

Metabolomics of World Trade Center-Lung Injury: a machine learning approach

George Crowley,¹ Sophia Kwon,¹ Syed Hissam Haider,¹ Erin J Caraher,¹ Rachel Lam,¹ David E St-Jules,² Mengling Liu,^{3,4} David J Prezant,^{5,6} Anna Nolan^{1,4,5}

To cite: Crowley G, Kwon S, Haider SH, *et al*. Metabolomics of World Trade Center-Lung Injury: a machine learning approach. *BMJ Open Resp Res* 2018;**5**:e000274. doi:10.1136/bmjresp-2017-000274

► Additional material is published online only. To view please visit the journal online (<http://dx.doi.org/10.1136/bmjresp-2017-000274>).

Received 27 December 2017
Revised 19 March 2018

ABSTRACT

Introduction Biomarkers of metabolic syndrome expressed soon after World Trade Center (WTC) exposure predict development of WTC Lung Injury (WTC-LI). The metabolome remains an untapped resource with potential to comprehensively characterise many aspects of WTC-LI. This case-control study identified a clinically relevant, robust subset of metabolic contributors of WTC-LI through comprehensive high-dimensional metabolic profiling and integration of machine learning techniques.

Methods Never-smoking, male, WTC-exposed firefighters with normal pre-9/11 lung function were segregated by post-9/11 lung function. Cases of WTC-LI (forced expiratory volume in 1s <lower limit of normal, n=15) and controls (n=15) were identified from previous cohorts. The metabolome of serum drawn within 6 months of 9/11 was quantified. Machine learning was used for dimension reduction to identify metabolites associated with WTC-LI.

Results 580 metabolites qualified for random forests (RF) analysis to identify a refined metabolite profile that yielded maximal class separation. RF of the refined profile correctly classified subjects with a 93.3% estimated success rate. 5 clusters of metabolites emerged within the refined profile. Prominent subpathways include known mediators of lung disease such as sphingolipids (elevated in cases of WTC-LI), and branched-chain amino acids (reduced in cases of WTC-LI). Principal component analysis of the refined profile explained 68.3% of variance in five components, demonstrating class separation.

Conclusion Analysis of the metabolome of WTC-exposed 9/11 rescue workers has identified biologically plausible pathways associated with loss of lung function. Since metabolites are proximal markers of disease processes, metabolites could capture the complexity of past exposures and better inform treatment. These pathways warrant further mechanistic research.

INTRODUCTION

Fire Department of New York (FDNY) rescue workers exposed to World Trade Center particulate matter (WTC-PM) developed lung disease similar to other exposed cohorts.¹⁻⁶ Characteristics of airflow obstruction predominated in symptomatic individuals that sought treatment.^{1,7,8} WTC-affected subjects continue

Key messages

- Can machine learning algorithms determine metabolomic signatures of World Trade Center Lung Injury (WTC-LI)?
- Machine learning techniques are an innovative, promising approach to capturing the systemic nature of the pathogenesis of particulate matter associated lung disease.
- We identified key pathways, such as sphingolipids and branched chain amino acids, that contribute to lung function loss after WTC particulate and toxin exposure.
- To learn how machine learning can be integral to identifying key mediators of lung function loss in this novel metabolomics investigation of WTC first responders.
- The metabolome may allow us to further identify biologically relevant therapeutic targets of particulate/toxin-associated lung disease.

to have their quality of life adversely impacted even 15 years after exposure.^{9,10}

Biomarkers of metabolic syndrome (MetSyn), vascular injury and inflammation predict developing World Trade Center Lung Injury (WTC-LI) (defined as forced expiratory volume in 1s percent predicted of normal ($FEV_{1, \%Pred}$) less than the lower limit of normal (LLN)).^{1,7,8,11} MetSyn is defined by inter-related risk factors of metabolic origin associated with end-organ disease and affects approximately one-third of the adult US population.^{12,13} A diagnosis of MetSyn is determined by the presence of at least three of the following comorbidities: abdominal obesity, insulin resistance, hypertriglyceridaemia, low high-density lipoprotein (HDL) and hypertension.¹² We adjusted to National Cholesterol Education Program Adult Treatment Panel III (NCEP ATP III) definitions of MetSyn for our cohort and diagnosed MetSyn as having three of five following criteria: systolic blood pressure (SBP) ≥ 130 mm Hg or diastolic BP (DBP) ≥ 85 mm Hg; HDL < 40 mg/



