Fall 12-7-2020

Metabolic Pathway Activation: An Analysis of the Effects of Traffic Related Air Pollution and Physical Activity

Micah Streiff

ABSTRACT

Metabolic Pathway Activation: An Analysis of the Effects of Traffic Related Air Pollution and Physical Activity

By

Micah Jamal Streiff - December 7, 2020

INTRODUCTION: Air pollution is a major public health concern. Reducing air pollution exposure can reduce the burden of disease and improve cardiovascular and respiratory health. Traffic related air pollution (TRAP) is a significant contributor to air pollution globally. TRAP and physical activity have a dynamic relationship with interacting and opposing effects that are not fully understood.

AIM: This study investigates the relationship between physical activity and exposure to TRAP by identifying the metabolic pathways associated with important TRAP components including Black Carbon (BC), Ozone (O_3), Fine Particulate Matter <2.5 μm ($PM_{2.5}$), and Particle Number Concentration (PNC).

METHODS: Saliva samples were collected from 57 study participants playing sports outdoors near major roadways in Atlanta, GA in 2016. Outdoor air pollution measurements were taken along with participants' physical activity level and breathing metrics. Liquid chromatography coupled with high resolution mass spectrometry was used to process the saliva samples. Mummichog, a pathway analysis tool, was used to identify the metabolic pathways activated.

RESULTS: Ninety-seven pathways were found to be activated as a result of exposure to TRAP and covariates. Acute air pollution exposure was found to be significant in activating metabolic pathways a total of 81 times, with 14 metabolic pathways that showed significant correlation with acute TRAP: Vitamin B6, Beta-Alanine metabolism, Pyruvate metabolism, Selenoamino acid metabolism, Drug metabolism pathway, Saturated fatty acids beta-oxidation, Fatty acid activation metabolism, Glyoxylate and dicarboxylate, TCA cycle metabolism, Vitamin B3 metabolism, Propanoate, Bile acid biosynthesis, Caffeine metabolism, Glyoxylate and dicarboxylate metabolism. One-day lagged $PM_{2.5}$ exposure has a higher number of pathways activated when compared to acute pollutant exposure.

DISCUSSION: The lower degree of acute exposure pathway activation may be due to insufficient time for changes in metabolic activity to be expressed in saliva compared to long-term exposure or ongoing biological processes. While acute exposure to air pollution may contribute to activating less metabolic pathways than long term exposure, heat index and other covariates all contribute to activating metabolic pathways in addition to the effects of static covariates. There is limited research in untargeted, high-resolution metabolomics data of the effects of air pollution in activating metabolic pathways, but this study offers a glimpse into this relationship for future studies to build upon.

METABOLIC PATHWAY ACTIVATION FROM TRAFFIC RELATED AIR POLLUTION AND PHYSICAL ACTIVITY

by

MICAH JAMAL STREIFF

B.S., GEORGIA INSTITUTE OF TECHNOLOGY
B.S., MOREHOUSE COLLEGE

A Thesis Submitted to the Graduate Faculty
of Georgia State University in Partial Fulfillment
of the
Requirements for the Degree

MASTER OF PUBLIC HEALTH

ATLANTA, GEORGIA
30303

APPROVAL PAGE

METABOLIC PATHWAY ACTIVATION FROM TRAFFIC RELATED AIR POLLUTION AND PHYSICAL ACTIVITY

by

MICAH JAMAL STREIFF

Approved:

Roby Greenwald, PhD
Committee Chair

Matt Hayat, PhD
Committee Member

Donghai Liang, PhD
Committee Member

Date
December 7, 2020

Acknowledgments

I would like to express my sincere gratitude and appreciation for my committee chair, Dr. Roby Greenwald for his guidance, encouragement and support during the course of my thesis experience. From supporting my initial interest during his Research Methods course to finding time to meet with me weekly during a pandemic, his time and effort have helped push me along the path to finding the intersection of my interests and abilities. I would also like to thank my committee members, Dr. Matthew Hayat and Dr. Donghai Liang, for contributing so generously of their time and contributions and for being so responsive to make finishing this process possible.

I would also like to thank all my professors in the School of Public Health who have inspired me with their knowledge and selfless contributions to this field and the education of myself and my classmates. Dr. Maya Sternberg, Dr. Karen Nielson, Dr. Roby Greenwald, Dr. Tesfayi Gebreselassie, Dr. Heather Bradley, Dr. Terry Pechacek, Dr. Mogus Ido, Dr. John Steward, Dr. Rachel Powel, Dr. Matthew Hayat, and Dr. Ruiyan Luo, thank you for your encouragement and guidance. All of the faculty and staff have been tremendously helpful, including but not limited to Ms. Jessica Pratt and Ms. Gina Sample who have supported me with navigating the past few semesters. Dr. Claire Spears and the whole mindful living lab Mindful Living Lab team have helped me to learn so much about research in academia and have provided a true model of collaboration and support. Dr. Huang, Dr. Hayat, Mrs. Pam Redmon and Mr. Vuong Do have also contributed to creating a strong learning environment and giving me an opportunity to gain experience and exposure to meaningful research.

Author's Statement Page

In presenting this thesis as a partial fulfillment of the requirements for an advanced degree from Georgia State University, I agree that the Library of the University shall make it available for inspection and circulation in accordance with its regulations governing materials of this type. I agree that permission to quote from, to copy from, or to publish this thesis may be granted by the author or, in his/her absence, by the professor under whose direction it was written, or in his/her absence, by the Associate Dean, School of Public Health. Such quoting, copying, or publishing must be solely for scholarly purposes and will not involve potential financial gain. It is understood that any copying from or publication of this dissertation which involves potential financial gain will not be allowed without written permission of the author.

Micah Jamal Streiff
Signature of Author

TABLE OF CONTENTS

1. INTRODUCTION

Air pollution is a major public health concern. Environmental exposures, including air pollution, play an important role in the development of non-communicable diseases. Reducing air pollution exposure can reduce the burden of disease associated with air pollution and improve cardiovascular and respiratory health. The global disease burden has grown over the past 25 years due largely to population aging, disease rates, and increasing air pollution (Cohen et al., 2017).

Outdoor air pollution exposure contributes to the large disease burden through cardiovascular and respiratory conditions according to compelling epidemiological and clinical research (Franklin, Brook, & Arden Pope, 2015). Symptoms associated with exposure to air pollution include coughing, tearing, difficulty breathing and angina to name a few (Schraufnagel et al., 2019). The World Health Organization (WHO) created air quality guidelines and established safe limits to exposure, of which less than 10% of the world population lived within as of 2016.

Traffic related air pollution (TRAP) is a significant contributor to air pollution globally. In the United State nearly 20% of the population lives near highly congested traffic areas. Some cities, such as Barcelona (Spain) have 96% of the population exposed to higher levels of TRAP, with New Delhi (India) and Paris (France) having over 60%. While the total air pollutant levels are still much greater in India than Spain, it is clear that a significant percentage of the population are still exposed to high levels of TRAP. Motor vehicles generate emissions which are the source of TRAP. Tailpipe emissions include gas-phase pollutants such as carbon monoxide, volatile organic compounds (VOCs) and oxides of nitrogen (which themselves are the

most important precursors of ozone, which is not directly emitted) as well as particulate pollutants such as black carbon particles from incomplete fuel combustion. Important traffic-related non-tailpipe emissions include metal-rich particulates that are generated by brake pad and tire wear (Rowangould, 2013).

TRAP and physical activity have a dynamic relationship with interacting and opposing effects that are not fully understood. There are many health benefits to physical activity, and in most cases these benefits outweigh the harmful effects of exposure to environmental pollutants (Fisher et al., 2016). Only in cases of extreme air pollution concentrations do benefits of physical activity not outweigh the harm caused by the increased air pollution inhaled dose (Tainio et al., 2016). This study seeks to investigate the relationship between physical activity and exposure to TRAP (BC, O_3, $PM_{2.5}$, PNC) by identifying the metabolic pathways activated.

2. REVIEW OF LITERATURE

2.1. BASICS OF AIR POLLUTION AND HEALTH AT THE POPULATION LEVEL

Air pollution, any substance in the air that may harm humans, animals, vegetation or materials (Kampa & Castanas, 2008), contributes to negative health effects in people and numerous adverse effects on the environment. Environmental exposures play an important role in the development of cancer, cardiovascular, respiratory, and other diseases (Cui et al., 2016) Reducing air pollution can reduce the burden of disease associated with air pollution and improve cardiovascular and respiratory health. The global disease burden has grown over the past 25 years due largely to population aging, disease rates, and increasing air pollution (Cohen et al., 2017).

Symptoms associated with exposure to air pollution include coughing, tearing, difficulty breathing and angina to name a few. Long-term effects may be more subtle, and people may remain unaware of how or to what extent other health and medical conditions are worsened. There are many biological mechanisms through which TRAP effects the body. Oxidative stress often leads to inflammation of the lungs as a result of TRAP exposure. A cause of oxidative stress is that more reactive oxygen species (ROS) are formed than can be removed by the body's defenses, which leads to cells and organs being harmed. This leads to adverse health effects "including decreased lung function, increased airway hyperactivity, pulmonary inflammation, damaged lungs and cell permeability." Oxidative stress has been quantified by measurement of ROS-modified molecules, which are indirect biomarkers of oxidative stress. Studying these biomarkers led to advances in understanding how "air pollution exposures are associated with oxidative stress and inflammation in humans" (Barthelemy, Sanchez, Miller, & Khreis, 2020).

The chemical composition and surface area of pollutants contribute to the amount of oxidative stress produced. The metabolism (which will be discussed later) can be affected in many ways by exposure to pollutants and different organs may also experience inflammation by particles that are directly exposed (Schraufnagel et al., 2019). One metabolite may be associated with multiple metabolic pathways, and multiple metabolites are associated with each metabolic pathway. A challenge in metabolomics lies in the chemical annotation of metabolic features.

Pollutants come from various sources and have a range of impacts on health and the environment. Primary gas-phase pollutants include sulfur dioxide (SO_2), nitrogen dioxide (NO_2)

and carbon monoxide (CO). Secondary pollutants such as O_3 and sulfate aerosol are formed in the atmosphere from the primary pollutants. Human health effects depend on the gases' "water solubility, concentration, ability to oxidize and a person's susceptibility to the pollutant" (Schraufnagel et al., 2019). Despite a large and growing body of knowledge around air pollution and health, the underlying mechanisms are still largely unknown and metabolomics studies involving human subjects are limited (Mu et al., 2019).

Pollutants that contribute significantly to air pollution include $PM_{2.5}$ and O_3. Outdoor fine particulate matter accounts for 4.2 million deaths and is the fifth leading risk factor for death according to the Global Burden of Disease Report. Exposure to O_3 caused an additional 250,000 or more deaths annually. Indoor air pollution accounted for 3.8 million deaths with the greatest affliction falling on vulnerable populations (WHO, 2018). According to Jeoung et al., a meta-analysis showed that "increases in $PM_{2.5}$ and PM_{10} were associated with risks of fatal and total coronary events, respectively (Cesaroni et al., 2014)" (Jeong et al., 2018).

Limited reductions in disease burden will occur in highly polluted countries unless $PM_{2.5}$ levels are significantly decreased. There is potential for substantial health benefits from exposure reduction (Cohen et al., 2017). Policy can contribute to the reduction of key sources of air pollution. Successful examples include policies that contribute to reduction through transport, urban planning, power generation, waste management, food and other industries (WHO, 2018).

Respiratory symptoms are detrimentally impacted by exposure to TRAP and other outdoor pollutants. There is strong epidemiological evidence indicating the relationship between air pollution and asthma exacerbation, respiratory morbidity and mortality in patients

with chronic obstructive pulmonary disease (COPD). Recent studies have also suggested that pollutants contribute to the development of both asthma and COPD. Given the considerable contribution that traffic emissions make to urban air pollution researchers have sought to characterize the relative toxicity of traffic-related PM2.5 pollutants (Kelly & Fussell, 2011).

