AWARD NUMBER: W81XWH-17-1-0533

TITLE: Metabolomics of Lead Exposure and Its Role in Respiratory Disease

PRINCIPAL INVESTIGATOR: Jessica Lasky-Su

CONTRACTING ORGANIZATION: Brigham and Women's Hospital
Boston MA 02129,

REPORT DATE: Oct 2019

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE	**2. REPORT TYPE**	**3. DATES COVERED**
Oct 2019	Annual Report	15 Sept 2018-14 Sept 2019

4. TITLE AND SUBTITLE

Metabolomics of Lead Exposure and Its Role in Respiratory Disease

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-17-1-0533

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Jessica Lasky-Su; Rachel Kelly

E-Mail: rejas@channing.harvard.edu; hprke@channing.harvard.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Channing Division of Network Medicine
Brigham and Women's Hospital, Harvard
Medical School

181 Longwood Avenue
Boston
MA 02129, USA

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

The overarching hypothesis of this proposal was that exposure to lead (Pb) during active military service is related to the observed high prevalence of poor lung health among veterans. We aimed to explore the mechanisms underlying this relationship utilizing metabolomics; i.e. the systematic profiling of small (<10kDa) metabolites in a biological sample, which will allow us to construct a causal pathway demonstrating the mechanistic and biological connections between Pb exposure and lung health. To achieve this aim we identified participants from the ongoing Normative Ageing Study of Veterans. We selected men with detailed histories on their exposure to Pb, with comprehensive data on long term lung health and with blood samples suitable for metabolomics profiling. During this reporting period, we identified 661 plasma samples from 464 veterans, which we shipped to Metabolon Inc. for metabolomic profiling using four LCMS platforms, enabling the broadest coverage of the metabolome possible. We applied QC and data processing pipelines to these data, and initiated the statistical analysis plan outlined in our proposal. To date, we have successfully identified a metabolomic profile associated with Pb exposure and a metabolomic profile associated with poor lung health. Our analyses, encompassing both frequentist and network approaches, suggested that the relationship between Pb and the lung is mediated, in part, by dysregulated Glycine, Serine and Threonine Metabolism; Histidine Metabolism; Leucine, Isoleucine and Valine Metabolism; Phospholipid Metabolism and Sphingolipid metabolism. These findings have been presented at both national and international conferences, and have resulted in two manuscripts pending submission, and additional planned publications.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
					USAMRMC
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**	Unclassified	55	**19b. TELEPHONE NUMBER** *(include area code)*
Unclassified	Unclassified	Unclassified			

Table of Contents

Page

1. Introduction

Poor respiratory health relating to environmental exposures during active service represents a significant public health burden for Military service personnel and Veterans. Lead (Pb) is commonly found on military bases in the form of fine particulate matter and is thought to adversely affect pulmonary function for many decades after exposure. However, a complete understanding of the effects of Pb exposure on the respiratory system remains to be fully elucidated, and further investigation is paramount. Metabolomics, the systematic profiling of all the small molecules in a biological system, represents a powerful tool to (i) increase mechanistic understanding of the pathogenesis of Pb exposure on the respiratory system, and (ii) identify biomarkers of 'toxic' levels of exposure. The overarching hypothesis of this project is that a metabolomic profile of Pb exposure can be identified and used to understand how Pb exposure and the resulting changes on the metabolome have a downstream impact on respiratory health. Therefore, this projects aims to determine the influence of Pb exposure on respiratory health through the integration of the metabolomic profiles of heavy metal exposure and of respiratory disease. This will allow for the identification of novel blood-based biomarkers associated with long-term toxic Pb exposure and respiratory health. These biomarkers will provide mechanistic insights into the pathogenic effects of Pb on the respiratory system and into the disease pathways involved, supporting the future development of novel therapeutics. This represents the first study to utilize metabolomics in the exploration of both short-term and long-term Pb exposure and its effect on respiratory disease.

2. Keywords

Respiratory disease; Metabolomics; Metabolome; Lead (Pb); Heavy metals; Biomarkers; Metals Toxicology; Omics

3. Accomplishments

3.a Major Goals of the Project

The original major goals of this project, as detailed in the **Statement of work (SOW)** in our proposal are show in **Table 1a-c.** These goals are delineated by aim and subdivided by tasks (major and sub), and the extent of their completion and completion dates where relevant, along with accompanying notes is provided. A detailed description of the methodology and findings to date is provided below. The corresponding table or figure describing the results in indicated in square brackets. We note that due to the nature of the analyses, rather than proceeding sequentially as we previously indicated in the **SOW**, it has proven more effective to perform a number of tasks concurrently. One notable issue was a delay in the HPRO approval that has now been resolved (*see **section 5a** for details*).

Table 1a: SOW and Progress in Accomplishing the Outline Goals (Specific Aim 1)

Specific Aim 1 To identify metabolomic profiles of lead exposure in a population of Veterans	Timeline	% completion	Completion Date	Comments
Major Task 1 Metabolomic profiling of 374 NAS individuals	Months			
Subtask 1: Selection and shipment of blood plasma	1	100%	11/16/17	Additional samples; *total 661*, were profiled following price negotiations with Metabolon [**Table 2**]
Subtask 2: Metabolomic profiling	2-3	100%	1/31/18	
Subtask 3: Data processing and Quality control procedures using Metabolon, Inc internal standards	2-3	100%	1/31/18	
Subtask 4: Data Quality Control pipeline at the Channing Laboratory	3	100%	3/31/18	
Milestone Achieved: Metabolomic dataset received ready for analysis	*3*	100%	3/31/18	[**Table 3 & Figure 1**]
HRPO regulatory review	1-3	100%	2/4/18	IRB concluded the research did not constitute human subjects research
Milestone Achieved: HRPO Approval	*3*	100%	2/4/18	
Major Task 2 Identification of the metabolome of Pb exposure				
Subtask 1: Creation of a composite measure of Pb exposure incorporating duration and level, based on pre-existing measures in bone, toenail, blood and urine	1	90%	ongoing	We focus primarily on blood levels of Pb Exposure, as a and utilize the additional measures in other bio samples to further explore longer term exposure [**Figures 3&4**]
Subtask 2: Linear regression models and sensitivity analyses, exploring different confounders, to identify differential metabolites and pathway enrichment analysis to identify the metabolomic pathways mapping to these metabolites associated with exposure	2-3	100%	ongoing	Analyses has been conducted for existing measures of lead exposure including whole blood at the time concurrent to metabolomic profiling and spirometry. Metabolite and pathways of interest were identified [**Tables 4-7; Figures 2-4**]
Subtask 3: Network approaches, including WGCNA to identify metabolomic networks	3-4	95%	ongoing	Metabolite networks of interest have been identified [**Figures 5&6**]
Milestone Achieved: Publication of a manuscript exploring the metabolome of Pb exposure	*5*	75%	ongoing	Systematic literature searches have been conducted, introduction, methods, results and discussion write ups are ongoing. Plan for manuscript to be submitted within the next two months

Table 1b: SOW and Progress in Accomplishing the Outline Goals (Specific Aim 2)

Specific Aim 2 To identify metabolomic profiles of pulmonary Function in a population of Veterans	Timeline	% completion	Completion Date	Comments
Major Task 1 To identify a metabolomic profile of pulmonary function within a population of Veterans				
Subtask 1: Linear regression and network analysis as described in Major task 2 applied to the outcome of pulmonary function, as denoted by FEV_1 and FVC	6-8	100%	ongoing	Analyses completed [**Tables 6,8, 10, 11, 12**] manuscript pending submission based on the metabolome of lung function [**Appendix 1**]
Subtask 2: Development and assessment of a metabolomic score based on the findings from subtask 1 that can be used to discriminate men by their degree of lung function	8-9	75%	ongoing	ROC curve analyses and sensitivity and specificity are used to assess score. To date we have explored generating scores based on the first principal component of the significant metabolites [**Table 9**]. Currently exploring alternative approaches for the development of the score, as well as complementary approaches to assess discriminatory ability.
Milestone Achieved: Publication of a manuscript exploring the metabolome pulmonary function	*9*	95%	ongoing	Manuscript due to be submitted to the journal 'Metabolites' (IF 3.303) in the next two weeks

Table 1c: SOW and Progress in Accomplishing the Outline Goals (Specific Aim 3)

Specific Aim 3 To determine the influence of Pb exposures on respiratory health through the integration of the metabolomics profiles of Pb exposure and pulmonary function	**Timeline**	**% completion**	**Completion Date**	**Comments**
Major Task 1: Identify metabolites, metabolite networks and metabolomics pathways along the causal pathway from Pb exposure to reduced pulmonary function	Months			
Subtask 1: Identify common differential metabolites and metabolomic pathways for Pb exposure and pulmonary lung function	10	75%	ongoing	Analyses complete for those relating to whole blood Pb exposure. Pending for other Pb measures, final selection of common metabolites will depend on comparison of robustness of the Pb models. [**Figure11; Table 13**]
Subtask 2: Utilize WGCNA to identify the elements of metabolic networks that are consistent for Pb and pulmonary lung function	10-11	25%	ongoing	Have looked at consistency between WGCNA generated modules; will next explore preservation analysis
Major Task 2: Construction of a biologically informative causal pathway				
Subtask 1: Utilize the results from Major Task 1 together with mediation analyses and structural equation modelling to construct a causal pathway from Pb exposure to respiratory outcome	12-14	0%	pending	
Milestone Achieved: Publication of findings	*15*	0%	pending	
Milestone Achieved: Publication discussing the applicability of metabolomics for constructing causal pathways	*16*	0%	pending	
Major Task 2: Biomarker Development				
Subtask 1: Utilizing the results from Aims 1-3 explore the potential development of biomarkers of exposure, outcome and intermediate biomarkers along the causal pathway that may be suitable for Therapeutic target	15-18	0%	pending	
Subtask 2: Search for suitable replication population(s)	18	100%	12/01/2018	We identified the European Prospective Investigation into Cancer – Norfolk subset (EPIC_Norfolk) as a replication population
Milestone Achieved: Publication of findings	*18*	0%	pending	
Milestone Achieved: Finalize plan and secure funding for future study to develop findings further	*18*	0%	pending	

3.b. Accomplishments of these goals

There have been no changes to the overall goals of this project as stated in the **SOW** (**Table 1a-c**),with the exception of specific Aim 3; Major task 2; subtask 2. Despite extensive searches we were unable to identify suitable collaborators with Pb exposure information, pulmonary function data and plasma metabolomic profiling. Consequently, we now limit our replication to our pulmonary function findings. We identified the European Prospective Investigation into Cancer – Norfolk subset (EPIC-Norfolk), as a suitable replication population for this aim, and currently have a manuscript pending submission based on the replicated findings. *See **section 3.b.2b** and **Appendix** for more details*. The major accomplishments by task are outlined below, methods are described and preliminary results and conclusions are presented.

3.b.1 AIM ONE: Major Task one: Metabolomic profiling of NAS individuals

There was an advantageous change in the number of samples we were able to profile following price negotiations with Metabolon Inc., who were sub-contracted to perform the metabolomics profiling analyses (*see **section 5.b** for further details*). Rather than the 374 samples we stated in our proposal, we were able to profile 661 samples from 464 men, including a number from the same men over a longer period of time (**Table 2**). This substantially increased our overall power, and specifically our power to explore changes in the metabolome over time. Subject selection was carefully performed to ensure we included both a range of lung function and range of Pb exposure, and we prioritized the selection of men with multiple longitudinal blood samples for profiling.

Metabolomic profiling was performed during the first reporting period, and is described in detail in our first report. In brief relative abundance of 1301 metabolites could be quantified in the 661 plasma samples (**Table 3**). The metabolic super pathways, as defined by metabolon, covered by these profiles are described in **Figure 1**.

Table 2: Baseline characteristics of 464 Men from the Normative Ageing Study with Metabolomic of Plasma Samples

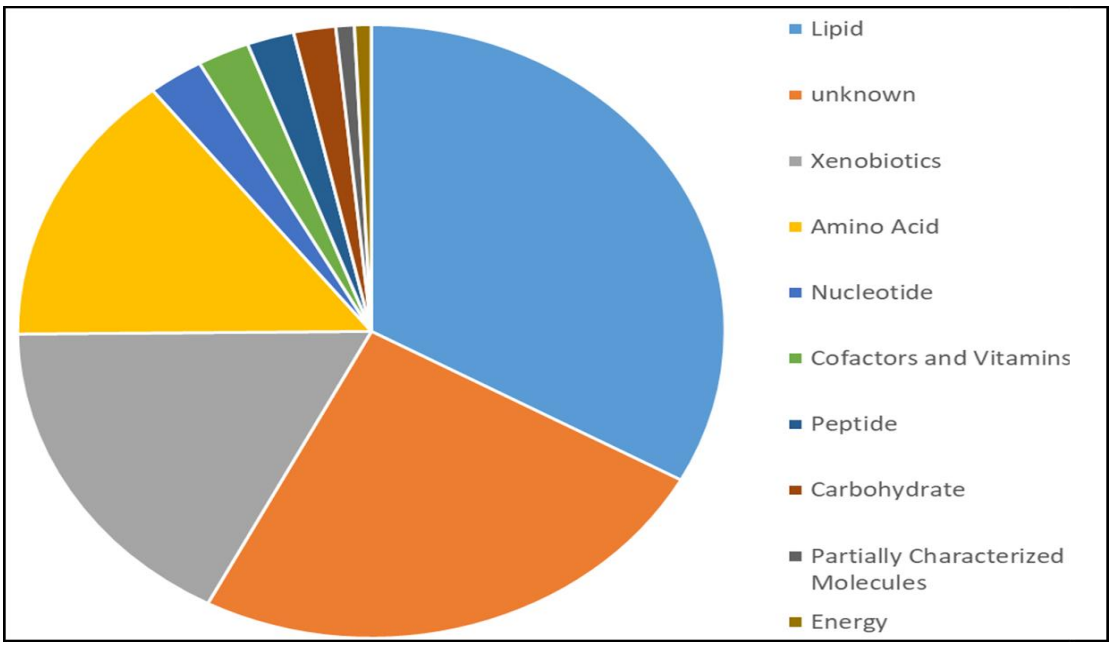
Characteristic		n=464 Men from the Normative Ageing Study
Age (yrs)	**mean [range]**	75 [57, 97]
BMI	**Underweight *n* (%)**	3 (0.7%)
	Normal *n* (%)	122 (26.3%)
	Overweight *n* (%)	249 (53.7%)
	Obese *n* (%)	90 (19.4%)
Race	**White *n* (%)**	456 (98.7%)
	Black *n* (%)	6 (1.3%)
	Other *n* (%)	2 (0.4%)
Smoking Status	**Never *n* (%)**	138 (29.9%)
	Regular Smoker *n* (%)	19 (4.1%)
	Former Smoker *n* (%)	307 (66.5%)
Asthma	**Yes *n* (%)**	15 (3.2%)
	Previous *n* (%)	16 (3.5%)
	No *n* (%)	433 (93.7%)
Forced Expiratory Volume in One Second (FEV_1, L)	**mean [range]**	3.39 [1.42, 5.85]
Forced Vital Capacity (FVC, L)	**mean [range]**	2.51 [0.80, 4.32]
FEV_1/FVC ratio	**mean [range]**	73.7% [36.1%, 92.5%]
Fresh Blood Lead (µg/mL)	**mean [range]**	3.46 [0.00, 29.00]
Second Blood Sample Available	**Yes *n* (%)**	169 (36.4%)
Third Blood Sample Available	**Yes *n* (%)**	28 (6.0%)

Table 3: Four Profiling Platforms Employed by Metabolon to Characterize 1301 Metabolites

Profiling Platform	Metabolites	n
LC/MS Negative	Metabolites that Ionize in the Negative Mode	688
LC/MS Polar	Polar Metabolites	70
LC/MS Positive Early	Metabolites that Ionize in the Positive Mode, elute early	286
LC/MS Positive Late	Metabolites that Ionize in the Positive Mode, elute late	257

LC/MS – Liquid Chromatography Mass Spectrometry

Figure 1: Metabolic Pathways Encompassed by the Metabolomic Profiles



3.b.1AIM ONE: Major Task 2 Identification of the metabolome of Pb exposure

Data analyses are complete and we are in the process of finalizing the initial manuscript. Due to the wealth of information on Pb exposure, we anticipate we will write an additional manuscript before the close of the project.

3.b.1a: Blood Pb Metabolome

The primary focus of this initial manuscript is blood Pb levels as measured by Zeeman background-corrected flameless atomic absorption and calibrated with National Institute of Standards and Technology Blood Pb Standard Reference Materials (PMID: 24905780) in the same samples used for metabolomic profiling. A total of 399 men were eligible for this analysis (**Table 4**). Pb levels measured in blood provide a good measure of short term exposure to lead. The majority of men (n=290, 72.7%) had low levels of Pb in their blood; 17 (4.3%) men had no trace of Pb in their blood. There was no difference in the baseline characteristics (age, height, weight, race or smoking status) of these 399 men by blood Pb level.

