

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 works (such as articles or books) all or part of this thesis or dissertation.

Signature:

Zhenjiang Li

Date

Application of high-resolution metabolomics in the CHDWB cohort to identify biological pathways perturbed by traffic-related air pollution

By

Zhenjiang Li
Master of Science in Public Health

Environmental Health

Stefanie Ebel Sarnat, ScD
Committee Chair

Donghai Liang, PhD
Committee Member

Application of high-resolution metabolomics in the CHDWB cohort to identify biological pathways perturbed by traffic-related air pollution

By

Zhenjiang Li

M.S.
Peking University
2017

Thesis Committee Chair: Stefanie Ebel Sarnat, ScD

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 Science in Public Health
in Environmental Health and Epidemiology
2019

Abstract

Application of high-resolution metabolomics in the CHDWB cohort to identify biological pathways perturbed by traffic-related air pollution

By Zhenjiang Li

Purpose

To identify metabolic perturbations associated with short-term exposures to ambient traffic-related air pollutants (TRAP), including carbon monoxide (CO), nitrogen dioxide (NO₂), ozone (O₃), fine particulate matter (PM_{2.5}), organic carbon (OC), and elemental carbon (EC) among a subset of participants in the Center for Health Discovery and Well-Being (CHDWB), a cohort of Emory University employees.

Methods

A cross-sectional study was conducted on baseline visits of 180 CHDWB participants whose plasma samples that had undergone untargeted high-resolution metabolomics (HRM) profiling using liquid chromatography-high-resolution mass spectrometry with positive and negative electrospray ionization modes. Ambient pollution concentrations were measured at an ambient monitor near downtown Atlanta and assigned to each participant according to their date of visit. Metabolic variations associated with pollution exposures were assessed following a metabolome-wide association study framework, considering both Tobit models and regular multiple linear regression models with adjustment of temporal covariates to identify significant metabolic features. Enriched biological pathways, i.e., those perturbed by pollution, were then identified by Mummichog.

Results

The study population was predominantly white (76.1%) and non-smokers (95.6%). All participants has at least a high school education. In total, 7,821 and 4,123 metabolic features were extracted from the plasma samples by the negative and positive ion mode runs, respectively. After removing features present in less than 10% of participants, 7,106 and 3,628 remains. Biological pathways enriched by metabolic features associated with the pollutants of interest primarily pertained to nucleotide metabolism (e.g., pyrimidine metabolism, which was associated with CO, NO₂, EC, and OC), lipid metabolism, and amino acid metabolism. NO₂ and EC were associated most consistently with these pathways.

Conclusions

We identified a range of ambient pollutants, including components of TRAP, associated with changes to the metabolic phenotype among the cohort. The results demonstrate the use of HRM as a viable platform for untargeted characterization of molecular mechanisms underlying exposure to TRAP.

Application of high-resolution metabolomics in the CHDWB cohort to identify biological pathways perturbed by traffic-related air pollution

By

Zhenjiang Li

M.S.
Peking University
2019

Thesis Committee Chair: Stefanie Ebel Sarnat, ScD

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 Science in Public Health
in Environmental Health and Epidemiology
2019

Acknowledgements

I want to express my sincere gratitude to my thesis advisor Dr. Stefanie Ebel Sarnat for the opportunity to do this research. The TRAPHIC study has been a perfect platform for me to learn new skills and hone my knowledge of epidemiology and environmental health. I cannot count how many times we have met for this study, and it is euphoric to recall how the present study set up step by step. Her careful guidance and support not only help me build a solid base of research skills but also fortifies my interest in research work.

I would also like to thank Dr. Donghai Liang. He has introduced me to the field of metabolomics in which I plan to dive deeper for my PhD program. Moreover, Donghai has supported me as a friend as well and shared his valuable experience with me regarding both academic and daily life. I am grateful to Dr. Dongni Ye for her support to set up the dataset for my thesis.

I am forever thankful to my friends and family for their encouragement.

Table of Contents

Introduction	1
Methods	4
Results	8
Discussion.....	9
Conclusions and Recommendations	15
References.....	17
Tables and Figures.....	26
Appendices	37

Introduction

Outdoor air pollution is an important environmental risk factor for human health all over the world (Lelieveld, Evans, Fnais, Giannadaki, & Pozzer, 2015), and traffic-related air pollution (TRAP) from motor vehicles is the main contributor in urban areas (Greenbaum & CANCER, 2013). TRAP contains large quantities of carbon monoxide (CO), nitrogen dioxide (NO₂), fine particulate matter [PM_{2.5}, with components such as elemental carbon (EC), organic carbon (OC), and metals] which are emitted directly from vehicles via combustion processes and tire and brake wear, along with ozone (O₃), a secondary by-product (Greenbaum & CANCER, 2013). The adverse health effects of TRAP and its main components has raised public concerns for decades.

There has been substantial research investigating the effects of TRAP on respiratory and cardiovascular diseases, and a comprehensive systematic review of the literature was published in 2010 by the Health Effects Institute, a nonprofit corporation funded by the U.S. Environmental Protection Agency and the worldwide motor vehicle industry (Health Effects Institute, 2010). In epidemiological studies, TRAP and its components were found to be associated with cardiopulmonary morbidity and mortality, with impacts on stroke, asthma exacerbation, impaired lung function, and non-asthmatic respiratory allergy. Although the review found that a number of animal studies indicated several mechanistic pathways, such as systemic inflammation and DNA damage, linking TRAP and its components with cardiovascular and respiratory events, the underlying mechanisms in human body were poorly depicted due to difficulties in extrapolating from animal evidence to humans (Health Effects Institute, 2010). Furthermore, the epidemiological literature pointed to detrimental effects beyond the cardiovascular and respiratory systems, with observed associations of TRAP and its components with outcomes such as type 2 diabetes, preterm birth, and low birth weight (Brauer et al., 2008; Brook, Jerreft, Brook, Bard, & Finkelstein, 2008; Kramer et al., 2010; Wilhelm et

al., 2011). Identifying biological mechanisms that regulate the effects of TRAP can inform causal inference and prevention, so it has been a priority to identify and characterize the internal biological pathways that link exposure to TRAP with clinical outcomes. Since the 2010 HEI review, progress has been made to better understand the nature of biological responses induced by exposure to TRAP. Systemic inflammatory markers, oxidative stress factors, and cell counts in blood have been the main endpoints measured in the more recent studies (Carvalho et al., 2018; Chiu et al., 2016; Golan et al., 2018; Jacobs et al., 2010; Krishnan et al., 2013; Kubesch et al., 2015; Sarnat et al., 2014; Zuurbier et al., 2011). However, results from these studies are inconsistent likely due in part to differences in study design and exposure measurement. In addition to these studies, some researchers have employed plasma circulating microRNAs, whole blood RNA, or mitochondrial abundance to identify changes in gene expression following exposure to TRAP; the findings indicate some molecular mechanisms involved in the pathogenesis of multiple diseases, such as breast and lung cancers, and cardiovascular diseases (Chu et al., 2016; Krauskopf et al., 2018; Zhong et al., 2016). Overall, research to date provides increasing evidence indicating that TRAP can induce diverse biological responses in the human body, however the underlying mechanisms remain primarily inconclusive.

High-resolution metabolomics (HRM) is an emerging quantitative method for comprehensive identification and quantitation of internal metabolites (e.g., present in a given biological media, such as blood or saliva). This method is providing new opportunities for epidemiologists to investigate the associations of external exposures with endogenous processes at the molecular level (Jones, Park, & Ziegler, 2012; Uppal et al., 2016). In most prior work, targeted methods have been used to identify and quantify a defined set of metabolites. In recent years, untargeted methods have been developed to maximize the detection of metabolite features in biological media and

research has leveraged these data for biomarker discovery or data-driven analysis of biological pathways perturbed due to some internal or external factor of interest (Uppal et al., 2016). For example, an untargeted metabolome-wide association study (MWAS) workflow has been developed, and used in a handful of studies to date, to distinguish statistically significant metabolites associated with the exposure to TRAP from tens of thousands of metabolic features detected (Jeong et al., 2018; Ladva et al., 2018; Liang et al., 2018; D. I. Walker et al., 2018). As part of the MWAS workflow, these studies have also employed methods to identify the biological pathways altered after exposure to TRAP (Jeong et al., 2018; Ladva et al., 2017; Ladva et al., 2018; Liang et al., 2018; Miller et al., 2016; Surowiec et al., 2016; van Veldhoven et al., 2019; Vlaanderen et al., 2017; D. I. Walker et al., 2018).

Prior work on HRM and TRAP has used either a cross-over design in which participants served as their own matched controls (Miller et al., 2016; Surowiec et al., 2016; van Veldhoven et al., 2019; D. I. Walker et al., 2018) or a panel study design (Ladva et al., 2017; Ladva et al., 2018; Liang et al., 2018; Vlaanderen et al., 2017); Jeong et al. conducted two nested case-control studies focusing on asthma and cardio-cerebrovascular cases in Switzerland and Italy (Jeong et al., 2018). These studies have all been limited to less than 60 participant samples due to cost and practicality. The small sample sizes may have caused studies to be underpowered to detect some air pollution effects, and small participant numbers is likely to have affected the generalizability of findings to date.

To expand on this growing body of research, we performed a cross-sectional study with a relatively large sample size of 180 participants. In the present study, we followed an MWAS workflow to identify the biological pathways perturbed by TRAP among participants at baseline in the Center for Health Discovery and Wellbeing (CHDWB) Cohort. The CHDWB Cohort at Emory University in Atlanta, Georgia, USA, was an

observational study designed to investigate the effects of clinical self-knowledge and health partner counseling (Tabassum et al., 2014). We applied HRM to 180 plasma samples collected at the baseline visit to obtain metabolic profiles and applied these data in epidemiologic analyses to identify metabolic features associated with short-term exposures to ambient CO, NO₂, O₃, PM_{2.5}, elementary carbon (EC), and organic carbon (OC). Pathway enrichment analysis was conducted to identify biological pathways associated with significant metabolic features.

Methods

Study Design

The present study was a cross-sectional design that included the baseline visits of a subset of participants in the CHDWB Cohort. Details of the cohort can be found elsewhere (Brigham, 2010; Rask, Brigham, & Johns, 2011; Tabassum et al., 2014). Briefly, the CHDWB Cohort was initiated in May 2008 and recruited employees of Emory University from 2008 to 2012. The participants were free of poorly controlled chronic diseases or acute illness at the time of recruitment. Basic demographics and plasma samples were collected during the clinical visit, along with tobacco and alcohol usage. A subset of 180 baseline visits were included in the current analysis, for participants whose plasma samples had previously undergone untargeted high-resolution metabolic profiling by the Clinical Biomarkers Laboratory at Emory University (Jin, Kang, & Yu, 2018). All participants provided informed consent, and the study was approved by the Emory University Institutional Review Board.

Exposure Assessment

For the study period (2008-2012), continuous measurements of CO, NO₂, O₃, PM_{2.5}, EC, and OC were made at the Jefferson St. monitor, an ambient monitoring site near downtown Atlanta. The details of site information and measure methods can be found elsewhere (Hansen et al., 2006). Briefly, CO was measured continuously with 1-min

resolution using a TEI Model 48S NDIR analyzer. NO₂ was not directly measured but converted photolytically from nitrogen monoxide (NO), and NO was measured continuously using a TEI Model 42ctl analyzer via chemiluminescence. O₃ was measured by UV-based ozone analyzer. All trace gases were aggregated to the daily level and reported as daily 1-hr maximum values (for CO and NO₂) or daily 8-hr maximum values (for O₃). PM_{2.5} was measured continuously with an R&P Model 1400 a/b tapered element oscillating microbalance (TEOM). OC and EC were measured using an R&P Model 5400 ambient particulate carbon monitor with 60-min resolution. PM_{2.5} and its components were reported as daily 24-hr averages. The daily concentration of each pollutant was assigned to each participant according to the date of their baseline visit. Daily meteorological data were obtained from the Atlanta Hartsfield-Jackson International Airport.

High-Resolution Metabolomics

Analysis of the biological samples was accomplished by randomizing the de-identified and blinded plasma samples into blocks of 20. Each analytical batch contained a pooled plasma sample that had been referenced against the NIST 1950 standard reference material. Sample preparation for mass spectral analyses included: 65 µL of plasma stored at -80°C was added to 130 µL of acetonitrile containing 3.5 µL mixture of 14 stable isotope standards (Soltow et al., 2013). After mixing and incubation at 4°C for 30 min, precipitated proteins were pelleted via centrifugation for 10 min at 4°C and 14,100 × g. Following protein precipitation, triplicate 10 mL aliquots were analyzed by reverse-phase C18 liquid chromatography with Fourier transform mass spectrometry (Dionex Ultimate 3000, Q-Exactive, Thermo Scientific, Waltham, MA) operated in either negative or positive electrospray ionization mode. Having data from both modes should provide a more comprehensive profile of metabolism than data extracted from only one mode; compounds are ionized with different ionization efficiency between the positive

and negative ion mode, which results in varying sensitivity and detection limits (Liigand et al., 2017). A mass-to-charge (m/z) scan range of 85 to 1275 was used. Data from each analytical run was saved in both the .RAW format and converted to the .mzML format using ProteoWizard. Then, metabolic profiles were extracted by apLCMS with modifications using the R package xMSanalyzer (Uppal et al., 2013).

Date Analysis

To prepare the HRM data for epidemiologic analysis, extracted metabolic features present in less than 10% of the participants were excluded. This was done in order to reduce the background noise generated by the measurement equipment (Alonso, Marsal, & Julia, 2015), and also as a means of focusing the analytic dataset to more common metabolites (e.g., endogenous metabolites that are common across people) in order to facilitate detection of biological pathways later in the analysis.

Associations of air pollution with metabolic features were estimated following an untargeted MWAS workflow. Specifically, for each air pollutant-feature pair, Tobit regression models were conducted, controlling for potential temporal confounders. Although we filtered the features to remove those with less than 10% presence across participants, the analytic dataset still contained a large number of features with missing values (i.e., not all of the remaining features were present in all 180 participant plasma samples). The Tobit model was therefore used, as it is devised for situations where the dependent variable is censored at a specific nonnegative value (McBee, 2010). The basic form of the model was:

$$\log(Y_i)^* = \beta_{i0} + \beta_{i1} \times \text{pollutant} + \beta_{i2} \times \text{year} + \beta_{i3} \times \text{season} + \beta_{i4} \times \text{weekday} + \beta_{i5} \\ \times \text{apparent_temp} + \beta_{i6} \times \text{apparent_temp}^2 + \varepsilon_i$$

$$\begin{cases} \log(Y_i) = \log(Y_i)^* & \text{if } \log(Y_i) > LOD \\ \log(Y_i) = LOD & \text{if } \log(Y_i) \leq LOD \end{cases}$$

where Y_i refers to the observed intensity of metabolic feature i , $\log(Y_i)^*$ refers to the latent log-transformed feature intensity; β_{i1} refers to the coefficient for the air pollutant, indicating the change in feature intensity for a one unit increase in pollution, holding all other covariates as constant. Year (3-level: 2008, 2009, >2009), season (4-level: spring, summer, fall, winter), and weekday (5-level) were determined by the date of baseline visit. *Apparent_temp* denotes the daily apparent temperature and was computed using the daily mean air temperature in combination with the daily mean dew point (Steadman, 1984). We introduced a linear term and a quadratic term of apparent temperature into the model to account for the nonlinear relationship between apparent temperature and dependent variable. Results were visualized using Manhattan plots that displayed results for all features by pollutant, with the retention time of metabolic feature i on the x-axis against the $-\log_{10}(p)$ for β_{i1} on the y-axis. All analyses were performed in R (v.3.5.1).

Biological pathways for significant metabolic features (i.e., meeting the $p = 0.025$ and $p = 0.05$ threshold in MWAS analyses for negative and positive ion modes, respectively) were identified using Mummichog (v.2.0.1) (Li, Dunlop, Jones, & Corwin, 2016). Different thresholds were implemented because the appropriate number (100-500) of features needed for Mummichog. A p -value for each pathway was generated by penalizing pathways with fewer significant metabolic features and assigning greater significance to pathways with more significant features. We excluded pathways with a p -value higher than 0.05 and those containing less than 3 significant metabolic features.