Although epidemiologic studies have shown associations of cardiovascular morbidity and mortality with ambient CO and NO_2 (Brook et al. 2010; Bhaskaran et al. 2009), concentrations of these gases have been considerably lower relative to levels causing effects in experimental models, and it is possible that these gases are serving as surrogates for other causal components from fossil fuel combustion. There is also some epidemiologic evidence that PM2.5 exposure significantly increases biomarkers of oxidative stress in blood, but the data are limited for susceptible populations exposed to urban air pollution. The epidemiologic evidence includes panel studies of healthy subjects with few repeated measures (Chuang, Chan, Su, Lee, & Tang, 2007; Liu et al., 2007; Vinzents et al., 2005). Oxidative stress responses to ROS and subsequent inflammation resulting from air pollution exposure may be one of these important mechanisms (Brook et al., 2010; Frampton, 2006; D. Liang et al., 2019; Sinharay et al., 2018; Xia, Kovochich, & Nel, 2006).

2.2. PHYSICAL ACTIVITY AND ITS RELATIONSHIP TO AIR AND HEALTH

There are many health benefits to physical activity, and in most cases these benefits outweigh the harmful effects of exposure to environmental pollutants. Leisure time physical activity has a protective effect for cardiovascular disease mortality and all-cause mortality (Kaplan, Strawbridge, Cohen, & Hungerford, 1996). According to a study in Denmark, the benefits of physical activity outweighed the harmful effects associated with increased exposure

to air pollution and the risk of asthma and COPD (Fisher et al., 2016). In cases of extreme air pollution concentrations, the benefits of physical activity were outweighed by the harm caused (Tainio et al., 2016). The complex interaction between the benefits of physical activity and harmful effects of exposure to air pollution are incompletely understood. This study seeks to examine this relationship by identifying the metabolic pathways associated with exposure to different pollutants.

Physical activity increases ventilation rate and augments pollutant inhalation, and it is therefore important to quantify ventilation rates and dose in order to better assess the effects of pollutants (Greenwald, Hayat, Barton, & Lopukhin, 2016). Pollution exposure has been shown to decrease exercise performance in athletes. One example is that exercise duration and oxygen consumption are reduced with exposure to CO. Moderate-to-vigorous physical activity (MVPA) is associated with reduction in mortality risk with increased benefits for older adults. Often the recommended amount of physical activity is not reached in one's daily life, especially for adults. Despite this, a recent study has shown that even a low dose of MVPA resulted in mortality risk reduction (Hupin et al., 2015).

2.3. METABOLOMICS BACKGROUND AND AIR POLLUTION STUDIES

High resolution metabolomics (HRM) is an emerging research tool for the analysis of metabolic pathways. The metabolome consists of all metabolites present within an organism and is influenced by genetics, epigenetics, environmental factors and the microbiome (Kennedy et al., 2018). Metabolomics aims to measure, identify and quantify a large number of metabolites in a biological system of interest and can be applied to many fields, including exercise, toxicology, nutrition, physiology, and clinical diagnostics (Kennedy et al., 2018;

Stanstrup et al., 2019). Metabolic networks trace a path of the relationships between small biological molecules (metabolites) and the enzymes (proteins) that interact with them to catalyze a biochemical reaction (Walhout, Vidal, & Dekker, 2013). Studying these pathways can help scientists understand disease progression and possibly identify and detect molecular changes that may shed light on diagnosis and therapy development. For example, scientists studying COVID-19 (SARS-CoV-2) have used metabolomics to improve the ability to assess which cases will likely become severe in an effort to improve treatment (Shen et al., 2020).

Mass spectrometry (MS) is an analytical technique used in metabolomics research to determine molecular mass and gives other important information to help identify metabolites. Mass spectrometry is over 100 years old and has continued to advance in sensitivity, throughput and range of applications since then, especially in recent years.

When combined with ion separation techniques, mass spectrometry can be used to analyze complex biological matrices and identify metabolites (Király, Dalmadiné Kiss, Vékey, Antal, & Ludányi, 2016). Chromatography is a technique used to separate a mixture by dissolving it in a fluid and passing it through a packed column at high pressure. As a result of physical and chemical interactions between the analytes present in the sample and the packing material in the column, different analytes will elute from the far end of the column at different times and will thus be separated before proceeding to the mass spectrometer. Liquid chromatography (LC) is a commonly used separation technique. Liquid chromatography-mass spectrometry (LC-MS) is the combination of the separation technique of liquid chromatography with the mass analysis of mass spectrometry.

Analytes in the sample are referred to as “features” and are labeled by their mass to charge ratio (m/z) and retention time (rTime) in the column; however, in most cases, their chemical identity is not definitively known. The chromatographic intensity of the features is determined based on peak area.

Saliva samples were the medium of choice for this study. Saliva has the advantages of being quick and easy-to-collect. Saliva collection is also an inexpensive and non-invasive approach to metabolite analysis. Saliva consists of primarily water, with about 1% consisting of “electrolytes, mucus, cellular debris, proteins, and small molecules” coming from the blood by passive diffusion (Bessonneau, Pawliszyn, & Rappaport, 2017).

The LC process is applied using two separation techniques: positive (hydrophilic-interaction liquid chromatography (HILIC)) and negative (reverse-phased liquid C18 chromatography(C18)). Identical aliquots of the same samples are analyzed in this process, and although these two methods utilize different technical columns and mobile phases, they both are analyzed using high-resolution mass spectrometry. Thus, the same metabolomic features are assessed by each column, although with different retention times for each feature.

2.4. AIR POLLUTION STUDIES

Investigating health response to air pollution exposure using HRM has been an increasingly active research area for the last decade. A 2017 study assessed the perturbation of the plasma metabolome in response to short term exposure to air pollution. 89 associations were identified between exposures to air pollutants and metabolic features. Eight pathways were predicted to have enriched pathway activity (Vlaanderen et al., 2017).

Another study from 2018 compared TRAP exposure and metabolic pathway perturbation on a college campus in Atlanta, GA between a dorm near a high traffic area and a dorm further away. The analysis resulted in a total of 20,766 metabolic features extracted from plasma samples and 29,013 features from saliva samples. The chemical identity of 10 metabolomic features associated with traffic pollutants were definitively identified including arginine, histidine, γ-linolenic acid, and hypoxanthine (Donghai Liang et al., 2018).

2.5. LIMITATIONS

Certainty of chemical identification and biological relevance in feature annotation vary and can be challenging to communicate concisely and accurately to readers (Goodacre et al., 2007; Schymanski et al., 2014). A pathway analysis tool (Mummichog, http://mummichog.org/) that leverages large data libraries with identified metabolites of known m/z ratios was used in this analysis. The mummichog analysis may not take the level of metabolite identification certainty and biological relevance into account in its own analysis. We are also not able to know if metabolic pathways are down or upregulated. Many search terms for metabolic pathways and exposures did not result in any findings in PubMed, which limits the ability to compare results with existing literature. This limitation is also a strength, which contributes to direction for future studies.

3. METHODS AND PROCEDURES

3.1. DATA COLLECTION

A shortened summary of the data collection process is provided in this section because his paper builds on a previous analysis. The complete study overview is provided in Elizabeth Finlon's thesis in the GSU Scholarworks website (Finlon, 2018). Study data was collected in

Atlanta, GA in 2016 by the Study of Air Pollution and Physical Activity (SAPPA). SAPPA was approved by the Georgia State University Institutional Review Board and the Emory University Institutional Review Board. Informed consent was provided by adults through written consent and minors provided written assent and parental consent.

Saliva samples were collected from a sample of fifty-seven persons, including adolescents and adults, over nine non-consecutive sampling days to be used for metabolic analysis. Forced Vital Capacity (FVC), the total volume air that can be exhaled from the lungs with ordinary effort, was measured at the beginning and end of the exposure period.

Participants strapped a biomonitoring device (BioHarness 3, Zephyr Technology Corp.) to their torso to monitor motion, heart and breathing rates. Motion data from the biomonitoring device was used to provide the activity value (ACT) for this study, which is the magnitude of 3-dimensional acceleration expressed as a fraction of standard gravity (G-force). Minute ventilation (the volumetric flow rate of respired air) was calculated according to an extensively validated model developed by Greenwald and Hayat. Minute ventilation multiplied by the ambient concentration of the pollutant gives the inhaled dose.

A 1mL saliva sample was given by each participant immediately before and after the approximately 2-hour practice. Participants reported their date of birth, sex, and race. Height, weight and BMI were measured.

Continuous measurements were taken for the ambient outdoor concentrations of Particulate Matter <2.5µm ($PM_{2.5}$), ozone (O_3), black carbon (BC), and particle number concentration (PNC). The cumulative inhaled dose was calculated as the sum of the one-minute doses during the exposure period. Finlon analyzed the pollutants listed above using four distinct

exposure metrics: ambient concentration, exposure (defined as “concentration X time” to account for exposure periods of different durations), cumulative inhaled dose, and maximum one-minute dose. Outdoor ambient concentrations were measured on-field during practice each day of the study. Participant exposure was calculated by multiplying the ambient concentration of each pollutant by the exposure time period. The sum of one-minute doses during an exposure period gives the cumulative inhaled dose. The average concentration of $PM_{2.5}$ or O_3 at the nearest state of Georgia air quality monitoring site (the United Avenue site) over the 24-hour period preceding the sports practice was used as a proxy for the 1-day lagged exposure.

3.2. ANALYSIS

Ultra-high resolution LC-MS was used to process saliva samples, resulting in a database of metabolic features with mass-to charge ratios, retention times, and intensities. The features are the intensity or relative abundance of the analyte as described in section 2.6. This collection of features are the changes to an individual’s metabolic profile from pre- to post-exposure. Two datasets were created, one for each column used during liquid chromatography, C18 and HILIC. A sample of the same saliva goes into each column, so the same features may likely appear in both datasets. After running features through quality control to filter for biologically relevant features, the C18 and HILIC datasets contained 7,608 and 8,293 features, respectively. Data were analyzed using RStudio version 3.3.3 packages lme4 and lmerTest.

A variable selection process was used to ensure exposure effects were correctly attributed. If important predictors and covariates are not included in a model then the estimated effect may appear larger than the true effect. Alternatively, if unimportant covariates

are included, then there is risk of overfitting the models. All exposures and pollutants were included in a data driven approach to identify the covariates with high predictive value of metabolic features. Many features are not positively identified in the metabolic analysis. This data driven approach can assess features whose change in intensity is related to model variables.

The backward stepwise model selection process began with the full model including all covariates and eliminated sequentially the terms demonstrating p-values greater than .05. This was achieved by producing frequency distribution histograms of the p-values for the association of the covariate in question with each of the ~15,000 features. At each step in the model selection process, the covariate with the fewest p-values <.05 was removed from the selection process, and the process was repeated. The full model contained race, sex, age, BMI, ACT (activity based on accelerometry), lagged exposure, heat index, and pollutant exposure metrics. For models examining ozone exposure metrics, lagged exposure was based on ambient ozone concentrations, but all other pollutants used lagged exposure based on ambient $PM_{2.5}$ from the United Avenue monitor of the Georgia Environmental Protection Division. Based on this analysis, the order of variable elimination was race followed by sex, BMI, age and then ACT while same-day pollutant exposure, lagged exposure, and heat index were retained in all models. The model selection process did not determine the most important metric or the most important pollutants or exposures in activating metabolic pathways. The model selection process determined which model variables to keep by identifying the variables that were most predictive of features. Histograms are shown below to illustrate this point. Heat index is an example of a variable that was decided to keep in all models after reviewing the histograms for

the distribution of adjusted p-values for heat index in the in each model. In Figure 3.1, the histogram for the distribution of adjusted p-values for heat index in the seven variable BC concentration model shows that heat index had a value of .05 or less in approximately 100 models, indicating it may have been a valuable contributor in predicting around 100 features.