Table 4: Baseline Characteristics of 399 Men included in the Metabolome of Pb Analysis Sratified by Whole blood Pb Level at Time of Metabolomic Profiling

		Whole Blood Lead (µg/dL)				
		0	**1-4**	**5-9**	**10-30**	
Characteristic at blood collection		**n=17**	**n=290**	**n=81**	**n=11**	**p-value**
Age (yrs)	mean [SD]	77.7 [6.0]	74.4 [6.6]	77.0 [7.0]	75.5 [7.6]	*0.105*
Height (Inches)	mean [SD]	68.7 [2.2]	68.5 [2.9]	68.2 [2.3]	67.3 [2.3]	*0.117*
Weight (Lbs)	mean [SD]	186.4 [24.4]	185.7 [34.0]	177.8 [26.9]	186.1 [20.6]	*0.151*
BMI	Underweight [<18.5 kg/m2] (%)	0 (0)	3 (1)	0 (0)	0 (0)	*0.276*
	Healthy weight [18.5-25 kg/m2] (%)	2 (11.8)	79 (27.2)	23 (28.4)	1 (9.1)	
	Overweight [25-30 kg/m2] (%)	11 (64.7)	134 (46.2)	47 (58)	7 (63.6)	
	Obese [>30 kg/m2] (%)	4 (23.5)	74 (25.5)	11 (13.6)	3 (27.3)	
Race	White *n* (%)	17 (100)	286 (98.6)	78 (96.3)	10 (90.9)	*0.067*
	Black *n* (%)	0 (0)	4 (1.4)	2 (2.5)	0 (0)	
	Other *n* (%)	0 (0)	0 (0)	1 (1.2)	1 (9.1)	
Smoking Status	Never *n* (%)	10 (58.8)	88 (30.3)	22 (27.2)	2 (18.2)	*0.238*
	Regular Smoker *n* (%)	0 (0)	12 (4.1)	5 (6.2)	0 (0)	
	Former Smoker *n* (%)	7 (41.2)	190 (65.5)	54 (66.7)	9 (81.8)	

To consider the association between Pb and metabolite levels in this population, we ran 1301 models for each metabolite fit separately, adjusting for age, height, weight, smoking status (never, former current) and race (White, Black, Other). A total of 241 metabolites (18.5%) were significantly associated with blood Pb levels at a confidence interval of 95%.

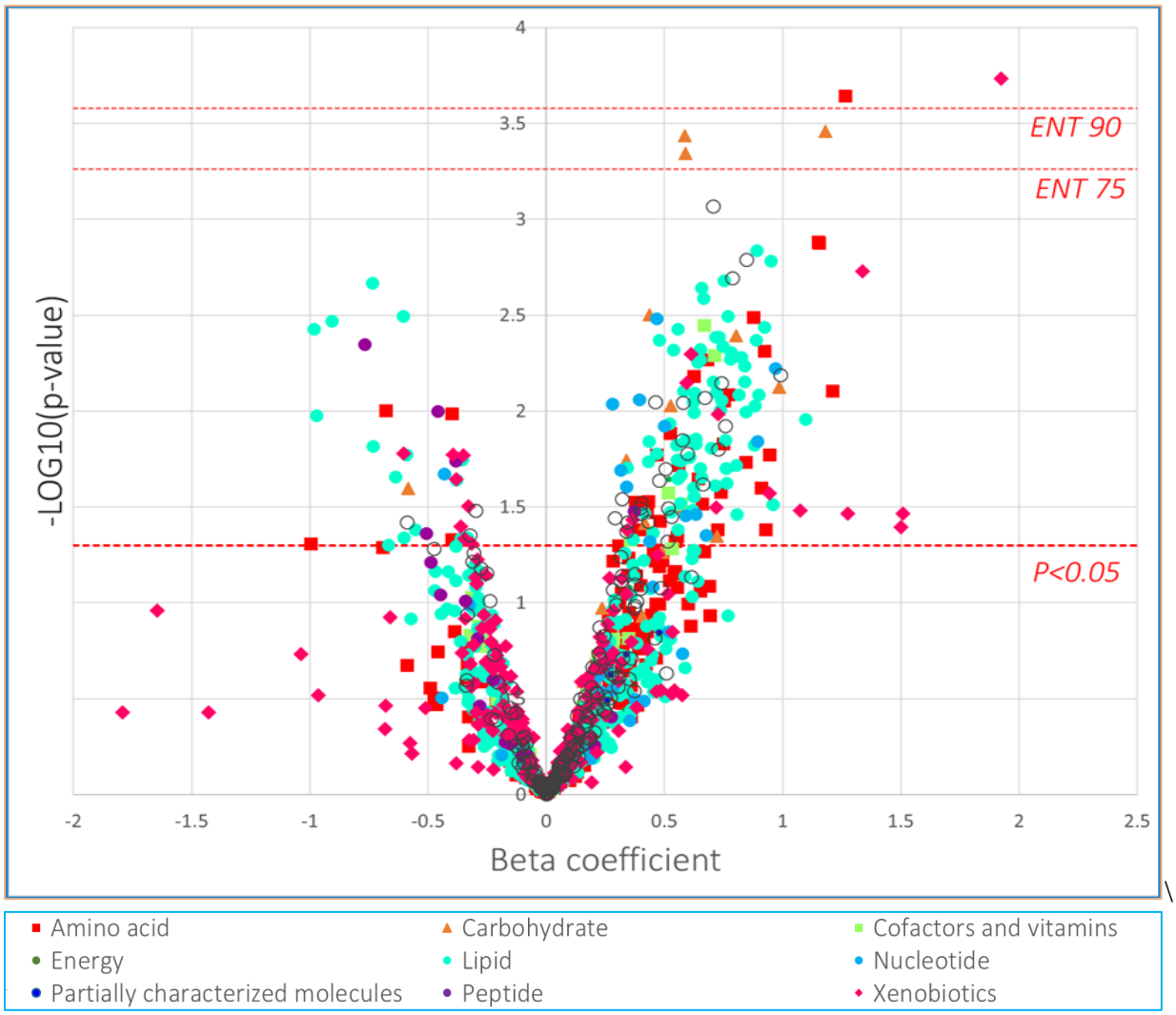
It should be noted that there are currently no consensus standards for multiple testing correction in metabolomics; methods applied to other 'omic' datatypes such as the Bonferroni correction, and even more liberal corrections, are considered too stringent for metabolomics data due to the high correlation of metabolites that are closely linked together through biological pathways. Therefore, in this work and our recent analyses of other populations (PMID: 30841573), we have pioneered the use of the 'Effective Number of Tests approach' (PMID: 22143225; PMID: 14997420), which was initially developed for genetic data; for use with metabolomic data. This method takes into account the presence of highly correlated metabolites mapping to the same biological pathway, using a principal components analysis (PCA) approach to identify the number of effective tests. We applied PCA to the 1301 metabolites and determined the number of components required to explain a given % of the variance in the data (i.e., the number of effective tests). The adjusted *p*-value threshold is then calculated as α/m where α denotes the nominal *p*-value threshold of 0.05, and m denotes the number of effective (i.e., independent) tests. For this analysis we explored a threshold of 75% and of 90% variance explained and found that five and two metabolites, respectively, retained significance (**Table 5**).

Table 5: Significance Thresholds Based on the 'Effective Number of Tests' approach to Account for Multiple Testing

Significance Threshold	% Variance Explained	Number of PCs	P-value	N (%) Significant Metabolites
α	-	-	<0.05	241 (18.5%)
ENT 75	75%	88	<5.68x10-5	2 (0.15%)
ENT 90	90%	187	<2.67x10-4	5 (0.38%)

The majority of the 241 significant associations were lipids; particularly acyl carnitines involved in fatty acid metabolism, which is known to be affected by Pb exposure (**Figure 2).** A large number of metabolites involved in amino acid metabolism were also among the significant findings. The top hits were dimethylglyine (β: 1.291, $p=1.60x10^{-4}$) and N-acetylneuraminate (β: 1.236, $p=1.69x10^{-4}$) which were at higher levels in those with higher blood Pb. The strongest inverse association was for 1-stearoyl-GPI (18:0) (β: -0.722, $p=2.66x10^{-3}$).

Figure 2: Metabolite-Blood Pb Associations According to Metabolite SuperPathway



Pathway analysis was performed with MetaboAnalyst v.4.0 (www.metaboanalyst.ca) to identify the KEGG defined (www.genome.jp/kegg/pathway.html) metabolomic pathways the 241 metabolites were enriched for. Pathway analysis extends and enhances the concept of metabolite set enrichment analysis by incorporating topology analysis. This evaluates the importance of a given metabolite based on its position within a pathway using graph theory, and therefore provides a more meaningful interpretation of list of differential metabolites. In this analysis, the hypergeometric test was specified for the over-representation analysis and relative-betweenness centrality was specified for the pathway topology analysis; all 1301 metabolites were input as the reference metabolome. The 241 significant metabolites were enriched for 15 metabolic pathways (**Table 6**). Among these 15 pathways, were a number that we hypothesized would be dysregulated with Pb exposure in our original grant application, including pyrimidine metabolism and alanine, aspartate and glutamate metabolism. In addition, sphingolipid metabolism, which plays a crucial role in multiple cellular functions as well as in lung health, was also identified. (this table also contains the results of the pathway analysis for pulmonary function, *see **section 3.b.2** for further details*).

Table 6: Metabolomic Pathway analysis of metabolites identified as being significantly dysregulated by degree of lung function and by blood Pb levels measured concurrently
A red square indicates that a pathway was significantly enriched among the metabolites identified as being significantly associated with the phenotype. It can therefore by hypothesized that this pathway is dysregulated with changes in the relevant phenotype

Metabolic Pathway	FEV1 (L)	FVC (L)	FEV1/FVC (%)	Fresh Blood Lead (µg/mL)
Alanine, aspartate and glutamate metabolism				■
Aminoacyl-tRNA biosynthesis	■	■	■	■
Arginine and proline metabolism	■	■		■
beta-Alanine metabolism				■
Caffeine metabolism			■	■
Cysteine and methionine metabolism		■		■
Galactose Metabolism	■			
Glycerophospholipid metabolism	■	■	■	■
Glycine, serine and threonine metabolism	■	■	■	■
Histidine metabolism	■	■		■
Lysine Degradation		■		
Nitrogen metabolism		■		■
Pantothenate and CoA biosynthesis				■
Purine Metabolism		■		
Pyrimidine metabolism		■		■
Sphingolipid metabolism		■	■	■
Taurine and hypotaurine metabolism		■	■	■
Valine, leucine and isoleucine biosynthesis	■	■	■	■

FEV1 – measure of how much air can be exhaled in one second following a deep inhalation
FVC - measurement of lung size (in liters) that represents the volume of air in the lungs that can be exhaled following a deep inhalation
FEV1/FVC – represents the percent of the lung size (FVC) that can be exhaled in one second

3.b.1b: Pb Metabolome in other biosamples

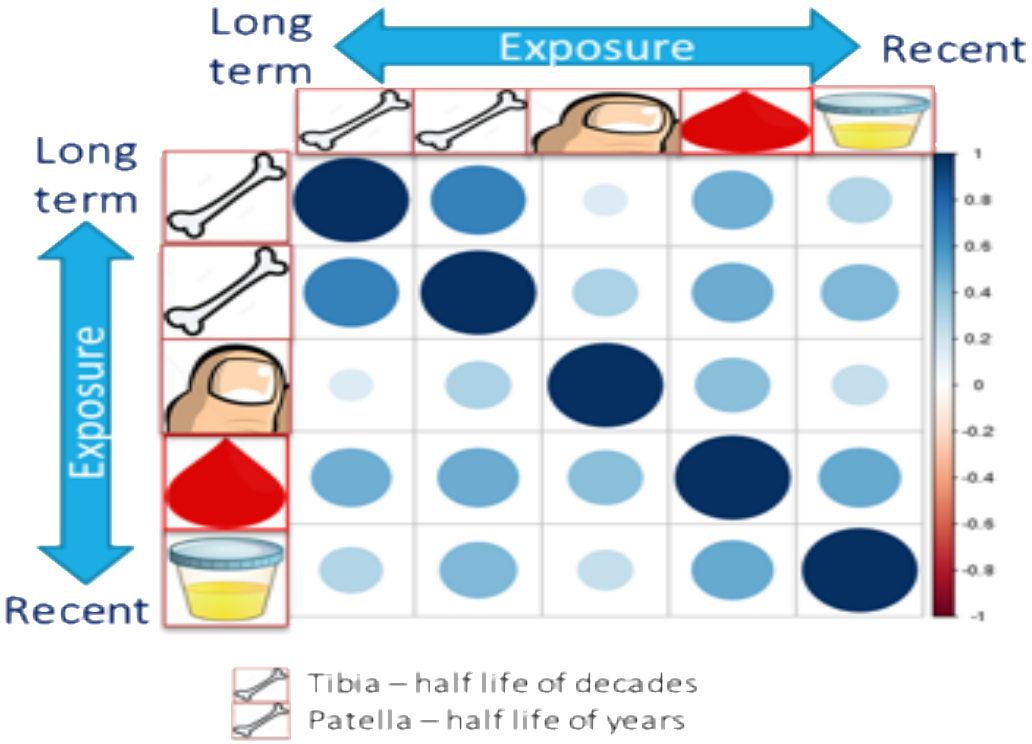
In addition to blood Pb levels, we had information on Pb in other biosamples which provide a more complete picture of long time Pb exposure (**Table 7**). (i) Long terms exposure measured in mid-tibia shaft (cortical bone: Pb half-life of many decades) and patella (primarily trabecular bone: Pb half-life of a few years) using cadmium-109 K-shell X-ray fluorescence spectroscopy (ii) Mid-term exposure measured in toenail by ICP-MS, and (iii) recent exposure measured in urine by inductively coupled plasma mass spectrometry. However, given the small available sample size the urine analyses will excluded from the final manuscript.

Table 7: Mean and Standard deviation of Pb Levels as measured in bone tonenail and urine

Measure	**Mean**	**Standard Deviation**
Tibia (µg/g)	1.51	*0.95*
Patella (µg/g)	20	*13.47*
Toenail (µg/g)	27.09	*19.37*
Urine (mg/dL)	0.67	*1.09*

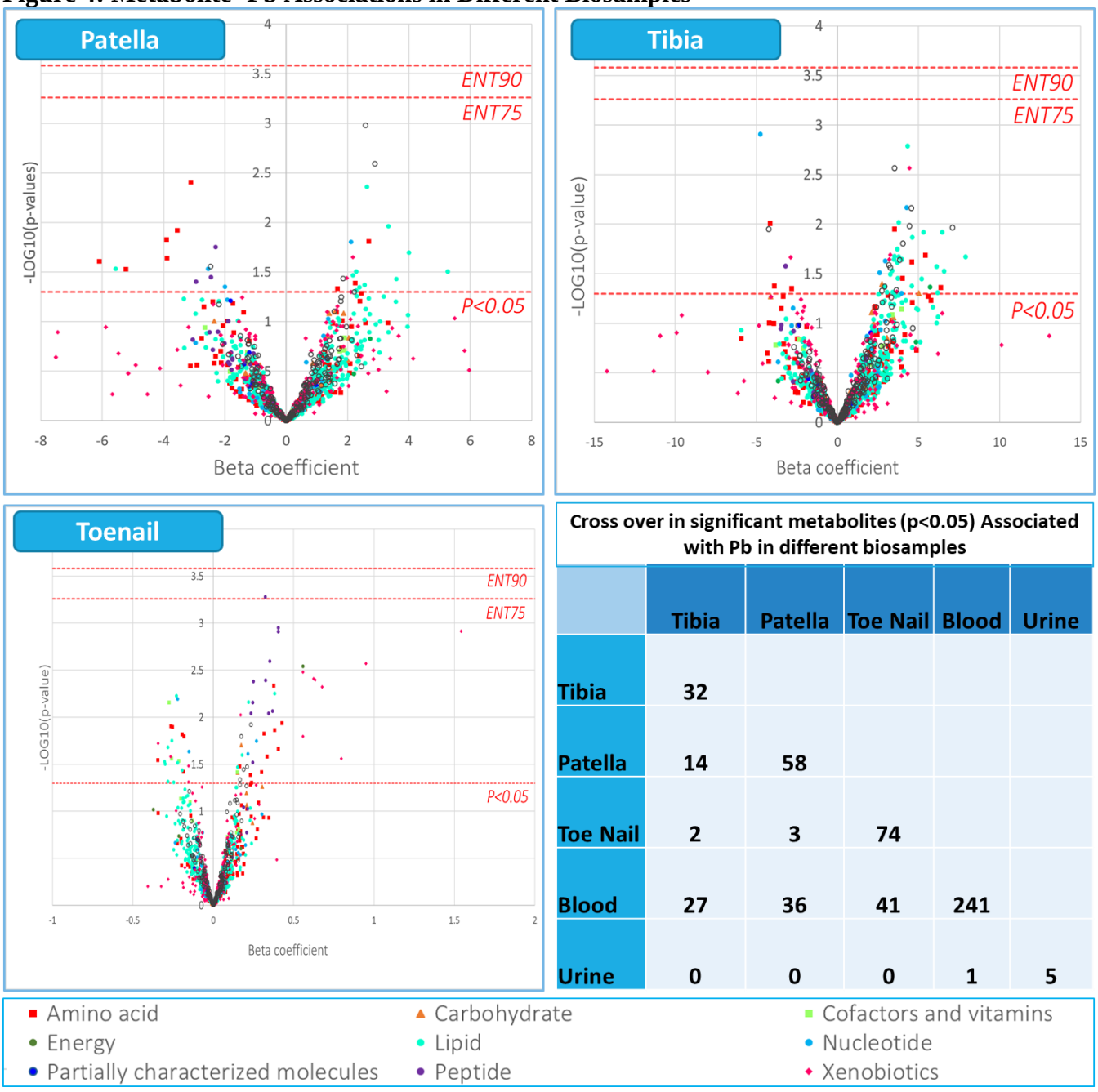
There was evidence of correlation between temporal measures of Pb (**Figure 3**). Tibia and Patella levels were highly correlated and blood Pb levels were correlated with these longer term exposures, indicating a consistency in an individual's exposure to Pb over time.