MWAS analyses (and pathway enrichment analysis) were also performed using regular multiple linear regression (MLR) models as a sensitivity analysis. For MLR analyses, missing values (i.e., metabolic features missing for a given participant) were assigned the half of the minimum feature intensity (also defined as the limit of detection) observed across all metabolic features in the dataset (note that the Tobit model, given its design, does not require such imputation). Otherwise, the MLR model was constructed

similar to the Tobit model, with the same covariate control. The performance of MLR models was compared to that of Tobit models based on the total number of significant metabolic features detected for each pollutant, and based on comparisons of the unique significant features detected by the models relative to each feature's percent presence among participants given the assumption that the significance of features with fewer missing values will be more robust across different statistical methods.

Results

Baseline information for the 180 participants is shown in Table 1. Three-quarters of the participants were over the age of 42, and 76.1% of them were white. Over half of participants had completed a graduate school or above. They predominantly were not current smokers (95.6%). Most baseline visits (95.0%) were conducted in 2008 and 2009, and none of them was over the weekend.

We extracted 7,821 metabolic features by negative ion mode and 4,123 by positive ion mode in plasma samples. After data filtering for removing features that were present in less than 10% of participants, the data contained 7,106 negative ion mode features and 3,628 positive ion mode features, respectively. We performed Tobit models for all pollutants of interest. Numbers of significant features for the pollutants are summarized in Table 2. Different significant levels were used for negative ion mode (p -value < 0.025) and positive ion mode (p -value < 0.05) due to the requirement of the appropriate number of significant metabolic features for Mummichog. Figure 1 to 6 depict the results for each pollutant in Manhattan plots displaying the $-\log_{10} p$ values of each metabolic feature against its retention time.

Pathway Enrichment Analysis

We further performed pathway enrichment analyses and examined whether the features associated with the pollutants of interest co-occurred as enriched metabolites within specific metabolic pathways. Nineteen biological pathways associated with at

least one pollutant were identified (Figures 7 and 8). The significant pathways mainly pertain to nucleotide metabolism, lipid metabolism, and amino acid metabolism. There was no overlap in identified pathways between results of negative ion mode and positive ion mode analyses, which may be due to the differences in features identified by each mode. Pyrimidine metabolism was associated with the most pollutants, including 1-day lag concentrations of NO₂, EC, and OC, and the moving average of 1-2 day lag concentrations of CO, NO₂, EC, and OC. The 1-day lag concentration of NO₂ was associated with seven biological pathways. The moving average concentration of EC was associated with nine pathways, six of which are involved in lipid metabolism.

Sensitivity Analysis

MLR models were performed to replace Tobit models as a sensitivity analysis. The results from MLR models are shown in Table S1 and Figures S1-S8 in the Appendix. The significant pathways identified with MLR models (Figures S7 and S8) were all contained within the results observed with Tobit models (Figures 7 and 8). We further compared the performance of these two modeling approaches types of models based on their identification of unique significant metabolic features (aligned on m/z values), and the distribution of those features with respect to their % presence among participants. As shown in Figures S9-S14, Tobit models identified more significant features with a presence of over 60% among participants than did MLR models; this was especially the case for features absent in only one or two participants (with a percentage of presence close to 100%).

Discussion

In the present study, we applied HRM to identify metabolic alterations associated with exposures to ambient CO, NO₂, O₃, PM_{2.5}, OC, and EC in an adult Emory University-based employee cohort using a cross-sectional study design. We identified several biological pathways that were associated with one or multiple pollutants. These

pathways are involved in various biochemical processes in the human body, including nucleotide metabolism (pyrimidine metabolism and purine metabolism), amino acid metabolism (alanine and aspartate metabolism, and tyrosine metabolism), lipid metabolism (C21-steroid hormone biosynthesis and metabolism, fatty acid biosynthesis, and fatty acid metabolism, and carnitine shuttle), and metabolism of cofactors and vitamins (vitamin A metabolism) (Kanehisa & Goto, 2000). These biochemical processes are responsible for maintaining homeostasis and wellbeing in humans.

Pyrimidine metabolism was associated with 1-day lag concentrations of NO_2 , EC, and OC, and the moving average of 1-2 day lag concentrations of CO, NO_2 , EC, and OC, respectively, among positive ion mode metabolic features. Purine metabolism was associated with 1-day lag concentration of OC. Pyrimidine and purine are heterocyclic aromatic organic compounds and serve as a critical part of DNA and RNA. DNA damage has been previously considered as a potential mechanism of adverse effects due to TRAP exposure (Baccarelli et al., 2009; Carvalho et al., 2018; Huang et al., 2012). For example, Carvalho et al. found that exfoliated buccal mucosa cells collected from professional motorcyclists presented a higher average frequency of micronuclei compared to those from office workers, which suggests elevated DNA damage associated with occupational exposure to TRAP (Carvalho et al., 2018). In that study, personal measures of NO_2 and O_3 had a strong positive correlation with micronuclei when analyzed individually (Carvalho et al., 2018). In Huang et al. (2012), the concentration of DNA strand breaks in blood samples collected from traffic conductors within half an hour after the work-shift was significantly higher than that among the office workers (Huang et al., 2012). In addition to DNA damage, gene expression changes due to TRAP exposure have also been studied. For example, blood DNA methylation was found to decrease in individuals with recent exposure to higher levels of ambient $\text{PM}_{2.5}$ and black carbon as well (Baccarelli et al., 2009). Krauskopf et al. identified changes of

gene expression profiles after exposure to TRAP by plasma circulating miRNA, which are involved in posttranscriptional regulation of gene expression (Krauskopf et al., 2018). And, Chu et al. employed gene expression network analyses based on whole blood RNA collected before and after work shift from trucking industry workers with regular exposures to TRAP and detected differentially expressed genes that have implicated a range of cellular responses and pathways such as oxidative stress responses and interferon-mediated inflammatory responses to viral infection (Chu et al., 2016). Pyrimidine and purine metabolism were also reported as significant pathways with TRAP exposure in previous studies using untargeted metabolomics (Jeong et al., 2018; Liang et al., 2018; D. I. Walker et al., 2018). Walker et al. found that pyrimidine metabolism and purine metabolism were associated with the long-term exposure to ultrafine particles (UFP) estimated by modeling in a community-based participatory cross-sectional study (D. I. Walker et al., 2018). Annual average exposure to NO₂ was also associated with pyrimidine metabolism among patients with adult-onset asthma or cardiovascular diseases compared with healthy controls (Jeong et al., 2018). Liang et al. reported multiple ambient air pollutants, including black carbon, nitric oxide, and PM_{2.5}, to be associated with purine metabolism among a panel of 54 college students living in dormitories located either near or far from a major highway (Liang et al., 2018). Our current results provide further evidence of NO₂ and PM inducing variation in nucleotide metabolism as a mechanism of action.

Lipid metabolism is central to the function of the human body; lipids serve not only as an energy source and building block for biological membranes, but they also serve as components of lipoproteins, fat-soluble vitamins, corticosteroid hormones, and mediators of electron transport (Gropper & Smith, 2012). We found that the moving average of 1-2 day lagged EC was associated with several pathways involved in fatty acid biosynthesis and metabolism, including carnitine shuttle, de novo fatty acid biosynthesis, fatty acid

activation, fatty acid oxidation, and omega-6 fatty acid metabolism. For NO_2 and $\text{PM}_{2.5}$, 1-day lag concentrations were associated with saturated fatty acid beta-oxidation. One-day lag concentrations of CO was associated with the carnitine shuttle pathway. Fatty acids (FAs) are a type of simple lipids transported from tissue to tissue by lipoproteins and crucial as an energy nutrient, structural elements of membranes, and signaling molecules (Gropper & Smith, 2012). Human studies that have investigated the effect of exposure to TRAP on blood lipid profiles are scarce. Chen et al. employed blood measurements of total cholesterol, high-density lipoproteins-cholesterol (HDL-C), low-density lipoproteins-C (LDL-C), and HDL-C-to-LDL-C among Mexican Americans and measured exposure to TRAP using NO_x (Chen et al., 2016). Although no significant association was observed between TRAP and lipid measurements, they reported that short-term exposure to ambient $\text{PM}_{2.5}$ was associated with a lower HDL-C-to-LDL-C ratio, and higher total cholesterol and LDL-C (Chen et al., 2016). Jiang et al. reported a significant higher level of LDL-C among subjects living within 50 m to the major road compared with those living more than 200 m, and no significance of total cholesterol and HDL-C were found (Jiang et al., 2016). Gouveia et al. compared the effect of controlled biodiesel exhaust exposure on circulating lipid metabolites to that of filtered air among healthy subjects using a randomized and double-blinded crossover study; the authors observed that the level of monohydroxy fatty acids was altered by biodiesel exhaust (Gouveia-Figueira et al., 2018). The potential protective effect of fish oil supplementation against pro-allergic sensitization effects of TRAP exposure was found among children who were randomized to fish oil supplementation or placebo (Hansell et al., 2018). In addition, an increasing body of evidence indicates that exposure to ambient PM can enhance the adverse effects on atherosclerotic processes, though the mechanisms underlying are undefined (Health Effects Institute, 2010). Atherosclerosis contributes a lot to cardiovascular diseases, and a large body of evidence has well established the

essential role of lipids and lipoproteins in atherosclerotic processes (Linton et al., 2019). Untargeted metabolomics allows researchers to characterize the metabolism profile of lipids more comprehensively, and previous MWAS studies also reported several lipid metabolism pathways (including fatty acid activation, de novo fatty acid biosynthesis, and carnitine shuttle) associated with air pollutants (Jeong et al., 2018; Miller et al., 2016; D. I. Walker et al., 2018). Overall, our current results support these prior findings.

Four significant pathways of amino acid metabolism were also identified in our study: tyrosine metabolism, histidine metabolism, tryptophan metabolism, and alanine and aspartate metabolism. Tyrosine, histidine, and tryptophan are aromatic amino acid and susceptible to the attack of reactive oxidative species (ROS) (Stadtman, 2006). ROS can mediate the conversion process of tyrosine residues to hydroxyl derivatives (for example, 3-nitrotyrosine), histidine residues to 2-oxohistidine and asparagine, and tryptophan residues to formyl-kynurenine and kynurenine (Stadtman, 2006). The oxidative modification of amino acid residues of proteins is involved in the etiology or progression of many diseases (Stadtman & Berlett, 1998). Moreover, 3-nitrotyrosine is used as a marker of oxidative stress, and Rossner reported a significantly higher level of 3-nitrotyrosine in plasma among bus drivers compared with healthy volunteers spending most daily times indoors (Khan et al., 1998; Rossner et al., 2007). Histidine is a well-known inflammatory agent involved in airway hyper-responsiveness (Juniper, Frith, & Hargreave, 1981). However, few studies have employed markers of protein oxidation to investigate the effect of TRAP on human body, and thus our current findings are unique and require confirmation by future studies.

To the best of our knowledge, Tobit models have not been used before in addressing missing values in mass spectrometry-based metabolomics data. As such, we used MLR models with missing value imputation of half the minimum feature intensity as a sensitivity analysis. Replacing missing values by the half of the minimum of non-missing

values is commonly used and provided by almost all statistical packages and online toolkits, although limitations exist (e.g., distorting distributions) (Wei et al., 2018). We compared the performance of Tobit models with that of MLR models with the expectation that features with a high percent presence would be identified by both models (e.g., for MLR, these features would have little imputation) and that as percent presence decreases, the Tobit models would be able to identify more features than MLR models. As shown in Tables 2 and S1, in general, Tobit models identified more significant features than MLR models. However, Figures S9 to S14 suggest an opposite trend to our expectation that a number of features with a high percent presence were significant only in Tobit models. There are two possible reasons. Firstly, we used half of the minimum feature intensity of the whole dataset for imputation for MLR models. For features with high abundance and a high percent presence among participants, the relatively low imputed values may have been outliers which would severely distort the association between these features and exposures. Replacing missing values with half of the feature-specific minimum is an alternative method in this situation that could be considered in future work. Secondly, the p -values of the features that were uniquely significant in Tobit models were close to the significant threshold as well in MLR models. The difference was not so dramatic as showed by the significance we observed. Further analysis can be constructed to verify these two explanations.

There are several limitations in the present study. First, metabolism exhibits diurnal variation. Diurnal changes of energy expenditure and intake distribution are associated with many factors, including time of day, amount of sleep, timing of meals, and light-dark cycles (Douglas I Walker, Go, Liu, Pennell, & Jones, 2016). Principal component analysis on averaged metabolic profiles according to time of sampling shows three time-of-day patterns, morning, afternoon, and night (Park et al., 2009). Diurnal changes may influence intra-individual variation in response to TRAP, so the time of sampling should

be considered in interpretation of results. . Second, the exposure assessment in the current study was relatively simple and may not have been sufficiently accurate to observe associations of interest. Exposures were measured at one ambient monitoring site located near downtown Atlanta and assigned to participants living across the city. So, the spatial gradients in day-to-day pollutant concentrations were not captured. Exposures of participants living near highways may have been underestimated. Moreover, due to the lack of information on daily activities, we were not able to account for factors that may have affect participants' daily exposures to ambient concentrations, such as time spent inside or outside. Finally, metabolic features were not annotated or identified in our study. This presents difficulties in the interpretation of results since with the pathway analysis that was conducted, it is not possible to determine which nodes in the metabolic pathways are primarily impacted and whether the biological processes are upregulated or downregulated, only that they are perturbed.

Conclusions and Recommendations

Despite these limitations, we identified a range of ambient pollutants, including components of TRAP, associated changes to the metabolic phenotype. The results demonstrate the use of HRM as a viable platform for untargeted characterization of molecular mechanisms underlying exposures to TRAP. The biological pathways identified are primarily involved in nucleotide metabolism, lipid metabolism, and amino acid metabolism. These results provide further evidence for the hypothesis that exposure to TRAP can induce biological effects on the human body via a range of mechanisms manifested by perturbed biological pathways, and oxidative stress is a plausible explanation due to the capacity of ROS to oxidize DNA, lipids, and proteins. Future work on HRM and air pollution in the CHDWB Cohort is planned that will utilize information from the repeated measures over the set of annual follow-up visits for each participant in the cohort. We anticipate that these data will provide a rich resource for

validating the underlying mechanisms associated with air pollution, when used in combination with more comprehensive exposure assessment and feature annotation (e.g., annotating compounds based upon physicochemical properties and/or spectral similarity with public spectral libraries) or identification (e.g., verifying compounds with authentic chemical standards analyzed under the same experimental conditions) (Sumner et al., 2007).