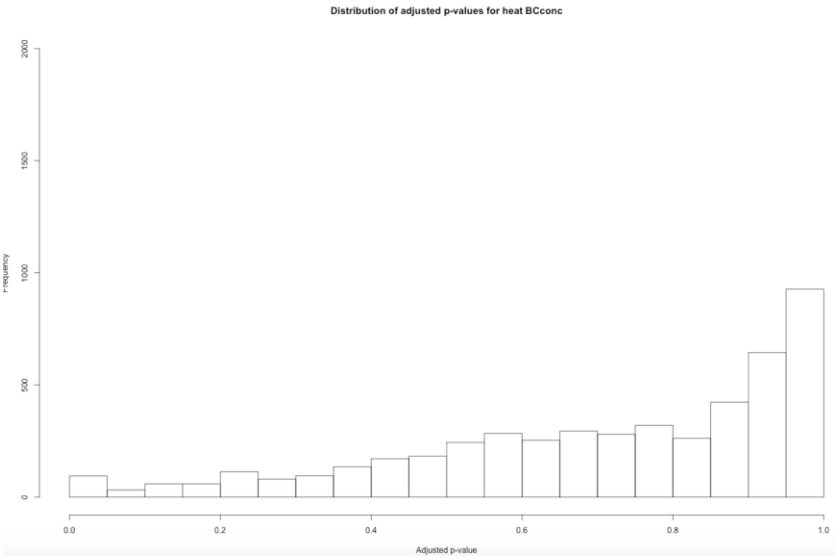

Figure 3.1: Histogram for the distribution of adjusted p-values for heat index in the seven variable BC concentration C18 model.

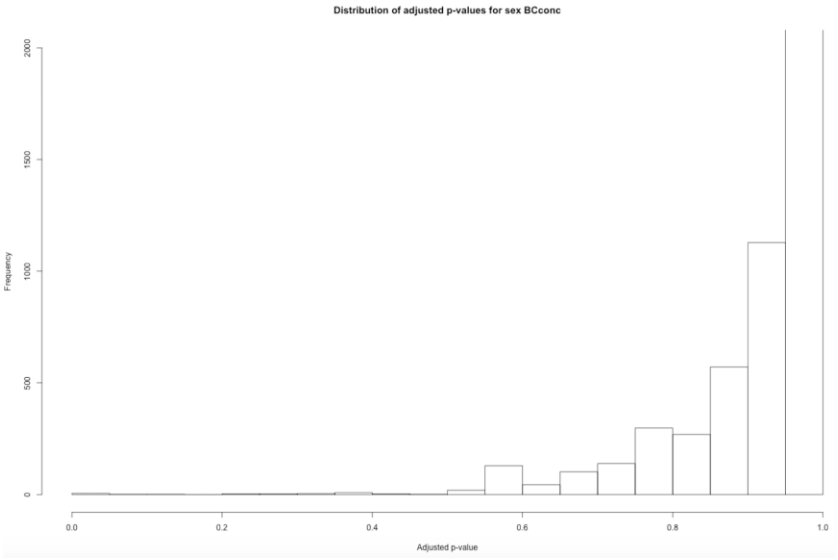

Figure 3.2: Histogram for the distribution of adjusted p-values for sex in the seven variable BC concentration C18 model.

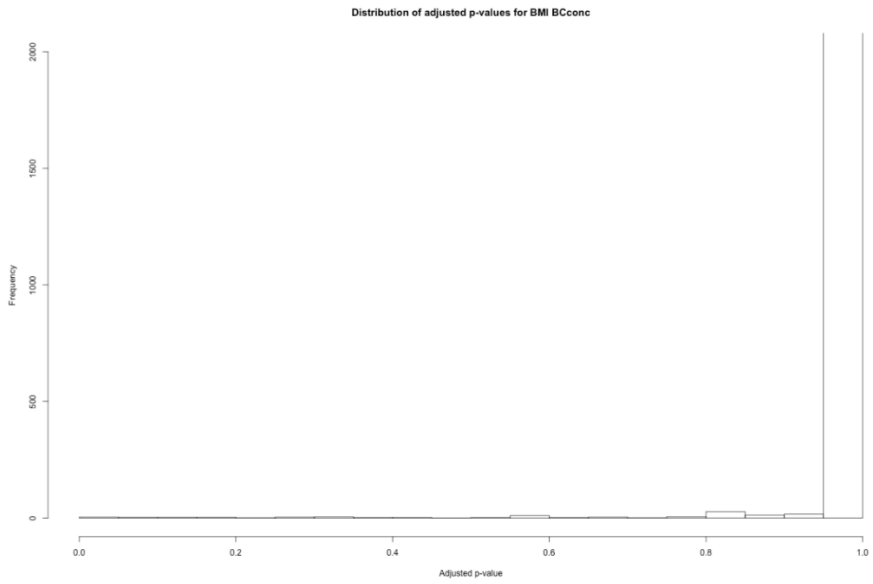

Figure 3.3: Histogram for the distribution of adjusted p-values for BMI in the seven variable BC concentration C18 model.

Comparing the histograms for the distribution of adjusted p-values for BMI and sex in the same seven variable BC concentration models reveals a much lower, almost negligible number of p-values < .05.

Repeated measures were taken in this experiment by collecting saliva before and after physical exercise. The preferred analytic choice for analyzing repeated measurements on individuals in a study is mixed or marginal models (Hayat, 2012). Mixed and marginal models account for the correlation within and between subjects. To account for the repeated measures for some subjects in the dataset, we included a random effect for subject in our models. The resulting linear mixed models used the following metrics: ambient outdoor concentration, exposure, cumulative inhaled dose, and maximum one-minute dose to analyze pollutants. BC, PNC, O_3, and $PM_{2.5}$ were included as predictors while controlling for covariates as described in

models below. This allowed us to identify associations between changes in pollution metrics and changes in feature intensity from before and after the acute exposure (practice).

A total of seven models per exposure metric were selected to determine which features were associated with each pollutant exposure metric while controlling for covariates. A total of 15,901 features (7,608 C18 & 8,293 HILIC), four pollutants, five metrics per pollutant and seven models per exposure metric led to analysis of over 2.23 million models (15,901 x 4 x 5 x 7).

There was a total of three distinct general linear mixed models with three variables each, which accounted for the effect of lag exposure, heat index or the pollutant while controlling for the other 2 covariates. The general linear mixed model was formulated as follows:

$$\Delta Feature_{ij} = \beta_0 + \beta_1 \Delta PollutantMetric_{ij} + \beta_2 \Delta HeatIndex_{ij} + \beta_3 \Delta Lag_{ij} + \gamma_{oj,} + \varepsilon_{ij}$$

where i=1,…,n (n=number of subjects)

j=1,…,n (sampling day)

$\Delta PollutantMetric_{ij}$ Is the pollutant metric of the ith subject on the jth sampling day

$\Delta HeatIndex$ Is the heat index value for the ith subject on the jth sampling day

ΔLag_{ij} Is the lag exposure for the ith subject on the jth sampling day

$\varepsilon_{ij} \sim N\left(0, {\delta_{day}}^2\right)$ ←variance for sampling day

$\gamma_{0j} \sim N\left(0, {\delta_{subject}}^2\right)$ ←variance for subject

Including heat index and lag exposure in the models helps in the analysis to account for the high correlation between some of the pollutants and exposures. For example, ozone and heat index are highly correlated with one another and their effects may be additive. If we do not include heat index in a model, it would appear as though the effect of heat index is being attributed to ozone. If we assume the change for heat index is zero since controlling for it in a model, then we are isolating it from this prediction. Models that include an acute exposure but

not lag (closely related to chronic) exposure may similarly not account for the effect of lag exposure that is correlated with the acute exposure of other pollutants. For this reason, heat index and lag exposure are included in all models along with the pollutant of interest.

What will follow are the changes for the fixed effects, but the random effects will remain the same as the 3-variable model. The four 4-variable models selected accounted for the same effects as the three-variable model with the addition of physical activity:

$$\Delta Feature_{ij} = \beta_0 + \beta_1 \Delta PollutantMetric_{ij} + \beta_2 \Delta HeatIndex_{ij} + \beta_3 \Delta Lag_{ij} + \beta_4 \Delta Act_{ij} + \gamma_{oj,} + \varepsilon_{ij}$$

where ΔAct_{ij} Is physical activity for the ith subject on the jth sampling day all other terms are identical to the previous model.

The activity variable and factors used to measure ventilation rate, and by extension, dose are highly correlated. For this reason, physical activity was added to account for this correlation.

Similarly, we included one 5-variable model, which accounted for the effect of age:

$$\Delta Feature_{ij} = \beta_0 + \beta_1 \Delta PollutantMetric_{ij} + \beta_2 \Delta HeatIndex_{ij} + \beta_3 \Delta Lag_{ij} + \beta_4 \Delta Act_{ij} + \beta_5 \Delta Age_{ij} + \gamma_{oj,} + \varepsilon_{ij}$$

where ΔAge_{ij} Is physical activity for the ith subject on the jth sampling day all other terms are identical to the previous model.

One 6-variable model was selected, which accounted for the effect of BMI:

$$\Delta Feature_{ij} = \beta_0 + \beta_1 \Delta PollutantMetric_{ij} + \beta_2 \Delta HeatIndex_{ij} + \beta_3 \Delta Lag_{ij} + \beta_4 \Delta Act_{ij} + \beta_5 \Delta Age_{ij} + \beta_6 \Delta BMI_{ij} + \gamma_{oj,} + \varepsilon_{ij}$$

where ΔBMI_{ij} Is BMI for the i[th] subject on the j[th] sampling day all other terms are identical to the previous model.

And a 7-variable model accounted for the effect of sex:

$$\Delta Feature_{ij} = \beta_0 + \beta_1 \Delta PollutantMetric_{ij} + \beta_2 \Delta HeatIndex_{ij} + \beta_3 \Delta Lag_{ij} + \beta_4 \Delta Act_{ij} + \beta_5 \Delta Age_{ij} + \beta_6 \Delta BMI_{ij} + \beta_7 \Delta Sex\gamma_{oj,} + \varepsilon_{ij}$$

where ΔSex_{ij} is sex for the i[th] subject on the j[th] sampling day all other terms are identical to the previous model.

The metabolome-wide association framework (MWAS) was used to run models and generate p-values for assessing the association between each pollution metric and feature intensity. With thousands of p-values being generated, the likelihood that a feature would be found significant by chance alone was high. The number of false discoveries would be much higher than would be allowed in order to make meaningful predictions. To account for this, the p-values for each model were adjusted for type I statistical errors using the Benjamini-Hochberg false discovery rate (FDR) at FDR <.05. Controlling the FDR served to decrease the number of excessive false positives.

Mummichog is a pathway analysis tool that leverages large data libraries with identified metabolites of known m/z ratios. Each FDR adjusted significant feature was processed using Mummichog version 1.0.3. Mummichog predicts the probability of the contribution of metabolic features to the activation of metabolic pathways in each model. The probability is a p-value assessing the association between the features and the pathways. Identified pathways were considered significant if they resulted in a Mummichog calculated p-value of <.05 and at least 2 overlapping nodes.