© Author(s) (or their employer(s)) 2018. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

Correspondence to

Dr Anna Nolan;
Anna.Nolan@med.nyu.edu

dL; triglycerides ≥ 150 mg/dL; insulin resistance as glucose ≥ 100 mg/dL; and body mass index (BMI) > 30 kg/m².

Our findings in the WTC-exposed FDNY cohort fit into a larger set of studies demonstrating the association of MetSyn, lipids and obstructive airways disease (OAD) after pollutant exposure.¹⁴ Prior studies have identified an association between metabolic mediators and disease severity following toxin exposure.^{15 16} Specifically, metal fume exposure secondary to welding has been linked to an increase in plasma unsaturated fatty acid (FA) levels. Consisting of particulates and metals (such as nickel, copper and iron), the welders' exposure is similar to the WTC exposure; however, there exist significant differences. The exposure at the WTC site was diverse. The destruction of the WTC complex pulverised 1.2 million tons of construction material.^{17 18} Extensive analysis of the particulates for size and composition showed that metals (vanadium, chromium, nickel, iron, silicon and mercury) supplemented common construction materials, such as powdered concrete, calcium carbonate and silicates.¹⁷ Additional toxins included fibrous glass, asbestos, components of jet fuel, fire retardants and dioxins.¹⁸ With exposure including high-intensity particulates spanning the ultrafine to coarse range and associated volatile compounds, our WTC cohort allows for novel metabolic observations.

Metabolic profiling (metabolomics), a comprehensive analytical technique that measures all small organic molecules detectable in biological fluids, provides a contemporaneous snapshot of the organism's physiology.¹⁹ The metabolome's assessment of low-molecular-weight compounds may prove the closest link to disease phenotype.²⁰ Furthermore, metabolomics provides a non-invasive functional genetics approach to understanding molecular complexity.²⁰

However, metabolomics uses a high-throughput platform and therefore optimising the analysis of this large data set is a primary challenge in 'omics research.²¹ A common approach, using significance value cut-offs in reducing the dimensionality of the data, subsumes the possibility of producing false discoveries, while discounting truly important compounds. For this reason, we turn to machine learning, which has the ability to analyse differential expression of individual metabolites, as well as consider how metabolites interact with each other to produce class separation. This may allow us to unmask metabolites that appear insignificant on their own, but reveal key phenotypic changes when analysed in tandem with other metabolites.

Machine learning methods include random forests (RF), neural networks, partial least squares and support vector machines. Our machine learning method of choice is RF because it is unbiased, performs well on datasets with sizes similar to ours, is non-parametric and rarely overfits a data set.²²

RF uses a large ensemble of decision trees to classify subjects, produces an unbiased estimate of model classification accuracy and measures each metabolite's

importance to model classification accuracy. This measure, called mean decrease accuracy, allows the researcher to narrow his focus to only those metabolites most important to class differentiation. We hypothesised that implementation of machine learning techniques on the metabolome of WTC-exposed FDNY firefighters may elucidate pathways of interest in progression to WTC-LI, reveal clustering patterns reflective of mechanistic action and enable discovery of bioactive lipid metabolites related to loss of lung function.