References

- Alonso, A., Marsal, S., & Julia, A. (2015). Analytical methods in untargeted metabolomics: state of the art in 2015. *Front Bioeng Biotechnol*, 3, 23. doi:10.3389/fbioe.2015.00023
- Baccarelli, A., Wright, R. O., Bollati, V., Tarantini, L., Litonjua, A. A., Suh, H. H., . . . Schwartz, J. (2009). Rapid DNA methylation changes after exposure to traffic particles. *Am J Respir Crit Care Med*, 179(7), 572-578. doi:10.1164/rccm.200807-1097OC
- Brauer, M., Lencar, C., Tamburic, L., Koehoorn, M., Demers, P., & Karr, C. (2008). A cohort study of traffic-related air pollution impacts on birth outcomes. *Environ Health Perspect*, 116(5), 680-686. doi:10.1289/ehp.10952
- Brigham, K. L. (2010). Predictive health: the imminent revolution in health care. *J Am Geriatr Soc*, 58 Suppl 2, S298-302. doi:10.1111/j.1532-5415.2010.03107.x
- Brook, R. D., Jerreft, M., Brook, J. R., Bard, R. L., & Finkelstein, M. M. (2008). The relationship between diabetes mellitus and traffic-related air pollution. *Journal of Occupational and Environmental Medicine*, 50(1), 32-38. doi:10.1097/JOM.0b013e31815dba70
- Carvalho, R. B., Carneiro, M. F. H., Barbosa, F., Batista, B. L., Simonetti, J., Amantea, S. L., & Rhoden, C. R. (2018). The impact of occupational exposure to traffic-related air pollution among professional motorcyclists from Porto Alegre, Brazil, and its association with genetic and oxidative damage. *Environmental Science and Pollution Research*, 25(19), 18620-18631. doi:10.1007/s11356-018-2007-1
- Chen, Z., Salam, M. T., Toledo-Corral, C., Watanabe, R. M., Xiang, A. H., Buchanan, T. A., . . . Gilliland, F. D. (2016). Ambient air pollutants have adverse effects on insulin and glucose homeostasis in Mexican Americans. *Diabetes Care*, 39(4), 547-554. doi:10.2337/dc15-1795

- Chiu, Y. H., Garshick, E., Hart, J. E., Spiegelman, D., Dockery, D. W., Smith, T. J., & Laden, F. (2016). Occupational vehicle-related particulate exposure and inflammatory markers in trucking industry workers. *Environ Res*, *148*, 310-317. doi:10.1016/j.envres.2016.04.008
- Chu, J. H., Hart, J. E., Chhabra, D., Garshick, E., Raby, B. A., & Laden, F. (2016). Gene expression network analyses in response to air pollution exposures in the trucking industry. *Environ Health*, *15*(1), 101. doi:10.1186/s12940-016-0187-z
- Golan, R., Ladva, C., Greenwald, R., Krall, J. R., Raysoni, A. U., Kewada, P., . . . Sarnat, J. (2018). Acute pulmonary and inflammatory response in young adults following a scripted car commute. *Air Quality Atmosphere and Health*, *11*(2), 123-136. doi:10.1007/s11869-017-0530-8
- Gouveia-Figueira, S., Karimpour, M., Bosson, J. A., Blomberg, A., Unosson, J., Sehlstedt, M., . . . Nording, M. L. (2018). Mass spectrometry profiling reveals altered plasma levels of monohydroxy fatty acids and related lipids in healthy humans after controlled exposure to biodiesel exhaust. *Anal Chim Acta*, *1018*, 62-69. doi:10.1016/j.aca.2018.02.032
- Greenbaum, D. S. J. A. P., & CANCER. (2013). . Sources of air pollution: gasoline and diesel engines. 49-62.
- Gropper, S. S., & Smith, J. L. (2012). *Advanced nutrition and human metabolism*: Cengage Learning.
- Hansell, A. L., Bakolis, I., Cowie, C. T., Belousova, E. G., Ng, K., Weber-Chrysochoou, C., . . . Marks, G. B. (2018). Childhood fish oil supplementation modifies associations between traffic related air pollution and allergic sensitisation. *Environ Health*, *17*(1), 27. doi:10.1186/s12940-018-0370-5

- Hansen, D. A., Edgerton, E., Hartsell, B., Jansen, J., Burge, H., Koutrakis, P., . . . Rasmussen, R. (2006). Air quality measurements for the aerosol research and inhalation epidemiology study. *J Air Waste Manag Assoc*, *56*(10), 1445-1458.
- Health Effects Institute. (2010). *Traffic-related air pollution: a critical review of the literature on emissions, exposure, and health effects*. Retrieved from <https://www.healtheffects.org/publication/traffic-related-air-pollution-critical-review-literature-emissions-exposure-and-health>
- Huang, H. B., Lai, C. H., Chen, G. W., Lin, Y. Y., Jaakkola, J. J., Liou, S. H., & Wang, S. L. (2012). Traffic-related air pollution and DNA damage: a longitudinal study in Taiwanese traffic conductors. *PLoS One*, *7*(5), e37412. doi:10.1371/journal.pone.0037412
- Jacobs, L., Nawrot, T. S., de Geus, B., Meeusen, R., Degraeuwe, B., Bernard, A., . . . Panis, L. I. (2010). Subclinical responses in healthy cyclists briefly exposed to traffic-related air pollution: an intervention study. *Environ Health*, *9*, 64. doi:10.1186/1476-069X-9-64
- Jeong, A., Fiorito, G., Keski-Rahkonen, P., Imboden, M., Kiss, A., Robinot, N., . . . Probst-Hensch, N. (2018). Perturbation of metabolic pathways mediates the association of air pollutants with asthma and cardiovascular diseases. *Environ Int*, *119*, 334-345. doi:10.1016/j.envint.2018.06.025
- Jiang, S., Bo, L., Gong, C. Y., Du, X. H., Kan, H. D., Xie, Y. Q., . . . Zhao, J. Z. (2016). Traffic-related air pollution is associated with cardio-metabolic biomarkers in general residents. *International Archives of Occupational and Environmental Health*, *89*(6), 911-921. doi:10.1007/s00420-016-1129-3
- Jin, Z., Kang, J., & Yu, T. (2018). Missing value imputation for LC-MS metabolomics data by incorporating metabolic network and adduct ion relations. *Bioinformatics*, *34*(9), 1555-1561. doi:10.1093/bioinformatics/btx816

- Jones, D. P., Park, Y., & Ziegler, T. R. (2012). Nutritional metabolomics: progress in addressing complexity in diet and health. *Annu Rev Nutr*, 32, 183-202.
doi:10.1146/annurev-nutr-072610-145159
- Juniper, E., Frith, P., & Hargreave, F. J. T. (1981). Airway responsiveness to histamine and methacholine: relationship to minimum treatment to control symptoms of asthma. 36(8), 575-579.
- Kanehisa, M., & Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*, 28(1), 27-30.
- Khan, J., Brennand, D. M., Bradley, N., Gao, B., Bruckdorfer, R., & Jacobs, M. (1998). 3-Nitrotyrosine in the proteins of human plasma determined by an ELISA method. *Biochem J*, 332 (Pt 3), 807-808.
- Kramer, U., Herder, C., Sugiri, D., Strassburger, K., Schikowski, T., Ranft, U., & Rathmann, W. (2010). Traffic-related air pollution and incident type 2 diabetes: results from the SALIA cohort study. *Environ Health Perspect*, 118(9), 1273-1279. doi:10.1289/ehp.0901689
- Krauskopf, J., Caiment, F., van Veldhoven, K., Chadeau-Hyam, M., Sinharay, R., Chung, K. F., . . . Kleinjans, J. C. (2018). The human circulating miRNome reflects multiple organ disease risks in association with short-term exposure to traffic-related air pollution. *Environ Int*, 113, 26-34.
doi:10.1016/j.envint.2018.01.014
- Krishnan, R. M., Sullivan, J. H., Carlsten, C., Wilkerson, H. W., Beyer, R. P., Bammler, T., . . . Kaufman, J. D. (2013). A randomized cross-over study of inhalation of diesel exhaust, hematological indices, and endothelial markers in humans. *Part Fibre Toxicol*, 10, 7. doi:10.1186/1743-8977-10-7
- Kubesch, N. J., de Nazelle, A., Westerdahl, D., Martinez, D., Carrasco-Turigas, G., Bouso, L., . . . Nieuwenhuijsen, M. J. (2015). Respiratory and inflammatory

responses to short-term exposure to traffic-related air pollution with and without moderate physical activity. *Occup Environ Med*, 72(4), 284-293.

doi:10.1136/oemed-2014-102106

Ladva, C. N., Golan, R., Greenwald, R., Yu, T., Sarnat, S. E., Flanders, W. D., . . .

Sarnat, J. A. (2017). Metabolomic profiles of plasma, exhaled breath condensate, and saliva are correlated with potential for air toxics detection. *J Breath Res*,

12(1), 016008. doi:10.1088/1752-7163/aa863c

Ladva, C. N., Golan, R., Liang, D., Greenwald, R., Walker, D. I., Uppal, K., . . . Sarnat, J.

A. (2018). Particulate metal exposures induce plasma metabolome changes in a commuter panel study. *PLoS One*, 13(9), e0203468.

doi:10.1371/journal.pone.0203468

Lelieveld, J., Evans, J. S., Fnais, M., Giannadaki, D., & Pozzer, A. (2015). The

contribution of outdoor air pollution sources to premature mortality on a global scale. *Nature*, 525(7569), 367-371. doi:10.1038/nature15371

Li, S., Dunlop, A. L., Jones, D. P., & Corwin, E. J. (2016). High-resolution metabolomics:

review of the field and implications for nursing science and the study of preterm birth. *Biol Res Nurs*, 18(1), 12-22. doi:10.1177/1099800415595463

Liang, D., Moutinho, J. L., Golan, R., Yu, T., Ladva, C. N., Niedzwiecki, M., . . . Sarnat, J.

A. (2018). Use of high-resolution metabolomics for the identification of metabolic signals associated with traffic-related air pollution. *Environ Int*, 120, 145-154.

doi:10.1016/j.envint.2018.07.044

Liigand, P., Kaupmees, K., Haav, K., Liigand, J., Leito, I., Girod, M., . . . Krueve, A.

(2017). Think negative: finding the best electrospray ionization/MS mode for your analyte. *Anal Chem*, 89(11), 5665-5668. doi:10.1021/acs.analchem.7b00096

- Linton, M. F., Yancey, P. G., Davies, S. S., Jerome, W. G., Linton, E. F., Song, W. L., . . . Vickers, K. C. (2019). The role of lipids and lipoproteins in atherosclerosis. In *Endotext [Internet]*: MDText. com, Inc.
- McBee, M. (2010). Modeling outcomes with floor or ceiling effects: an introduction to the Tobit model. *Gifted Child Quarterly*, *54*(4), 314-320.
doi:10.1177/0016986210379095
- Miller, D. B., Ghio, A. J., Karoly, E. D., Bell, L. N., Snow, S. J., Madden, M. C., . . . Kodavanti, U. P. (2016). Ozone exposure increases circulating stress hormones and lipid metabolites in humans. *Am J Respir Crit Care Med*, *193*(12), 1382-1391. doi:10.1164/rccm.201508-1599OC
- Park, Y., Kim, S. B., Wang, B., Blanco, R. A., Le, N. A., Wu, S., . . . Jones, D. P. (2009). Individual variation in macronutrient regulation measured by proton magnetic resonance spectroscopy of human plasma. *Am J Physiol Regul Integr Comp Physiol*, *297*(1), R202-209. doi:10.1152/ajpregu.90757.2008
- Rask, K. J., Brigham, K. L., & Johns, M. M. E. (2011). Integrating comparative effectiveness research programs into predictive health: A unique role for Academic Health Centers. *Academic Medicine*, *86*(6), 718-723.
doi:10.1097/ACM.0b013e318217ea6c
- Rossner, P., Jr., Svecova, V., Milcova, A., Lnenickova, Z., Solansky, I., Santella, R. M., & Sram, R. J. (2007). Oxidative and nitrosative stress markers in bus drivers. *Mutat Res*, *617*(1-2), 23-32. doi:10.1016/j.mrfmmm.2006.11.033
- Sarnat, J. A., Golan, R., Greenwald, R., Raysoni, A. U., Kewada, P., Winquist, A., . . . Yip, F. (2014). Exposure to traffic pollution, acute inflammation and autonomic response in a panel of car commuters. *Environ Res*, *133*, 66-76.
doi:10.1016/j.envres.2014.05.004

- Soltow, Q. A., Strobel, F. H., Mansfield, K. G., Wachtman, L., Park, Y., & Jones, D. P. (2013). High-performance metabolic profiling with dual chromatography-Fourier-transform mass spectrometry (DC-FTMS) for study of the exposome. *Metabolomics*, 9(1 Suppl), S132-S143. doi:10.1007/s11306-011-0332-1
- Stadtman, E. R. (2006). Protein oxidation and aging. *Free Radic Res*, 40(12), 1250-1258. doi:10.1080/10715760600918142
- Stadtman, E. R., & Berlett, B. S. (1998). Reactive oxygen-mediated protein oxidation in aging and disease. *Drug Metab Rev*, 30(2), 225-243. doi:10.3109/03602539808996310
- Steadman, R. G. (1984). A Universal scale of apparent temperature. *Journal of Climate and Applied Meteorology*, 23(12), 1674-1687. doi:Doi 10.1175/1520-0450(1984)023<1674:Ausoat>2.0.Co;2
- Sumner, L. W., Amberg, A., Barrett, D., Beale, M. H., Beger, R., Daykin, C. A., . . . Viant, M. R. (2007). Proposed minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). *Metabolomics*, 3(3), 211-221. doi:10.1007/s11306-007-0082-2
- Surowiec, I., Karimpour, M., Gouveia-Figueira, S., Wu, J., Unosson, J., Bosson, J. A., . . . Nording, M. L. (2016). Multi-platform metabolomics assays for human lung lavage fluids in an air pollution exposure study. *Anal Bioanal Chem*, 408(17), 4751-4764. doi:10.1007/s00216-016-9566-0
- Tabassum, R., Cunningham, L., Stephens, E. H., Sturdivant, K., Martin, G. S., Brigham, K. L., & Gibson, G. (2014). A Longitudinal study of health Improvement in the Atlanta CHDWB Wellness Cohort. *J Pers Med*, 4(4), 489-507. doi:10.3390/jpm4040489
- Uppal, K., Soltow, Q. A., Strobel, F. H., Pittard, W. S., Gernert, K. M., Yu, T., & Jones, D. P. (2013). xMSanalyzer: automated pipeline for improved feature detection and

- downstream analysis of large-scale, non-targeted metabolomics data. *BMC Bioinformatics*, 14, 15. doi:10.1186/1471-2105-14-15
- Uppal, K., Walker, D. I., Liu, K., Li, S., Go, Y. M., & Jones, D. P. (2016). Computational metabolomics: A framework for the million metabolome. *Chem Res Toxicol*, 29(12), 1956-1975. doi:10.1021/acs.chemrestox.6b00179
- van Veldhoven, K., Kiss, A., Keski-Rahkonen, P., Robinot, N., Scalbert, A., Cullinan, P., . . . Vineis, P. (2019). Impact of short-term traffic-related air pollution on the metabolome - Results from two metabolome-wide experimental studies. *Environ Int*, 123, 124-131. doi:10.1016/j.envint.2018.11.034
- Vlaanderen, J. J., Janssen, N. A., Hoek, G., Keski-Rahkonen, P., Barupal, D. K., Cassee, F. R., . . . Vermeulen, R. C. H. (2017). The impact of ambient air pollution on the human blood metabolome. *Environ Res*, 156, 341-348. doi:10.1016/j.envres.2017.03.042
- Walker, D. I., Go, Y.-M., Liu, K., Pennell, K. D., & Jones, D. P. (2016). Population screening for biological and environmental properties of the human metabolic phenotype: implications for personalized medicine. In *Metabolic Phenotyping in Personalized and Public Healthcare* (pp. 167-211): Elsevier.
- Walker, D. I., Lane, K. J., Liu, K., Uppal, K., Patton, A. P., Durant, J. L., . . . Pennell, K. D. (2018). Metabolomic assessment of exposure to near-highway ultrafine particles. *J Expo Sci Environ Epidemiol*. doi:10.1038/s41370-018-0102-5
- Wei, R., Wang, J., Su, M., Jia, E., Chen, S., Chen, T., & Ni, Y. (2018). Missing value imputation approach for mass spectrometry-based metabolomics data. *Sci Rep*, 8(1), 663. doi:10.1038/s41598-017-19120-0
- Wilhelm, M., Ghosh, J. K., Su, J., Cockburn, M., Jerrett, M., & Ritz, B. (2011). Traffic-related air toxics and preterm birth: a population-based case-control study in Los

Angeles County, California. *Environ Health*, 10, 89. doi:10.1186/1476-069X-10-89

Zhong, J., Cayir, A., Trevisi, L., Sanchez-Guerra, M., Lin, X., Peng, C., . . . Baccarelli, A. (2016). Traffic-related air pollution, blood pressure, and adaptive response of mitochondrial abundance. *Circulation*, 133(4), 378-387.

doi:10.1161/circulationaha.115.018802

Zuurbier, M., Hoek, G., Oldenwening, M., Meliefste, K., Krop, E., van den Hazel, P., & Brunekreef, B. (2011). In-traffic air pollution exposure and CC16, blood coagulation, and inflammation markers in healthy adults. *Environ Health Perspect*, 119(10), 1384-1389. doi:10.1289/ehp.1003151

Tables and Figures

Table 1. Participant characteristics and the temporal characteristics of baseline visits

Characteristics	Number	Proportion (%)
Age, years [median (Q1-Q3)]	51.0 (42-57)	
BMI, lb/in ² [median (Q1-Q3)]	26.4 (23.6-29.7)	
Race/ethnicity		
White	137	76.1
Black	34	18.9
Other races ^a	9	5.0
Gender		
Female	113	62.8
Male	67	37.2
Marital status		
Married	117	65.0
Other statuses ^b	63	35.0
Income		
0-50,000	17	10.1
50,000-100,000	44	26.0
100,000-200,000	58	34.3
200,000+	50	29.6
Missing	11	
Education		
College and high school	78	43.3
Graduate school and above	102	56.7
Smoking status		
Non-smoker	172	95.6
Current smoker	8	4.4
Drinking status		
Yes	139	77.2
No	41	22.8
Year of visit		
2008	76	42.2
2009	95	52.8
Over 2009	9	5.0

Weekday of visit		
Monday	39	21.7
Tuesday	37	20.6
Wednesday	40	22.2
Thursday	30	16.7
Friday	34	18.9
Season of visit		
Spring	29	16.1
Summer	63	35.0
Autumn	44	24.4
Winter	44	24.4

^a Other races includes American Indian or Alaskan Native and Asian.