Understanding how model results compare to an understanding of biological processes helps to explore the validity of results. We hypothesized that cumulative inhaled dose normalized by body mass is the most biologically relevant exposure metric for exposures with systemic effects. We also recognized the possibility that inhaled dose normalized by lung volume may be biologically relevant for exposures with localized, point-of-contact respiratory effects. To test these hypotheses, we compared the mummichog p-value for each metric to see if the p-value was decreasing in value from concentration to dose to doseKG (dose normalized by body mass) to doseFVC (dose normalized by lung volume). Many of the mummichog models failed to converge to return a result. The only model with results for every pollutant is the doseKG model. For this reason, we used the doseKG models to examine positive association between pathways and exposures. Without this reduction in the number of models, the results will be skewed to show many more BC models because of the mummichog results. Analysis of these skewed data may lead to conclusions, which give undue weight to covariates which show up in more models. Some variables activated more pathways than others because some variables were in more models than other variables were. For example, initial analysis without this reduction in models led to the false relatively high number of times sex, age and BMI were the most common predictors in the activation of many pathways.

4. RESULTS

The mummichog analysis resulted in a dataset including p-values for 200 models. Mummichog analyzes a total of 119 pathways. 22 of the mummichog pathways did not have a p-value less than .05 for any of the seven variables (covariates), resulting in 97 pathways to analyze for significant predictors. Each exposure metric was analyzed to determine the most

likely variable responsible for the activation of each pathway. This was done by comparing each of the p values for the ten models described above with doseKG as the metric for pollutant measurement. If a model resulted in a p-value below .05, the predictor in this model was determined to be the significant contributor to the activation of this pathway. For example, if the BC doseKG 7 variable model including sex resulted in a p-value less than .05 then, sex was determined to be the significant exposure contributing to the activation of the pathway. Each exposure term had as many as four variables identified. A script was created in R to loop through this analysis for every model and pathway to minimize error and reduce the analysis time.

Two models showed most pathways with decreasing p-value from concentration to dose to doseKG: BC_sex_7 variable C18 and BC_Lag_4 variable_HILIC models. Of all the pathways assessed by mummichog, many metabolites pertain to biological processes unrelated to pollution exposure. We assessed all the pollutants and covariates sequentially, and then compared these results to determine which predictors are most frequently associated with each pathway.

The doseKG models were selected in order to analyze results in a comparative way. Using all models doesn't allow for an accurate comparison because some exposure terms are considered as main effects in multiple models, resulting in skewed results. From the doseKG models, each exposure term was analyzed to determine significant pathways activated by each. The three variable models were used for the exposures represented as the main effect in multiple models. The exposures with multiple models representing main effects include heat index, lag and the pollutant. The total number of times an exposure term was identified as a

significant contributor in activating a pathway was compiled and represented in Tables 1-18. These tables display the pathways activated by each pollutant model with significant exposure variables shown. Poll is short for pollutant in these tables. Therefore, BC Poll in Table 4.1 means BC pollutant. For BC Poll, the doseKG BC pollutant model identified BC as a significant contributor leading to the activating the pathways marked (Bile acid biosynthesis, De novo fatty acid biosynthesis, Glycerophospholipid metabolism, etc.).

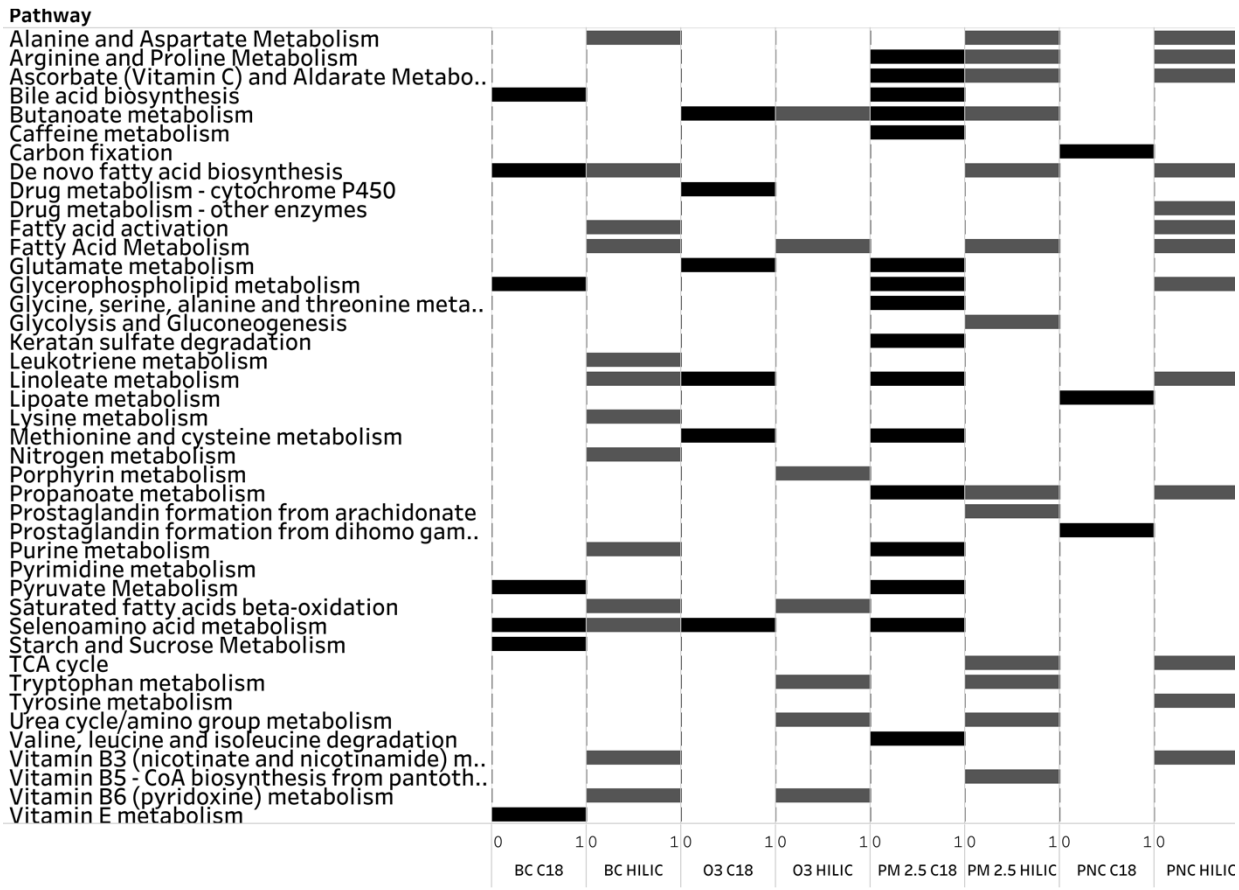

Table 4.1: Pathways (predictors) activated by pollutants (BC, O_3, $PM_{2.5}$*, PNC) while controlling for pollutant dose normalized by body mass, heat index and lagged exposure using the C18 column and HILIC columns. C18 pollutants are in black and HILIC pollutants are gray.*

Table 4.2: Pathways (predictors) activated by BC, Heat Index, Lagged Exposure, Physical Activity, Age, BMI and Sex while controlling for BC dose normalized by body mass, heat index and lagged exposure using the C18 column. Sex is controlling for everything to the left of it – put pollutant first. In this way models build from left to right. This model failed to produce a result for heat index.

BC	Heat Index	Lag	Activity	Age	BMI	Sex

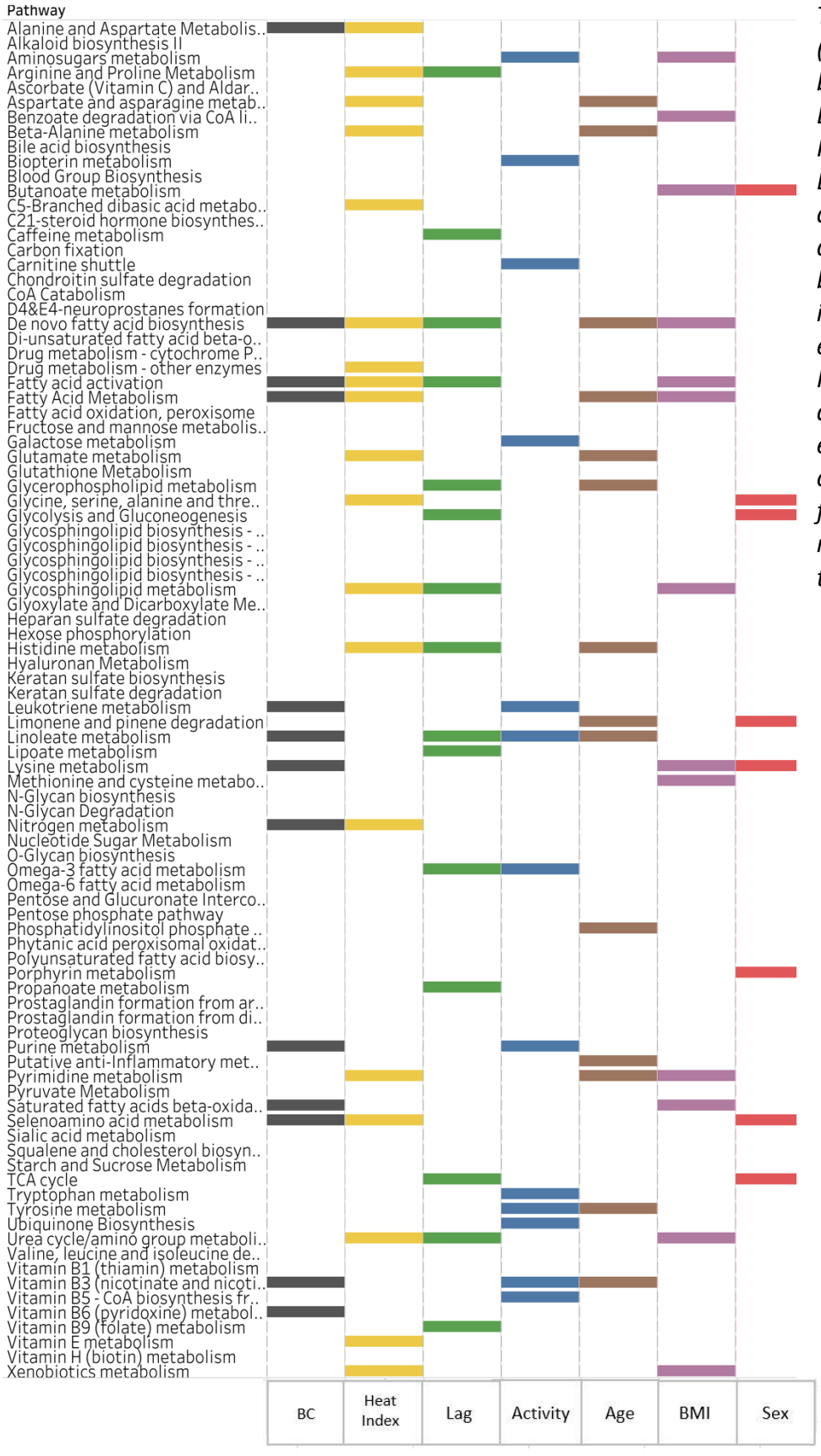

Table 4.3: Pathways (predictors) activated by BC, Heat Index, Lagged Exposure, Physical Activity, Age, BMI and Sex while controlling for BC dose normalized by body mass, heat index and lagged exposure using the HILIC column. Sex is controlling for everything to the left of it – put pollutant first. In this way models build from left to right.