Figure 3: Correlation between temporal measures of Pb exposure



Accordingly, both long term and more short term measures of Pb exposure primarily affect metabolites involves in lipid and amino acid metabolism; and a number of metabolites, including acylcarnitines, were associated with multiple measures of lead exposure (**Figure 4**).

Figure 4: Metabolite- Pb Associations in Different Biosamples



Cross over in significant metabolites (p<0.05) Associated with Pb in different biosamples

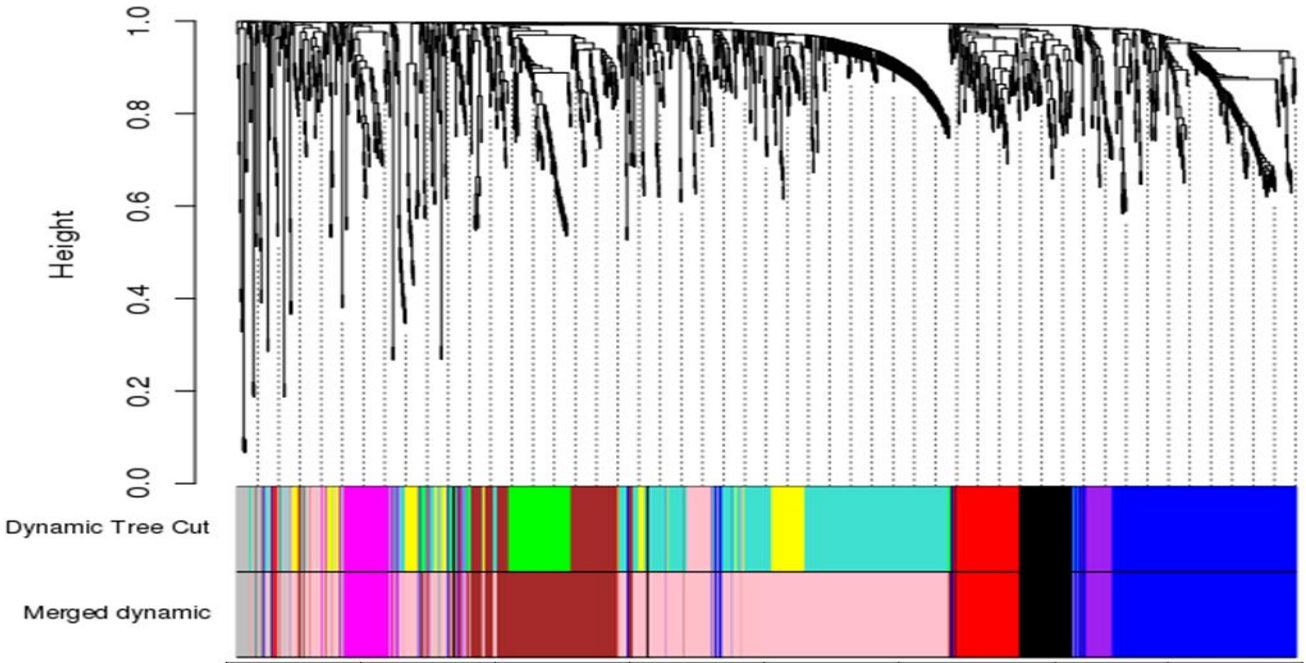
	Tibia	Patella	Toe Nail	Blood	Urine
Tibia	32				
Patella	14	58			
Toe Nail	2	3	74		
Blood	27	36	41	241	
Urine	0	0	0	1	5

We additionally employed a network approach to identify metabolic networks, rather than single metabolites associated with Pb exposure (subtask 3). Network approaches move away from reductionist methodologies to combine systems biology and network science, providing a holistic

methodology to better understand biology through the identification and investigation of non-linear relationships and networks of interacting components. Weighted Gene Correlation Network Analysis (WGCNA, horvath.genetics.ucla.edu/html/CoexpressionNetwork/Rpackages/WGCNA/index) is a network method for identifying clusters or 'modules' or highly correlated variables (such as metabolites) that are likely to be co-regulated, or working together in biologically coherent fashion. A module can then be summarized as a single unit, which can be correlated with phenotypes of interest. WGCNA was used to identify metabolomic network modules within the baseline samples based on correlation patterns. The correlation matrix quantifies interconnectedness between metabolites and assigns them to co-expression modules. Features that did not show high enough co-expression metrics with any module were excluded from further analysis (they are assigned to a redundant grey module). Highly correlated modules were then merged using a cut height (i.e. the *Euclidean* distance between clusters) of 0.6; chosen using an iterative process to identify an optimal number of adequately sized modules for analysis. Modules were summarized by an eigenvector (based on the first principal component of each module) for each participant, then associations between the modules and Pb exposures were explored.

The module assignments for the 1301 metabolites as measured in 399 samples included in the Pb analysis are shown in **Figure 5.** After merging the highly correlated clusters there were eight modules, which are assigned colors by the R package by default.

Figure 5: Cluster Dendrogram showing assignment of 1301 metabolites to eight merged dynamic modules

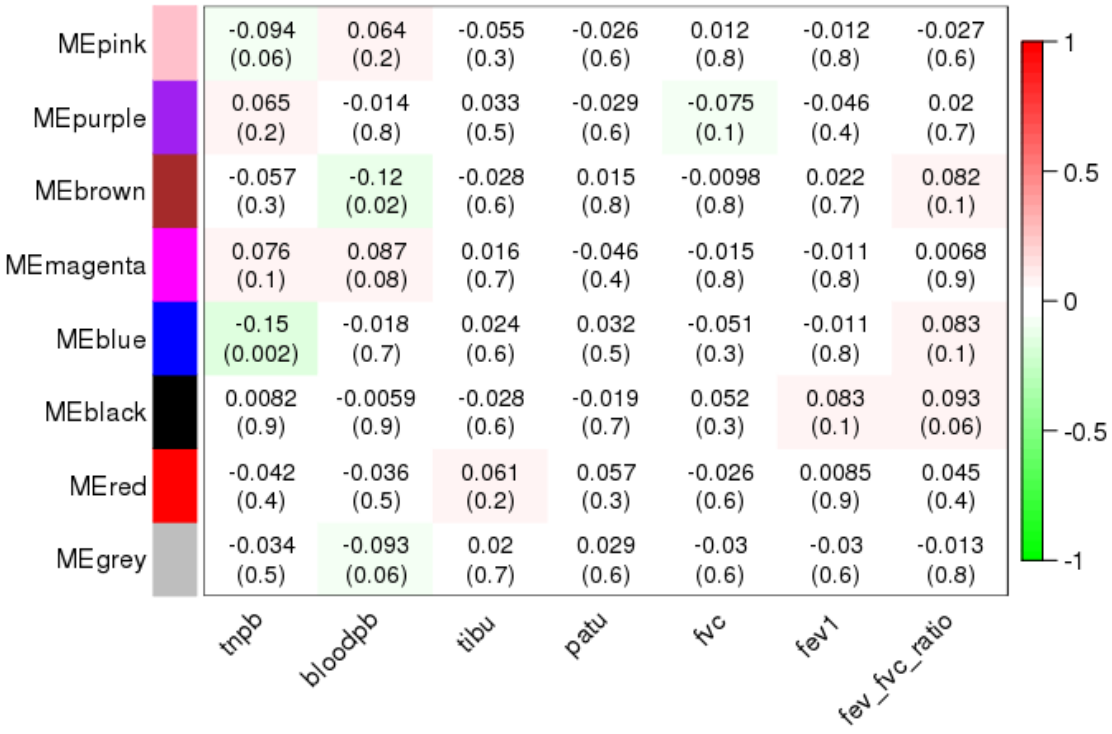


pink	blue	brown	red	black	magenta	grey	purple
509	261	210	87	71	62	56	39

NB. The grey module includes all those metabolites that did not show high enough correlation with any other metabolite to be included in a module. This grey module is redundant and is excluded from the analysis

Figure 6 shows the relationship between the modules and the variables of interest. There is little evidence that the metabolite modules were associated with bone measures of Pb. However, there were two modules of interest ($p<0.01$) for blood Pb; the brown and the magenta, which was also showed some evidence of correlation with toenail Pb. The blue module was also highly significantly correlated with Pb exposure as assessed in toenails. For the modules sigof interest, the 'hubs' were identified. Hubs are the features that are most highly connected within a module, and therefore drive module formation. WGCNA computes a module-membership value and associated p-value for each feature within a module, which is a measure of how connected or co-expressed that feature is with others within the same module. Features with a module-membership p-value that retained significance after Bonferroni-correction were considered to be hubs. For the brown, magenta and bluemodules there were 156, 53 and 217 distinct hub metabolites, respectively. The brown module was driven by long chain fatty acids and polyunsaturated fatty acids; the magenta module by steroids, specifically androgenic and pregnenolone steroids and the blue module by amino acids and nucelotides.

Figure 6: Heatmap describing the relationships between the module eigengenes and the traits of interest



Each cell shows the correlation coefficient and (p-value)

All analyses were conducted in R version 3.5.0 and all statistical tests were two-sided.

3.b.2 AIM TWO: Major Task 1: To identify a metabolomic profile of pulmonary function within a population of Veterans

3.b.2a Results from the Normative Aging Study

Lung function was assessed according to spirometry which is performed by deeply inhaling and forcefully exhaling into a spirometer. We used the three most commonly employed measures of spirometry; (i) forced vital capacity (**FVC**);a measurement of lung size (in liters) that represents the volume of air in the lungs that can be exhaled following a deep inhalation; (ii) forced expiratory volume-one second (**FEV_1);** a measure of how much air can be exhaled in one second following a deep inhalation; and (iii) **FEV_1/ FVC** ratio which represents the percent of the lung size (FVC) that can be exhaled in one second. As we also had multiple longitudinal spirometric measures from the same men, we employed a mixed model including race, smoking status, and BMI as fixed effects (subtask 1). A total of 464 men (**Table 2**) were included in these analyses. These results were reported in detail in the progress report from our first reporting period and are briefly summarized here. There were 368 (28.3%) metabolites associated with FEV_1, 189 (14.5%) associated with FVC and 425 (32.7%) associated with FEV1/FVC ratio at a p value<0.05 (**Table 8**). For both FEV_1 and FEV_1/FVC ratio, a large proportion of these metabolites were robust to correction for multiple testing. There was substantial consistency in the metabolites identified in the significant end-points as would be expected given the close relationships between these outcomes (*more information on the top hits of most interest for these outcomes is provided in the AIM THREE section*). The significantly enriched pathways, which can be hypothesized to be dysregulated with changes in FEV_1, FVC and FEV_1/FVC ratio are shown in **Table 6**. Again, similar pathways were identified for the three outcomes; Aminoacyl-tRNA biosynthesis; glycerophospholipid metabolism; Glycine, Serine and Threonine metabolism; Histidine metabolism; and Valine, leucine and isoleucine biosynthesis were all significant across the outcomes, suggesting dysregulation in these pathways may be a cause or an effect of decreased lung function.

Table 8: Metabolites Associated with Three Measures of Lung Function in a Population of Veterans

Significance Threshold	**FEV1 (L)**		**FVC (L)**		**FEV1/FVC (%)**	
	n	**%**	**n**	**%**	**n**	**%**
p<0.05	368	28.3%	189	14.5%	425	32.7%
FDR corrected p<0.05	182	14.0%	4	0.3%	208	16.0%

As described in our previous report, we further generated a metabolomic score that can be used to discriminate men by their degree of lung function (subtask 2), as defined by FEV_1/FVC ratio. In a clinical setting FVC, FEV_1 and FEV_1/FVC ratio are compared to reference values based on healthy individuals with normal lung function, to determine the degree of lung function of the patient. The normal value for the FEV_1/FVC ratio is 70% or above, with a lower measured value corresponding to a more severe lung abnormality. In this population, 93 (20%) men had a ratio ≤70%, while 371 (80%) had a ratio >70%.We compared two different models for the discrimination of an FEV_1/FVC ratio above and below 70% using receiver operator characteristic (ROC) cures and the corresponding area under the curves (AUC); Model 1: a summary score, generated by taking the first principal component of the 189 plasma metabolites that were significantly associated with FEV_1/FVC ratio; and Model 2: Levels of the four metabolites that

were robust to FDR correction. The sensitivity and specificity were computed based on the optimal cut-off to maximize sensitivity and specificity weighting both equally, as determined using the 'ROCR' package in R. The summary score Model 1 had moderate discriminatory ability in a Receiver Operator Characteristic (ROC) curve analysis (AUC: 0.602 (95% CI 0.552, 0.652) while the Model 2 based on four metabolites demonstrated a marginal, and non-significant, improvement in the AUC (0.631 (95% CI 0.581, 0.680). Similarly the sensitivity was identical for the two models (55.3%) but the specificity was slightly higher for model 2 (**Table 9**). Although encouraging, these findings demonstrate the need to develop alternative methods of generating metabolite biomarkers of lung function, which we will continue to focus on during the next and final reporting period.

Table 9: AUCs, Sensitivity and Specificity for two models for the prediction of an FEV_1/FVC ratio below 70%

Classifier	AUC (95% CI)	Performance compared to Model 1	Sensitivity	Specificity
Model 1: Metabolite Summary Score	0.602 (0.552, 0.652)		55.3%	62.0%
Model 2: Metabolite levels	0.631 (0.581, 0.680)	*p=0.312*	55.3%	62.8%

We utilized the same metabolites modules generated using an unsupervised approach in *AIM ONE: Major Task 2; subtask 3*, (**Figure 5**) and determined the relationship between these metabolite modules and the three measures of lung function (**Figure 6**). However, few strong associations were observed and further work is required to explore these findings. (*Further details on the network results are presented in the AIM three section*).

3.b.2b Manuscript: "Metabolomics of Lung Function: Evidence from two cohorts"

Through dissemination of our initial findings to the wider metabolomics community, we identified a suitable replication cohort for our *AIM TWO* analysis; The European Prospective Investigation into Cancer- Norfolk Population (EPIC-Norfolk) led by Claudia Langenberg of the University of Cambridge. EPIC Norfolk includes 10,460 participants (4,868 males [46.5%] and 5,592 females [53.5%]) with a mean age of 59.7 years [SD 9.0] who had both metabolomics profiling and spirometric measures of FEV_1 and FEV_1/FVC ratio. Replication of key findings remains a bottle-neck in metabolomics research, and therefore the ability to validate our results in a comparable population represents a significant strength of these analyses. Given the larger sample size of EPIC-Norfolk, it was determined that this would act as the discovery population, and the NAS as the replication population. These results have now been finalized in a manuscript that is due to be submitted to the journal Metabolites (Impact Factor: 3.3).

The final version of this manuscript, which includes extensive details on study populations, rational, methods, results and a discussion of these findings is attached to this report (***Appendix 1***). The main results are summarized in brief here.

To be entirely concordant with the analyses run in EPIC-Norfolk, a subset of men from the NAS (n=439) were used for these analyses, analyses were based on a single time point for each man, and only FEV_1 and FEV_1/FVC ratio were considered as endpoints. Therefore there are some small differences in the NAS results included in the final manuscript as compared to those reported in *section **3.b.2a.***

A total of 698 metabolites were measured in both the EPIC-Norfolk and the NAS populations**.** Ten FEV_1 associated metabolites were associated with FEV1 in EPIC Norfolk and replicated in the NAS population, with consistent directions of effect and a replication p-value threshold of <0.05 (**Table 10**). These were primarily amino acids and lipids, including several omega-3 fatty acids: DHA 22:6n3, DPA 22:5n3, and EPA 20:5n3.

Table 10. Replicated FEV_1 metabolites

METABOLITE	PATHWAY	SUBPATHWAY	EPIC-NORFOLK (N=1002)				NAS (N=1140)			
			Estimated effect (β)	*p*	Standard error	Confidence interval (lower, upper)	Estimated effect (β)	*p*	Standard error	Confidence interval (lower, upper)
3-HYDROXYISOBUTYRATE	Amino Acid	Leucine, Isoleucine and Valine Metabolism	1.85	$4.57x10^{-05}$	0.46	(0.96, 2.74)	0.15	0.017	0.06	(0.03, 0.27)
3-METHYL-2-OXOBUTYRATE	Amino Acid	Leucine, Isoleucine and Valine Metabolism	1.28	0.005	0.46	(0.39, 2.18)	0.16	0.014	0.07	(0.03, 0.29)
4-METHYL-2-OXOPENTANOATE*	Amino Acid	Leucine, Isoleucine and Valine Metabolism	2.35	$7.61x10^{-07}$	0.47	(1.42, 3.28)	0.14	0.034	0.07	(0.01, 0.27)
DHA 22:6N3*	Lipid	Polyunsaturated Fatty Acid (n3 and n6)	2.85	$5.37x10^{-10}$	0.46	(1.95, 3.75)	0.11	0.017	0.05	(0.02, 0.20)
DPA 22:5N3*	Lipid	Polyunsaturated Fatty Acid (n3 and n6)	2.06	$4.84x10^{-06}$	0.45	1.18, 2.95)	0.11	0.026	0.05	(0.01, 0.20)
EPA 20:5N3	Lipid	Polyunsaturated Fatty Acid (n3 and n6)	1.82	$5.80x10^{-05}$	0.45	(0.93, 2.70)	0.09	0.038	0.04	(0.01, 0.18)
GLUCOSE	Lipid	Polyunsaturated Fatty Acid (n3 and n6)	-1.28	$4.40x10^{-03}$	0.45	(-2.17, -0.39)	-0.08	0.034	0.04	(-0.15, -0.01)
PICOLINATE	Lipid	Polyunsaturated Fatty Acid (n3 and n6)	1.32	0.041	0.65	(0.06, 2.58)	0.08	0.024	0.04	(0.01, 0.16)
STACHYDRINE*	Xenobiotics	Food Component/ Plant	2.65	$5.13x10^{-09}$	0.45	(1.76, 3.54)	0.07	0.043	0.03	($2x10^{-3}$, 0.13)

*Indicates significance retained after a Bonferroni correction on EPIC-Norfolk data.