^b Other statuses includes single, divorced, widowed, separated, and partnered.

Table 2. Number of significant metabolic features by negative and positive ion mode associated with lag 1 day (i.e., previous day) and the moving average of lag 1-2 days pollution from Tobit models.

Total number of features extracted	Lag*	Negative mode ^a	Positive mode ^b
CO	L1	194	216
	MA	225	173
NO ₂	L1	354	338
	MA	283	280
O ₃	L1	311	245
	MA	190	247
PM _{2.5}	L1	337	259
	MA	287	244
EC	L1	264	208
	MA	330	230
OC	L1	322	251
	MA	236	254

*L1, the exposure at lag 1 day; MA, the moving average of exposure at lag 1-2 days; EC, elemental carbon, OC, organic carbon.

^a Metabolic features were statistically significant with p-values less than 0.025.

^b Metabolic features were statistically significant with p-values less than 0.05.

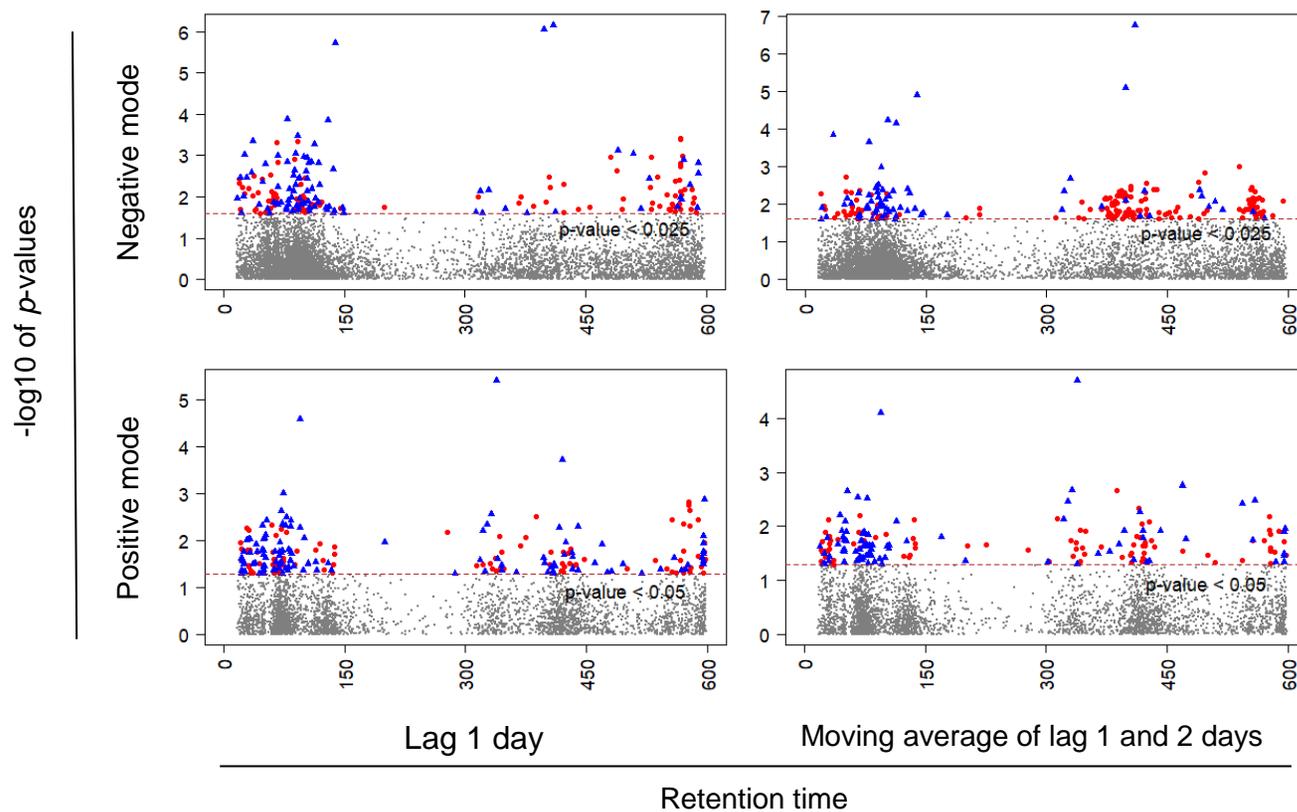


Figure 1. Manhattan plots of associations between log-transformed metabolic feature intensity and CO from Tobit models. X-axis denotes the retention time (in seconds), Y-axis denotes the negative log₁₀ of the *p*-values calculated from the Tobit model. Blue triangles and red circles denote negative and positive associations, respectively.

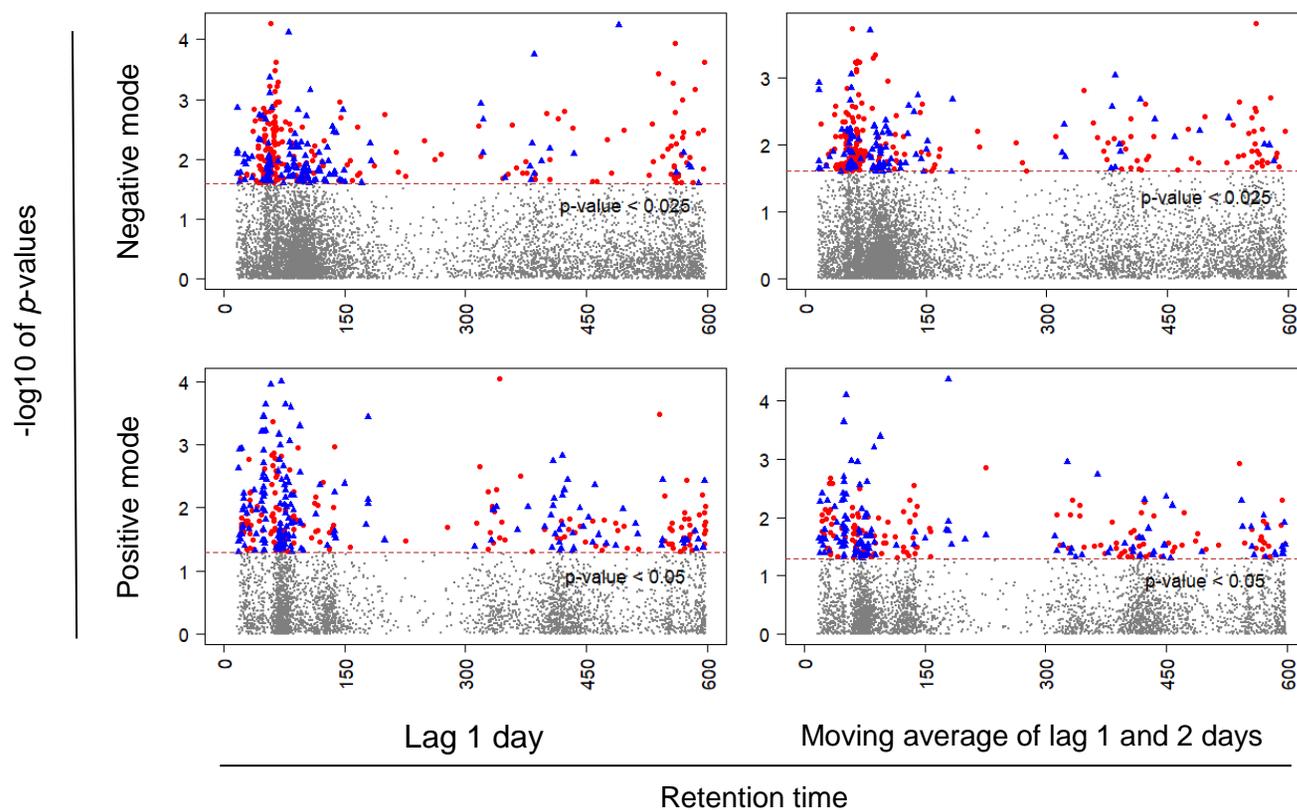


Figure 2. Manhattan plots of associations between log-transformed metabolic feature intensity and NO_2 from Tobit models. X-axis denotes the retention time (in seconds), Y-axis denotes the negative \log_{10} of the p -values calculated from the Tobit model. Blue triangles and red circles denote negative and positive associations, respectively.

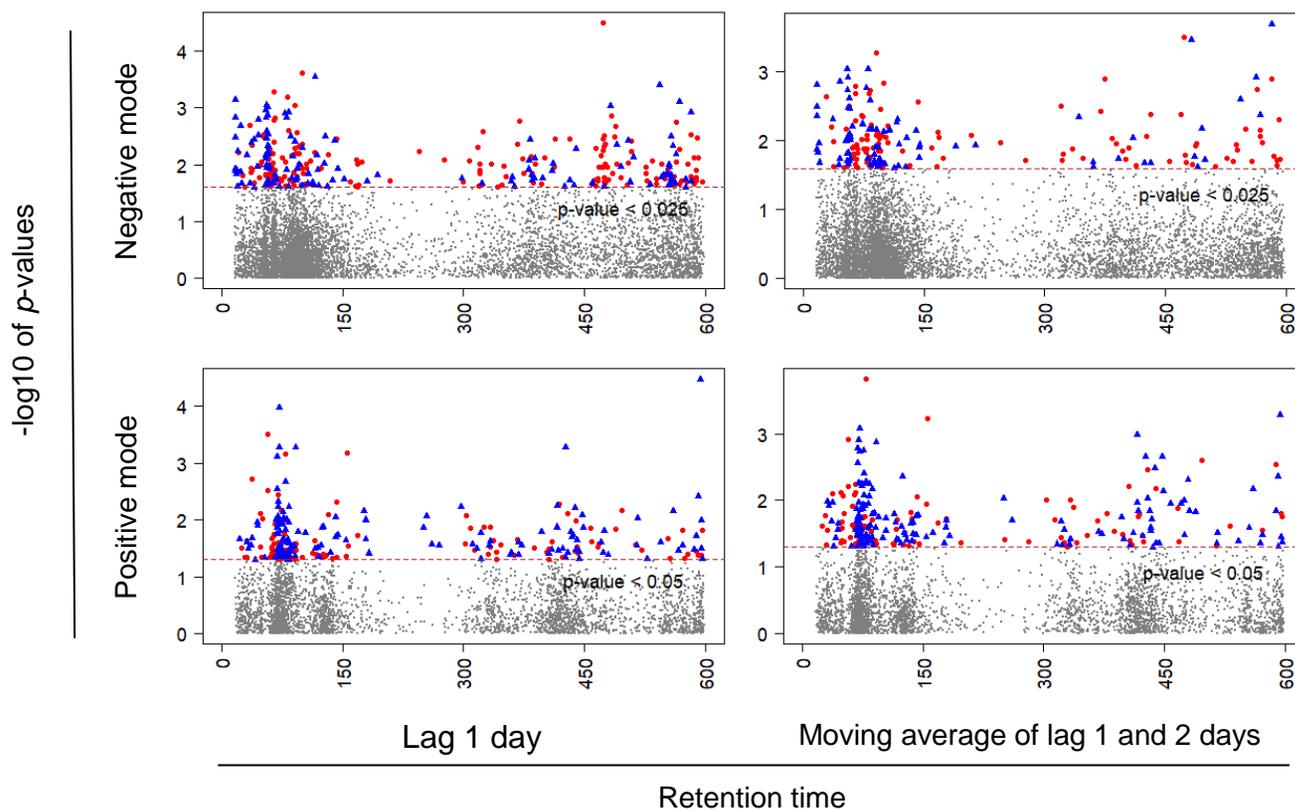


Figure 3. Manhattan plots of associations between log-transformed metabolic feature intensity and O_3 from Tobit models. X-axis denotes the retention time (in seconds), Y-axis denotes the negative log10 of the p -values calculated from the Tobit model. Blue triangles and red circles denote negative and positive associations, respectively.

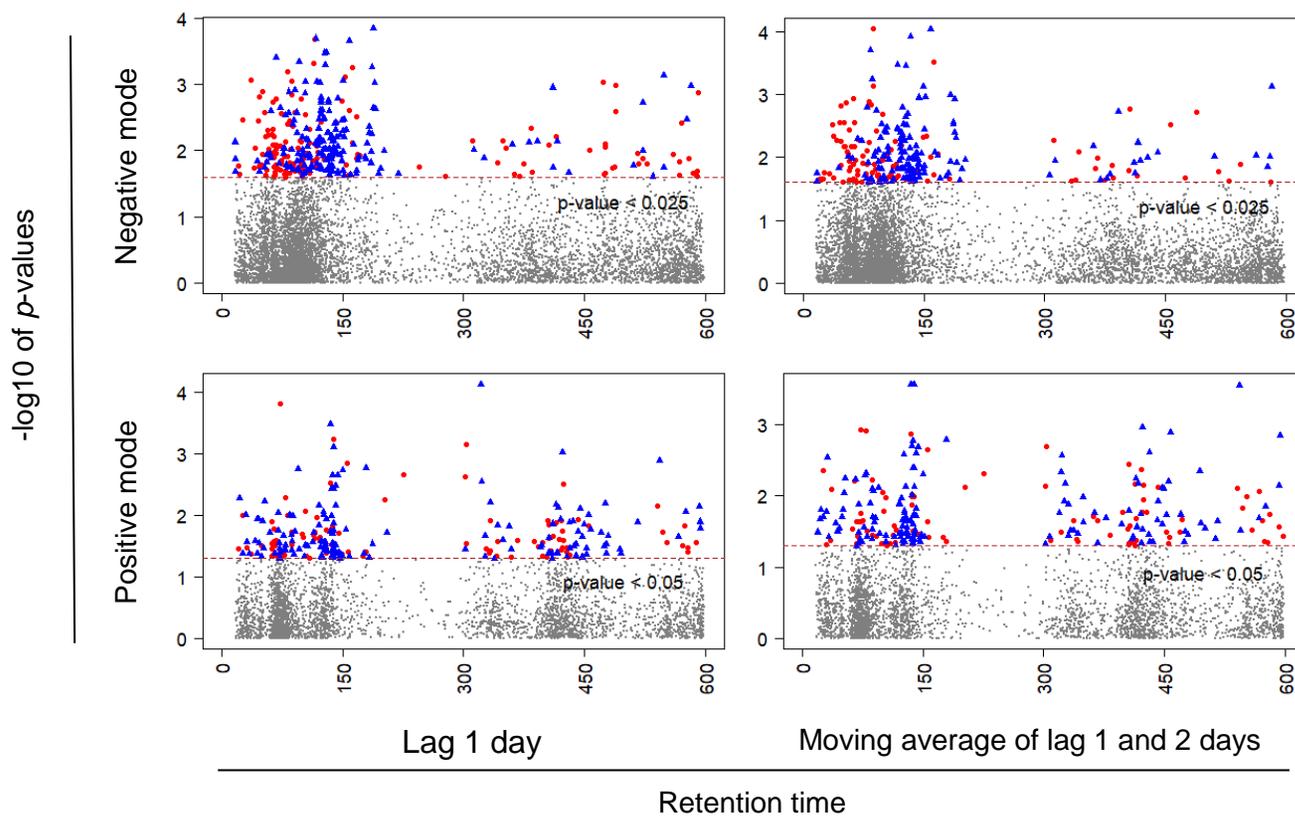


Figure 4. Manhattan plots of associations between log-transformed metabolic feature intensity and $PM_{2.5}$ from Tobit models. X-axis denotes the retention time (in seconds), Y-axis denotes the negative log10 of the p -values calculated from the Tobit model. Blue triangles and red circles denote negative and positive associations, respectively.