METHODS

Study design

Cases and controls were drawn from symptomatic subjects referred for subspecialty pulmonary examination (SPE) between 10/1/2001 and 3/10/2008, and underwent specialised pulmonary function testing as previously described.^{1 7 8} The parent cohort consisted of cases of WTC-LI (n=96) defined as $FEV_{1, \%Pred} < LLN$ at SPE and controls (n=127) selected as previously described.^{1 7 8} Subjects were chosen for untargeted metabolome assessment by application of a further set of inclusion criteria, including stable case/control assignment as ensured by most recent spirometry measures from annual health physical, and no chronic sinusitis diagnosis. Based on these criteria, 15 cases of WTC-LI were available from the previously studied group, and a 1:1 ratio of cases of WTC-LI to randomly selected controls was determined (figure 1).^{7 8} Specifically, cases of WTC-LI were defined as having $FEV_{1, \%Pred} < LLN$ as defined by National Health and Nutrition Examination Survey III and control subjects as $FEV_{1, \%Pred} \geq LLN$, as previously described.^{1 23}

Demographics

Demographic and clinical data were obtained from the WTC medical monitoring and treatment programme (MMTP). Exposure intensity is categorised as per the FDNY-WTC Exposure Intensity Index and is based on first arrival time at the WTC site, as previously described.^{5 24 25} Specifically, subjects are considered to have high exposure if they arrived during the morning of 11 September 2001, intermediate exposure if they arrived in the afternoon of 11 September 2001 and low exposure if they arrived on 12 September 2001. Duration is defined as the time in months spent at the WTC site performing rescue and recovery efforts. BMI was obtained at two time points: MMTP enrolment and SPE.

Metabolomics

Serum collected within 200 days after 9/11/2001 was processed and stored as previously described.^{1 7 8 11 26} Aliquots of serum were maintained at -80°C until processing and metabolomic quantification (see online supplemental methods). The bioinformatics system consisted of four major components: the Laboratory Information Management System, data extraction

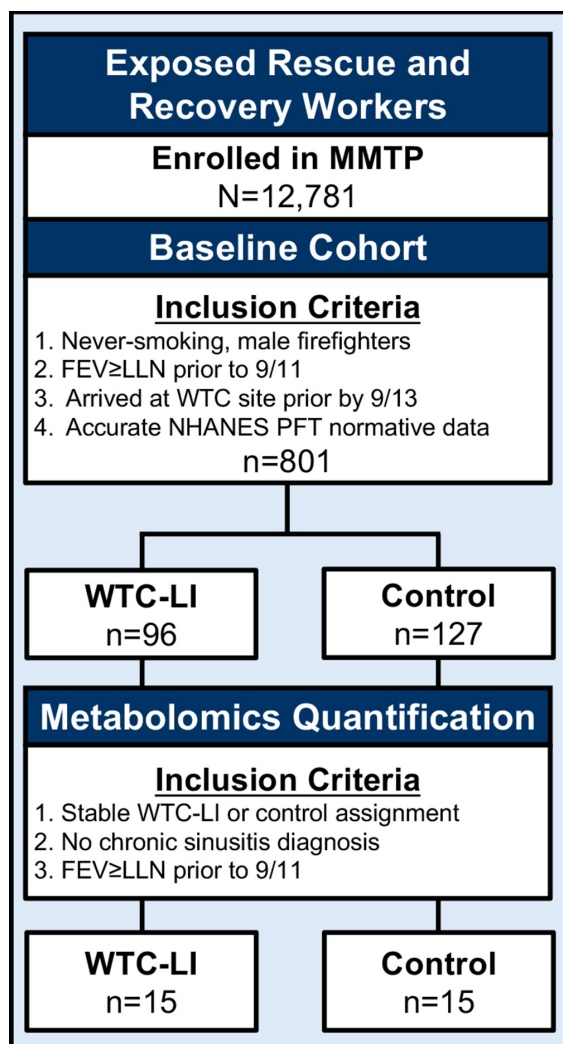


Figure 1 Study design. Metabolome assessment cases of World Trade Center Lung Injury (WTC-LI) and controls were selected. FEV, forced expiratory volume; LLN, lower limit of normal; MMTP, monitoring and treatment programme; NHANES PFT, National Health and Nutrition Examination Survey Pulmonary Function Test.

software, peak identification software and data processing tools for quality control, and compound identification. Quality control and curation were designed to ensure accurate and consistent identification of true chemical entities, and to remove those representing system artefacts, misassignments and background noise. Compounds were matched to library entries of retention index, mass and spectral data.^{27–30} Qualified metabolites were those detected in 80% or more of subjects per group with a relative SD of 15% or greater.³¹ In qualified metabolites, missing data were imputed using the minimum observed value of each compound.