Four FEV_1/FVC associated metabolites identified in EPIC-Norfolk replicated in NAS (**Table 11**), including two amino acids, one carbohydrate, and one xenobiotic.

Table 11. Replicated FEV_1/FVC metabolites

METABOLITE	PATHWAY	SUBPATHWAY	EPIC-NORFOLK (N=1002)				NAS (N=1140)			
			Estimated effect (β)	*p*	Standard error	Confidence interval (lower, upper)	Estimated effect (β)	*p*	Standard error	Confidence interval (lower, upper)
2-PIPERIDINONE	Xenobiotic	Food Component/ Plant	-0.003	$9.96x10^{-04}$	$1.1x10^{-03}$	($-5.6x10^{-3}$, $1.4x10^{-3}$)	0.01	0.039	$4.8x10^{-3}$	($5x10^{-3}$, 0.02)
CYSTEINE-GLUTATHIONE DISULFIDE*	Amino Acid	Glutathione Metabolism	-0.008	$1.07x10^{-12}$	$1.1x10^{-03}$	($-9.8x10^{-3}$, $-5.6x10^{-3}$)	-0.01	0.017	$4.9x10^{-3}$	(-0.02, $-2x10^{-3}$)
THREONINE	Amino acid	Glycine, Serine and Threonine Metabolism	-0.003	0.009	$1.1x10^{-03}$	($-4.9x10^{-3}$, $-7.2x10^{-4}$)	-0.03	0.012	$1.2x10^{-2}$	(-0.05, $-7x10^{-3}$)
XYLOSE	Carbohydrate	Pentose Metabolism	-0.006	$5.28x10^{-04}$	$1.6x10^{-03}$	($-8.8x10^{-3}$, $-2.4x10^{-3}$)	-0.01	0.037	$5.8x10^{-3}$	(-0.02, $-7x10^{-4}$)

*Indicates significance retained after a Bonferroni correction on EPIC-Norfolk data.

As NAS is a male-only population while EPIC-Norfolk includes both males and females, we additionally ran the analysis in EPIC-Norfolk restricted to males only, and then attempted to replicate these findings in NAS. Seven metabolites replicated FEV_1, including two that were not among the significant hits in the total population analyses: 2-hydroxystearate and dihomo-linoleate 20:2n6. While, three metabolites replicated between the EPIC-Norfolk male only analysis and NAS (**Table 12**) for FEV_1/FVC, cysteine-glutathione disulfide, threonine, and xylose. All of which were also significant in the EPIC-Norfolk total population.

Table 12. Replicated FEV_1 metabolites based on males only

METABOLITE	PATHWAY	SUBPATHWAY	EPIC-NORFOLK MALES (N=994)				NAS (N=1140)			
			Estimated effect (β)	*p*	Standard error	Confidence interval (lower, upper)	Estimated effect (β)	*p*	Standard error	Confidence interval (lower, upper)
2-HYDROXYSTEARATE	Lipid	Fatty Acid, Monohydroxy	0.02	0.028	$7.6x10^{-3}$	($1.8x10^{-3}$, 0.03)	0.16	0.026	$7.3x10^{-2}$	(0.02,0.31)
3-HYDROXYISOBUTYRATE	Amino Acid	Leucine, Isoleucine and Valine Metabolism	0.02	0.029	$7.8x10^{-3}$	($1.8x10^{-3}$, 0.03)	0.15	0.017	$6.1x10^{-2}$	(0.03, 0.27)
4-METHYL-2-OXOPENTANOATE	Amino Acid	Leucine, Isoleucine and Valine Metabolism	0.03	0.001	$8.3x10^{-3}$	(0.01, 0.04)	0.14	0.034	$6.6x10^{-2}$	(0.01, 0.27)
DIHOMO-LINOLEATE (20:2N6)	Lipid	Polyunsaturated Fatty Acid (n3 and n6)	0.02	0.035	$8.0x10^{-3}$	(-0.01, 0.01)	0.11	0.037	$5.3x10^{-2}$	($6.9x10^{-3}$, 0.22)
DHA 22:6N3*	Lipid	Polyunsaturated Fatty Acid (n3 and n6)	0.03	$1.57x10^{-05}$	$7.8x10_{-3}$	(0.02, 0.05)	0.11	0.017	$4.8x10^{-2}$	(0.02, 0.20)
DPA 22:5N3	Lipid	Polyunsaturated Fatty Acid (n3 and n6)	0.03	$1.84x10^{-04}$	$8.0x10^{-3}$	(0.01, 0.05)	0.11	0.026	$4.8x10^{-2}$	(0.01, 0.20)
STACHYDRINE	Xenobiotics	Food Component/Plant	0.03	$1.53x10^{-04}$	$7.6x10^{-3}$	(0.01, 0.04)	0.07	0.044	$3.3x10^{-2}$	($2x10^{-3}$, 0.13)

*Indicates significance retained after a Bonferroni correction on EPIC-Norfolk data.

In this study, we identified and replicated blood metabolites that were associated with reduced lung function. Our replicated hits were characterized by omega 3 fatty acids, which have a biologically plausible relationship with respiratory function, as well as several amino acids.

There was strong biological rationale for many of our hits. For example we validated the positive association between levels of 3-hydroxyisobutyrate and improved FEV_1 function in our two independent cohorts, which is in agreement with previous work showing that levels of this metabolite are lower among those with emphysema and COPD, as compared to controls. Interestingly, we also saw common pathways, such as glycine, serine and threonine metabolism, between lung function in this study and previously reported studies of COPD which is characterized by a decreased FEV_1. We observed a positive relationship between stachydrine, produced by the intake of foods and juices high in citrus, and $FEV_{1,}$ which is supportive of previous work linking higher plasma levels of stachydrine to a decreased risk of asthma. There is also strong biological rationale for the role of cysteine-glutathione, for which we validated a negative relationship with FEV_1/FVC in the total and male-only analyses. Cysteine-glutathione is

formed upon the oxidative stress of glutathione and it has long been suspected that disturbances in oxidation/reduction (redox) reactions are a risk factor for respiratory disease. Airway oxidative stress is linked to worsened disease severity, reduced lung function, and epigenetic changes that decrease airway responsiveness to steroids.

This study, made possible by the support of this grant is first to identify metabolites significantly associated with FEV_1 lung function parameters, in an attempt to better understand the metabolome of respiratory health and to identify potential biomarkers that can be used for pulmonary care.

AIM THREE: Major Task 1: Identify metabolites, metabolite networks and metabolomics pathways along the causal pathway from Pb exposure to reduced pulmonary function

This Aim is ongoing. Using the findings from Aims 1 and 2, we compared the individual metabolites and metabolite profiles (based on the WGCNA generated modules) that associated with both Pb exposure and lung function, and which may therefore lie along the causal pathway. **Figure 11**, shows the crossover between the metabolites identified as significant for FEV_1, FVC and blood Pb exposure. In total, 120 metabolites were significantly associated with at least one measure of lung function and with Pb; 72 metabolites were associated with all three indices. Some of the most biologically interesting metabolites for FEV_1 among these 72 are shown in **Table 13**.

Figure 11: Venn diagram showing the crossover between significant metabolites for two measures of lung function and for blood measures of Pb

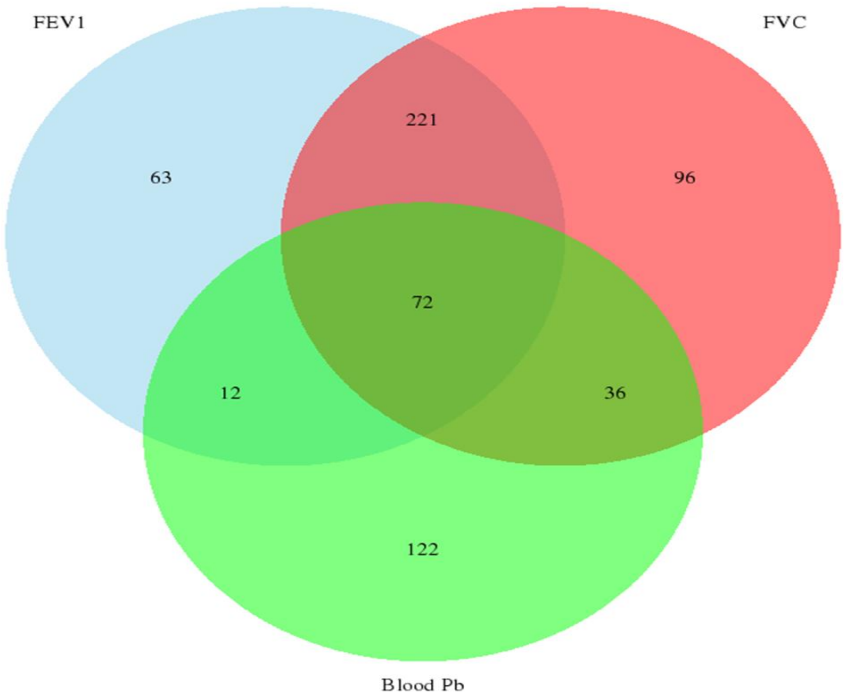


Table 13: Selection of the top metabolite hits associated with FEV1, FVC and with blood Pb

METABOLITE	SUPER PATHWAY	SUB PATHWAY	FEV1 (L)		Fresh Blood Lead (µg/mL)	
			β	p-value	β	p-value
hydroxyasparagine**	Amino Acid	Alanine and Aspartate Metabolism	-0.25	$2.5x10^{-4}$	1.58	$3.8x10^{-5}$
N-acetylalanine	Amino Acid	Alanine and Aspartate Metabolism	-0.28	0.001	1.74	$1.3x10^{-4}$
N-acetylserine	Amino Acid	Glycine, Serine and Threonine Metabolism	-0.28	$7.9x10^{-5}$	1.38	$3.4x10^{-4}$
serine	Amino Acid	Glycine, Serine and Threonine Metabolism	0.23	0.005	-1.12	0.010
N-acetylthreonine	Amino Acid	Glycine, Serine and Threonine Metabolism	-0.18	0.007	0.93	0.014
histidine	Amino Acid	Histidine Metabolism	0.23	0.014	-1.63	0.001
1-methylhistidine	Amino Acid	Histidine Metabolism	-0.12	0.031	0.62	0.034
leucine	Amino Acid	Leucine, Isoleucine and Valine Metabolism	0.31	0.002	-1.61	0.001
alpha-hydroxyisocaproate	Amino Acid	Leucine, Isoleucine and Valine Metabolism	0.17	$8.5x10^{-5}$	-0.61	0.008
choline phosphate	Lipid	Phospholipid Metabolism	0.11	0.019	-0.93	$1.3x10^{-4}$
phosphoethanolamine	Lipid	Phospholipid Metabolism	0.11	0.007	-0.76	$2.5x10^{-4}$
sphingadienine	Lipid	Sphingolipid Synthesis	0.08	0.005	-0.48	0.002
sphinganine	Lipid	Sphingolipid Synthesis	0.09	0.039	-0.66	0.002
sphinganine-1-phosphate	Lipid	Sphingolipid Synthesis	0.14	0.005	-0.66	0.008
sphingosine	Lipid	Sphingosines	0.09	0.020	-0.61	0.004
sphingosine 1-phosphate	Lipid	Sphingosines	0.16	0.003	-0.66	0.020

The identified metabolites were primarily amino acids and lipids, and in keeping with the results of the pathway analysis show in **Table 5**, they were specifically involved in the pathways of Alanine and Aspartate Metabolism; Glycine, Serine and Threonine Metabolism; Histidine Metabolism; Leucine, Isoleucine and Valine Metabolism; Phospholipid Metabolism and Sphingolipid synthesis. Interestingly, it was uniformly observed that metabolites that were inversely associated with lung function – i.e. those which were increased with decreased lung function, were positively associated with blood Pb. In others words, these results supported our hypothesis that increased blood levels lead to decreased lung function, and that metabolites may mediate this association and help us to underlying that biological mechanisms underlying it. These are novel findings in a human population, which are supported by experimental evidence in the literature.

To robustly test such a hypothesis requires more formal mediation and structural equation analyses, as outlined in the SOW, *AIM THREE; Major task two*. *AIM THREE; Major task two* and *Major Task three* are in currently in development. We anticipate this work will result in an additional manuscript. Given these methods are relevant for multiple research questions beyond the scope of this project, we will continue to work on this data in the next reporting period and beyond utilizing other funding sources.

All analyses were conducted in R version 3.5.0 and all statistical tests were two-sided.

3.c. Opportunities for training and professional Development Provided by the Project

Throughout this second reporting period, Dr. Lasky-Su continued to closely mentor Dr. Kelly. This grant has also been instrumental in providing Dr. Kelly with the opportunity, time and appropriate dataset to mentor early-stage investigators. Specifically, Dr. Kelly acted as the primary mentor for Ms. Haley Bayne, a summer intern who is an undergraduate student at Northeastern University, who under the supervision and guidance of Dr. Kelly performed the statistical analysis for the 'metabolome of pulmonary function (*AIM TWO*) manuscript. Ms. Bayne was interning at the Channing though the Northeastern Cooperative Education Program (https://careers.northeastern.edu/cooperative-education/); and provided her with the opportunity to experience a real life research environment, foster her analytical, statistical and coding skills and to present her work at an International Conference (*see* ***section 3.d***). Ms. Bayne has since decided to pursue a career in academic scientific research.

Furthermore, this grant has helped to propel Dr. Kelly from a postdoc to a junior faculty member in the Channing Division of Network Medicine, as she utilized the grant writing and project management skills developed throughout this period to be successfully awarded a K01 training grant (K01 HL146980) also focused on metabolomics and respiratory disease.

Finally, the dissemination of these results, specifically those related to the metabolome of Pb" led to Dr. Kelly being invited to be on the Editorial Board of a special issue of the journal Frontiers in Public Health titled "Metabolomics and Exposome".

3.d. Dissemination of results to communities of interes

Multiple presentations relating to this proposal, including theoretical study design and preliminary findings have been presented to colleagues in the fields of metabolomics and respiratory health within the Channing Division of Network Medicine. These findings have also been presented to the wider National and International Community as follows:

"Metabolomics of Lead Exposure and Its Role in Respiratory Disease among Veterans"
Research, Innovation, and Scholarship Expo (RISE) 2019;
Boston MA, April 4th 2019
Presenter: Haley Bayne

"Metabolomics of Lead Exposure and Its Role in Reduced Lung Function Among Veterans"
Advances in Clinical Lung Research
Boston MA, May 8th 2019
Presenter: Rachel Kelly

"The Metabolomics of Lead Exposure and Its Role in Respiratory Disease"
American Thoracic Society 2019 International Conference
Dallas TX, May 17-22nd 2019
Presenter: Rachel Kelly

"Metabolomics of Poor Respiratory Health"
15th Annual Conference of the Metabolomics Society
The Hauge, The Netherlands, June 23-27th 2019
Presenter: Haley Bayne

Furthermore, results from this project will form part of Dr. Lasky-Su's invited plenary talk at the first ever Metabolomics Association of North America Conference, in Atlanta in November 2019. Dependent on findings we also anticipate submitting updated and novel results to the American Thoracic Society and International Metabolomics Societies 2020 Annual conferences.

3.e. Plans for the next reporting period to accomplish these goals

During the next reporting period, we plan to finalize the publication of our first manuscript "**Metabolomics of Lung Function: Evidence from two cohorts",** which is due to be submitted to the journal Metabolites (Impact Factor 3.3); and our second manuscript **"The metabolome of Pb exposure"** which we intend to submit to an environmental health themed journal. As noted in **section 3.d** we also aim to present our novel findings at pertinent conferences in 2020.

Concurrently, we aim to complete the remaining goals and their subtasks, culminating in further publications (**Table1c**) and dissemination of our findings to the wider community. As part of this effort we will continue to search for collaborators with suitable replication populations particularly for the Pb analyses. Furthermore, throughout this award period we have been furthering our involvement in the International Metabolomics community; Dr. Jessica Lasky-Su is chair of the Consortium of Metabolomics Studies (COMETS) committee and a board member of the Metabolomics Society. Dr. Kelly is a steering committee member of COMETS, an executive committee member of the ATS Genetic and Genomic Section, representing metabolomics, and together they are co-leading the first international metabolomics meta-analysis within COMETS with a focus on body mass index. We will continue to work with the

metabolomics community to develop novel methodologies for the analysis of metabolomics data, to ensure can successfully complete our aims and drive forward the field.

4. Impact

4.a. Impact on the development of the principal discipline of the project

The principal discipline of this project is the field of metabolomics. The overarching goal is to construct a casual pathway between Pb exposure and poor respiratory health in a cohort of veterans, with the hypothesis that this pathway is mediated through measurable metabolomic pathways. To do so requires the development of novel analytical and statistical techniques with a focus on network methodology. We have, and will continue to work closely with our bioinformatics and network scientist colleagues to develop the appropriate methodologies. These techniques will then be applicable to a wide range of projects in the field of metabolomics, where the construction of causal pathways mediated by metabolites is a key, but as yet unachieved, goal for many.

4.b. Impact on other disciplines

In *Specific aim 1* (**Table 1a**), we propose to identify the metabolome of past Pb exposure that takes both duration and intensity into account. Our initial findings have been very promising in this area and, as such, support the utility of metabolomics in the disciplines of exposure science and exposure biomarker development. Exploring respiratory disease through metabolomics is more established, however these results also add to the discipline of respiratory health, though (i) the addition of novel literature in an underrepresented population, and (ii) supportive mechanistic evidence for a link between Pb exposure and poor respiratory health.