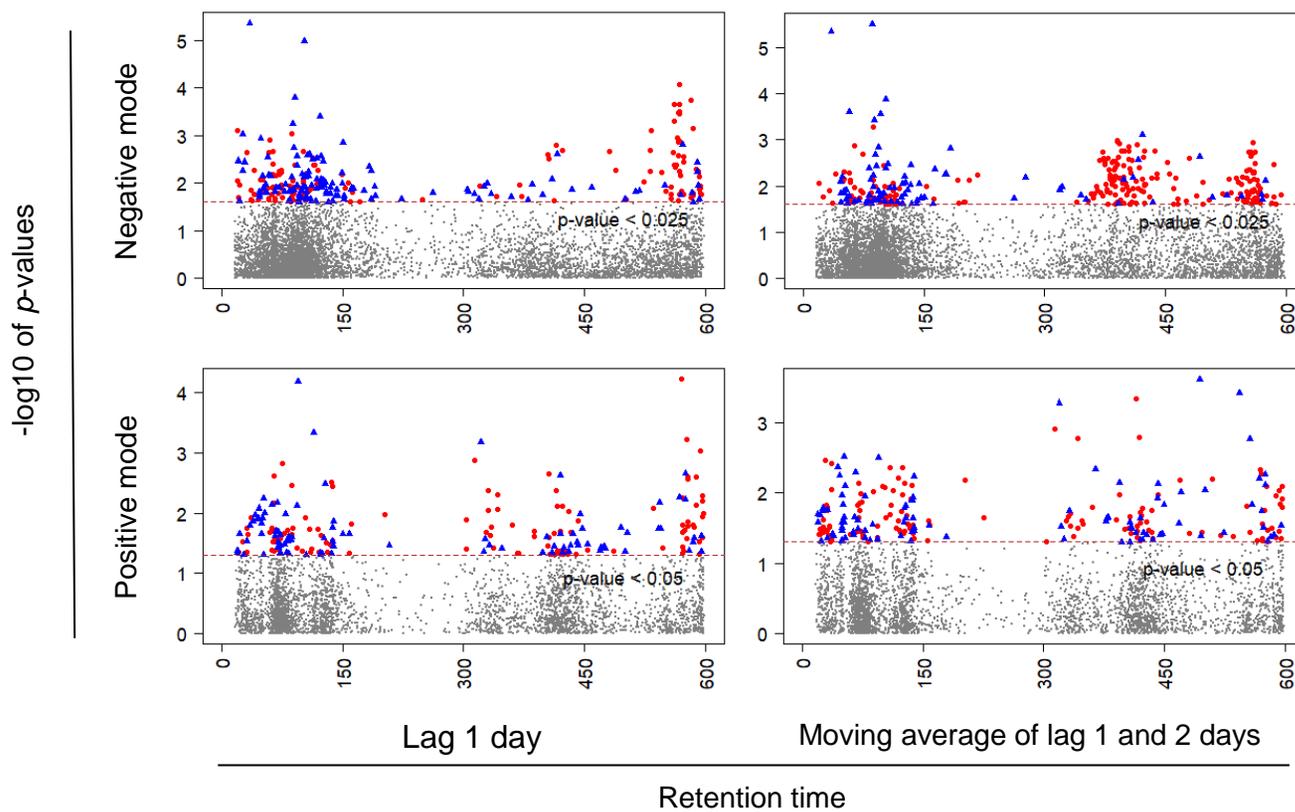


Figure 5. Manhattan plots of associations between log-transformed metabolic feature intensity and EC from Tobit models. X-axis denotes the retention time (in seconds), Y-axis denotes the negative log10 of the p -values calculated from Tobit model. Blue triangles and red circles denote negative and positive associations, respectively.

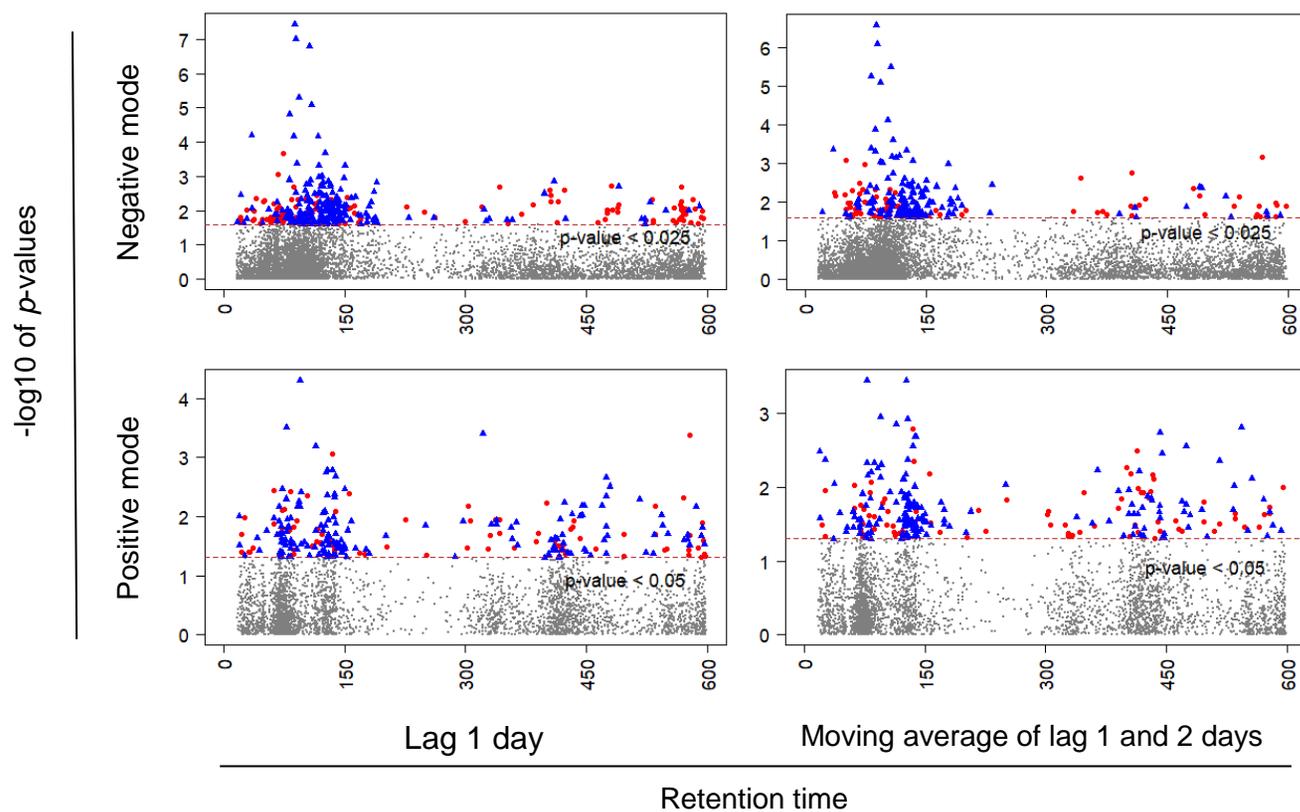


Figure 6. Manhattan plots of associations between log-transformed metabolic feature intensity and OC from Tobit models. X-axis denotes the retention time (in seconds), Y-axis denotes the negative log₁₀ of the *p*-values calculated from the Tobit model. Blue triangles and red circles denote negative and positive associations, respectively.

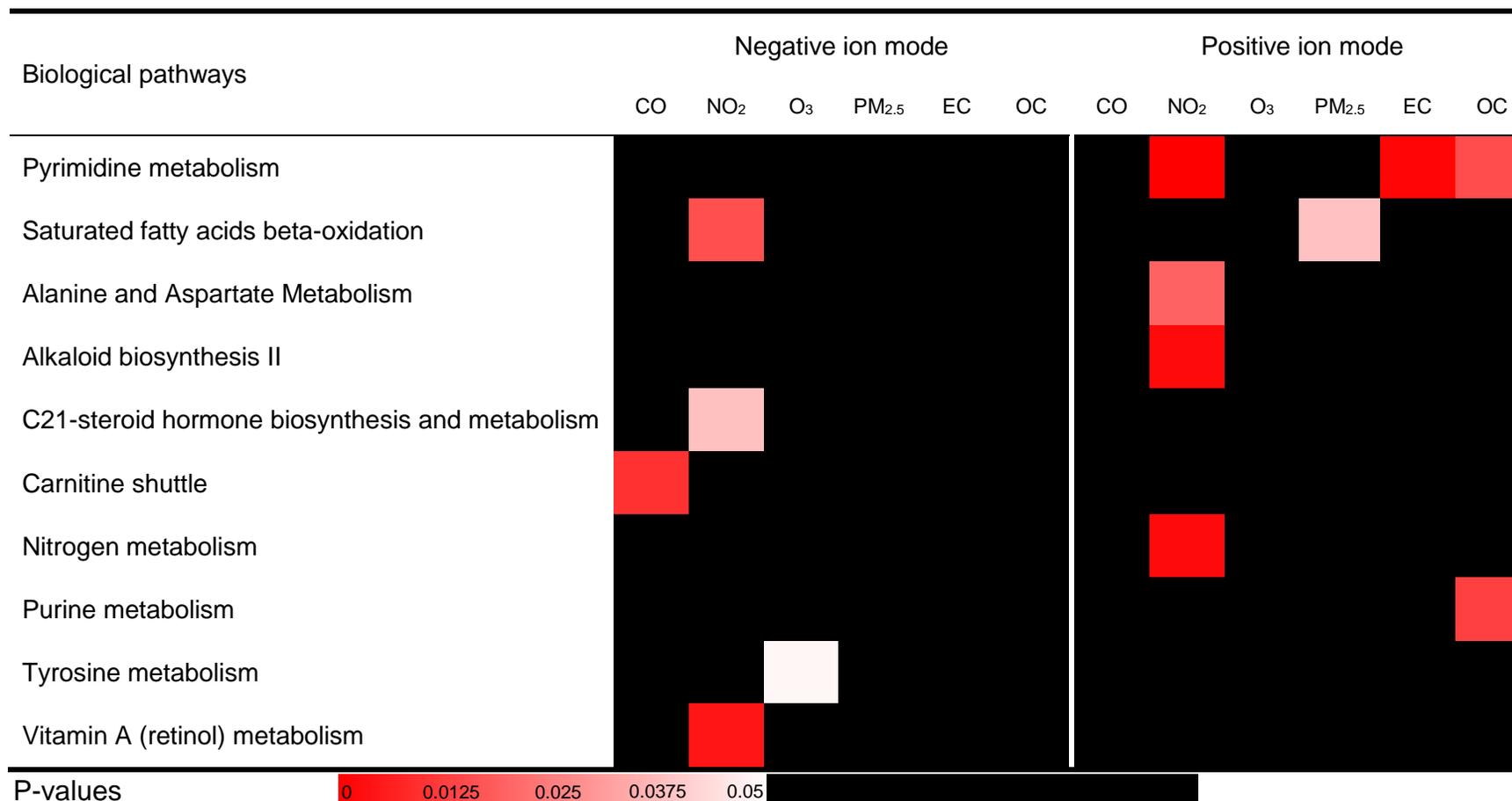


Figure 7. Metabolic pathways associated with 1-day lag pollution from Tobit models. Cells are shaded according to the magnitude of p-values derived from Tobit models. Pathways enriched by more than 2 annotated significant metabolic features were included and ordered according to the total number of the significant associations between pathways and pollutants by either negative or positive ion modes.

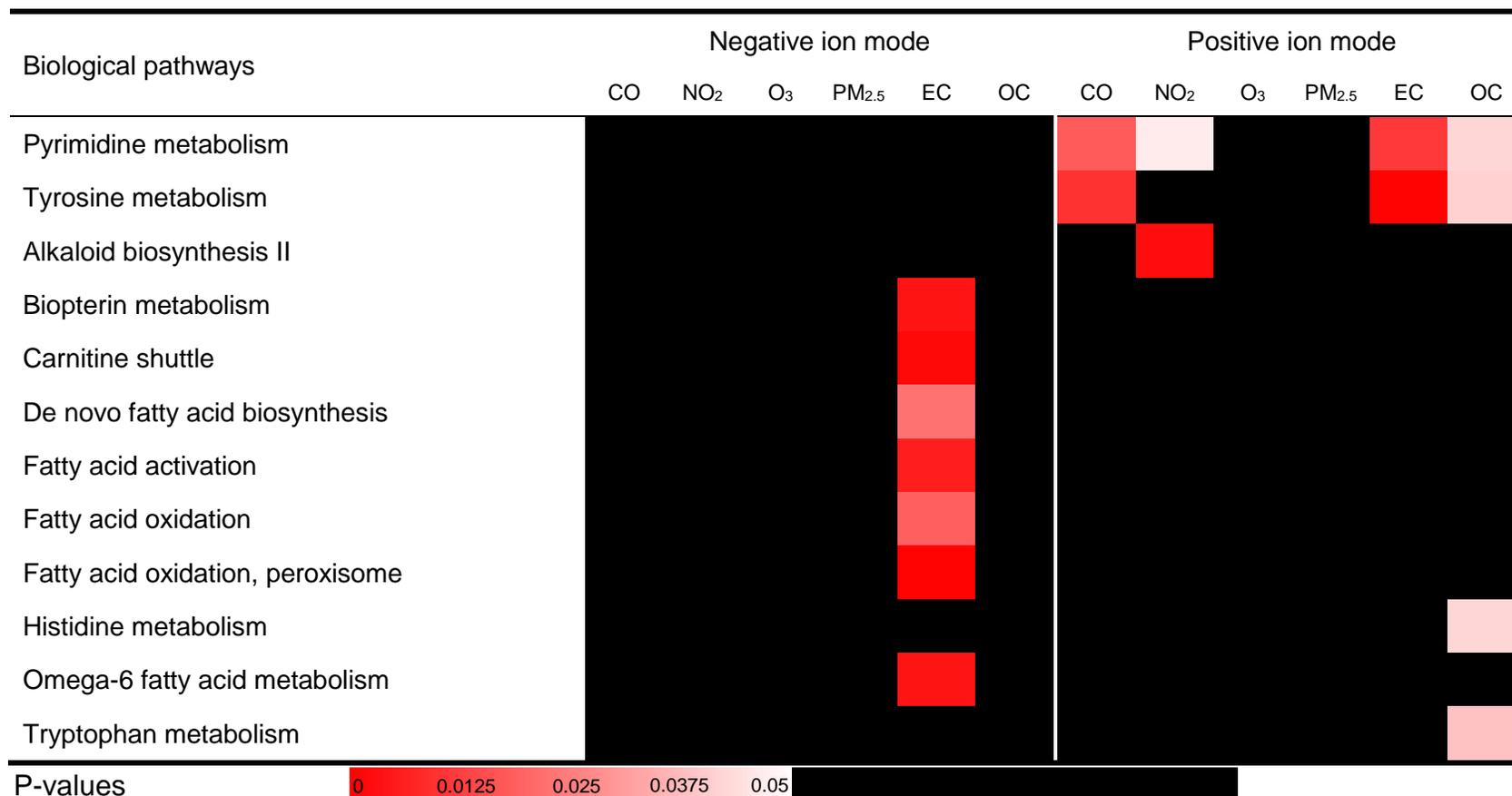


Figure 8. Metabolic pathways associated with the moving average of lag 1-2 days pollution from Tobit models. Cells are shaded according to the magnitude of p-values derived from Tobit models. Pathways enriched by more than 2 annotated significant metabolic features were included and ordered according to the total number of the significant associations between pathways and pollutants by either negative or positive ion modes.

