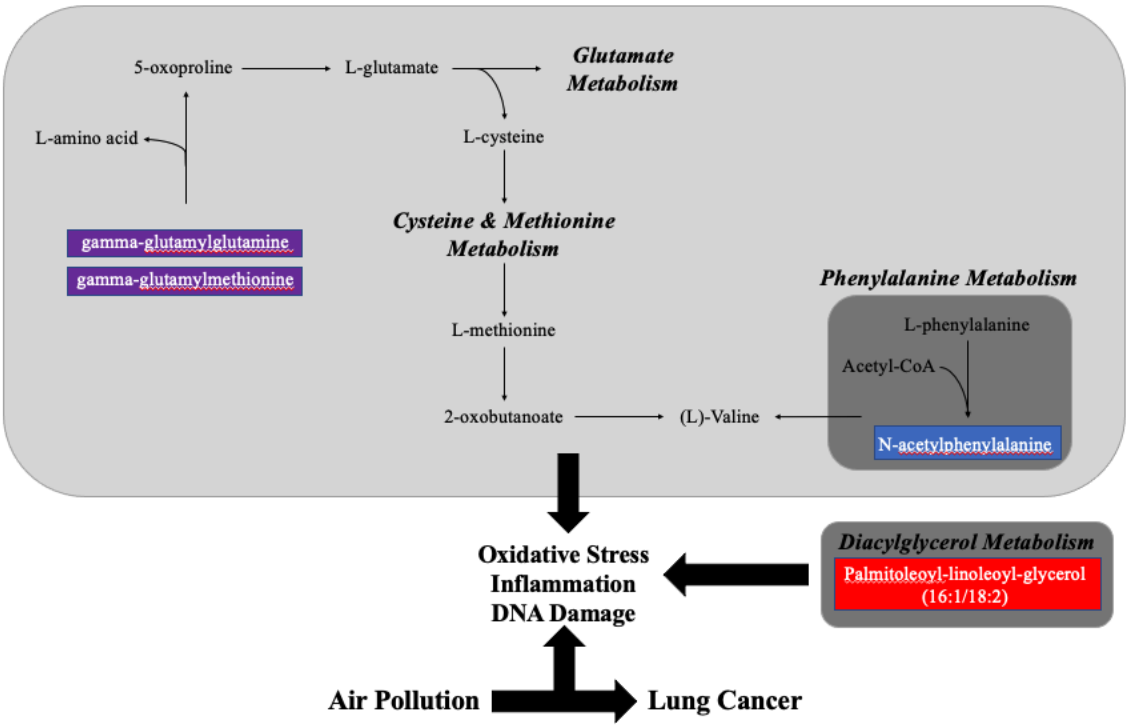


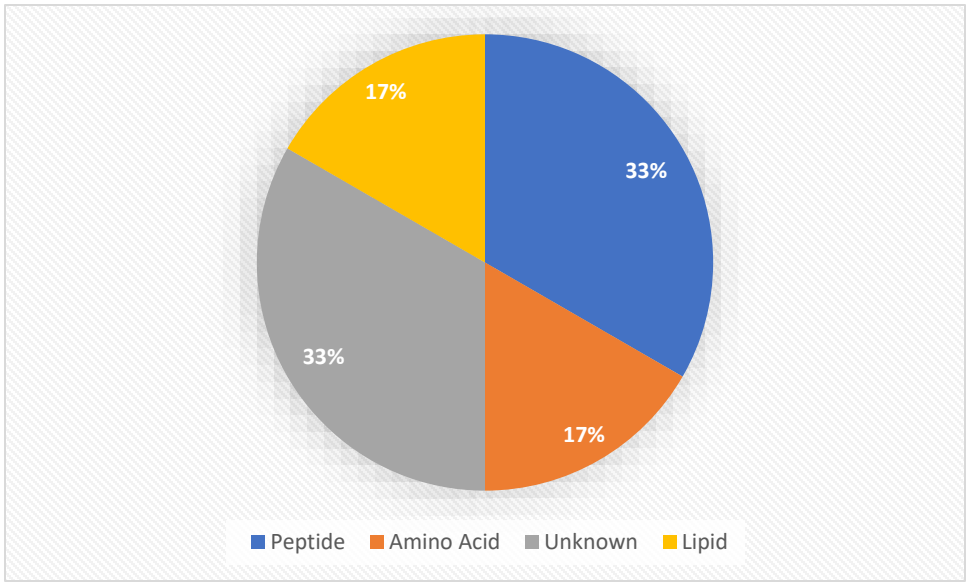
Figure 2 Mechanistic figure depicting the metabolic pathways that lead to biological changes in the association between air pollution and lung cancer.