As described in detail in section **6.e.,** throughout the reporting period we have also built up new collaborations with existing NAS researchers, who are keen to utilize the generated metabolomics dataset to explore research questions within their own disciplines, namely healthy ageing, the health impacts of air pollution and the biology of cognitive function.

4.c. Impact on technology transfer

Nothing to report

4.d. Impact on society beyond science and technology

To date, we have demonstrated evidence to support our hypothesis of a link between Pb exposure and respiratory health, mediated through dysregulated metabolism, as outlined in our proposal, the results of this study will be of great importance to the growing population of who were exposed to Pb during active service in the Gulf, Iraq and Afghanistan, as well as to the wider US population in whom Pb exposure still widespread. The findings will support **(i) Prevention:** by confirming that Pb exposure causes reduced lung function, thereby supporting improved regulations and safeguards pertaining to exposure in the military **(ii) Early intervention** through the development of markers in the blood that can be used to identify the Pb exposure at the critical level that can damage respiratory health **(iii) Treatment of Pb induced pulmonary dysfunction** through the identification of molecules that are affected by Pb exposure and influence respiratory health, which can then be targeted by novel drugs and therapies.

5. Changes/Problems

5.a Changes in approach and reasons for change

As outlined in our previous report, there was a six month delay in the receipt of HRPO approval, and therefore in the start of the project. It was ultimately concluded that "the activities conducted by the Partners investigators do not constitute human subjects research". As such no new IRB approval was required, and the project could proceed. This delay did not result in any changes to our approach or analytical plans, and has not influenced this previous reporting period or the overall success of this project.

Additionally, when exploring the additional available biosamples it was determined that the number of men with available measures of urine Pb was too low to be statistically meaningful. Therefore, we dropped urine from our analysis. Given that urine represents a recent biomarker of Pb exposure, and that we had long, medium and short term markers of Pb exposure in the form of bone, toenail and urine levels, which are more likely to be relevant to the pathogenicity of Pb on the respiratory system, we do not believe that the lack of urine samples will influence the overall success of this project.

5.b. Changes that had a significant impact on expenditures

As outlined in our previous report, due to price negotiations with Metabolon Inc., who performed the metabolomic profiling, we were able to profile 661 samples, rather than the anticipated 374, this increased the power and impact of our proposal. *Further details in Accomplishments. Section 3.a*

5.c. Significant changes in use or care of human subjects, vertebrate animals, biohazards, and/or select agents

Nothing to report

6. Products, Inventions, Patent Applications, and/or Licenses

6.a Publications, conference papers, and presentations

One manuscript, based on Aim TWO has is due to be submitted to the journal "Metabolites" In the next two weeks (attached here as an appendix) "Metabolomics of Lung Function: Evidence from two cohorts"

An additional manuscript based on AIM THREE is currently being finalized, and we anticipate will be submitted within the next two months.

We plan on at least two additional manuscripts arising from this work, including one focused on *AIM THREE*, as well as manuscripts resulting from our collaborations with other NAS investigators (***section 6.e***.)

This work has been presented at four National and International conferences to date (*see* ***section 3.d****, for details*), with additional presentations pending (MANA 2019, Dr. Lasky-Su) and planned (ATS 2020 and the International Metabolomics Society 2020)

6.b. Websites or Other Internet Sites

Nothing to report

6.c. Technologies or techniques

Novel statistical techniques under development (see *Impact* ***section 4.a***)

6.d. Inventions, patent applications, and/or licenses

Nothing to report

6.e Other Products

Comprehensive Metabolomic Dataset. The generation of a database of 661 longitudinal plasma samples with metabolomic profiling from 464 military veterans with measures of Pb exposure and lung function, provides an invaluable research resource beyond the scope of this current project. This database will allow us to address a large number of research questions and provide a valuable validation population for an ongoing project exploring the link between the metabolome and body mass index. Further, we have been in discussion with other collaborators from the Normative Ageing Study regarding a number of projects, exploring the (i) metabolome of Aging; (ii)The metabolome of Pb and its impact on cognitive function and (iii) The impact of air pollution on the metabolome. We anticipate we will develop these collaborations further within the next reporting period, and that they will result in additional manuscripts and conference presentations, as well as further projects levering the available metabolomics data.

7. Participants & Other Collaborating Organizations

7.a. Individuals who have worked on this project

Name:	*Jessica Lasky-Su*
Project Role:	*PI*
Researcher Identifier (e.g. ORCID ID):	*0000-0001-6236-4705*
Nearest person month worked:	*1.20 person months*
Contribution to Project:	*Dr. Lasky-Su is the PI of this project and has overseen all aspects and in particular the management of the budget and working with Metabolon to negotiate metabolomic profiling prices. Dr. Lasky-Su is working directly with Dr. Kelly on the data analysis and the preparation of manuscripts, and is helping to disseminate the findings via her plenary talk at the Metabolomics Society of North America (MANA) inaugural conference in November 2019*
Funding Support:	*This DOD award*

▪

Name:	*Rachel Kelly*
Project Role:	*Co-PI*
Researcher Identifier (e.g. ORCID ID):	*0000-0003-3023-1822*
Nearest person month worked:	*4.00 person months*
Contribution to Project:	*Dr. Kelly is the co-PI of this project. She worked with collaborators at the Normative Ageing study to collect the samples, applied the QC and data processing to the data and is now leading the data analysis and manuscript preparation. Dr Kelly was selected by the Environmental and Population Health Assembly of the American Thoracic Society to give an oral presentation of these findings to the to the American Thoracic Society Annual conference in Dallas in May 2019.*
Funding Support:	*This DOD award*

Name:	*Haley Bayne*
Project Role:	*Intern*
Researcher Identifier (e.g. ORCID ID):	*bayne.h@husky.neu.edu*
Nearest person month worked:	*3 person months*
Contribution to Project:	*Ms. Bayne was a summer intern under the supervision of Dr.'s Kelly and Lasky-Su who performed statistical analyses for AIMs 1 and 2, which formed the basis of the first publication. Additionally Ms. Bayne assisted with drafting this manuscript and the pending Pb manuscript. Ms. Bayne presented this work at both the RISE Expo, and at the International Metabolomics Society meeting.*
Funding Support:	*Northeastern Cooperative Education Program*

Name:	*Olga Faybushevich*
Project Role:	*Project Coordinator*
Researcher Identifier (e.g. ORCID ID):	*N/A*
Nearest person month worked:	*0.90 person months*
Contribution to Project:	*Ms. Faybushevich is responsible for fiscal oversight of the project managing all grant accounting, purchasing and reporting. Duties also include handling all project related correspondence, organizing teleconferences of internal and external collaborators, preparing slides for project presentations, travel arrangements associated with site visits and annual project meeting*
Funding Support:	*This DOD award*

7.b. Has there been a change in the active other support of the PD/PI(s) or senior/key personnel since the last reporting period?

Changes in Other Support since last reporting period:

Jessica A. LASKY-SU, Sc.D., M.S.

ENDED GRANTS: Administered by Brigham and Women's Hospital

1 P01 HL114501-01A1 (Choi) 09/15/13 – 06/30/19 0.12 calendar
NIH/NHLBI $101,165
Distinct And Overlapping Pathways Of Fibrosis And Emphysema In Cigarette Smokers
Core B: The Respiratory Computational Discovery Core

We will facilitate streamlined QC assessment of all experimental data with generation of formatted reports and integrate desperate data sources into merged working environments, with relevant annotation & metadata.
Aim 1: To serve as the central repository for all experimental data generated from the program projects;
Aim 2: To facilitate streamlined QC assessment of all experimental data with generation of formatted reports; Aim 3: To integrate desperate data sources into merged working environments, with relevant annotation & metadata;
Aim 4: To perform standardized data preprocessing and normalization:
Aim 5:To perform all statistical and bioinformatics analysis, and generate formatted results tables and figures for investigator review and manuscript preparation;
Aim 6: To archive all experimental data and results on a firewall and password protected internal network

Program Official: Punturieri, Antonello
punturieria@nhlbi.nih.gov ; 301-435-0202
National Heart, Lung, and Blood Institute
Building 31, Room 5A52
31 Center Drive MSC 2486
Bethesda, MD 20892

NEW ACTIVE GRANTS: Administered by Brigham and Women's Hospital

R01 HL141826 (Lasky-Su) 04/01/18 – 03/31/23 1.44 calendar
NIH/NHLBI $499,570
Mechanistic insights into asthma pathogenesis through the integration of asthma genes, risk exposures, and metabolomics
The overarching hypothesis of this proposal is that the sphingolipid and eicosanoid pathways are important in asthma pathogenesis and may enlighten the mechanisms through which asthma genes (e.g. ORMDL3, FADS) and prenatal early life exposures (vitamin D and n-3 PUFAs) operate to cause or prevent asthma.
AIM 1: We hypothesize that the maternal metabolome and prenatal vitamin D and n-3 PUFA supplements have a direct impact on the child metabolome, both broadly on a metabolic pathway level and directly on specific metabolites, to alter asthma risk.
AIM 2: We hypothesize that the sphingolipid pathway is critical to asthma development and that genetic variants in *ORMDL3* and prenatal vitamin D supplementation modify this effect.
AIM 3: We hypothesize that pro-inflammatory AA eicosanoids and anti-inflammatory eicosanoids (eicosapentaenoic acid (20:5n–3, EPA) and docosahexaenoic acid (22:6n–3, DHA), are critical to asthma development and that genetic variants in *FADS*, and prenatal n-3 PUFAs supplementation modify this effect.
National Heart, Lung, and Blood Institute
Building 31, Room 5A52
31 Center Drive MSC 2486
Bethesda, MD 20892

R01 AR075117 (Paik) 06/01/19-05/31/24 0.36 calendar
NIH/NIAMS $439,610
Acid-Base Status as a Novel Risk Factor for Fractures
This proposal's goal is to investigate the relation between acid-base status, assessed through dietary acid load, plasma bicarbonate level and plasma metabolites, and risk of incident fracture. We hypothesize that perturbations in acid-base status through diet-dependent and independent mechanisms resulting in increased acidosis will be associated with higher fracture risk. Role: Co-Investigator
Specific Aims:
Aim 1: Dietary Acid Load and Risk of Fractures in Women and Men
Aim 2: Plasma Bicarbonate Level and Risk of Incident Hip Fracture in Women and Men
Aim 3: Plasma Metabolites Associated with Acid-Base Status and Hip Fracture Risk in Women and Men
Program Official: Kristy Nicks
Email: Kristy.Nicks@nih.gov
Phone: (301) 594-5055
National Institute of Arthritis and Musculoskeletal and Skin Diseases
1 AMS Circle
Bethesda, MD 20892-3675

Rachel KELLY, Ph.D., M.P.H.

PAST GRANTS:

W81XWH-17-1-0533 (Lasky-Su / Kelly) 09/15/3017-03/14/3020 (NCE) 1.20 calendar
DOD – USA MED RESEARCH ACQ ACTIVITY
Metabolomics of lead exposure and its role in respiratory disease

Military service personnel and Veterans suffer disproportionately from poor lung health. The aim of this project was to use blood based metabolomic profiling of a large, well-characterized population of Veterans with comprehensive clinical and environmental exposure measurements, to understand and quantify how Pb exposure during active service affects respiratory health in those who have served in the military.
AIM 1: To identify metabolomic profiles of Pb exposure in a population of Veterans.
AIM 2: To identify metabolomic profiles of respiratory health within a population of Veterans
AIM 3: To determine the influence of Pb exposure on respiratory health through the integration of the metabolomics profiles of heavy metal exposure and respiratory disease

R01 HL141826 (Lasky-Su) 04/01/18 – 03/31/23 1.20 calendar
NIH/NHLBI $499,570

Mechanistic insights into asthma pathogenesis through the integration of asthma genes, risk exposures, and metabolomics

The overarching hypothesis of this proposal is that the sphingolipid and eicosanoid pathways are important in asthma pathogenesis and may enlighten the mechanisms through which asthma genes (e.g. ORMDL3, FADS) and prenatal early life exposures (vitamin D and n-3 PUFAs) operate to cause or prevent asthma.
AIM 1: We hypothesize that the maternal metabolome and prenatal vitamin D and n-3 PUFA supplements have a direct impact on the child metabolome, both broadly on a metabolic pathway level and directly on specific metabolites, to alter asthma risk.
AIM 2: We hypothesize that the sphingolipid pathway is critical to asthma development and that genetic variants in *ORMDL3* and prenatal vitamin D supplementation modify this effect.
AIM 3: We hypothesize that pro-inflammatory AA eicosanoids and anti-inflammatory eicosanoids (eicosapentaenoic acid (20:5n–3, EPA) and docosahexaenoic acid (22:6n–3, DHA), are critical to asthma development and that genetic variants in *FADS*, and prenatal n-3 PUFAs supplementation modify this effect.

National Heart, Lung, and Blood Institute
Building 31, Room 5A52
31 Center Drive MSC 2486
Bethesda, MD 20892

5 R01 HL123915-02 (Lasky-Su) 08/01/14 – 05/31/20 6.00 calendar
NIH/NHLBI $497,269

Integrative Metabolomics of Asthma Severity

We propose the use of 1,500 well-characterized asthmatics from two well-established cohorts to identify metabolic contributors of asthma exacerbations through comprehensive metabolic profiling and the integration with relevant clinical, environmental, and genome-wide SNP and gene expression data.
Aim 1: To identify metabolites associated with asthma severity and exacerbations in both untargeted and candidate approaches;
Aim 2: To integrate metabolomics data with genome-wide genetic (i.e. SNP) and genomic (i.e. gene expression) data;
Aim 3: To identify a metabolomic signature for asthma severity and exacerbations using Bayesian Networks through the integration of environmental, clinical, genetic, genomic, and metabolomics data

Program Official: Patricia Noel
Email: noelp@nhlbi.nih.gov Phone: (301) 435-0202
National Heart, Lung, and Blood Institute
Building 31, Room 5A52
31 Center Drive MSC 2486

Bethesda, MD 20892

2 P50 CA090381-11A1 (Loda) 09/23/2013 – 06/30/2019 1.20 calendar
NIH/NCI
SPORE in Prostate Cancer
The overall goal of this project is to elucidate the underlying links between obesity and lethal disease among men with incident prostate cancer who were participants in the Health Professionsals Follow-Up Study. We are proposing to investigate specific pathways associated with obesity and integrate anthropometric data, molecular features in prostatic tumor and stroma and circulation biomarkers measured in pre-diagnostic blood samples with cancer outcomes. We propose 4 Projects which address critical problems in prostate cancer and have translational components including
1) understanding the mechanism by which obesity impacts prostate cancer mortality
2) dissecting the clinical heterogeneity of Gleason 7 prostate cancer
3) understanding the role that genomic change has on resistance to primary androgen deprivation therapy and on aggressive prostate cancer and finally
4) understanding the androgen receptor based mechanisms of resistance in castration resistant disease.
The four projects are supported by three Cores - an Administrative Core, a Biostatistics and Computational Biology Core, and a Tissue and Pathology Core. We also have a highly successful Career Development Program that selects talented physician scientists and mentors them to independence as well as a Developmental Projects Program that generates new ideas for the SPORE in the future.

Program Official: Julia T. Arnold
Email: jarnold@mail.nih.gov
National Cancer Institute
Building 9609 MSC 9760
9609 Medical Center Drive
Bethesda, MD 20892-9760

ACTIVE GRANTS: Administered by Brigham and Women's Hospital

1K01 HL146980-01 (Kelly) 05/01/2019-05/01/2024 12.00 calendar
NIH/NHLBI
Multi-omic Endptyping of Asthma
The heterogeneous nature of asthma; a common chronic condition worldwide, is not captured by current clinical management guidelines leading to suboptimal treatment approaches for many asthmatics. This grant utilizes a novel and innovative approach to classifying individuals with asthma into endotypes based on their underlying pathobiological mechanisms, as defined by their omic-profiles. The findings will pave the way for the development of biomarkers, targeted therapeutics and personalized approaches, leading to a better quality of life for individuals with asthma.
Program Official: Xenia Tigno
Email: tignoxt@mail.nih.gov
Phone: 301-435-0202
National Heart, Lung, and Blood Institute
Building 31, Room 5A52
31 Center Drive MSC 2486
Bethesda, MD 20892

PENDING GRANTS

NONE

7.c. What other organizations were involved as partners?

Nothing to Report

8. Special Reporting Requirements – N/A.

Nothing to Report

9. Appendix

metabolites

MDPI

Article

Metabolomics of Lung Function: Evidence from two cohorts

Rachel S Kelly [1α*], Haley Bayne [1α], Isobel Stewart[2], Avron Spiro III[3], David Sparrow[3], Pantel Vokonas[3], Scott Weiss[1], Augusto A. Litonjua[4], Claudia Langenberg[2α], Jessica A. Lasky-Su[1,α]

[1] Channing Division of Network Medicine, Brigham and Women's Hospital Harvard Medical School, Boston MA, USA
[2] MRC Epidemiology Unit, University of Cambridge, School of Clinical Medicine, Institute of Metabolic Science, Cambridge, UK
[3] Massachusetts Veterans Epidemiology Research and Information Center, VA Boston Health Care System, Boston MA, USA
[4] Division of Pediatric Pulmonary Medicine, University of Rochester Medical Center, Rochester, NY USA
[α] These authors contributed equally to this manuscript
* Correspondence: hprke@channing.harvard.edu; Tel.: +18572718276

Abstract: The mechanisms underlying the decline of the respiratory system with age are not fully understood. This study aims to identify and validate the plasma metabolome of lung function in adulthood utilizing two independent cohorts: the European Prospective Investigation into Cancer - Norfolk (EPIC-Norfolk, n=10,460) and the VA Normative Aging Study (NAS, n=439. We employed linear regression to identify metabolites associated with forced expiratory volume in one second (FEV_1) and the ratio of FEV_1 to forced vital capacity (FEV_1/FVC). Seven FEV_1 associated ($p<0.05$) metabolites (2-hydroxystearate, 3-hydroxyisobutyrate, 4-methyl-2-oxopentanoate, dihomo-linoleate (20:2n6), DHA 22:6n3, DPA 22:5n3, and stachydrine) replicated in both cohorts. Three metabolites (cysteine-glutathione disulfide, threonine and xylose) replicated for FEV_1/FVC. We additionally identified a number of metabolites that replicated in males only, including 2-hydroxystearate and dihomo-linoleate 20:2n6. The validation of metabolites associated with respiratory function can help to better understand mechanisms of respiratory health and may assist in the development of biomarkers.