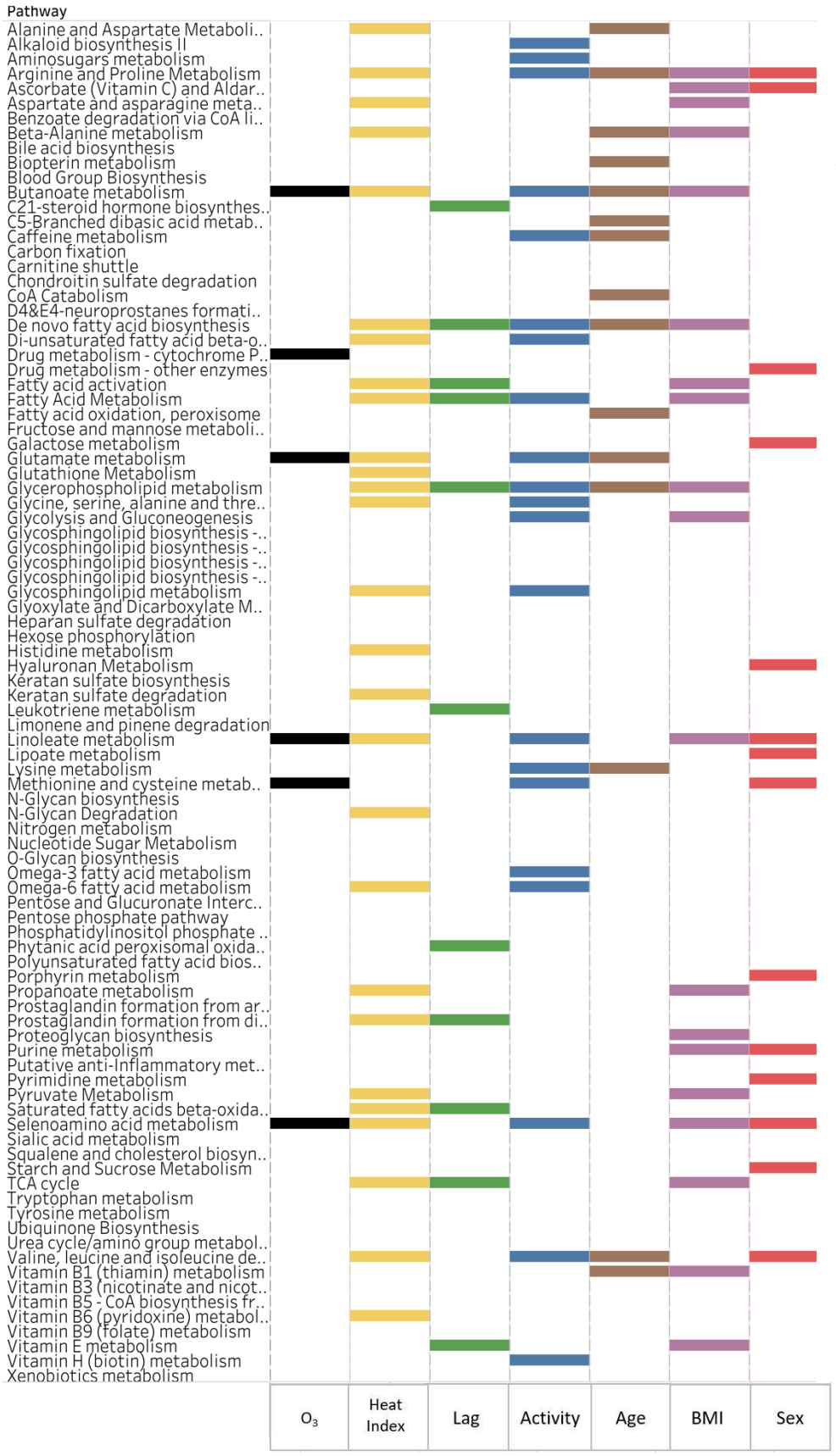

Table 4.4: Pathways (predictors) activated by O_3, Heat Index, Lagged Exposure, Physical Activity, Age, BMI and Sex while controlling for O_3 dose normalized by body mass, heat index and lagged exposure using the C18 column. Sex is controlling for everything to the left of it – put pollutant first. In this way models build from left to right.

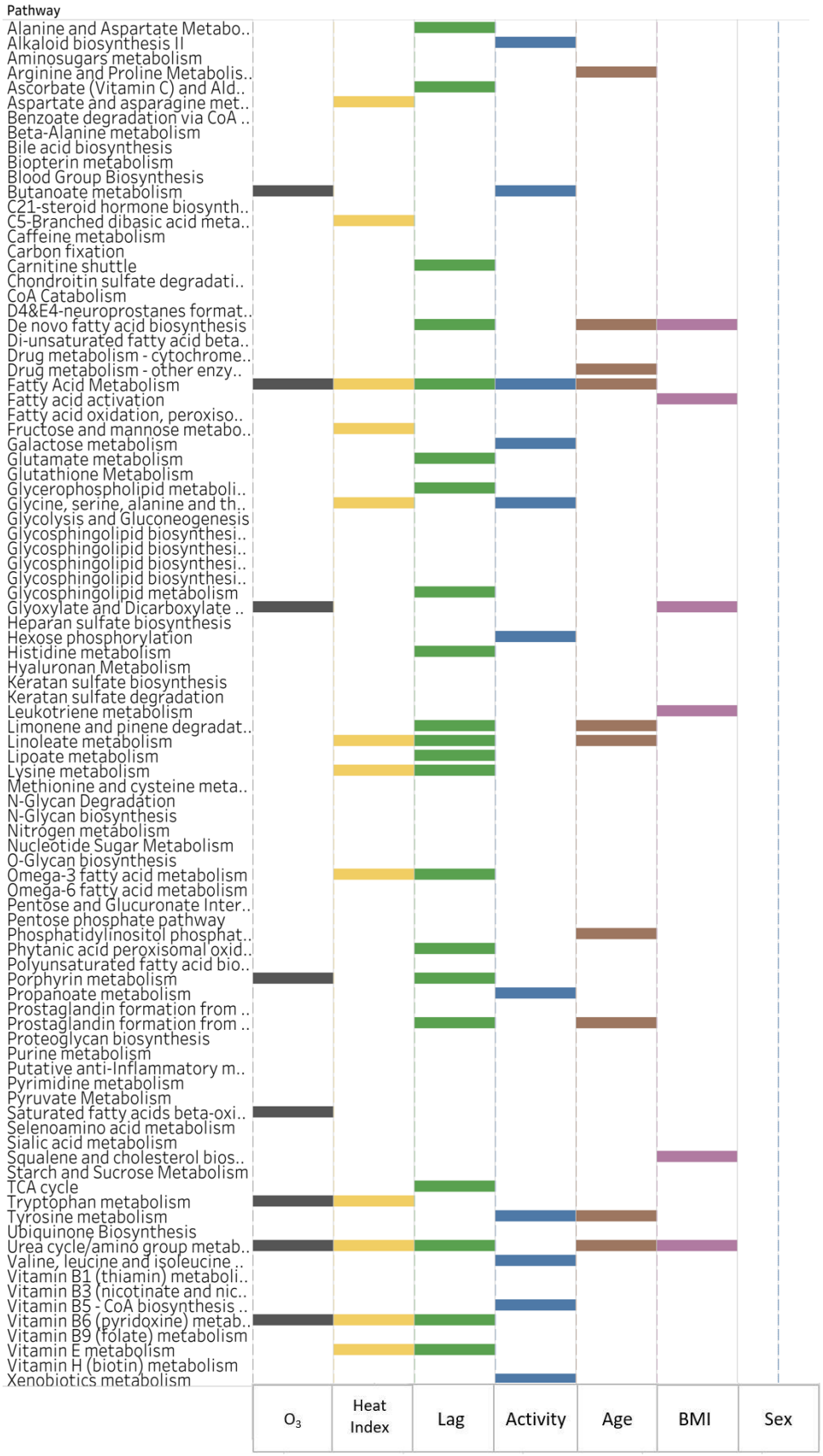

Table 4.5: Pathways (predictors) activated by O_3, Heat Index, Lagged Exposure, Physical Activity, Age, BMI and Sex while controlling for O_3 dose normalized by body mass, heat index and lagged exposure using the HILIC column. Sex is controlling for everything to the left of it – put pollutant first. In this way models build from left to right. This model fails to produce a result for sex.

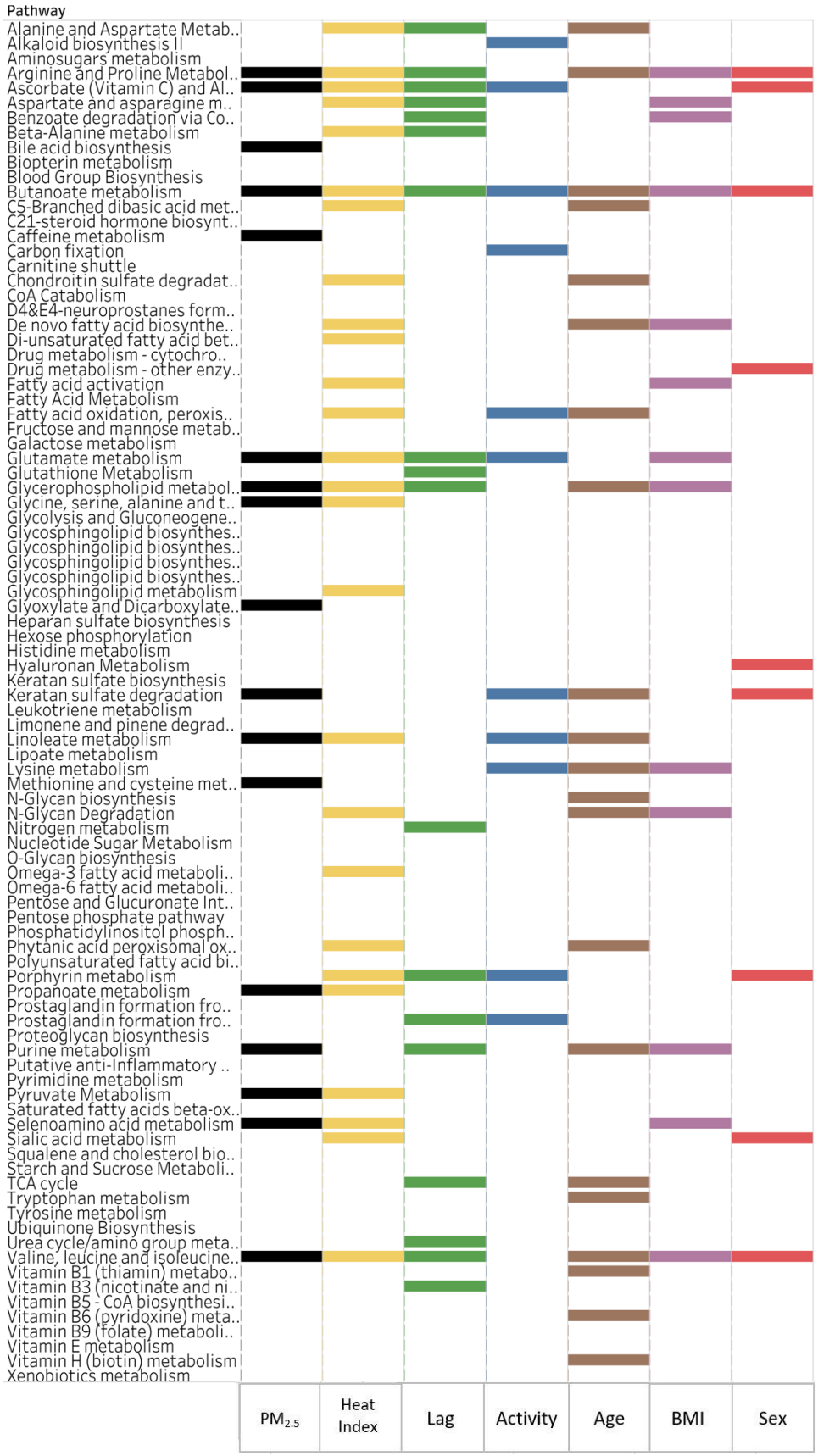

Table 4.6: Pathways (predictors) activated by $PM_{2.5}$*, Heat Index, Lagged Exposure, Physical Activity, Age, BMI and Sex while controlling for* $PM_{2.5}$ *dose normalized by body mass, heat index and lagged exposure using the C18 column. Sex is controlling for everything to the left of it – put pollutant first. In this way models build from left to right.*