Appendices

Table S1. Number of significant metabolic features by negative and positive ion mode associated with lag 1 day (i.e., previous day) and the moving average of lag 1-2 days pollution from multiple linear regression models.

Total number of features detected	Lag*	Negative mode ^a	Positive mode ^b
CO	L1	218	211
	MA	177	166
NO ₂	L1	338	307
	MA	270	240
O ₃	L1	283	214
	MA	171	227
PM2.5	L1	321	237
	MA	267	235
EC	L1	248	182
	MA	272	215
OC	L1	307	229
	MA	212	241

*L1, the exposure at lag 1 day; MA, the moving average of exposure at lag 1-2 days.

^a Metabolic features were statistically significant with p-values less than 0.025.

^b Metabolic features were statistically significant with p-values less than 0.05.

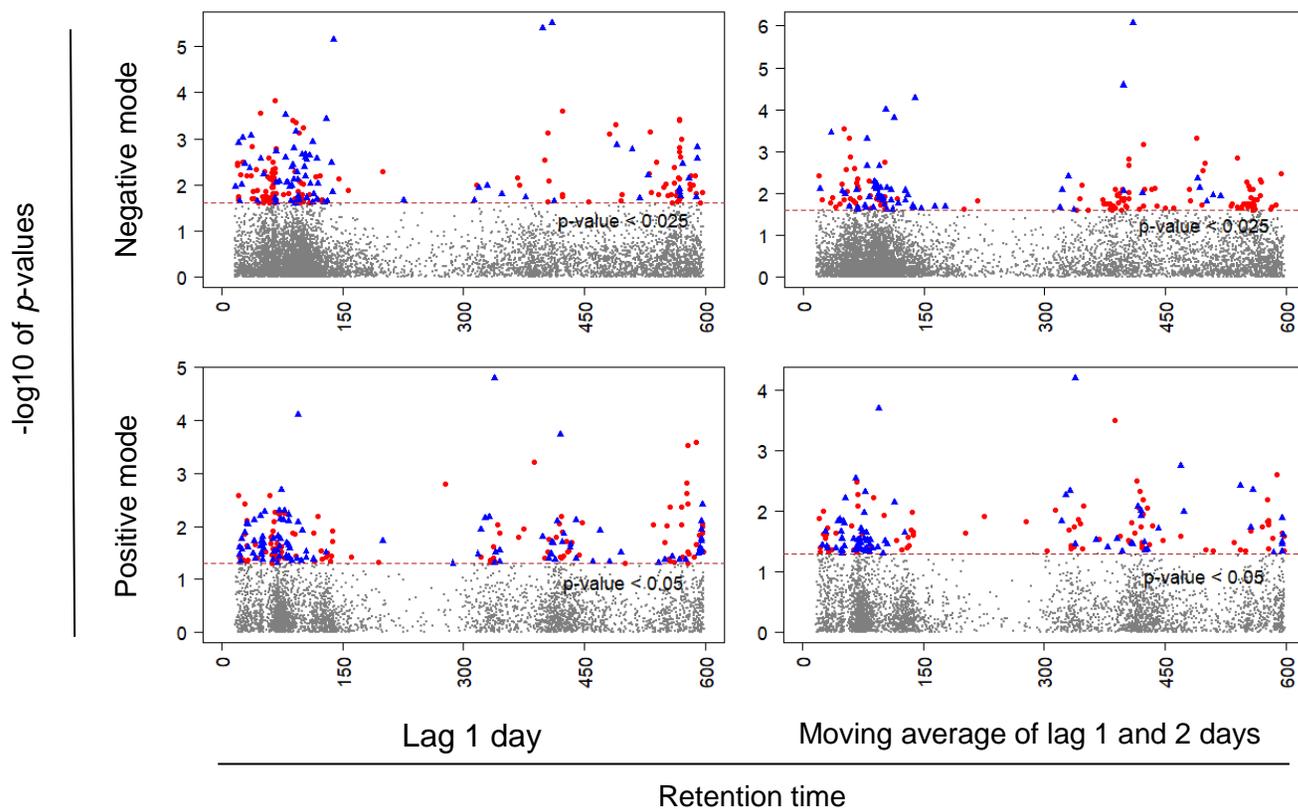


Figure S1. Manhattan plots of associations between log-transformed metabolic feature intensity and CO from multiple linear regression models. X-axis denotes the retention time (in seconds); Y-axis denotes the negative log10 of the p -values calculated from the multiple linear regression model. Blue triangles and red circles denote negative and positive associations, respectively.

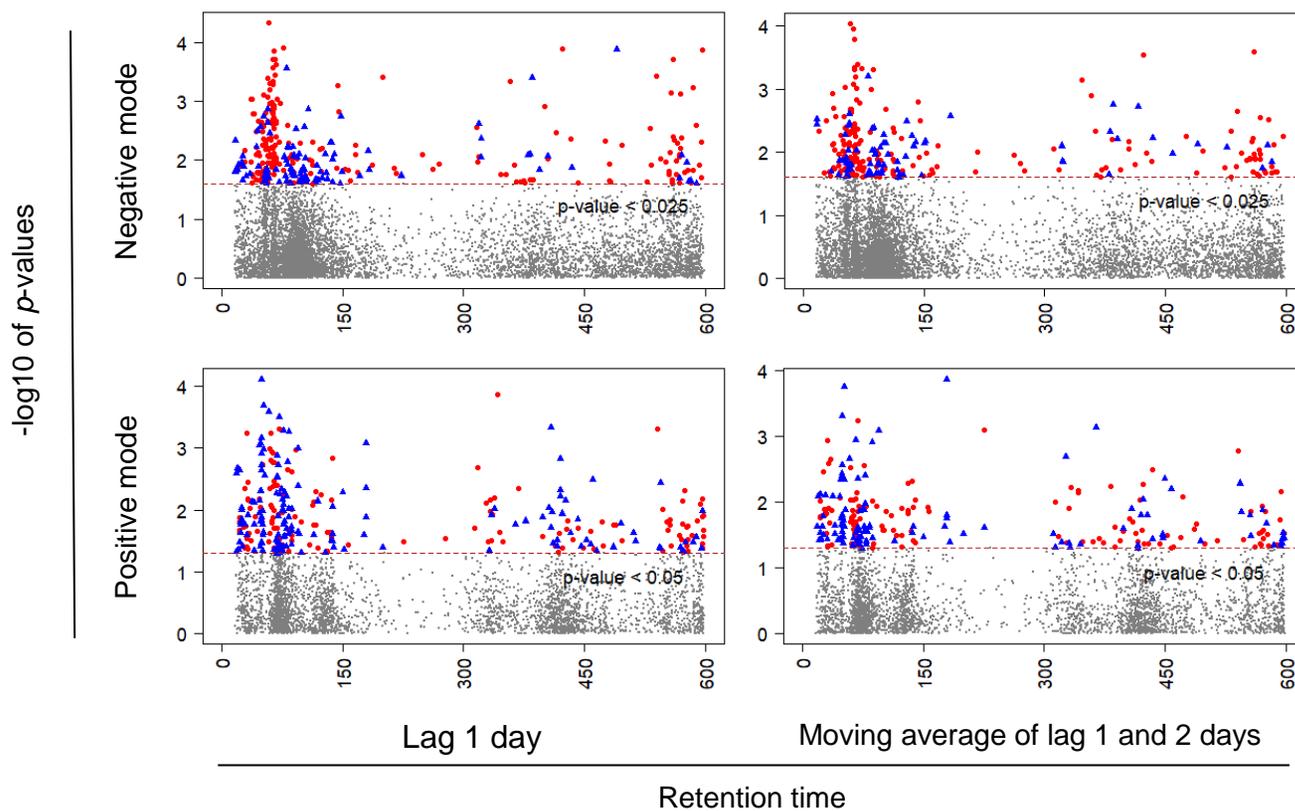


Figure S2. Manhattan plots of associations between log-transformed metabolic feature intensity and NO_2 from multiple linear regression models. X-axis denotes the retention time (in seconds); Y-axis denotes the negative log₁₀ of the p -values calculated from the multiple linear regression model. Blue triangles and red circles denote negative and positive associations, respectively.

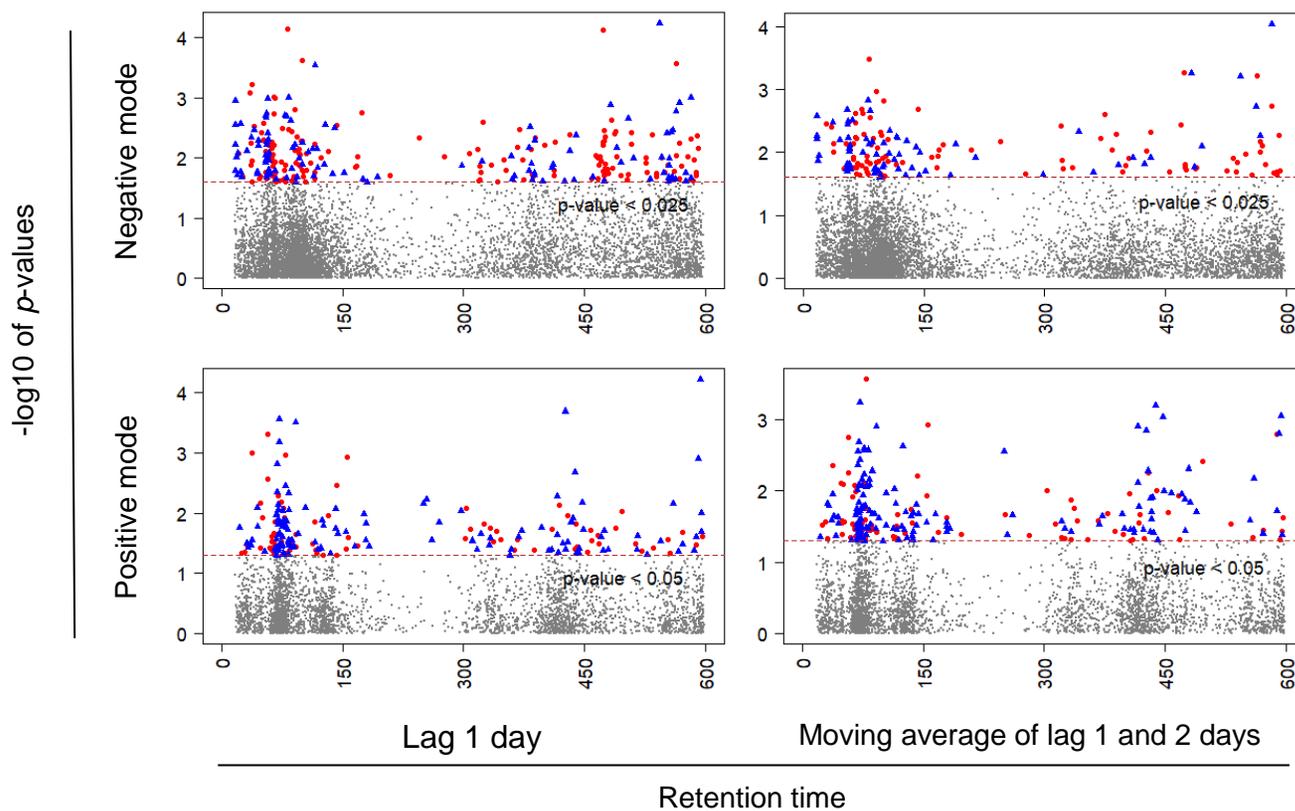


Figure S3. Manhattan plots of associations between log-transformed metabolic feature intensity and O_3 from multiple linear regression models. X-axis denotes the retention time (in seconds); Y-axis denotes the negative log₁₀ of the p -values calculated from the multiple linear regression model. Blue triangles and red circles denote negative and positive associations, respectively.

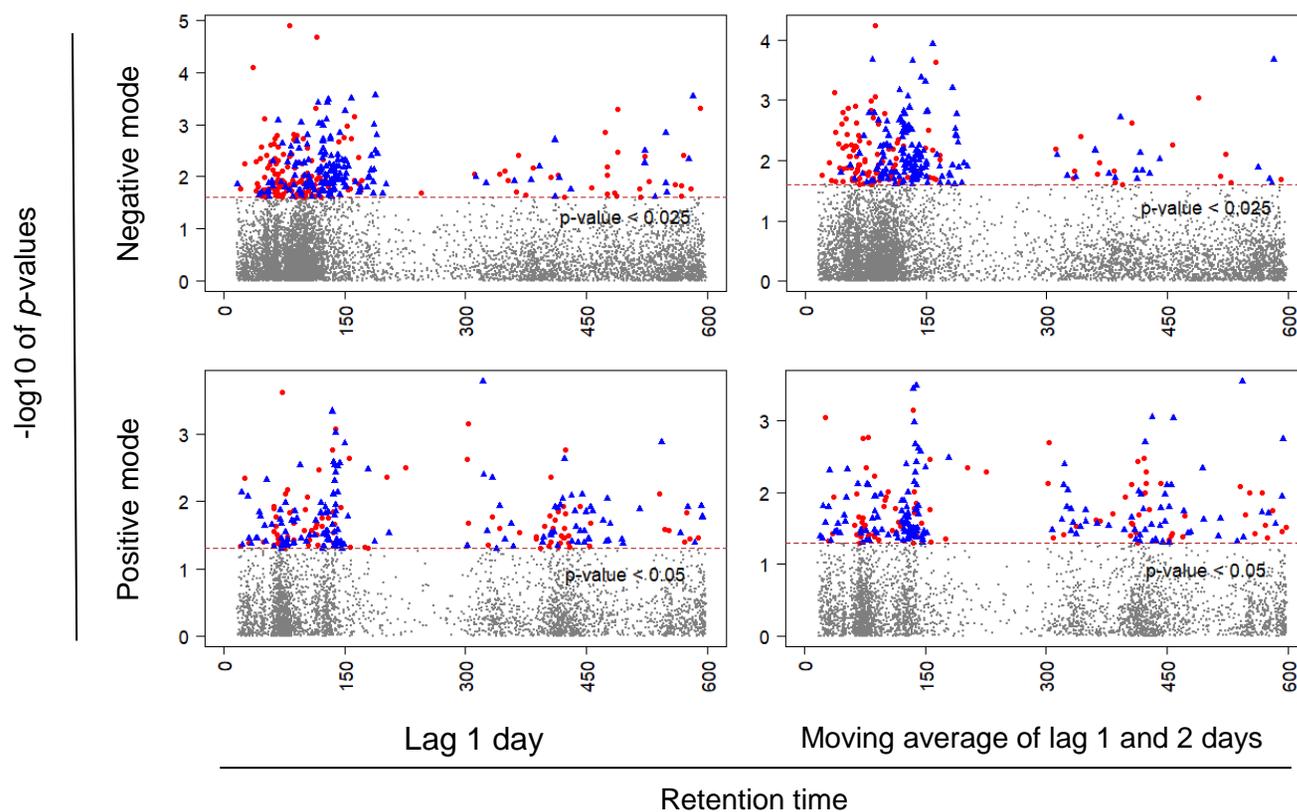


Figure S4. Manhattan plots of associations between log-transformed metabolic feature intensity and PM_{2.5} from multiple linear regression models. X-axis denotes the retention time (in seconds); Y-axis denotes the negative log₁₀ of the *p*-values calculated from the multiple linear regression model. Blue triangles and red circles denote negative and positive associations, respectively.