Keywords: metabolomics; lung function; FEV_1; European Prospective Investigation into Cancer - Norfolk (EPIC-Norfolk); VA Normative Aging Study (NAS); spirometry; Omega-3 fatty acids

1. Introduction

The functionality of the respiratory system, which shows anatomical, physiological, and immunological changes over time, is proven to decline progressively with age, resulting in significant morbidity and mortality [1].

Metabolomics, the systematic profiling of the small molecules in a biological system [2, 3], represents a powerful tool to increase the understanding of the mechanisms of respiratory health. Metabolomics provides a downstream 'snapshot' of the status of a biological system reflecting current phenotype as well as upstream genetic and environmental influences. As such, metabolomics is ideally suited to examine alterations in biological pathways that accompany phenotypic changes [4]. A number of studies have successfully utilized metabolomics to explore pulmonary diseases such as asthma and chronic obstructive pulmonary disease (COPD) [5, 6], suggesting that there are measurable and biologically informative changes in the metabolome that reflect changes in the respiratory system over time.

Despite these promising findings, existing metabolomic studies of respiratory health have been limited by sample size and by a lack of validation of significant findings [5, 6], or have focused on child rather than adult populations [7]. Furthermore, to our knowledge, no studies have specifically investigated FEV_1 or FEV_1/FVC ratio, two of the most important indicators of lung health and pulmonary function in adults. Consequently, much remains to be learned about the mechanisms underlying respiratory decline with age, and metabolomic profiling of lung function in adult and ageing populations warrants further investigation [8].

In this current study, we aim to identify and validate metabolites associated with lung function, as defined by forced expiratory values (FEV_1 and FEV_1/FVC), in two independent adult populations: the European Prospective Investigation into Cancer - Norfolk (EPIC-Norfolk) cohort [9] and the VA Normative Aging Study (NAS) [10] in order to provide an increased mechanistic understanding of respiratory health and to support the potential future development of novel biomarkers of lung function.

2. Results

2.1 Study population

In EPIC-Norfolk, the mean age of all 10460 participants at sample collection for metabolomic profiling was 59.73 years (SD = 8.95; **Table 1**), and the population was predominantly White (99.3%). Many participants were former smokers (42.5%), with only 10.9% reporting current smoking at the time of blood sample collection. The mean FEV_1 score for the total population was 2.52 L (SD = 0.72), while the mean FEV_1/FVC score for the total population was 0.82 (SD = 0.11).

The distribution of these baseline characteristics was largely consistent between males and females within EPIC-Norfolk. The mean age of male participants was 60.14 years (SD = 8.97) and were 99.7% white, while the mean age of female participants was 59.37 (SD = 8.91) and 99.4% were white. However, a larger proportion of males were former or current smokers. The mean FEV_1 score was higher among males at 2.93 L (SD = 0.72) compared to 2.16 L (SD = 0.51) for females, but the mean FEV_1/FVC scores were similar with 0.81 (SD = 0.12) for males and 0.82 (SD = 0.10) for females.

Table 1. Baseline characteristics of included participants in EPIC.

	VARIABLE INFO	TOTAL POPULATION	MALES	FEMALES
N		10,460	4,868 (46.5%)	5,592 (53.5%)
AGE	Mean (SD)	59.73 (8.95)	60.14 (8.97)	59.37 (8.91)
ETHNICITY	White (%)	10,392 (99.3%)	4,832 (99.7%)	5,560 (99.4%)
	Black (%)	9 (0.09%)	5 ($1.0x10^{-3}$%)	4 ($7.2x10^{-4}$%)
	Other (%)	8 ($8.6x10^{-4}$%)	12 ($2.5x10^{-3}$%)	4 ($7.2x10^{-4}$%)
FEV_1 (L)	Mean (SD)	2.52 (0.72)	2.93 (0.72)	2.16 (0.51)
FEV_1/FVC	Mean (SD)	0.82 (0.11)	0.81 (0.12)	0.82 (0.10)
BMI (KG/M^2)	Mean (SD)	26.18 (3.70)	26.35 (3.12)	26.04 (4.14)
SMOKE STATUS	Current (%)	1,145 (10.9%)	573 (11.8%)	572 (10.2%)
	Former (%)	4,448 (42.5%)	2,689 (55.2%)	1,759 (31.5%)
	Never (%)	4,867 (46.5%)	1,606 (33.0%)	3,261 (58.3%)

In the replication population, NAS, the mean age of the participants at sample collection for metabolomic profiling was older than in EPIC-Norfolk: 75.96 years (SD = 6.64; range = 40 years; **Table S1**) and included only males. Again, the population was predominantly White (98.3%). A greater percentage of participants were former smokers in NAS (64.4%) compared to EPIC-Norfolk (the total population, and the male only population), with only 4% reporting current smoking at the time of blood sample collection. The mean FEV_1 score in NAS was 2.50 (SD = 0.60) and the FEV_1/FVC ratio was 0.74 (SD = 0.43; **Table S1**), which were lower than the those observed in EPIC-Norfolk for both the total population and for men only

1002 metabolites passed Quality Control (QC) and data processing in EPIC-Norfolk, while 981 metabolites passed QC in NAS. There were 698 metabolites in common between the two populations.

2.2. Metabolome of lung function: Findings in EPIC-Norfolk and NAS

2.2.A. Total population

Of the 1002 metabolites that passed QC and data processing, 487 (48.6%) were significantly ($p<0.05$) associated with FEV_1 in EPIC-Norfolk after adjustment for sex, age, body mass index (BMI), smoking, height and asthma status (self-reported previous doctor's diagnosis; **Figure 1**). A nominal significance level of $p<0.05$ was employed, due to the lack of agreed standards for multiple testing, the highly correlated nature of metabolites and due to the presence of a validation population which provided us with the opportunity to determine true positive findings.
The significant metabolites were mainly comprised of lipids, xenobiotics and amino acids (**Figure S1**); 51% of significant metabolites were inversely associated with FEV_1, and 49% were positively associated. The top hits for FEV_1 in EPIC-Norfolk included: threonate ($\beta = 4.91$, $p = 3.57x10^{-25}$) and ergothioneine ($\beta = 4.72$, $p = 1.62x10^{-24}$).

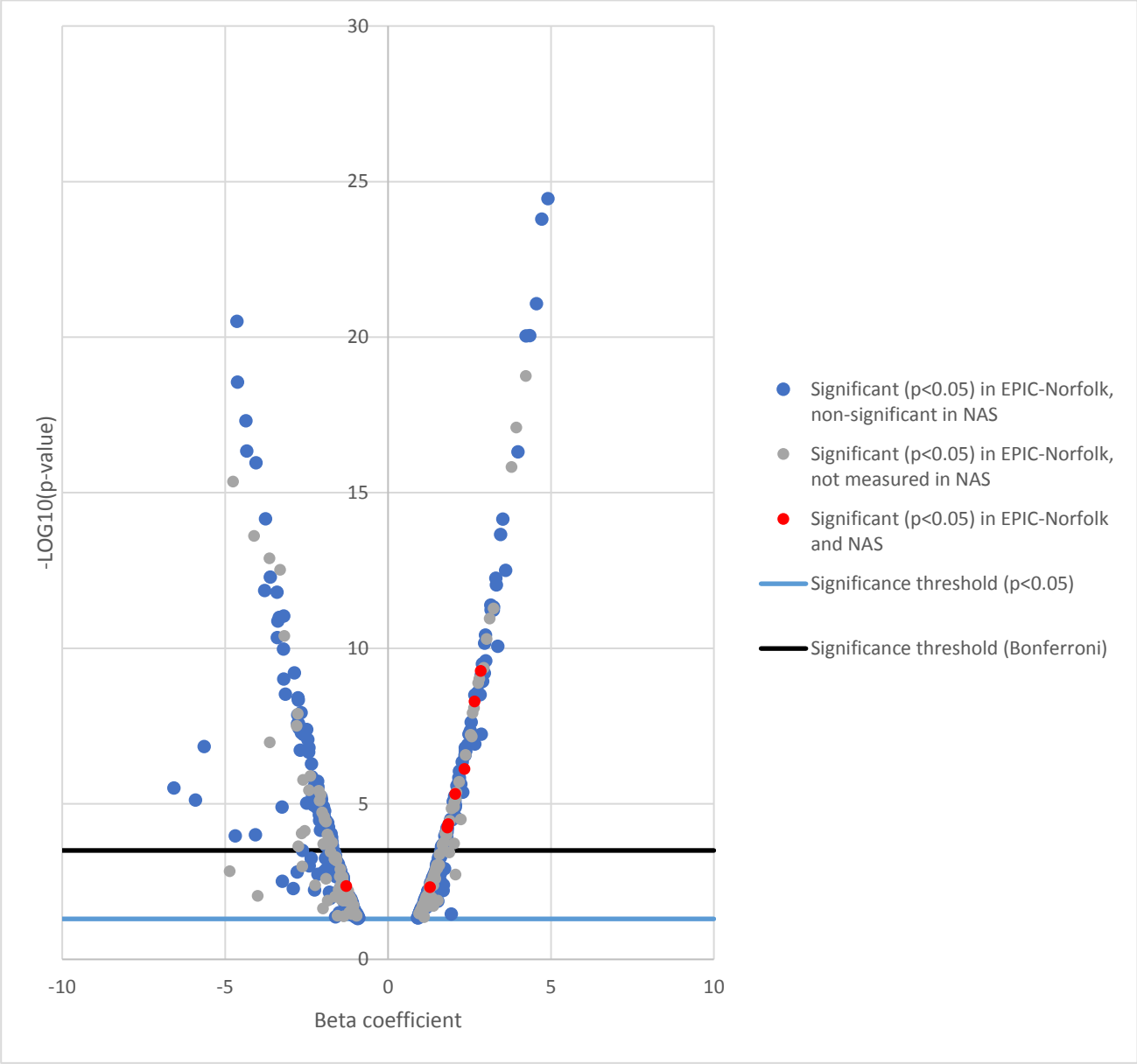


Figure 1. Volcano plot demonstrating the FEV_1-metabolite associations in EPIC-Norfolk and NAS.

Three hundred and six (30.5%) metabolites were found to be significant (p<0.05) for FEV_1/FVC in EPIC-Norfolk, most (n=207, 68%) of which were inversely correlated with FEV_1/FVC (**Figure 2**). These metabolites were also mainly lipids, xenobiotics and amino acids (**Figure S1**). Lactate (β = -0.014, $p = 2.19X10^{-44}$), 5-oxoproline (β = -0.014, $p = 3.33x10^{-38}$), sphingosine-1-phosphate (β = -0.013, $p = 1.81^{-37}$), and inosine (β = 0.011, $p = 7.09X10^{-24}$) were among the top hits.

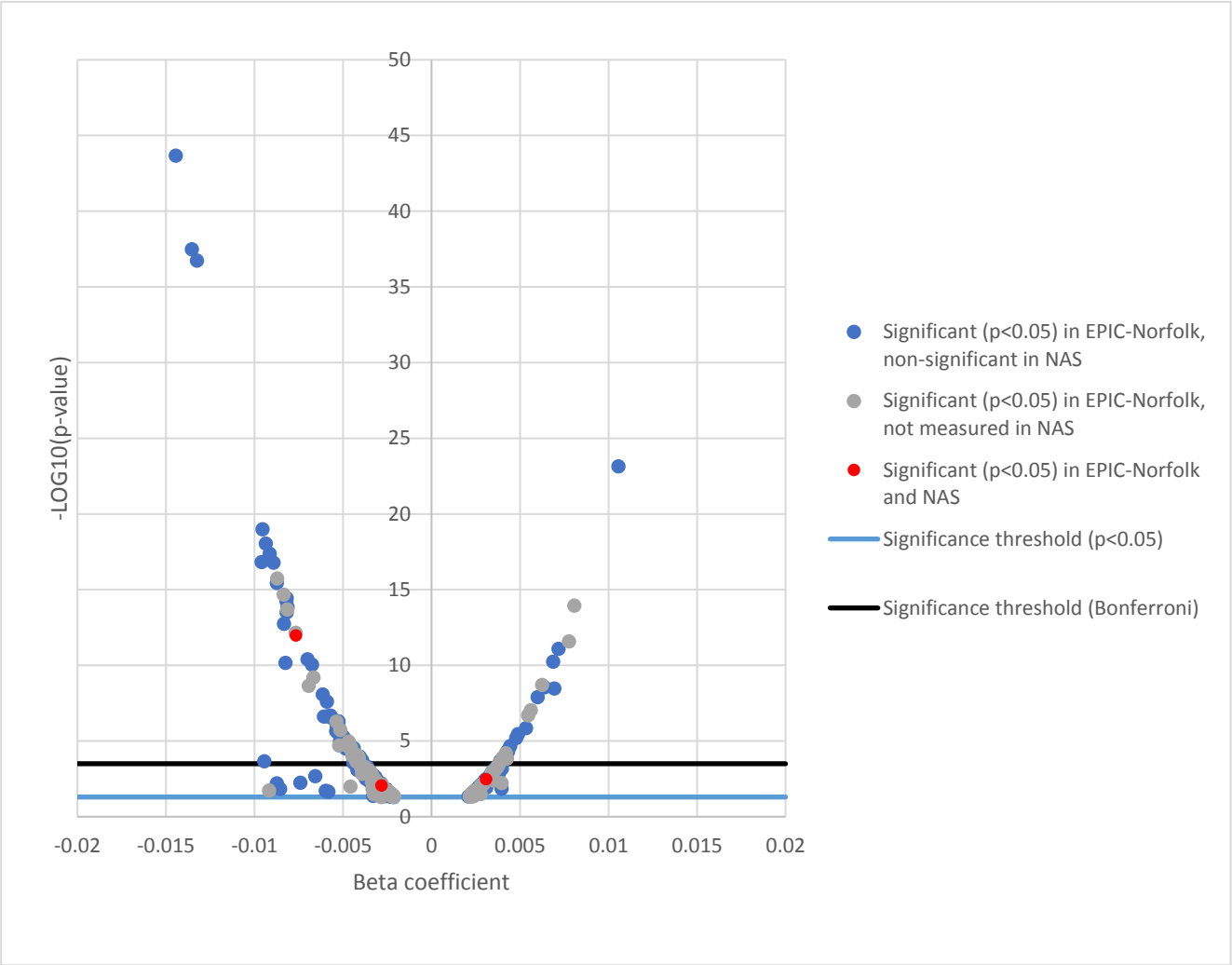


Figure 2. Volcano plot demonstrating the FEV_1/FVC-metabolite associations in EPIC-Norfolk and NAS.

In total, 141 metabolites were associated with both FEV_1 and FEV_1/FVC in EPIC-Norfolk (**Figure 3**), the largest proportion of which were amino acids (36 metabolites; 25.5%), followed by lipids (20, 14.2%), which included one omega 3 fatty acid: DPA (n3 DPA; 22:5n3).

In the NAS population, 36/1140 (3.2%) metabolites were associated with FEV_1 at $p<0.05$ (**Table S2**), while 33 (2.9%) metabolites were significant for FEV_1/FVC (**Table S3**). As in EPIC-Norfolk, these metabolites were mainly lipids, xenobiotics and amino acids (**Figure S2**), and the majority of associations were inverse.

Ten FEV_1 associated metabolites replicated between EPIC-Norfolk and NAS with consistent directions of effect and a replication p-value threshold of <0.05 (**Table 2**). These were primarily amino acids and lipids, including several omega-3 fatty acids: DHA 22:6n3, DPA 22:5n3, and EPA 20:5n3.

Table 2. Nominally significant metabolites, replicated in EPIC-Norfolk and NAS, for FEV_1.