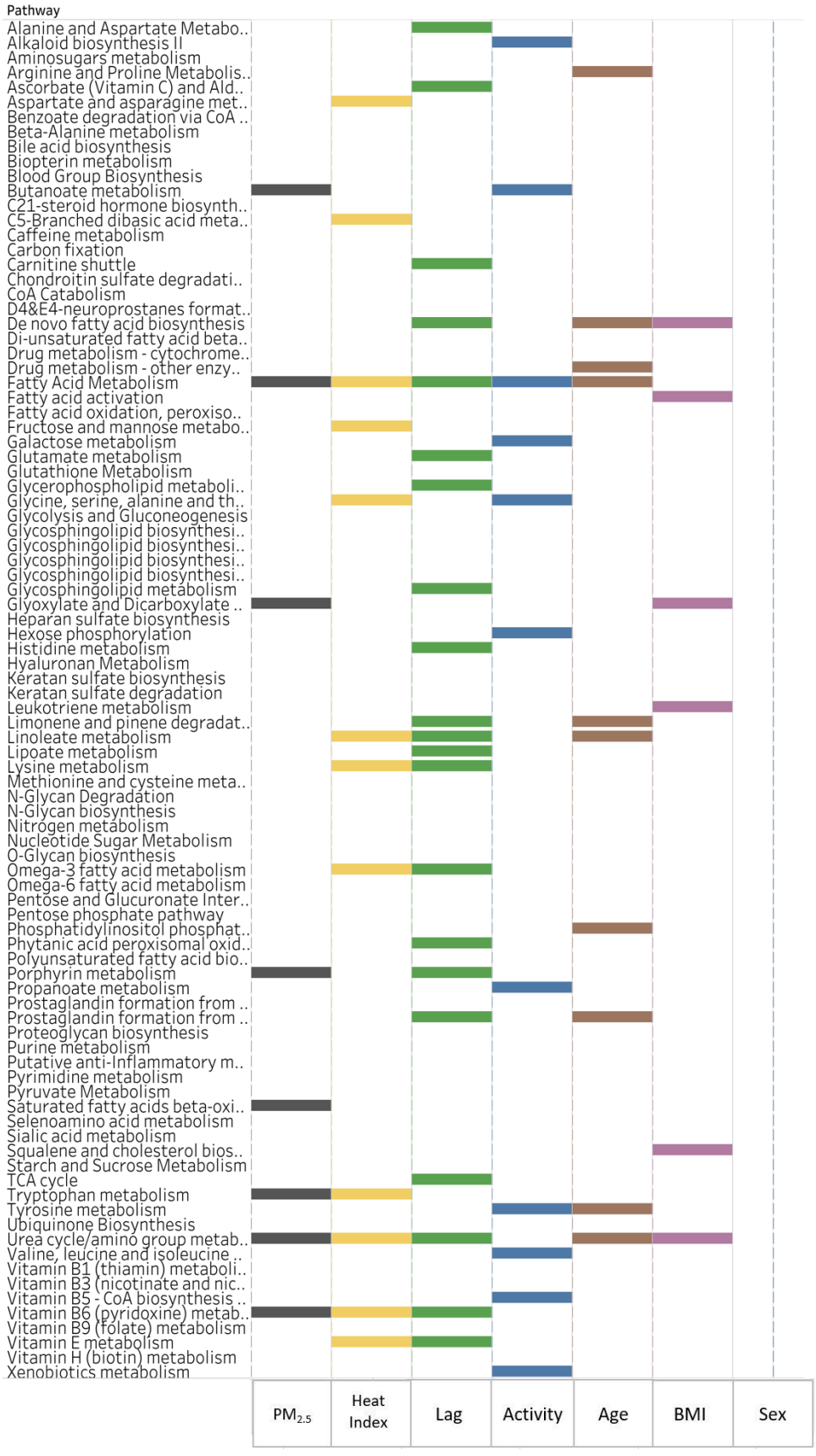

Table 4.7: Pathways (predictors) activated by $PM_{2.5}$*, Heat Index, Lagged Exposure, Physical Activity, Age, BMI and Sex while controlling for* $PM_{2.5}$ *dose normalized by body mass, heat index and lagged exposure using the HILIC column. Sex is controlling for everything to the left of it – put pollutant first. In this way models build from left to right.*

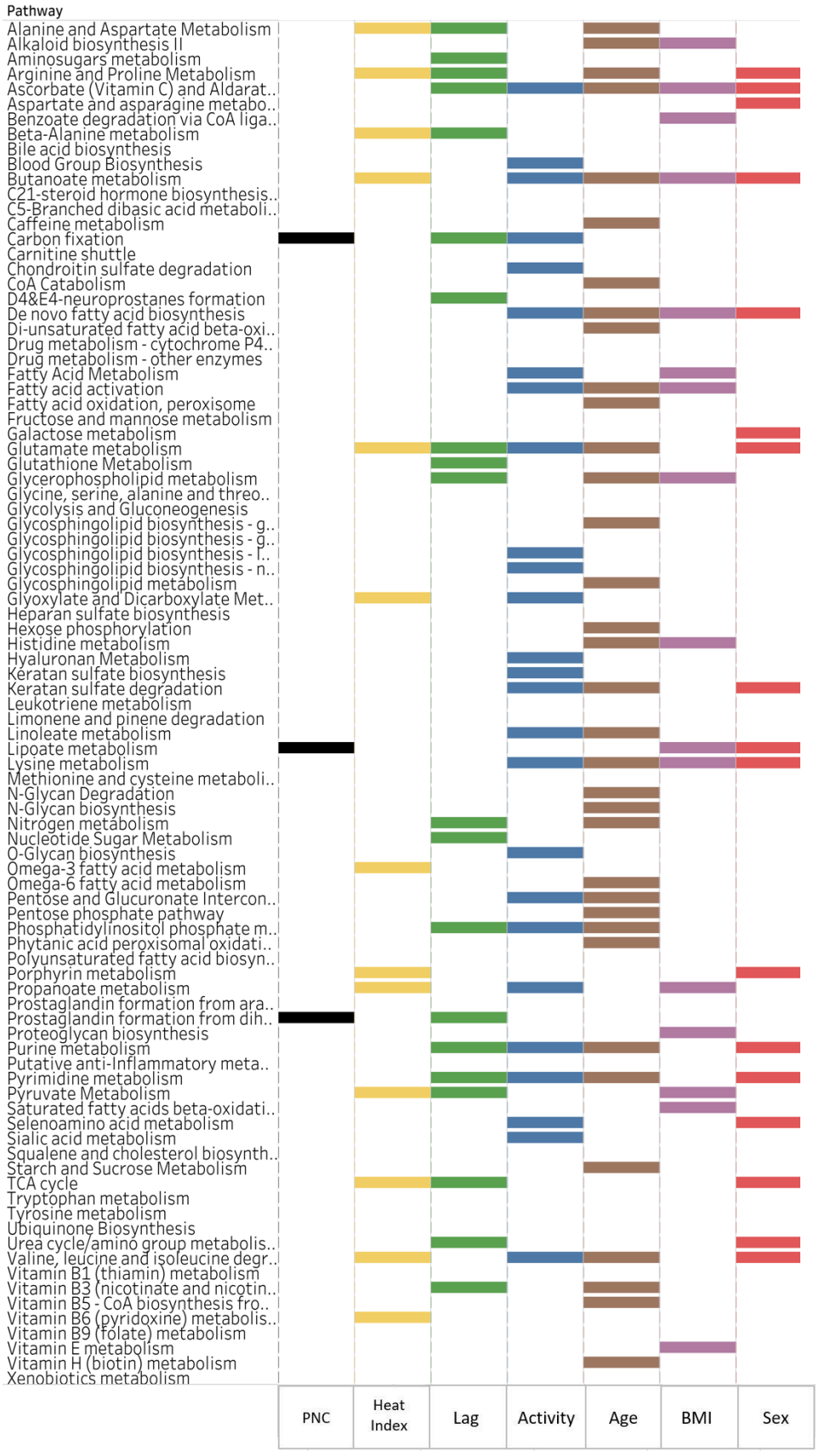

Table 4.8: Pathways (predictors) activated by PNC, Heat Index, Lagged Exposure, Physical Activity, Age, BMI and Sex while controlling for PNC dose normalized by body mass, heat index and lagged exposure using the C18 column. Sex is controlling for everything to the left of it – put pollutant first. In this way models build from left to right.

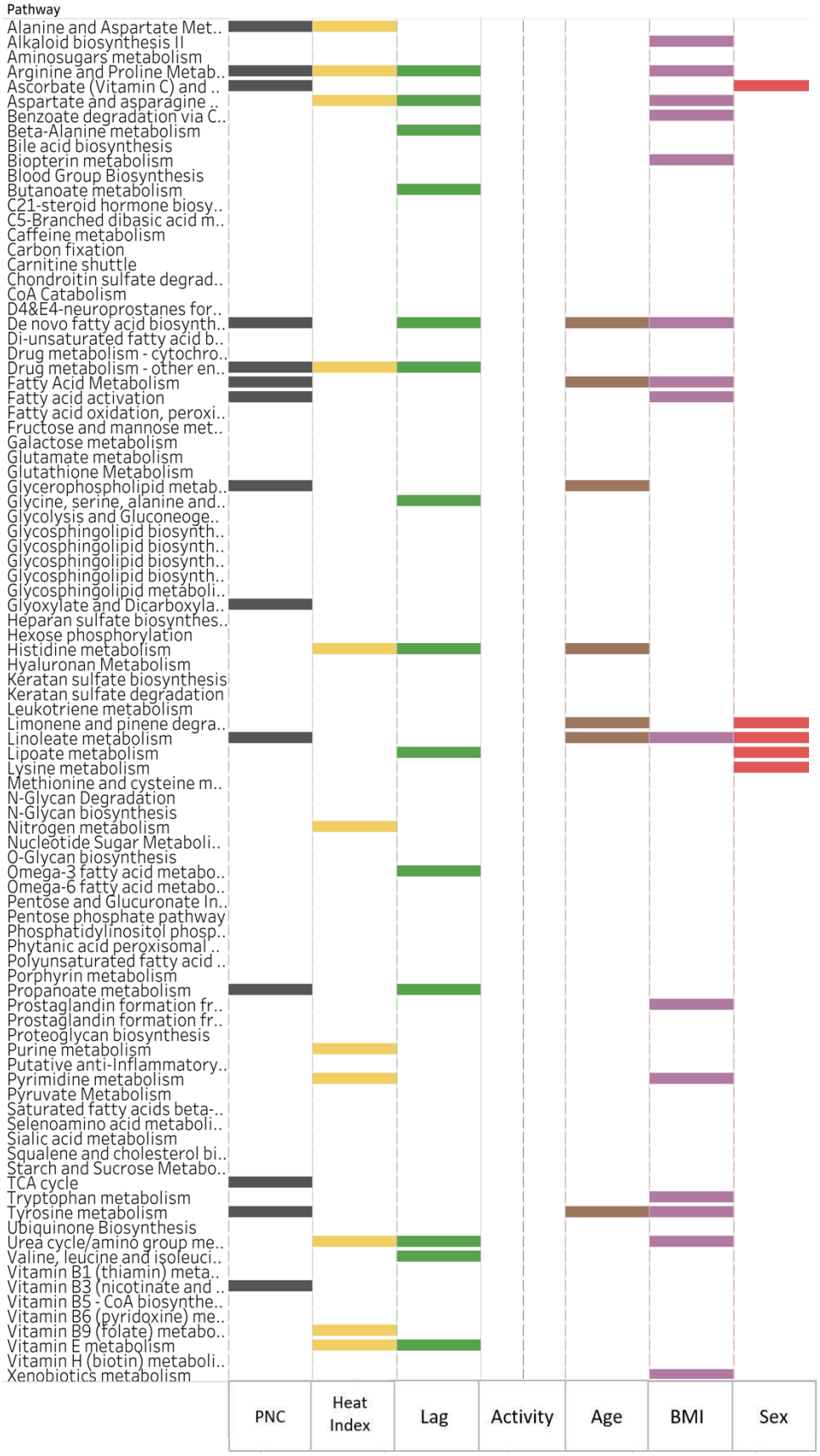

Table 4.9: Pathways (predictors) activated by PNC, Heat Index, Lagged Exposure, Physical Activity, Age, BMI and Sex while controlling for PNC dose normalized by body mass, heat index and lagged exposure using the HILIC column. Sex is controlling for everything to the left of it – put pollutant first. In this way models build from left to right. This model fails to produce a result for physical activity.