Database management and statistics

SPSS V.23 (IBM) was used for data storage and handling. Continuous variables were expressed as median and IQR. Non-parametric Mann-Whitney U test was used to

compare continuous data. For categorical data, count and proportions were used to summarise and Pearson χ^2 was used for comparison.

Machine learning, dimension reduction and pattern recognition

Curated data of the qualified profile were subjected to RF (randomForest Package R V.3.4.3, R-Project). A refined metabolite profile was developed by including the top 5% of metabolites important to class separation as measured by mean decrease accuracy, a measure of the decrease in classification accuracy of the model should the given metabolite be removed. RF was rerun with the refined profile to evaluate the classification accuracy produced by the refined profile.²² Ten replicates of each RF model were run to assure stability of the model. The best RF model was identified by the lowest estimated out-of-bag error rate of classification, a permutation-based assessment of internal model validity. RF models consisted of 10^6 and 10^3 trees for the qualified and refined profiles, respectively, with $n = \sqrt{\text{number of metabolites in profile}}$ sampled with replacement at each node.

Principal component analysis (PCA) (SPSS V.23, IBM) was implemented to view the case/control separation in qualified metabolites compared with the machine learning-generated refined profile. PCA is a tool used in exploratory data analysis, allows for dimension reduction while preserving the variance of the data and can be used in the construction of predictive models. To perform PCA, each attribute was mean-centred, normalised and projected onto the eigenbasis of the correlation matrix of all attributes. The number of components retained was determined based on analysis of the scree plot. A determination of the variance explained by PCA was quantified as the summation of the per cent total variance explained by the components identified by scree plot. Loading weights of metabolites on the principal components were plotted to identify potential relationships between correlated metabolites. We first used the qualified profile of 580 metabolites in the 30 subjects labelled by case/control status, then applied this method to the refined metabolite profile identified by RF.

Additionally, unsupervised two-way hierarchical clustering was performed on the refined profile's correlation matrix using Pearson correlation as a distance metric and average linkage (Cluster V.3.0, Java Treeview, Eisen Lab).

RESULTS

Demographics

Parent cohort

Cases of WTC-LI (n=96) were similar to controls (n=127) in the parent cohort in age on 9/11, clinically available lipids (triglycerides, HDL, low-density lipoprotein (LDL)), SBP and DBP. FEV_{1, %Pred}, forced vital capacity per cent predicted of normal (FVC_{%Pred}), FEV₁/FVC and duration of time (months) spent at WTC site were significantly decreased in WTC-LI cases compared with