METABOLITE	PATHWAY	SUBPATHWAY	EPIC-NORFOLK (N=1002)				NAS (N=1140)			
			Estimated effect (β)	*p*	Standard error	Confidence interval (lower, upper)	Estimated effect (β)	*p*	Standard error	Confidence interval (lower, upper)
3-HYDROXYISOBUTYRATE	Amino Acid	Leucine, Isoleucine and Valine Metabolism	1.85	$4.57x10^{-05}$	0.46	(0.96, 2.74)	0.15	0.017	0.06	(0.03, 0.27)
3-METHYL-2-OXOBUTYRATE	Amino Acid	Leucine, Isoleucine and Valine Metabolism	1.28	0.005	0.46	(0.39, 2.18)	0.16	0.014	0.07	(0.03, 0.29)
4-METHYL-2-OXOPENTANOATE*	Amino Acid	Leucine, Isoleucine and Valine Metabolism	2.35	$7.61x10^{-07}$	0.47	(1.42, 3.28)	0.14	0.034	0.07	(0.01, 0.27)
DHA 22:6N3*	Lipid	Polyunsaturated Fatty Acid (n3 and n6)	2.85	$5.37x10^{-10}$	0.46	(1.95, 3.75)	0.11	0.017	0.05	(0.02, 0.20)
DPA 22:5N3*	Lipid	Polyunsaturated Fatty Acid (n3 and n6)	2.06	$4.84x10^{-06}$	0.45	1.18, 2.95)	0.11	0.026	0.05	(0.01, 0.20)
EPA 20:5N3	Lipid	Polyunsaturated Fatty Acid (n3 and n6)	1.82	$5.80x10^{-05}$	0.45	(0.93, 2.70)	0.09	0.038	0.04	(0.01, 0.18)
GLUCOSE	Lipid	Polyunsaturated Fatty Acid (n3 and n6)	-1.28	$4.40x10^{-03}$	0.45	(-2.17, -0.39)	-0.08	0.034	0.04	(-0.15, -0.01)
PICOLINATE	Lipid	Polyunsaturated Fatty Acid (n3 and n6)	1.32	0.041	0.65	(0.06, 2.58)	0.08	0.024	0.04	(0.01, 0.16)
STACHYDRINE*	Xenobiotics	Food Component/ Plant	2.65	$5.13x10^{-09}$	0.45	(1.76, 3.54)	0.07	0.043	0.03	($2x10^{-3}$, 0.13)

*Indicates significance retained after a Bonferroni correction on EPIC-Norfolk data.

Four FEV_1/FVC associated metabolites identified in EPIC-Norfolk replicated in NAS (**Table 3**), including two amino acids, one carbohydrate, and one xenobiotic.

147 **Table 3.** Nominally significant metabolites, replicated in EPIC-Norfolk and NAS, for FEV_1/FVC.
148

METABOLITE	PATHWAY	SUBPATHWAY	EPIC-NORFOLK (N=1002)				NAS (N=1140)			
			Estimated effect (β)	*p*	Standard error	Confidence interval (lower, upper)	Estimated effect (β)	*p*	Standard error	Confidence interval (lower, upper)
2-PIPERIDINONE	Xenobiotic	Food Component/ Plant	-0.003	$9.96x10^{-04}$	$1.1x10^{-03}$	($-5.6x10^{-3}$, $1.4x10^{-3}$)	0.01	0.039	$4.8x10^{-3}$	($5x10^{-3}$, 0.02)
CYSTEINE-GLUTATHIONE DISULFIDE*	Amino Acid	Glutathione Metabolism	-0.008	$1.07x10^{-12}$	$1.1x10^{-03}$	($-9.8x10^{-3}$, $-5.6x10^{-3}$)	-0.01	0.017	$4.9x10^{-3}$	(-0.02, $-2x10^{-3}$)
THREONINE	Amino acid	Glycine, Serine and Threonine Metabolism	-0.003	0.009	$1.1x10^{-03}$	($-4.9x10^{-3}$, $-7.2x10^{-4}$)	-0.03	0.012	$1.2x10^{-2}$	(-0.05, $-7x10^{-3}$)
XYLOSE	Carbohydrate	Pentose Metabolism	-0.006	$5.28x10^{-04}$	$1.6x10^{-03}$	($-8.8x10^{-3}$, $-2.4x10^{-3}$)	-0.01	0.037	$5.8x10^{-3}$	(-0.02, $-7x10^{-4}$)

149 *Indicates significance retained after a Bonferroni correction on EPIC-Norfolk data.
150
151 *2.2.B. Male only*
152
153 While EPIC-Norfolk has male and female subjects, NAS is an only male cohort. For this reason,
154 we additionally looked at the results when restricting to only the males from EPIC-Norfolk. A total
155 of 994 metabolites passed QC in the male only population.
156 Among the male EPIC-Norfolk population, 367 (36.9%) metabolites were significantly ($p<0.05$)
157 associated with FEV_1, including 337 of the metabolites that were significant in the total population
158 (**Figure 3)**. Two-hundred and seventy-five (27.7%) metabolites were found to be nominally
159 significant ($p<0.05$) for FEV_1/FVC in EPIC-Norfolk males, of which 225 were significant in the total
160 population.
161

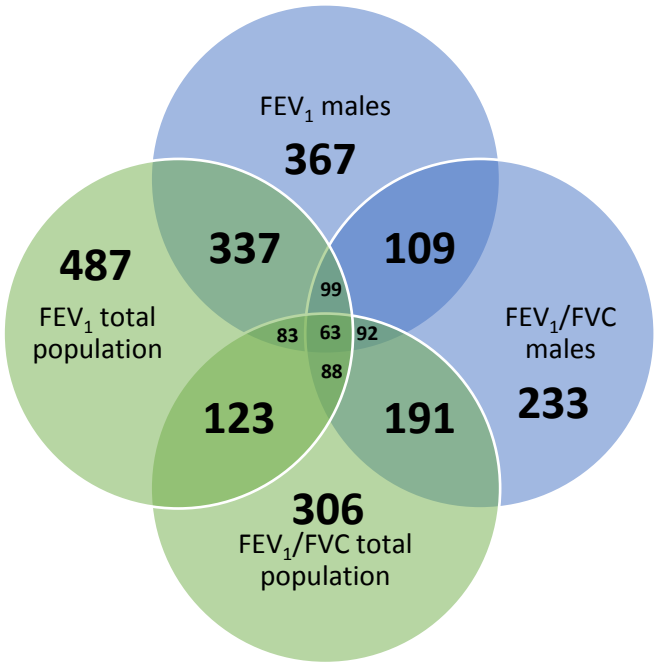


Figure 3. Overlap of nominally significant metabolites in the EPIC-Norfolk total population and in the male only analyses for FEV_1 and FEV_1/FVC.

Seven metabolites replicated between the EPIC-Norfolk male only analysis and NAS for FEV_1 (**Table 4**), including two that were not among the significant hits in the total population analyses: 2-hydroxystearate and dihomo-linoleate 20:2n6.

Table 4. Nominally significant metabolites, replicated in EPIC-Norfolk males and NAS, for FEV_1.

METABOLITE	PATHWAY	SUBPATHWAY	EPIC-NORFOLK MALES (N=994)				NAS (N=1140)			
			Estimated effect (β)	*p*	Standard error	Confidence interval (lower, upper)	Estimated effect (β)	*p*	Standard error	Confidence interval (lower, upper)
2-HYDROXYSTEARATE	Lipid	Fatty Acid, Monohydroxy	0.02	0.028	$7.6x10^{-3}$	($1.8x10^{-3}$, 0.03)	0.16	0.026	$7.3x10^{-2}$	(0.02,0.31)
3-HYDROXYISOBUTYRATE	Amino Acid	Leucine, Isoleucine and Valine Metabolism	0.02	0.029	$7.8x10^{-3}$	($1.8x10^{-3}$, 0.03)	0.15	0.017	$6.1x10^{-2}$	(0.03, 0.27)
4-METHYL-2-OXOPENTANOATE	Amino Acid	Leucine, Isoleucine and Valine Metabolism	0.03	0.001	$8.3x10^{-3}$	(0.01, 0.04)	0.14	0.034	$6.6x10^{-2}$	(0.01, 0.27)
DIHOMO-LINOLEATE (20:2N6)	Lipid	Polyunsaturated Fatty Acid (n3 and n6)	0.02	0.035	$8.0x10^{-3}$	(-0.01, 0.01)	0.11	0.037	$5.3x10^{-2}$	($6.9x10^{-3}$, 0.22)
DHA 22:6N3*	Lipid	Polyunsaturated Fatty Acid (n3 and n6)	0.03	$1.57x10^{-05}$	$7.8x10^{-3}$	(0.02, 0.05)	0.11	0.017	$4.8x10^{-2}$	(0.02, 0.20)
DPA 22:5N3	Lipid	Polyunsaturated Fatty Acid (n3 and n6)	0.03	$1.84x10^{-04}$	$8.0x10^{-3}$	(0.01, 0.05)	0.11	0.026	$4.8x10^{-2}$	(0.01, 0.20)
STACHYDRINE	Xenobiotics	Food Component/Plant	0.03	$1.53x10^{-04}$	$7.6x10^{-3}$	(0.01, 0.04)	0.07	0.044	$3.3x10^{-2}$	($2x10^{-3}$, 0.13)

*Indicates significance retained after a Bonferroni correction on EPIC-Norfolk data.

Three metabolites replicated between the EPIC-Norfolk male only analysis and NAS (**Table 5**) for FEV_1/FVC, cysteine-glutathione disulfide, threonine, and xylose. All of which were also significant in the EPIC-Norfolk total population.

Table 5. Nominally significant metabolites, replicated in EPIC-Norfolk males and NAS, for FEV_1/FVC.

METABOLITE	PATHWAY	SUBPATHWAY	EPIC-NORFOLK MALES (N=994)				NAS (N=1140)			
			Estimated effect (β)	*p*	Standard error	Confidence interval (lower, upper)	Estimated effect (β)	*p*	Standard error	Confidence interval (lower, upper)
CYSTEINE-GLUTATHIONE DISULFIDE*	Amino Acid	Glutathione Metabolism	-0.009	$6.10x10^{-07}$	$1.7x10^{-3}$	(-0.01, $-5x10^{-3}$)	-0.01	0.017	$4.9x10^{-3}$	(-0.02, $-2x10^{-3}$)
THREONINE	Amino acid	Glycine, Serine and Threonine Metabolism	-0.005	0.007	$1.8x10^{-3}$	($-8.5x10^{-3}$, $1.4x10^{-5}$)	-0.03	0.012	$1.2x10^{-2}$	(-0.05, $-7x10^{-3}$)
XYLOSE	Carbohydrate	Pentose Metabolism	-0.005	0.033	$2.5x10^{-3}$	(-0.01, $4x10^{-4}$)	-0.01	0.037	$5.8x10^{-3}$	(-0.02, $-7x10^{-4}$)

*Indicates significance retained after a Bonferroni correction on EPIC-Norfolk data.

2.3. Metabolome of lung function: Results in EPIC-Norfolk and NAS accounting for multiple testing

2.3.A. Total population

Of the 487 nominally significant metabolites for FEV_1 in EPIC-Norfolk, 168 (16.8%) metabolites were robust to Bonferroni correction [11], which imposed a p-value threshold of $4.99x10^{-5}$. Of these metabolites, four FEV_1-associated metabolites replicated in NAS, including 4-methyl-2-oxopentanoate, DHA, stachydrine, and DPA (**Table 3**).

Of the 306 nominally significant metabolites for FEV_1/FVC in EPIC-Norfolk, 71 (7.1%) metabolites retained significance after Bonferroni correction, of which one replicated in NAS: cysteine-glutathione disulfide (**Table 4**).

2.3.B. Male only

Seventy-six (7.6%) FEV_1-associated metabolites were robust to Bonferroni correction in EPIC-Norfolk males; of which one replicated in NAS: DHA. Thirty-five (3.5%) FEV_1/FVC-associated metabolites were robust to Bonferroni correction in EPIC-Norfolk males, of which only cysteine-glutathione disulfide replicated in NAS.

3. Discussion

Negative respiratory outcomes in older populations represent a growing problem globally. Aging lungs face decreased recoil rates, weaker supporting muscles, and a tightened chest wall, leading to reduced expiratory rates [12, 13]. Poor respiratory health in aging populations is associated with comorbidities, such as hypertension and heart failure [14-16]. As such, there is a need for an increased understanding of the underlying mechanisms of poor respiratory health, which could aid with the development of preventative or therapeutic strategies.

In this study, we identified metabolites measured in the circulating blood that were associated with reduced lung function in a large population-based cohort including over 10,460 participants. We were then able to replicate a number of our top hits in an independent population of males. Our replicated hits were characterized by omega 3 fatty acids, which have a biologically plausible relationship with respiratory function, as well as several amino acids.

We validated the positive association between levels of 3-hydroxyisobutyrate and improved FEV_1 (lung) function in our two independent cohorts, which is in agreement with previous work showing that levels of this metabolite are lower among those with emphysema and COPD, as compared to controls [17]. Interestingly, we also saw common pathways, such as glycine, serine and threonine metabolism, between lung function in this study and previously reported studies of COPD[18] which is characterized by a decreased FEV_1. We observed a positive relationship between stachydrine, produced by the intake of foods and juices high in citrus, and FEV_1, which is supportive of previous work linking higher plasma levels of stachydrine to a decreased risk of childhood asthma [19]. There is also strong biological rationale for the role of cysteine-glutathione, for which we validated a negative relationship with FEV_1/FVC in the total and male-only analyses. Cysteine-glutathione is formed upon the oxidative stress of glutathione [20] and it has long been suspected that disturbances in oxidation/reduction (redox) reactions are a risk factor for respiratory disease [21, 22]. Airway oxidative stress is linked to worsened disease severity, reduced lung function, and epigenetic changes that decrease airway responsiveness to steroids [23]. These metabolite associations did not appear to be gender dependent, as they were found both in the total EPIC-population and when we restricted to males only.

Interestingly, 2-hydroxystearate, which replicated only in our male populations and may therefore be gender dependent, was found to be significant in a clustering analysis used to identify World Trade Center lung injury-related metabolites in a study that was also based only on males [24].

Finally, there were three omega-3 fatty acids that were significant and replicated in our all population analysis: DPA, DHA and EPA. DPA and DHA retained nominal significance in male only analyses. Two of these three metabolites (DHA and EPA) are also known anti-inflammatory metabolites [25] and have been inversely correlated with the incidence of asthma [26]. DHA has demonstrated protective effects on lung function in adults and has been associated with asthma incidence [27] and atopy [28] in children. Furthermore, EPA is known to have a negative relationship with oxidative stress, suggesting that it may be effective in preventing rapid loss of lung function [29]. It is hypothesized that these metabolites alleviate inflammatory-related respiratory conditions, leading to improved respiratory health outcomes [15]. It is also suggested that these metabolites are critical to respiratory development and play a large role in the

pathogenesis of respiratory disease [30]. On a biological level, DHA and EPA can join the membrane phospholipids of effector cells. This causes changes in membrane fluidity, which can affect lipid rafts essential for immune cell activation [16]. Our results, demonstrating that levels of these metabolites are lower among those with reduced lung capacity, provides further evidence for the protective role of omega-3 fatty acids in pulmonary function.

Our study design has a number of limitations. There were differences in sample collection methods and data pre-processing for these two independent cohorts. Also, while all NAS participants are male, the EPIC-Norfolk cohort is comprised of both males and females. Despite this, we replicated a number of biologically relevant metabolites between the two populations. We subsequently stratified by gender and determined that a majority of FEV_1 and FEV_1/FVC associated metabolites do not appear to be gender dependent, while others, such as 2-hydroxystearate and dihomo-linoleate 20:2n6, may influence lung function in males only. Further work is required to disentangle these gender specific effects including replication in a female population. EPIC-Norfolk has a lower median age than NAS, which may also have influenced our findings. In both studies, the majority of participants are Caucasian; therefore, our results may not be generalizable to other races.

There is currently no gold-standard approach for correction for multiple testing in metabolomics analyses, due to the highly correlated nature of metabolites within co-regulated and redundant pathways. Therefore, we report both nominally significant and Bonferroni corrected results. Our results show that a number of our nominally significant findings were not robust to this correction, however we do note that Bonferroni is widely acknowledged to be too stringent for metabolomics data [31]. Finally, we had incomplete metabolome coverage in both populations, and there was incomplete crossover in terms of measured metabolites between the two cohorts, which also influenced our ability to replicate findings.

Nevertheless, our study holds power due to its large sample size and the usage of discovery/validation cohorts in order to identify metabolites significantly associated with lung function.

4. Materials and Methods

4.1. Study Population

4.1.1. Discovery population

European Prospective Investigation into Cancer-Norfolk (EPIC-Norfolk) [9] is an ongoing prospective cohort that represents one of the study sites within the wider EPIC project [32]. Between 1993 and 1997, 25,639 men and women aged 39-79 years attended baseline clinical examination for EPIC-Norfolk. These participants resided in Norfolk, England, an area that has little outward migration and is mainly served by one hospital. Information was collected via questionnaire on lifestyle variables and dietary habits, anthropometric measurements were performed, and blood samples were taken, from which plasma, serum, red cells and buffy coat fractions were separated and aliquoted for long-term storage in liquid nitrogen at minus 175 degrees Celsius. Participants were not requested to fast prior to blood sampling and were largely unfasted. All other methods have been described in detail previously [33].

All individual participants included in the study provided signed informed consent. The study was approved by the Norwich Local Ethics Committee (REC Ref. 98CN01).

4.1.2. Replication population

The Department of Veterans Affairs (formerly known as the US Veterans Administration) Normative Aging Study (NAS) is a longitudinal study of aging based in Boston, Massachusetts. The NAS recruited 2,280 men aged 21-80 between the years of 1961 and 1970 who were free of known chronic disease [10]. Participants have routine physical examinations and laboratory tests every 3-5 years. At study visits, a sample of venous blood was obtained from each participant after fasting. The blood samples were spun at 3000 revolutions per minute (RPMs) for 15 minutes. Subsequently, serum and plasma samples were placed in 1.8ml Nunc tubes for long term storage at –80 degrees Celsius. Self-reported information regarding medical history, smoking history, dietary intake and other health-related inquiries were also collected throughout follow-up. For the current study, we selected a subset of 439 men from the NAS with spirometric measures and concurrent blood samples available for metabolomic profiling. NAS received written consent from all study participants at each visit, as well as approval from the Review Boards of all institutions involved in data collection/analysis.

NAS is an ongoing study, to date 190 subjects remain alive and are actively participating in follow-up. Those who are still active participants attend data collection follow-ups every 2-3 years; and we continue to collect morbidity and mortality data for all subjects up until the time of death.