Not all models converged to a solution in the mummichog analysis. Tables 4.2, 4.5 and 4.9 show models that did not provide results for some variables.

5. DISCUSSION

In many models, covariates activated more pathways than the measured acute pollution exposure metrics. For example, in Table 4.4, heat index is significantly associated with 27 metabolic pathways whereas BC dose normalized to body mass is only associated with six. The lower degree of acute exposure pathway activation may be due to the limited amount of time (approximately two hours) between pre-exposure and post-exposure saliva samples, especially in comparison to longer-term exposures or ongoing biological processes, which may have been ongoing before the experiment was conducted. For example, age, BMI, and sex are static variables that do not change during the exposure period, and hence pathways associated with these variables would not need to be altered by a relatively brief exposure time. Similarly, heat index is measured during the exposure period, but it is likely to be strongly correlated with heat index earlier that same day (or even the preceding day) so that changes in metabolomic features could reflect hours-long or days-long processes rather than those that occur during the acute exposure period. According to Fang et.al, prior $PM_{2.5}$ exposure is a stronger predictor of metabolic pathways compared to acute exposure (Fang, Cassidy, & Christiani, 2010). This is consistent with the findings from this analysis. The acute exposure of BC, O_3, $PM_{2.5}$ and PNC contribute to the activation of each pathway less than the longer-term exposure from Heat Index and $PM_{2.5}$.

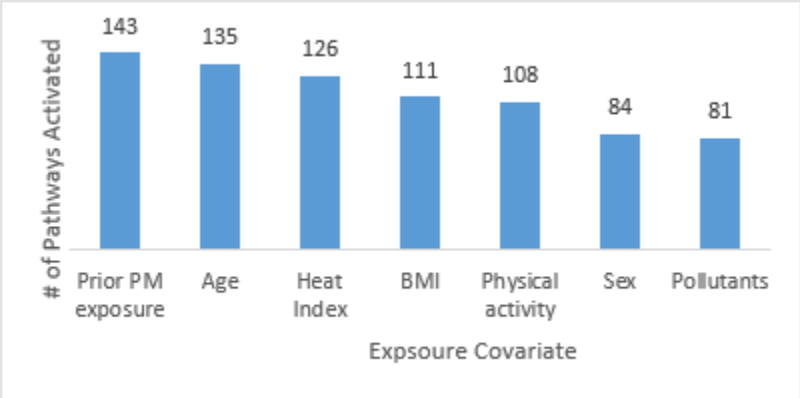

Table 5.1: Total number of pathways activated by covariates (Prior lagged exposure ($PM_{2.5}$ *for all pollutants except* O_3*, which used prior* O_3 *exposure)) exposure, age, heat index, BMI, physical activity, sex) and pollutants (BC,* O_3*,* $PM_{2.5}$*, PNC) with HILIC and C18 models for dose normalized by body weight.*

Table 5.1 shows the total number of pathways activated by the covariates and pollutants in the dose normalized by body weight models. Prior $PM_{2.5}$ exposure has a higher number of pathways activated when compared to acute pollutant exposure, which aligns with Fang et.al's research findings. The most important variables (listed from most important to least important) in activating pathways identified from the model selection process were heat index, prior $PM_{2.5}$ exposure, physical activity, age, BMI and then sex. According to Table 5.1, the covariates with the greatest number of pathways activated in descending order are prior $PM_{2.5}$ exposure, age, heat index, BMI, physical activity and sex.

A PubMed search was automated using an R script with the ReadPubMed package. Search terms were created for each pathway and the following exposure terms: heat index, physical activity and air pollution. The pathway and exposure term were both put in parenthesis to search for exact terms in PubMed. Many search terms did not result in any findings in PubMed. This doesn't mean there is no association between these pathways and exposures, but there may not be any research exploring these associations and can be seen as a limitation

of this study. Alternatively, a high number of articles found in this search do not necessarily mean there is an association as further research is needed. Articles were reviewed and associations found were recorded in Table 5.2. Of the search of 119 pathways, 56 pathways included search results with physical activity. Of these 56 pathways, 42 pathways included physical activity as a significant contributor in at least one model. Air pollution matched with 46 pathways in the PubMed search, of which 31 demonstrated significant contributors in at least one model. Cells highlighted in green in Table 5.2 indicate that the association identified was found to be consistent with the literature review. An empty cell means there were no results from the PubMed search. Other pathways that did not have an association identified with at least one of these three exposures were not included here.

pathway	Activity	Air Pollution	Heat Index		Activity	Air Pollution	Heat Index
1- and 2-Methylnaphthalene degradation		2		Linoleate metabolism		1	
3-oxo-10R-octadecatrienoate beta-oxidation	38	6		Methionine and cysteine metabo	4	3	
Alanine and Aspartate Metabolism	113	14		Nitrogen metabolism	3	7	
Alkaloid biosynthesis II	4	1		Pentose phosphate pathway	3	2	
Androgen and estrogen biosynthesis and metal	82	2		Porphyrin metabolism		4	
Arachidonic acid metabolism	4	1		Prostaglandin formation from ara	0	2	
Arginine and Proline Metabolism	1			Prostaglandin formation from dil	1		
Aspartate and asparagine metabolism		1		Purine metabolism	3	1	
Blood Group Biosynthesis	2676	170		Pyrimidine metabolism		1	
C21-steroid hormone biosynthesis and metabo	348	30		R Group Synthesis	989	49	
Caffeine metabolism	2			ROS Detoxification	0	1	
Carbon fixation		19		Saturated fatty acids beta-oxidati	24	2	
Carnitine shuttle	1			Sphingolipid metabolism	3		
CoA Catabolism	19			Squalene and cholesterol biosynthesis		1	
D4&E4-neuroprostanes formation	19	6	14	TCA cycle	16	3	
Dimethyl-branched-chain fatty acid mitochon	10	2		Tryptophan metabolism	7	3	
Electron transport chain	36	9		Tyrosine metabolism		1	
Fatty Acid Metabolism	48	9		Valine, leucine and isoleucine deg	14		
Fatty acid oxidation	111	4		Vitamin A (retinol) metabolism	111	12	
Fatty acid oxidation, peroxisome	30			Vitamin B1 (thiamin) metabolism	18	1	
Galactose metabolism	1			Vitamin B12 (cyanocobalamin) m	66	5	
Geraniol degradation		1		Vitamin B2 (riboflavin) metabolis	36	3	
Glutamate metabolism		1		Vitamin B6 (pyridoxine) metabol	7	1	
Glutathione Metabolism		3		Vitamin B9 (folate) metabolism	121	10	
Glycerolipid metabolism		1		Vitamin D	1861	102	1
Hyaluronan Metabolism	1			Vitamin D3 (cholecalciferol) meta	156	6	
Leukotriene metabolism		1		Vitamin E metabolism	1	1	
Limonene and pinene degradation		6		Vitamin H (biotin) metabolism	3	2	

Table 5.2 – PubMed search results for pathways identified mummichog analysis. Search term included pathway and either physical activity, air pollution or heat index, each of which was in quotes.

The following pathways were activated by the short-term exposure to pollutants in the dose normalized by body weight model, with none of the covariates being found significant in activating the same pathway. BC was found to be significant in activating the Vitamin B6 pathway using the HILIC column. Beta-Alanine metabolism (amino acid metabolism), pyruvate metabolism (carbohydrate metabolism) and selenoamino acid metabolism pathways were activated by BC using the C18 column. O_3 was found to be significant in activating the drug metabolism pathway (xenobiotics biodegradation and metabolism) using the C18 column and the saturated fatty acids beta-oxidation (lipid metabolism) in the HILIC column. PNC was found to be significant in activating the fatty acid activation metabolism (global metabolism), glyoxylate and dicarboxylate metabolism (carbohydrate metabolism), TCA cycle metabolism (carbohydrate metabolism) and Vitamin B3 metabolism using the HILIC column. $PM_{2.5}$ was found to be significant in activating the propanoate metabolism (carbohydrate metabolism) using the HILIC column. $PM_{2.5}$ was found to be significant in activating the bile acid biosynthesis (lipid metabolism), caffeine metabolism (biosynthesis of secondary metabolites), glyoxylate and dicarboxylate metabolism (carbohydrate metabolism), and methionine and cysteine metabolism (amino acid metabolism) using the C18 column.

A PubMed search of these pathways and their relationship to the pollutants that activated them found $PM_{2.5}$ to be known to activate the bile acid biosynthesis metabolism (lipid metabolism) and the methionine cysteine metabolism (amino acid metabolism). Untargeted metabolomic analysis demonstrated over 300 metabolites of LO2 cells dysregulated after exposure to $PM_{2.5}$, with cholesterol and bile acid metabolism being significantly dysregulated (Duan et al., 2020). According to Shi et.al., the metabolic pathways activated when mice were

exposed to $PM_{2.5}$ include: glycine, serine and threonine metabolism, aminoacyl-tRNA biosynthesis, cysteine and methionine metabolism, alanine, aspartate and glutamate metabolism, methane metabolism, linoleic acid metabolism and valine, and leucine and isoleucine biosynthesis, which are related to liver metabolism (Shi, Han, Mao, Fan, & Jin, 2019). The glycine, serine and threonine metabolism and alanine, aspartate and glutamate metabolism were both identified as pathways in this study being activated by $PM_{2.5}$ as well.

Air pollution and physical activity have a complicated relationship that is difficult to disentangle. While acute exposure to air pollution may contribute to activating less metabolic pathways than long term exposure, heat index and other covariates, BC, O_3, PNC and $PM_{2.5}$ all contribute to activating metabolic pathways in addition to the effects of covariates. Acute air pollution exposure was found to be significant in activating metabolic pathways a total of 81 times. We identified 14 metabolic pathways that showed significant correlation with acute air pollution exposure through at least one pollutant (BC, O_3, $PM_{2.5}$, PNC) with no other covariates contributing to activating that pathway. There is limited research in untargeted, high-resolution metabolomics data of the effects of air pollution in activating metabolic pathways. This research offers a glimpse of the relationships between air pollution in both acute and long-term exposures and the pathways activated by each. Future studies may build on this research to further explore the air pollution factors that contribute to activating different metabolic pathways and their effects on health.