Table 1 Clinical measures, biomarker prevalence and model definition

Measure	Parent cohort		Metabolomics subcohort	
	Controls n=127	WTC-LI n=96	Controls n=15	WTC-LI n=15
PFT at SPE				
FEV _{1, % Pred} [†]	93 (85–99)	72 (66–75)	92 (90–98)	73 (70–75)
FVC _{% Pred} ^{*†}	96 (89–103)	79 (72–85)	97 (95–100)	78 (75–88)
FEV ₁ /FVC ^{*†}	76 (73–80)	72 (65–77)	75 (71–82)	71 (64–76)
BMI (kg/m²)				
MMTP entry ^{*†‡}	28 (26–30)	29 (27–31)	26 (25–27)	29 (26–31)
SPE ^{*†‡}	29 (27–31)	30 (28–34)	26 (24–28)	30 (28–31)
Age on 9/11 (years)	41 (37–45)	40 (36–45)	42 (38–46)	39 (37–46)
Exposure N (%)				
Low [*]	13 (10%)	21 (22%)	1 (7%)	4 (27%)
Intermediate [*]	85 (67%)	46 (48%)	11 (73%)	8 (53%)
High [*]	29 (23%)	29 (30%)	3 (20%)	3 (20%)
Duration (months) [*]	3 (1–5)	1 (1–4)	2 (1–5)	3 (1–5)
Lipids (mg/dL)				
Triglycerides	164 (98–238)	157 (107–243)	126 (99–237)	133 (110–201)
HDL [§]	47 (40–55)	43 (38–54)	48 (45–57)	42 (35–57)
LDL [¶]	131 (104–157)	131 (108–153)	134 (100–144)	147 (115–162)
Heart rate ^{*†‡§} (beats/min)	72 (66–76)	74 (69–80)	66 (64–70)	75 (68–79)
BP (mm Hg)				
Systolic ^{†‡§}	114 (108–124)	120 (110–128)	110 (100–112)	119 (108–129)
Diastolic ^{†‡§}	70 (70–80)	75 (70–81)	70 (60–72)	71 (66–80)

Values in median (IQR) or N (%) as indicated. Significant by Mann-Whitney U or χ^2 between:

*127 vs 96.

†15 vs 15.

‡Controls 127 vs 15.

§Data available for 14 subcohort WTC-LI cases.

¶Data available for 13 subcohort WTC-LI cases; all comparisons between cases 96 vs 15 were insignificant.

BMI, body mass index; BP, blood pressure; FEV_{1, %Pred}, pred forced expiratory volume in 1 s per cent predicted of normal; FVC_{%Pred}, pred forced vital capacity per cent predicted of normal; HDL, high-density lipoprotein; LDL, low-density lipoprotein; MMTP, medical monitoring and treatment programme; PFT, Pulmonary Function Test; SPE, subspecialty pulmonary examination; WTC-LI, World Trade Center Lung Injury.

controls, while BMI (at MMTP and SPE), exposure intensity, and heart rate (HR) were significantly increased in WTC-LI cases compared with controls in the parent cohort (table 1).

Metabolomics subcohort

Cases of WTC-LI (n=15) were similar to controls (n=15) in age on 9/11, exposure intensity, duration, lipids (triglycerides, HDL, LDL), leucocyte subtype percentages (neutrophil, lymphocyte, monocyte, basophil and eosinophil), serum levels of glucose, sodium, chloride, potassium, uric acid, total protein and DBP. Cases of WTC-LI (n=15) had significantly decreased FEV_{1, %Pred}, FVC_{%Pred} and FEV₁/FVC at SPE compared with controls (n=15) and also significantly increased BMI at both MMTP and SPE. Lastly, cases of WTC-LI had significantly increased SBP and HR (table 1).

Overall, cases of WTC-LI with metabolome (n=15) assessed were similar to the controls in the parent cohort (n=127) in spirometry, BMI, age on 9/11, exposure intensity, lipids (triglycerides, HDL, LDL), blood pressure, leucocyte subtype percentages, serum levels of glucose, sodium, chloride, potassium, uric acid and total protein (table 1).

Controls with metabolome assessed (n=15) and controls in the parent cohort (n=127) also were similar in clinically available lipids, leucocyte subtype percentages, serum levels of glucose, sodium, chloride, potassium, uric acid and total protein. Controls (n=15) differed from their parent cohort (n=127) in BMI at MMTP and SPE, HR, SBP and DBP; however, controls consistently have significantly lower BMI at MMTP and SPE compared with cases of WTC-LI, and the same trend remains true for HR, SBP and DBP. The cases of WTC-LI with metabolome