Supplemental Materials

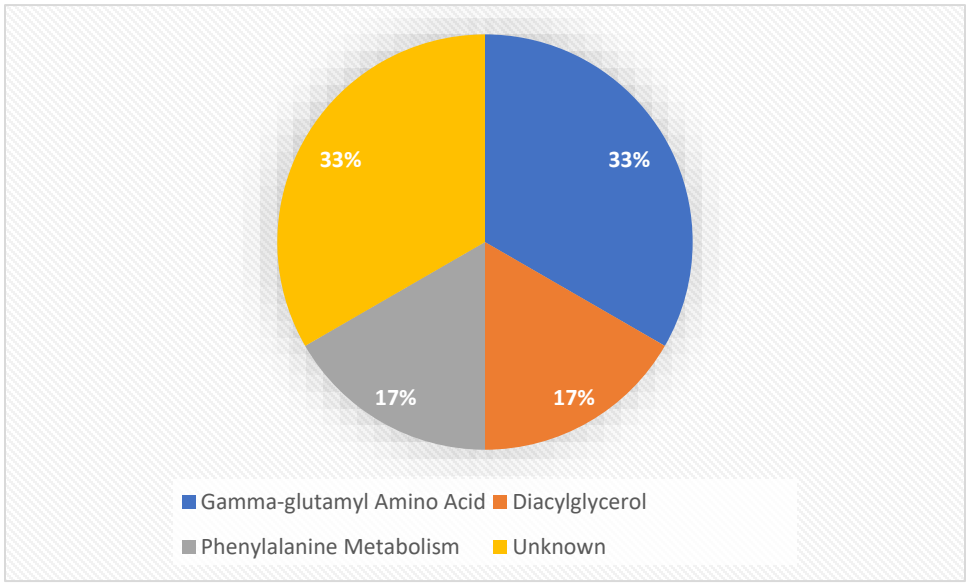
Supplement 1 Air pollution assessments of participants based on residential address at time of blood draw (n = 1,360)

Air pollutant assessments	Mean (SD)	25th	50th	75th	Max
NO_2 [a] (ppb)	12.1 (5.9)	7.7	11.2	15.3	37.8
O_3 [b] (ppb)	48.2 (6.5)	44.4	47.1	51.5	68.4
CO[c] (ppm)	0.4 (0.2)	0.30	0.39	0.50	1.3
SO_2[d] (ppm)	3.1 (1.8)	1.7	2.6	4.1	11.8
PM_{10}[e] ($\mu g/m^3$)	20.9 (5.9)	17.1	20.4	23.2	51.4
$PM_{2.5}$[f] ($\mu g/m^3$)	12.2 (3.2)	9.9	11.9	14.0	25.2

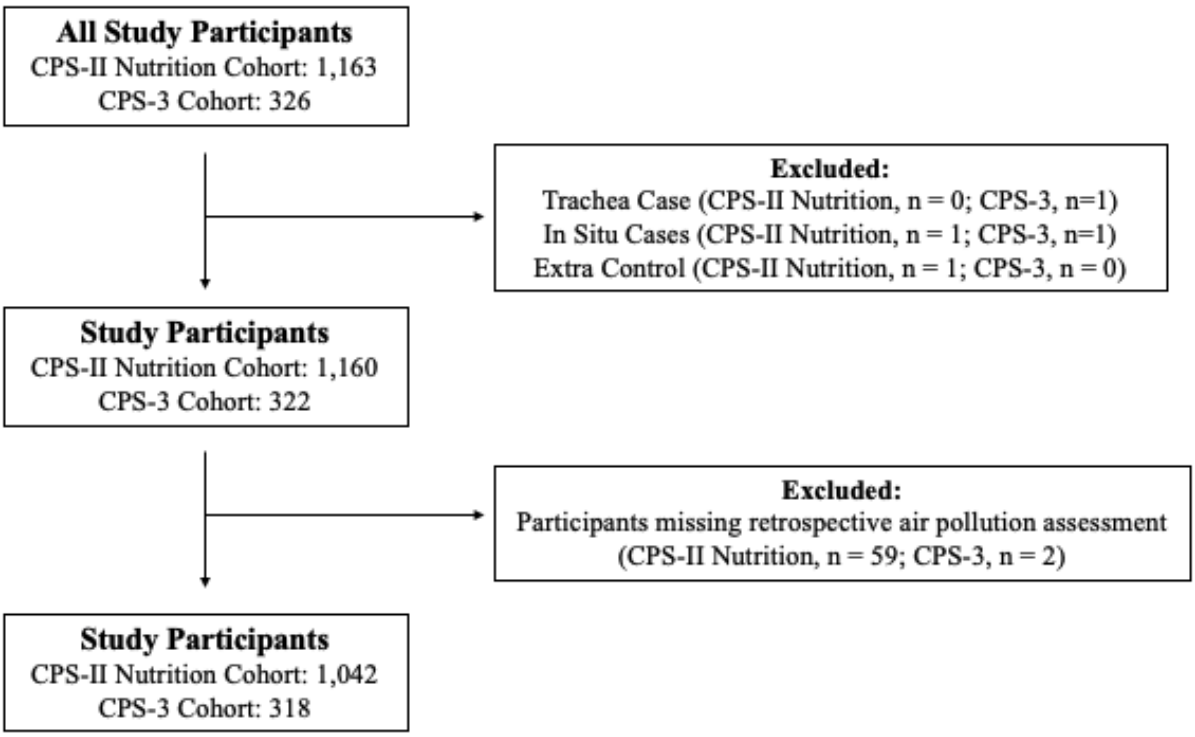
A.



B.



Supplement 2 Pathway breakdown for combined significant metabolites in meet-in-the-middle (MITM) and high dimensional mediation analysis (HDMA) A. Super pathway breakdown, B. Sub pathway breakdown



Supplement 3 Summary of exclusion criteria for study participants included in final analysis. All participants that were matched to a participant that was excluded was also included from the final analysis.