REFERENCES

Barthelemy, J., Sanchez, K., Miller, M. R., & Khreis, H. (2020). New Opportunities to Mitigate the Burden of Disease Caused by Traffic Related Air Pollution: Antioxidant-Rich Diets and Supplements. *Int J Environ Res Public Health, 17*(2). doi:10.3390/ijerph17020630

Bessonneau, V., Pawliszyn, J., & Rappaport, S. M. (2017). The Saliva Exposome for Monitoring of Individuals' Health Trajectories. *Environ Health Perspect, 125*(7), 077014. doi:10.1289/ehp1011

Brook, R. D., Rajagopalan, S., Pope, C. A., 3rd, Brook, J. R., Bhatnagar, A., Diez-Roux, A. V., . . . Kaufman, J. D. (2010). Particulate matter air pollution and cardiovascular disease: An update to the scientific statement from the American Heart Association. *Circulation, 121*(21), 2331-2378. doi:10.1161/CIR.0b013e3181dbece1

Chuang, K. J., Chan, C. C., Su, T. C., Lee, C. T., & Tang, C. S. (2007). The effect of urban air pollution on inflammation, oxidative stress, coagulation, and autonomic dysfunction in young adults. *Am J Respir Crit Care Med, 176*(4), 370-376. doi:10.1164/rccm.200611-1627OC

Cohen, A. J., Brauer, M., Burnett, R., Anderson, H. R., Frostad, J., Estep, K., . . . Forouzanfar, M. H. (2017). Estimates and 25-year trends of the global burden of disease attributable to ambient air pollution: an analysis of data from the Global Burden of Diseases Study 2015. *Lancet, 389*(10082), 1907-1918. doi:10.1016/s0140-6736(17)30505-6

Cui, Y., Balshaw, D. M., Kwok, R. K., Thompson, C. L., Collman, G. W., & Birnbaum, L. S. (2016). The Exposome: Embracing the Complexity for Discovery in Environmental Health. *Environ Health Perspect, 124*(8), A137-140. doi:10.1289/ehp412

Duan, X., Zheng, L., Zhang, X., Wang, B., Xiao, M., Zhao, W., . . . Sui, G. (2020). A Membrane-free Liver-Gut-on-Chip Platform for the Assessment on Dysregulated Mechanisms of Cholesterol and Bile Acid Metabolism Induced by PM(2.5). *ACS Sens*. doi:10.1021/acssensors.0c01524

Fang, S. C., Cassidy, A., & Christiani, D. C. (2010). A systematic review of occupational exposure to particulate matter and cardiovascular disease. *International journal of environmental research and public health, 7*(4), 1773-1806. doi:10.3390/ijerph7041773

Finlon, E. (2018). *Use of High-Resolution Metabolomics to Assess Changes in Metabolic Pathways Associated with Acute Exposure to Air Pollutants.* (MPH). Georgia State University, Retrieved from https://scholarworks.gsu.edu/iph_theses/580

Fisher, J. E., Loft, S., Ulrik, C. S., Raaschou-Nielsen, O., Hertel, O., Tjønneland, A., . . . Andersen, Z. J. (2016). Physical Activity, Air Pollution, and the Risk of Asthma and Chronic Obstructive Pulmonary Disease. *Am J Respir Crit Care Med, 194*(7), 855-865. doi:10.1164/rccm.201510-2036OC

Frampton, M. W. (2006). Inflammation and airborne particles. *Clin Occup Environ Med, 5*(4), 797-815. doi:10.1016/j.coem.2006.07.006

Franklin, B. A., Brook, R., & Arden Pope, C., 3rd. (2015). Air pollution and cardiovascular disease. *Curr Probl Cardiol, 40*(5), 207-238. doi:10.1016/j.cpcardiol.2015.01.003

Goodacre, R., Broadhurst, D., Smilde, A. K., Kristal, B. S., Baker, J. D., Beger, R., . . . Wulfert, F. (2007). Proposed minimum reporting standards for data analysis in metabolomics. *Metabolomics, 3*(3), 231-241. doi:10.1007/s11306-007-0081-3

Greenwald, R., Hayat, M. J., Barton, J., & Lopukhin, A. (2016). A Novel Method for Quantifying the Inhaled Dose of Air Pollutants Based on Heart Rate, Breathing Rate and Forced Vital Capacity. *PLoS One, 11*(1), e0147578. doi:10.1371/journal.pone.0147578

Hayat, M. J., Hedlin, H. (2012). Modern Statistical Modeling Approaches For Analyzing Repeated-Measures Data. *Lippincott Williams & Wilkins*.

Hupin, D., Roche, F., Gremeaux, V., Chatard, J. C., Oriol, M., Gaspoz, J. M., . . . Edouard, P. (2015). Even a low-dose of moderate-to-vigorous physical activity reduces mortality by 22% in adults aged ≥60 years: a systematic review and meta-analysis. *Br J Sports Med, 49*(19), 1262-1267. doi:10.1136/bjsports-2014-094306

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. *Environment International, 119*, 334-345. doi:https://doi.org/10.1016/j.envint.2018.06.025

Kampa, M., & Castanas, E. (2008). Human health effects of air pollution. *Environ Pollut, 151*(2), 362-367. doi:10.1016/j.envpol.2007.06.012

Kaplan, G. A., Strawbridge, W. J., Cohen, R. D., & Hungerford, L. R. (1996). Natural history of leisure-time physical activity and its correlates: associations with mortality from all causes and cardiovascular disease over 28 years. *Am J Epidemiol, 144*(8), 793-797. doi:10.1093/oxfordjournals.aje.a009003

Kelly, F. J., & Fussell, J. C. (2011). Air pollution and airway disease. *Clin Exp Allergy, 41*(8), 1059-1071. doi:10.1111/j.1365-2222.2011.03776.x

Kennedy, A. D., Wittmann, B. M., Evans, A. M., Miller, L. A. D., Toal, D. R., Lonergan, S., . . . Pappan, K. L. (2018). Metabolomics in the clinic: A review of the shared and unique features of untargeted metabolomics for clinical research and clinical testing. *J Mass Spectrom, 53*(11), 1143-1154. doi:10.1002/jms.4292

Király, M., Dalmadiné Kiss, B., Vékey, K., Antal, I., & Ludányi, K. (2016). [Mass spectrometry: past and present]. *Acta Pharm Hung, 86*(1), 3-11.

Liang, D., Ladva, C. N., Golan, R., Yu, T., Walker, D. I., Sarnat, S. E., . . . Sarnat, J. A. (2019). Perturbations of the arginine metabolome following exposures to traffic-related air pollution in a panel of commuters with and without asthma. *Environ Int, 127*, 503-513. doi:10.1016/j.envint.2019.04.003

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. *Environment International, 120*, 145-154. doi:https://doi.org/10.1016/j.envint.2018.07.044

Liu, L., Ruddy, T. D., Dalipaj, M., Szyszkowicz, M., You, H., Poon, R., . . . Dales, R. (2007). Influence of personal exposure to particulate air pollution on cardiovascular physiology and biomarkers of inflammation and oxidative stress in subjects with diabetes. *J Occup Environ Med, 49*(3), 258-265. doi:10.1097/JOM.0b013e31803220ef

Mu, L., Niu, Z., Blair, R. H., Yu, H., Browne, R. W., Bonner, M. R., . . . Swanson, M. (2019). Metabolomics Profiling before, during, and after the Beijing Olympics: A Panel Study of Within-Individual Differences during Periods of High and Low Air Pollution. *Environ Health Perspect, 127*(5), 57010. doi:10.1289/ehp3705

Rowangould, G. M. (2013). A census of the US near-roadway population: Public health and environmental justice considerations. *Transportation Research Part D: Transport and Environment, 25*, 59-67. doi:https://doi.org/10.1016/j.trd.2013.08.003

Schraufnagel, D. E., Balmes, J. R., Cowl, C. T., De Matteis, S., Jung, S. H., Mortimer, K., . . . Wuebbles, D. J. (2019). Air Pollution and Noncommunicable Diseases: A Review by the Forum of International Respiratory Societies' Environmental Committee, Part 1: The Damaging Effects of Air Pollution. *Chest, 155*(2), 409-416. doi:10.1016/j.chest.2018.10.042

Schymanski, E. L., Jeon, J., Gulde, R., Fenner, K., Ruff, M., Singer, H. P., & Hollender, J. (2014). Identifying small molecules via high resolution mass spectrometry: communicating confidence. *Environ Sci Technol, 48*(4), 2097-2098. doi:10.1021/es5002105

Shen, B., Yi, X., Sun, Y., Bi, X., Du, J., Zhang, C., . . . Guo, T. (2020). Proteomic and Metabolomic Characterization of COVID-19 Patient Sera. *Cell, 182*(1), 59-72.e15. doi:10.1016/j.cell.2020.05.032

Shi, C., Han, X., Mao, X., Fan, C., & Jin, M. (2019). Metabolic profiling of liver tissues in mice after instillation of fine particulate matter. *Sci Total Environ, 696*, 133974. doi:10.1016/j.scitotenv.2019.133974

Sinharay, R., Gong, J., Barratt, B., Ohman-Strickland, P., Ernst, S., Kelly, F. J., . . . Chung, K. F. (2018). Respiratory and cardiovascular responses to walking down a traffic-polluted road compared with walking in a traffic-free area in participants aged 60 years and older with chronic lung or heart disease and age-matched healthy controls: a randomised, crossover study. *Lancet, 391*(10118), 339-349. doi:10.1016/s0140-6736(17)32643-0

Stanstrup, J., Broeckling, C. D., Helmus, R., Hoffmann, N., Mathé, E., Naake, T., . . . Neumann, S. (2019). The metaRbolomics Toolbox in Bioconductor and beyond. *Metabolites, 9*(10). doi:10.3390/metabo9100200

Tainio, M., de Nazelle, A. J., Götschi, T., Kahlmeier, S., Rojas-Rueda, D., Nieuwenhuijsen, M. J., . . . Woodcock, J. (2016). Can air pollution negate the health benefits of cycling and walking? *Prev Med, 87*, 233-236. doi:10.1016/j.ypmed.2016.02.002

Vinzents, P. S., Møller, P., Sørensen, M., Knudsen, L. E., Hertel, O., Jensen, F. P., . . . Loft, S. (2005). Personal exposure to ultrafine particles and oxidative DNA damage. *Environ Health Perspect, 113*(11), 1485-1490. doi:10.1289/ehp.7562

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

Walhout, A. J. M., Vidal, M., & Dekker, J. (2013). Front Matter. In A. J. M. Walhout, M. Vidal, & J. Dekker (Eds.), *Handbook of Systems Biology* (pp. iii). San Diego: Academic Press.

WHO. (2018). Ambient (outdoor) air pollution. Retrieved from https://www.who.int/news-room/fact-sheets/detail/ambient-(outdoor)-air-quality-and-health

Xia, T., Kovochich, M., & Nel, A. (2006). The role of reactive oxygen species and oxidative stress in mediating particulate matter injury. *Clin Occup Environ Med, 5*(4), 817-836. doi:10.1016/j.coem.2006.07.005

LIST OF TABLES AND FIGURES

and lagged exposure using the HILIC column. Sex is controlling for everything to the left of it – put pollutant first. In this way models build from left to right.

Table 4.8: Pathways (predictors) activated by PNC, Heat Index, Lagged Exposure, Physical Activity, Age, BMI and Sex while controlling for PNC dose normalized by body mass, heat index and lagged exposure using the C18 column. Sex is controlling for everything to the left of it – put pollutant first. In this way models build from left to right.

Table 4.9: Pathways (predictors) activated by PNC, Heat Index, Lagged Exposure, Physical Activity, Age, BMI and Sex while controlling for PNC dose normalized by body mass, heat index and lagged exposure using the HILIC column. Sex is controlling for everything to the left of it – put pollutant first. In this way models build from left to right. This model fails to produce a result for physical activity.

Table 5.1: Total number of pathways activated by covariates (Prior lagged exposure ($PM_{2.5}$ *for all pollutants except* O_3*, which used prior* O_3 *exposure)) exposure, age, heat index, BMI, physical activity, sex) and pollutants (BC,* O_3*,* $PM_{2.5}$*, PNC) with HILIC and C18 models for dose normalized by body weight.*

Table 5.2 – PubMed search results for pathways identified mummichog analysis. Search term included pathway and either physical activity, air pollution or heat index, each of which was in quotes.