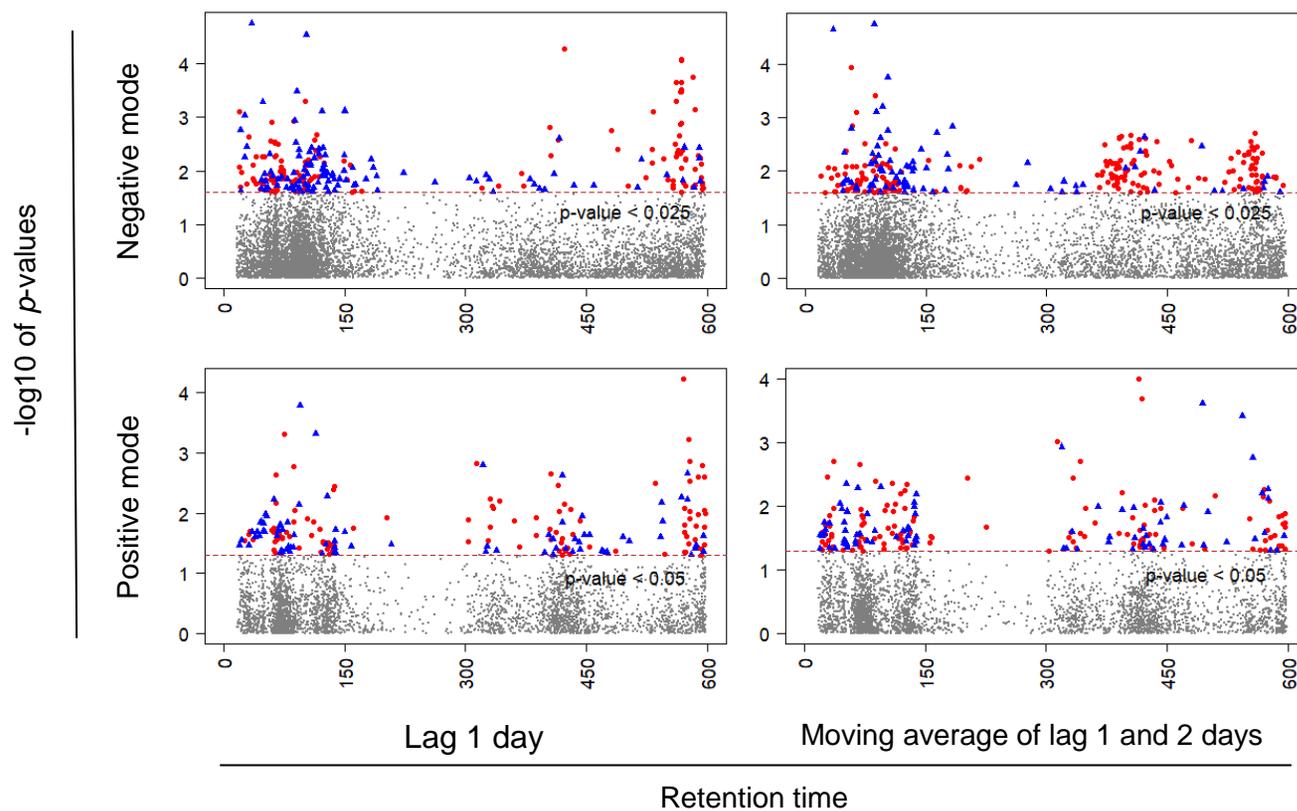


Figure S5. Manhattan plots of associations between log-transformed metabolic feature intensity and EC from multiple linear regression models. X-axis denote the retention time (in seconds); Y-axis denotes the negative log₁₀ of the *p*-values calculated from the multiple linear regression model. Blue triangles and red circles denote negative and positive associations, respectively.

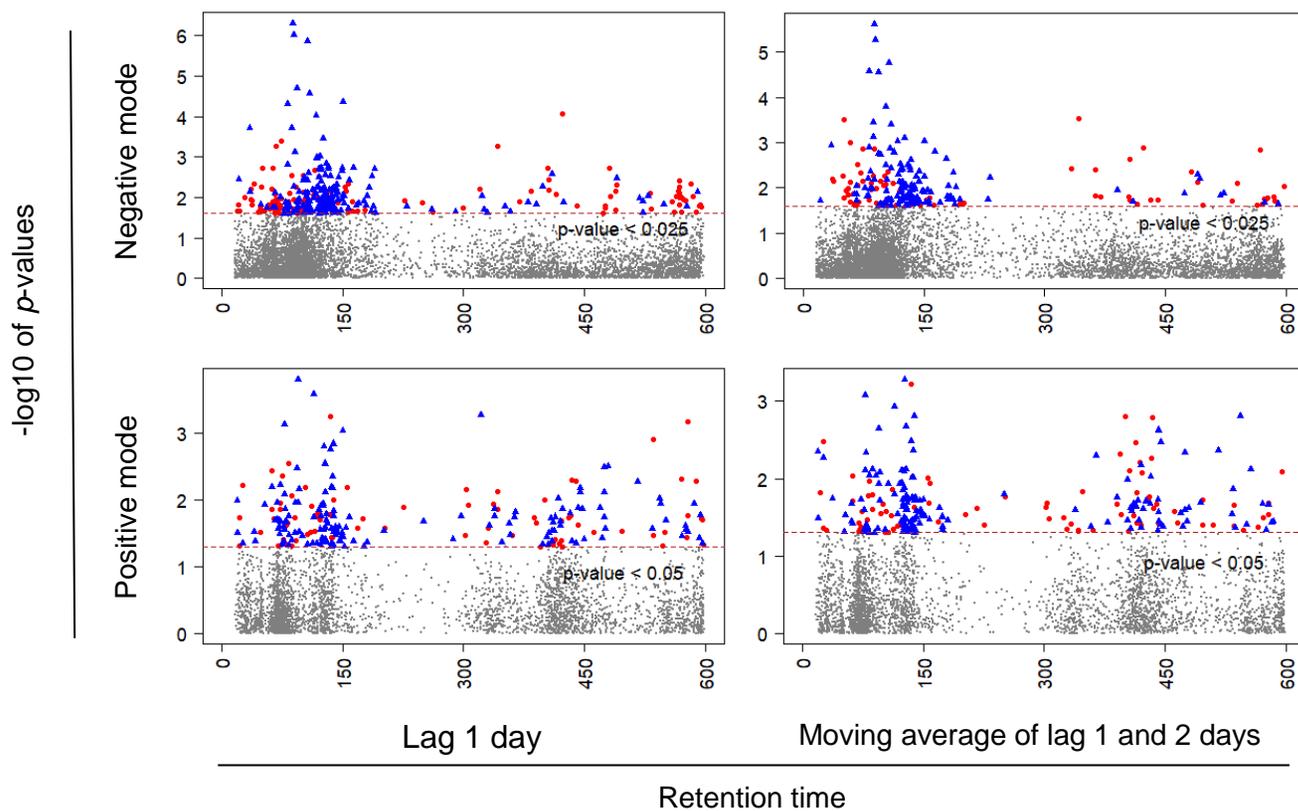


Figure S6. Manhattan plots of associations between log-transformed metabolic feature intensity and OC from multiple linear regression models. X-axis denotes the retention time (in seconds); Y-axis denotes the negative log₁₀ of the *p*-values calculated from the multiple linear regression model. Blue triangles and red circles denote negative and positive associations, respectively.

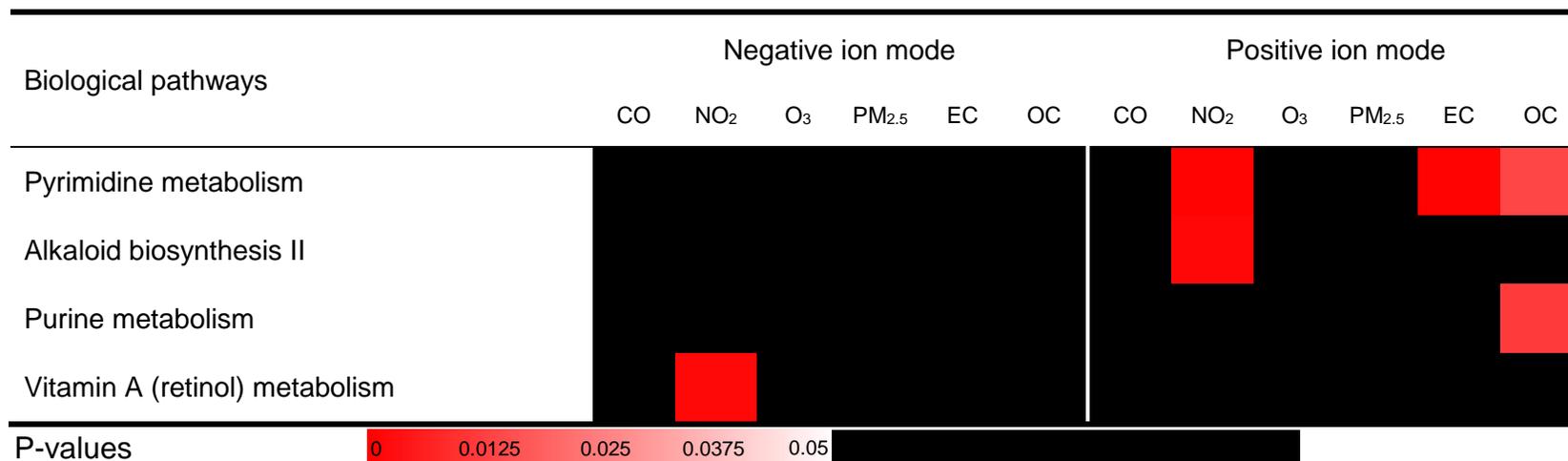


Figure S7. Metabolic pathways associated with 1-day lag pollution from multiple linear regression models. Cells were shaded according to the magnitude of p-values derived from multiple linear regression models. Pathways enriched by more than 2 annotated significant metabolic features were included and ordered according to the total number of the significant associations between pathways and pollutants by either negative or positive ion modes.

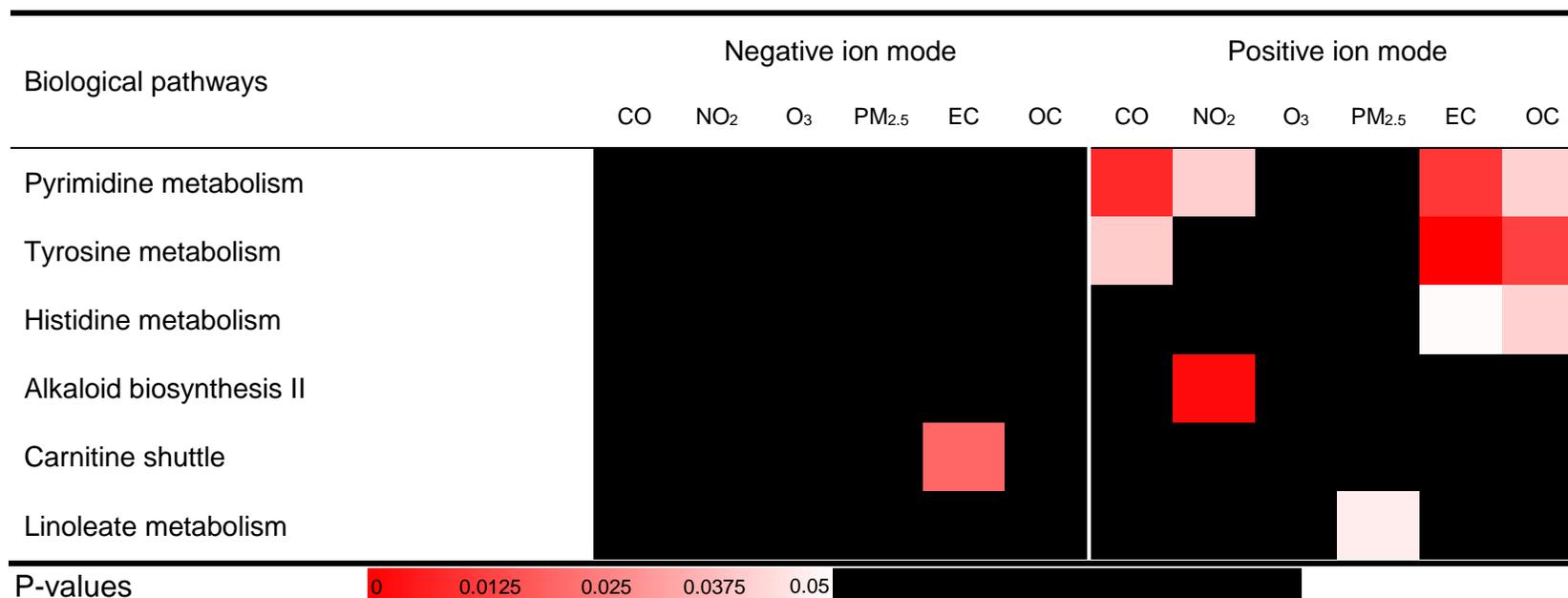


Figure S8. Metabolic pathways associated with the moving average of lag 1-2 days pollution from multiple linear regression models. Cells were shaded according to the magnitude of p-values derived from multiple linear regression models. Pathways enriched by more than 2 annotated significant metabolic features were included and ordered according to the total number of the significant associations between pathways and pollutants by either negative or positive ion modes.

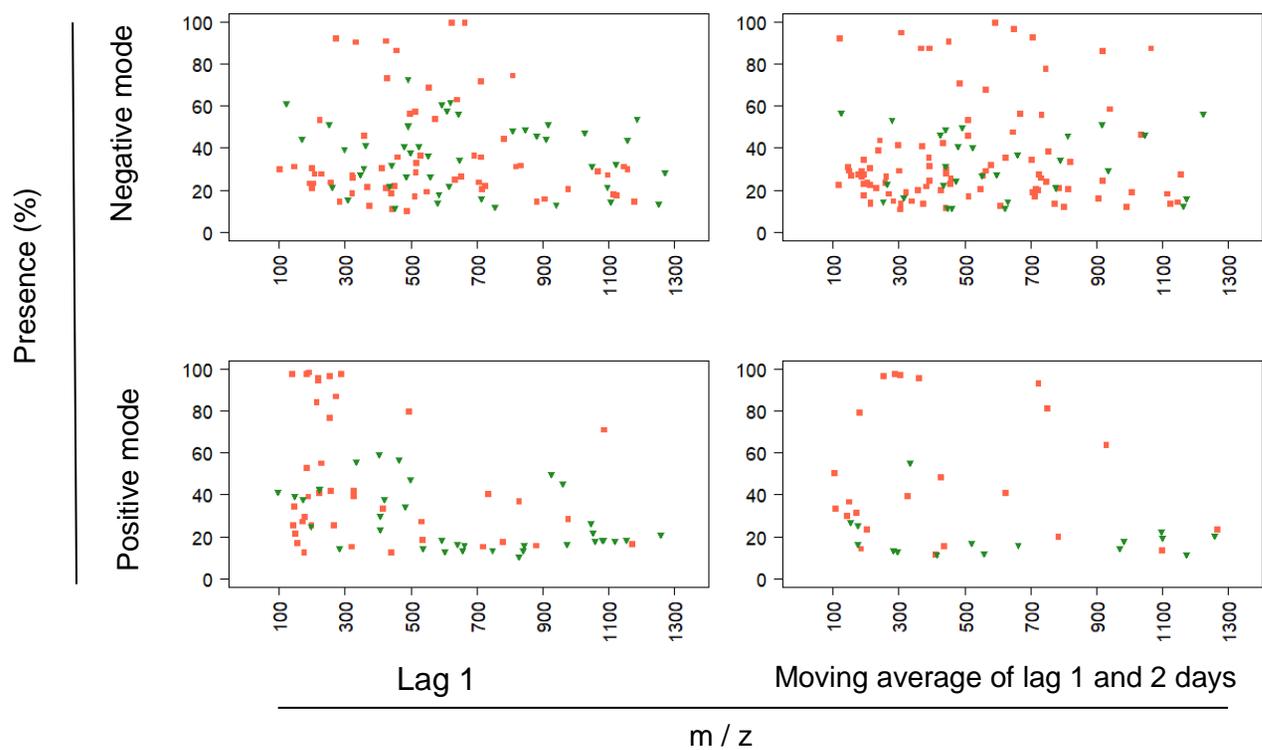


Figure S9. The % presence among all participants by mass to charge (m/z) ratio of unique significant metabolic features associated with CO from Tobit (red squares) and multiple linear regression (green triangles) models. X-axis denotes the m/z ratio, Y-axis denotes the percent presence of the feature among the 180 participants.