4.2. Spirometric Measures

4.2.1. EPIC-Norfolk: Spirometric measures

Spirometry, including FEV_1 and FVC were measured using an electronic turbine spirometer (Micro Medical Instruments, Rochester, UK) [9]. Measurements were taken twice, and the better of the two measures was analyzed [34].

4.2.2. NAS: Spirometric measures

Methods for conducting spirometry have been described previously [35]. Spirometry was repeated up to a maximum of eight spirograms, allowing at least three acceptable spirograms, at least 2 of which were reproducible with FEV_1 and FEV_1/FVC measurements within 5% of each spirogram; the best of these 2 values was selected from a given encounter. Acceptability of spirograms was judged according to American Thoracic Society standards [36, 37]. All spirometric values are pre-bronchodilator.

4.3. Metabolomic Profiling

In both studies, metabolomic profiling was conducted by Metabolon Inc. (Durham, NC, USA) using four non-targeted Liquid Chromatography Couple Mass Spectroscopy (LCMS), platforms, enabling the broadest coverage of the metabolome. The methods have been described in detail previously [38]. In short, metabolites were identified by their mass-to-charge ratio (m/z), retention time (rt), and through a comparison to a library of purified known standards. Metabolites were matched to standards using LC-MS peak areas. Peaks were quantified using area-under-the-curve.

In EPIC-Norfolk, measurements were made in two batches each consisting of approximately 6,000 quasi-randomly selected individuals. Individuals with self-reported previous doctor's diagnosis of bronchitis and/or emphysema at baseline were excluded. Only metabolites measured in both batch samplings were included in analyses.

Data were processed according to the standard quality control pipeline for each cohort. In EPIC-Norfolk, metabolite measures were median normalized across run days (with medians set to 1), without imputation of missing values. Within each measurement batch, individuals with high levels of metabolite missingness were excluded. Metabolite measures were natural log transformed, windsorized at 5 SD and standardized ($\mu = 0$, SD = 1). In NAS, metabolite intensities were log transformed and *pareto* scaled. Those metabolites with a variance of 0 were excluded from further analysis (n=144). Minimum scaling imputation was practiced by assigning missing/unquantified values half the lowest value across all samples for each metabolite.

4.4. Statistical Analysis

Analyses were conducted initially in EPIC-Norfolk, and then in NAS. Linear regression models were run for each metabolite independently to identify those significantly associated with FEV_1 and FEV_1/FVC after adjustment for age, body mass index (BMI), and smoking. Sex was also included as a covariate in non-stratified EPIC-Norfolk models. Smoking was defined as: never, regular or quit (no subjects self-identified as an occasional smoker). Height and asthma status (self-reported previous doctor's diagnosis) were also included as covariates in EPIC-Norfolk models.

EPIC-Norfolk analyses were performed individually within each measurement batch, applying a minimum sample size threshold of >30 observations.

A fixed effects inverse variance weighted meta-analysis was performed to pool the results of two measurement batches. Metabolites were matched between measurement batches by the Metabolon-assigned "Chemical ID" or by metabolite name for those without chemical identification.

5. Conclusions

Metabolites associated with lung function parameters FEV_1 and FEV_1/FVC were identified in EPIC-Norfolk and validated in NAS using multiple linear regression models. Overall, ten metabolites were significantly associated with FEV_1 in EPIC-Norfolk and replicated in NAS with nominal significance, while four metabolites were significantly associated with FEV_1/FVC and replicated with nominal significance. After stratifying by gender, seven metabolites were significantly associated with FEV_1 in EPIC-Norfolk males and replicated in NAS with nominal significance, while three metabolites were significantly associated with FEV_1/FVC and replicated with nominal significance. Our study is first to identify metabolites significantly associated with FEV_1 lung function parameters, in an attempt to better understand the metabolome of respiratory health and to identify potential biomarkers that can be used for pulmonary care.

Supplementary Materials: The following are available online at www.mdpi.com/xxx/s1, **Table S1**: Baseline characteristics of included participants in NAS., **Table S2**: Number of metabolites observed at various nominal *p*-value cutoffs for FEV_1 in NAS, **Table S3**: Number of metabolites observed at various nominal *p*-value cutoffs for FEV_1/FVC in NAS. **Figure S1**: Significant metabolites by pathway in EPIC. **Figure S2**: Significant metabolites by pathway in NAS.

Table S1. Baseline characteristics of included participants in NAS.

	VARIABLE INFO	NUMBER (%), MEAN (SD)
N		439
AGE		75.12 (6.66)
ETHNICITY	White	431 (98.2%)
	Black	6 (1.4%)
	Other	2 (0.5%)
FEV_1 (L)		2.50 (0.60)
FEV_1/FVC		0.74 (0.43)
BMI (KG/M^2)		27.61 (4.12)
SMOKE STATUS	Regular[1]	19 (4.3%)
	Former	288 (65.6%)
	Never	132 (30.1%)

[1]*There were no subjects who self-identified as occasional smokers.*

Table S2. Number of metabolites observed at various nominal *p*-value cutoffs for FEV_1 in NAS.

Significance	N = 1140 metabolites (%)
P<0.05	36 (3.2)
P<0.01	8 (0.7)
P<0.001*	1 (0.09)

**No metabolites retained significance after Bonferroni correction.*

Table S3. Number of metabolites observed at various nominal *p*-value cutoffs for FEV_1/FVC in NAS.

Significance	N = 1140 metabolites (%)
P<0.05	33 (2.9)
P<0.01*	6 (0.5)

**No metabolites retained significance after Bonferroni correction.*

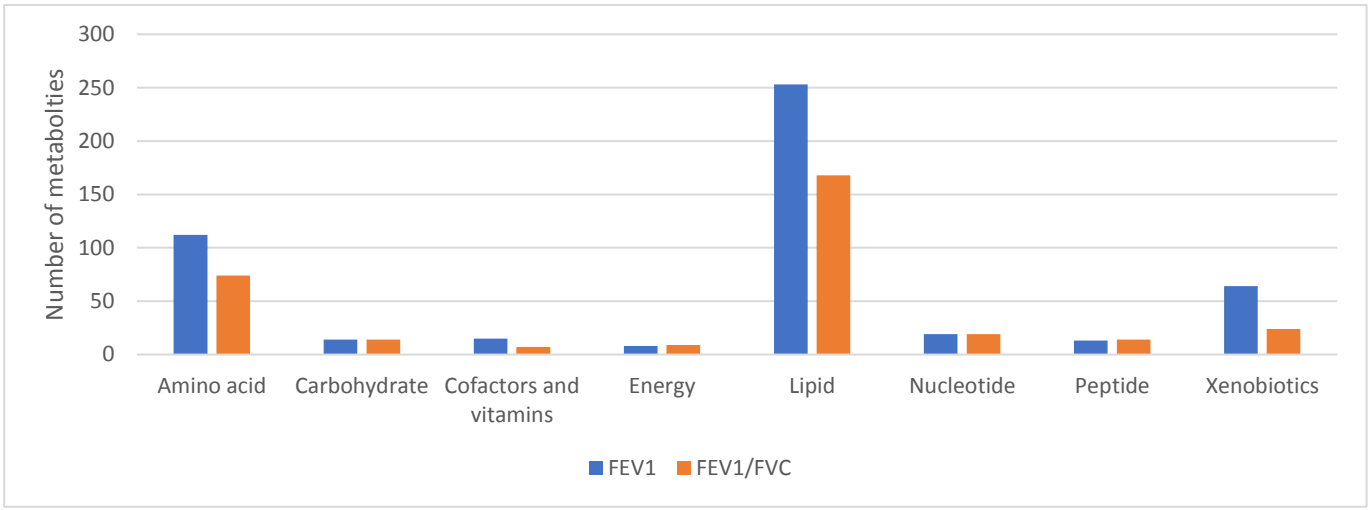


Figure S1. Significant metabolites by pathway in EPIC.

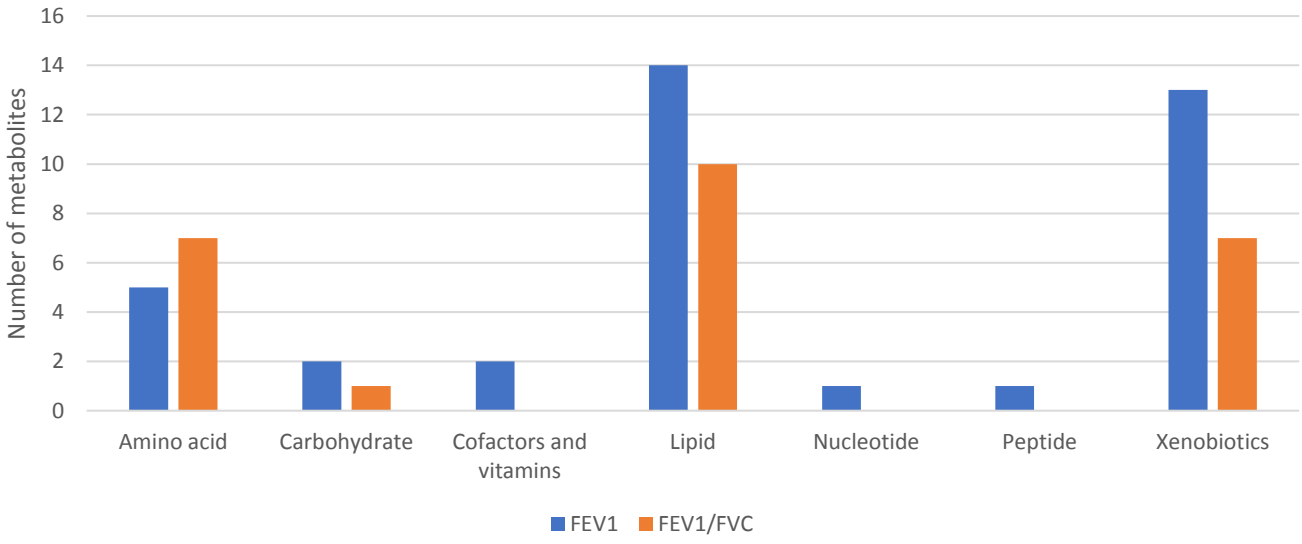


Figure S2. Significant metabolites by pathway in NAS.

Funding: Effort from PK and STW is supported by P01HL132825. Effort for RK and JALS is supported by R01HL123915 and W81XWH-17-1-0533. RK is additionally supported by K01HL146980-01. Efforts for JALS is also supported by R01HL141826. AS was supported by a Senior Research Career Scientist Award from the Clinical Science R&D Service, US Department of Veterans Affairs

Acknowledgments: A special thanks to Jody Sylvia (Brigham and Women's Hospital) for her mentorship and Elaine Dibbs (The United States Department of Veterans Affairs (VA)) for her data contributions. We thank the participants of both EPIC-Norfolk and NAS. We would also like to thank the Northeastern University Cooperative Education Program for enabling the work of HB.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Sharma, G. and J. Goodwin, *Effect of aging on respiratory system physiology and immunology.* Clin Interv Aging, 2006. **1**(3): p. 253-60.
2. Baker, M.G., et al., *The Use of Metabolomics to Identify Biological Signatures of Manganese Exposure.* Ann Work Expo Health, 2017. **61**(4): p. 406-415.
3. Patti, G.J., O. Yanes, and G. Siuzdak, *Innovation: Metabolomics: the apogee of the omics trilogy.* Nat Rev Mol Cell Biol, 2012. **13**(4): p. 263-9.
4. Fiehn, O., *Metabolomics--the link between genotypes and phenotypes.* Plant Mol Biol, 2002. **48**(1-2): p. 155-71.
5. Stringer, K.A., et al., *Metabolomics and Its Application to Acute Lung Diseases.* Front Immunol, 2016. **7**: p. 44.
6. Kelly, R.S., et al., *Asthma Metabolomics and the Potential for Integrative Omics in Research and the Clinic.* Chest, 2017. **151**(2): p. 262-277.
7. Kelly, R.S., et al., *Metabolomic profiling of lung function in Costa-Rican children with asthma.* Biochim Biophys Acta Mol Basis Dis, 2017. **1863**(6): p. 1590-1595.
8. Zuo, L., B.K. Pannell, and Z. Liu, *Characterization and redox mechanism of asthma in the elderly.* Oncotarget, 2016. **7**(18): p. 25010-21.
9. Day, N., et al., *EPIC-Norfolk: study design and characteristics of the cohort. European Prospective Investigation of Cancer.* Br J Cancer, 1999. **80 Suppl 1**: p. 95-103.
10. Spiro, A.I.B., R. , *The Normative Aging Study*, in *The Encyclopedia of Aging*. 2001, Springer Press: New York. p. 744-46.
11. Bender, R. and S. Lange, *Adjusting for multiple testing--when and how?* J Clin Epidemiol, 2001. **54**(4): p. 343-9.
12. Campbell, E.J. and S.S. Lefrak, *How aging affects the structure and function of the respiratory system.* Geriatrics, 1978. **33**(6): p. 68-74.
13. Lalley, P.M., *The aging respiratory system--pulmonary structure, function and neural control.* Respir Physiol Neurobiol, 2013. **187**(3): p. 199-210.
14. Melani, A.S., *Management of asthma in the elderly patient.* Clin Interv Aging, 2013. **8**: p. 913-22.
15. Battaglia, S., et al., *Asthma in the elderly: a different disease?* Breathe (Sheff), 2016. **12**(1): p. 18-28.

16. Divo, M., et al., *Comorbidities and risk of mortality in patients with chronic obstructive pulmonary disease.* Am J Respir Crit Care Med, 2012. **186**(2): p. 155-61.
17. Nobakht, M.G.B.F., et al., *The metabolomics of airway diseases, including COPD, asthma and cystic fibrosis.* Biomarkers, 2015. **20**(1): p. 5-16.
18. Ubhi, B.K., et al., *Metabolic profiling detects biomarkers of protein degradation in COPD patients.* Eur Respir J, 2012. **40**(2): p. 345-55.
19. Kelly, R.S., et al., *Partial Least Squares Discriminant Analysis and Bayesian Networks for Metabolomic Prediction of Childhood Asthma.* Metabolites, 2018. **8**(4).
20. Andrisic, L., et al., *Short overview on metabolomics approach to study pathophysiology of oxidative stress in cancer.* Redox Biol, 2018. **14**: p. 47-58.
21. Fitzpatrick, A.M., et al., *Children with severe asthma have unique oxidative stress-associated metabolomic profiles.* J Allergy Clin Immunol, 2014. **133**(1): p. 258-61 e1-8.
22. Kirkham, P.A. and P.J. Barnes, *Oxidative stress in COPD.* Chest, 2013. **144**(1): p. 266-273.
23. Holguin, F., *Oxidative stress in airway diseases.* Ann Am Thorac Soc, 2013. **10 Suppl**: p. S150-7.
24. Crowley, G., et al., *Metabolomics of World Trade Center-Lung Injury: a machine learning approach.* BMJ Open Respir Res, 2018. **5**(1): p. e000274.
25. Fabian, C.J., B.F. Kimler, and S.D. Hursting, *Omega-3 fatty acids for breast cancer prevention and survivorship.* Breast Cancer Res, 2015. **17**: p. 62.
26. Venter, C., et al., *EAACI position paper: Influence of dietary fatty acids on asthma, food allergy, and atopic dermatitis.* Allergy, 2019.
27. Durack, J., et al., *Delayed gut microbiota development in high-risk for asthma infants is temporarily modifiable by Lactobacillus supplementation.* Nat Commun, 2018. **9**(1): p. 707.
28. Miller, D.B., et al., *Ozone Exposure Increases Circulating Stress Hormones and Lipid Metabolites in Humans.* Am J Respir Crit Care Med, 2016. **193**(12): p. 1382-91.
29. Petersen, H., *Support Vector Machine Identifies a Set of Metabolites in Plasma that Predict Rapid FEV1 Decline in Humans and Mice.* Proceedings, 2016.
30. Kachroo, P., et al., *Fish oil supplementation during pregnancy is protective against asthma/wheeze in offspring.* J Allergy Clin Immunol Pract, 2019.
31. Vinaixa, M., et al., *A Guideline to Univariate Statistical Analysis for LC/MS-Based Untargeted Metabolomics-Derived Data.* Metabolites, 2012. **2**(4): p. 775-95.
32. Riboli, E., *Nutrition and cancer: background and rationale of the European Prospective Investigation into Cancer and Nutrition (EPIC).* Ann Oncol, 1992. **3**(10): p. 783-91.
33. Myint, P.K., et al., *Respiratory function and self-reported functional health: EPIC-Norfolk population study.* Eur Respir J, 2005. **26**(3): p. 494-502.
34. Sargeant, L.A., A. Jaeckel, and N.J. Wareham, *Interaction of vitamin C with the relation between smoking and obstructive airways disease in EPIC Norfolk. European Prospective Investigation into Cancer and Nutrition.* Eur Respir J, 2000. **16**(3): p. 397-403.
35. Sparrow, D., et al., *The relationship of nonspecific bronchial responsiveness to the occurrence of respiratory symptoms and decreased levels of pulmonary function. The Normative Aging Study.* Am Rev Respir Dis, 1987. **135**(6): p. 1255-60.
36. *Standardization of Spirometry.* American Review of Respiratory Disease, 1987. **136**: p. 1285-98.
37. *Standardization of Spirometry.* American Review of Respiratory Disease, 1994. **152**: p. 1107-36.

38. Evans, A.B., B; Mitchell, M; Liu, Q; Stewart, S; Dai, H; Dehaven, C; Miller, L, *High Resolution Mass Spectrometry Improves Data Quantity and Quality as Compared to Unit Mass Resolution Mass Spectrometry in High-Throughput Profiling Metabolomics.* Metabolomics, 2014. **4**(132).