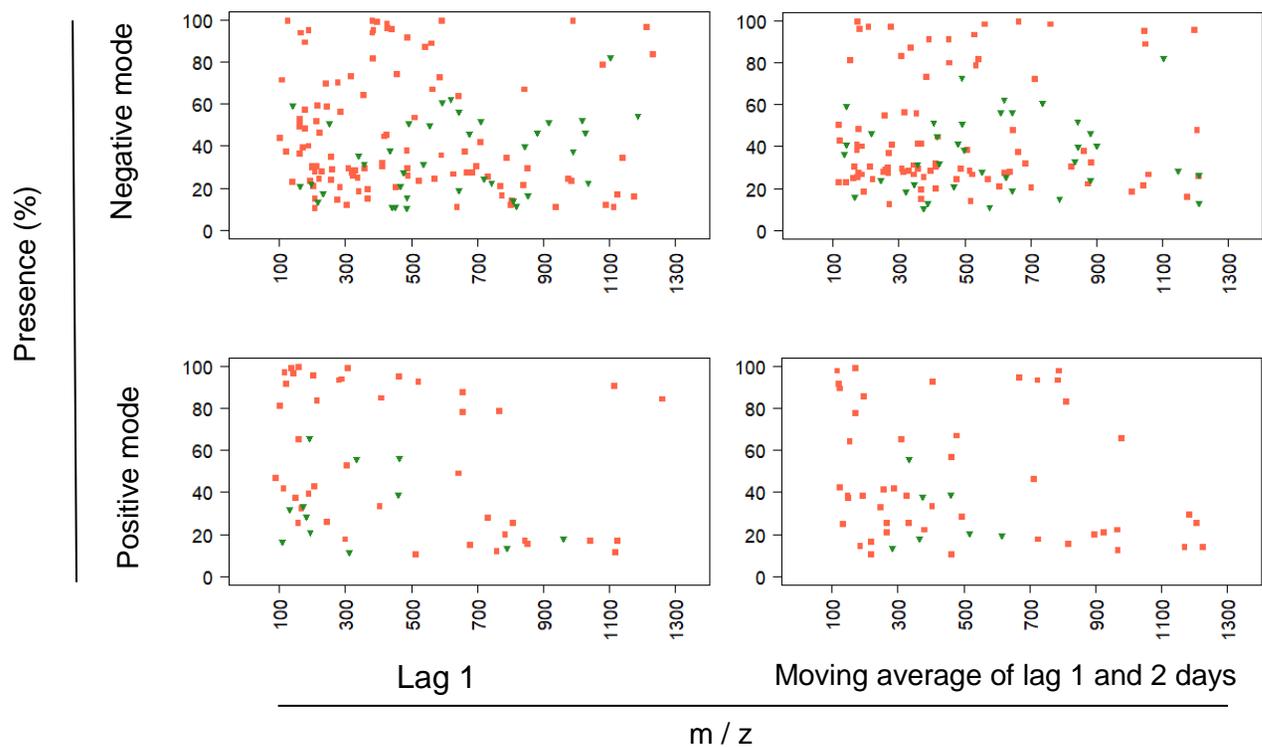


Figure S10. The % presence among all participants by mass to charge (m/z) ratio of unique significant metabolic features associated with NO_2 from Tobit (red squares) and multiple linear regression (green triangles) models. X-axis denotes the m/z ratio, Y-axis denotes the percent presence of the feature among the 180 participants.

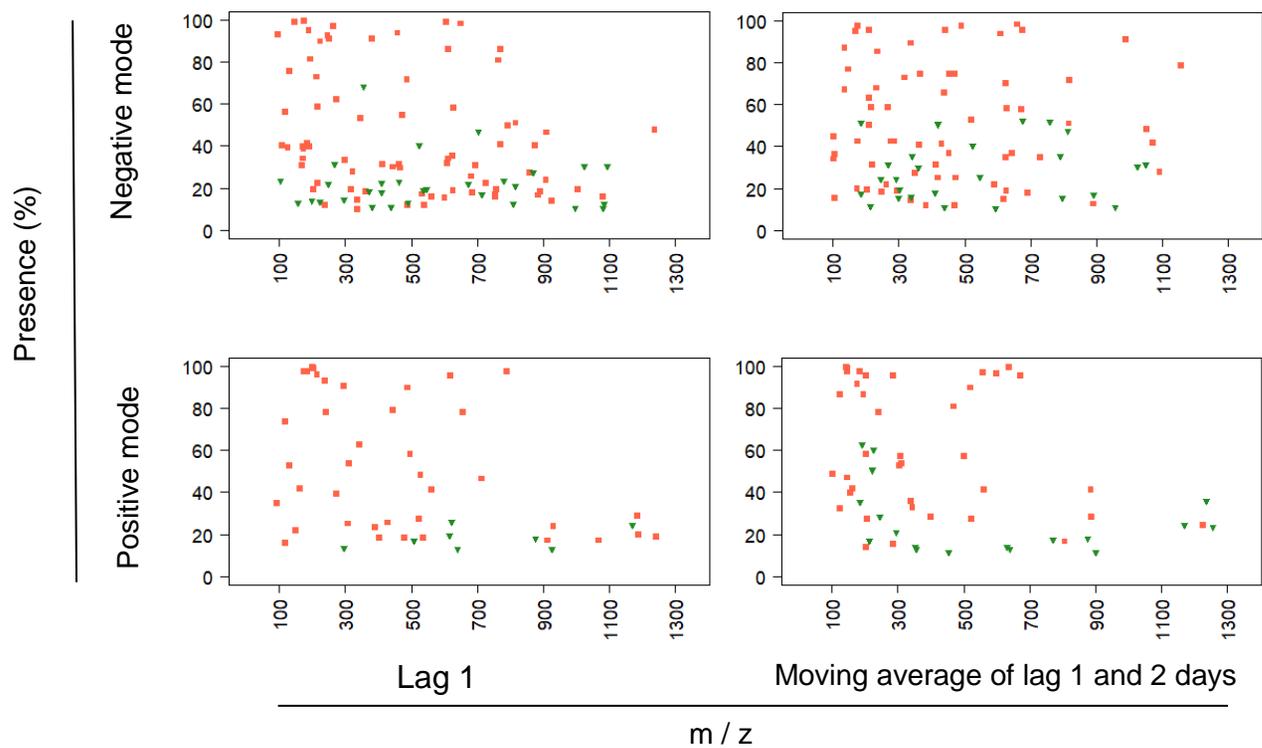


Figure S11. The % presence among all participants by mass to charge (m/z) ratio of unique significant metabolic features associated with O_3 from Tobit (red squares) and multiple linear regression (green triangles) models. X-axis denotes the m/z ratio, Y-axis denotes the percent presence of the feature among the 180 participants.

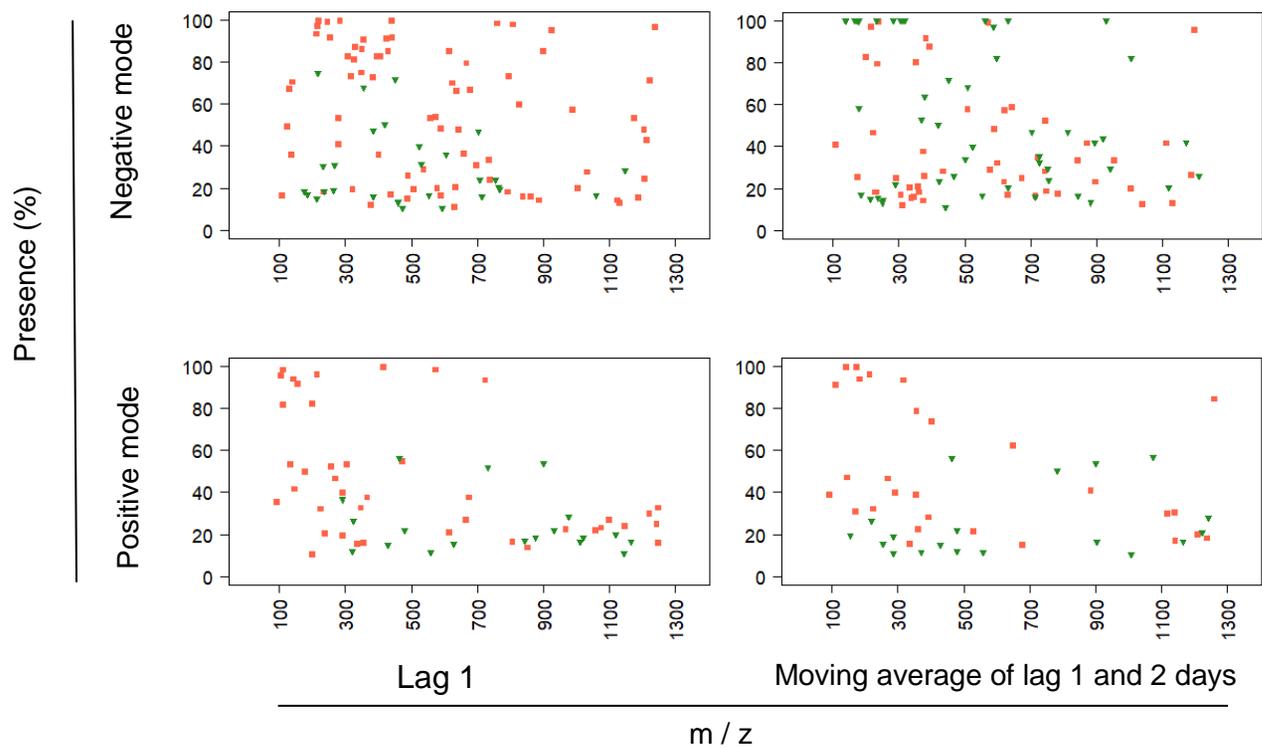


Figure S12. The % presence among all participants by mass to charge (m/z) ratio of unique significant metabolic features associated with PM_{2.5} from Tobit (red squares) and multiple linear regression (green triangles) models. X-axis denotes the m/z ratio, Y-axis denotes the percent presence of the feature among the 180 participants.

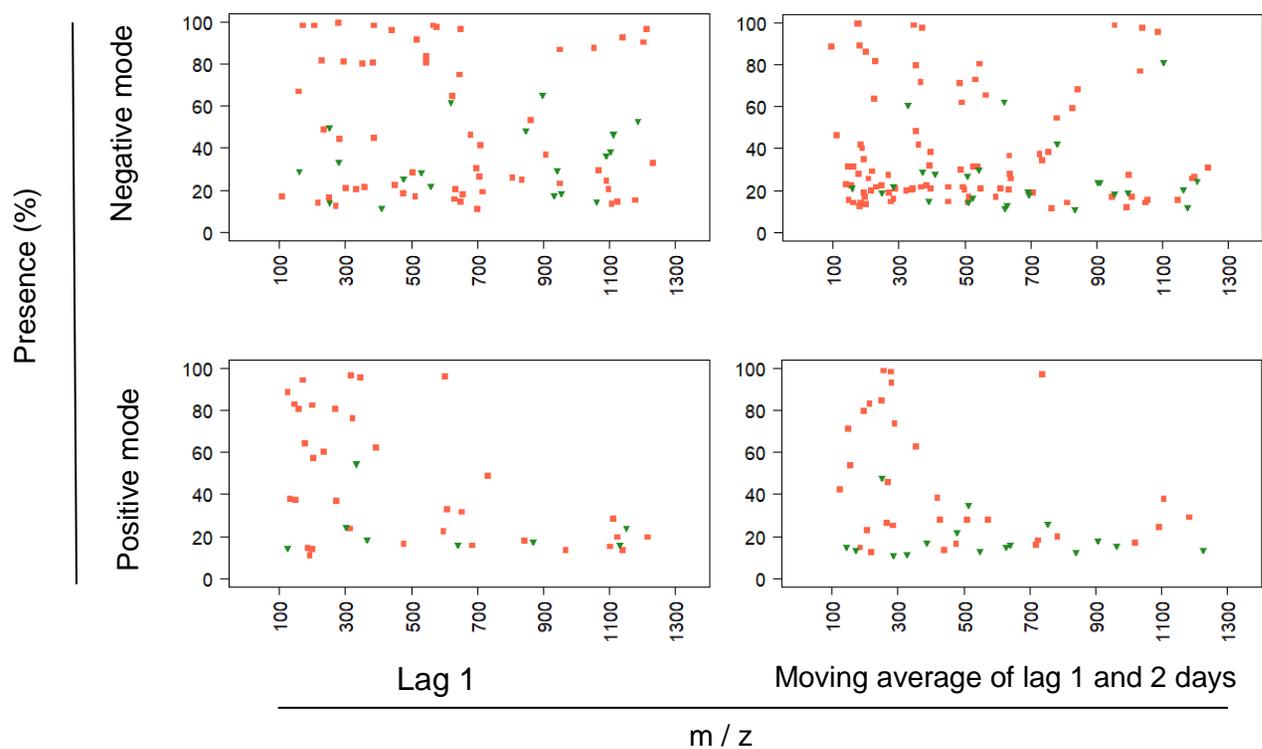


Figure S13. The % presence among all participants by mass to charge (m/z) ratio of unique significant metabolic features associated with EC from Tobit (red squares) and multiple linear regression (green triangles) models. X-axis denotes the m/z ratio, Y-axis denotes the percent presence of the feature among the 180 participants.

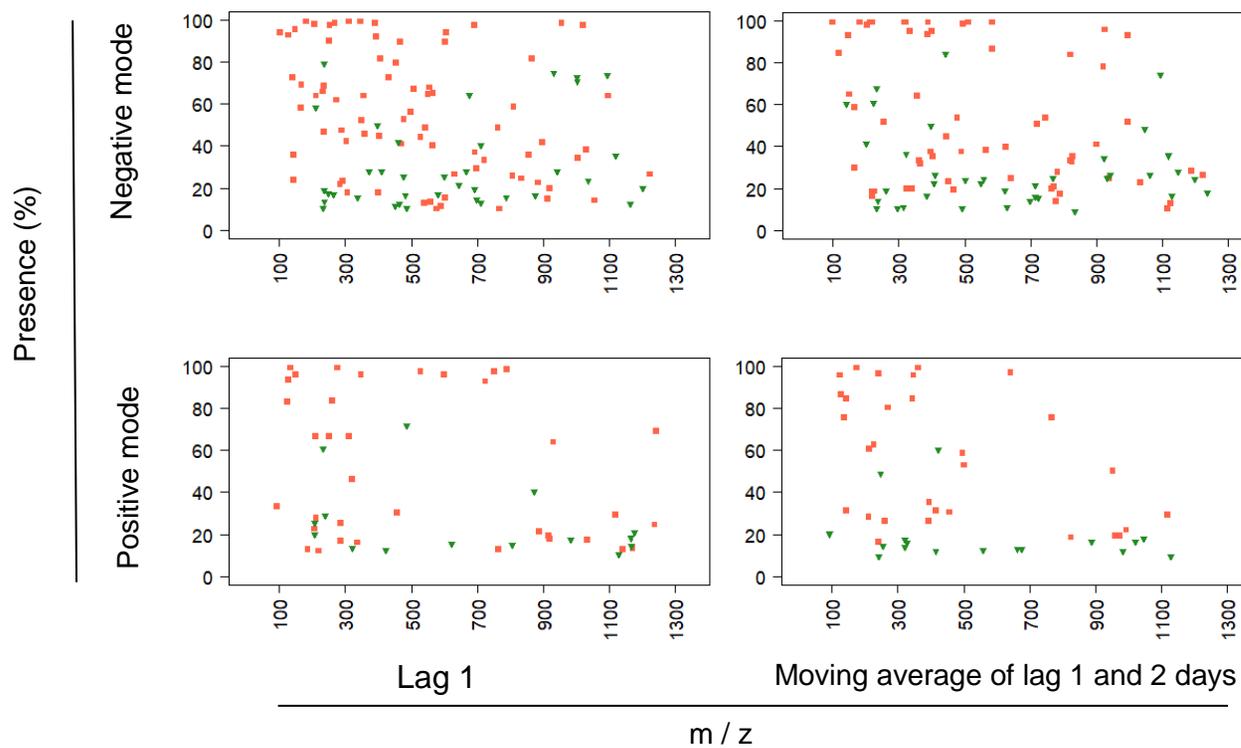


Figure S14. The % presence among all participants by mass to charge (m/z) ratio of unique significant metabolic features associated with OC from Tobit (red squares) and multiple linear regression (green triangles) models. X-axis denotes the m/z ratio, Y-axis denotes the percent presence of the feature among the 180 participants.