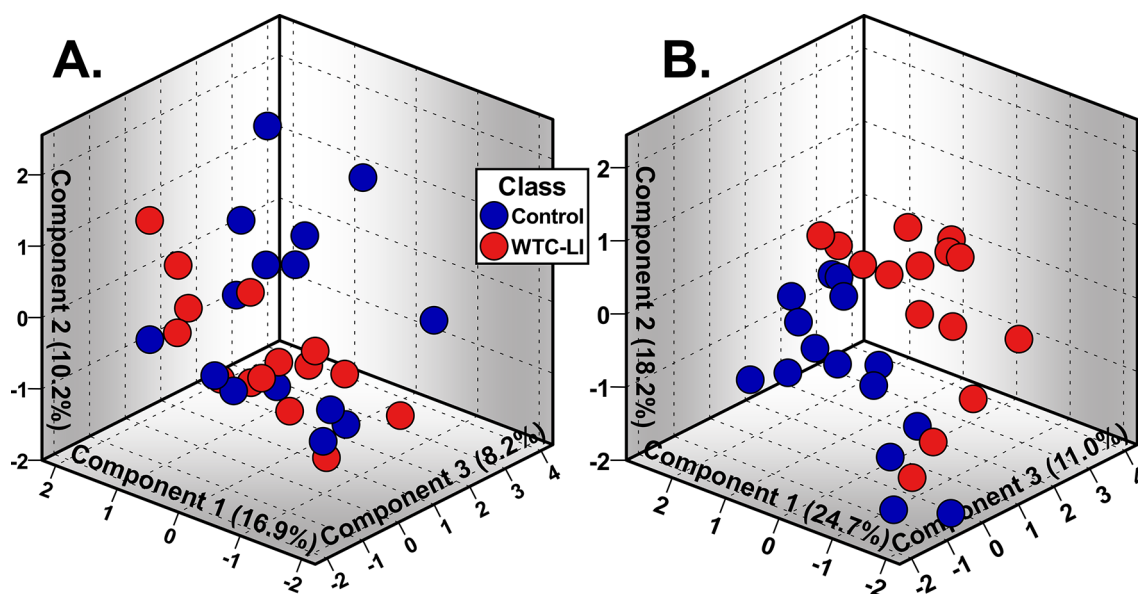


Figure 2 Demonstration of model optimisation: principal component analysis (PCA) scores plot. (A) PCA of the qualified profile reveals heterogeneity in the data. (B) PCA of the refined profile demonstrates improved class separation produced by the refined profile compared with initial PCA (A). WTC-LI, World Trade Center Lung Injury.

assessed and cases in the parent cohort were similar in all of the described clinical characteristics and biochemical measurements (table 1).

Metabolomics

Of 765 metabolites detected, 580 qualified, were included in further analysis and are summarised, (online supplementary table S1). Initial PCA of the qualified profile revealed heterogeneity and overlap in metabolite expression levels between cases of WTC-LI and controls (figure 2A) and ill-defined metabolite clustering (online supplementary figure S1). Qualified metabolites were subjected to RF to reduce the dimensionality of the data by discovering those metabolites most relevant to our outcome. Mean decrease accuracy was measured for every metabolite to assess variable importance, and the top 5% were included in the refined metabolite profile (figure 3 and online supplementary table S2). RF of the refined metabolite profile achieved a 6.7% out-of-bag estimated error rate (93.3% estimated accuracy) (figure 3).

PCA of the refined metabolite profile captured 68.3% of variance in the five components retained based on examination of the scree plot (online supplementary figure S2) and demonstrated improved class separation compared with the initial PCA (figure 2B).

In PCA, we see that features of the metabolome years prediagnosis can separate cases of WTC-LI from controls. Thereafter, we can begin to examine how these features may mechanistically interact.

Characteristics of the refined profile

The PCA loading weights plot of the refined profile and correlation heatmap suggest metabolite associations through clustering patterns (figure 4A, B). Overall, we identified five clusters (C1–5) of metabolites based on the

correlation heatmap (figure 4B). These clusters included metabolites from the following pathways: amino acids, carbohydrates, cofactors/vitamins, lipids, nucleotides, peptides and xenobiotics.

The *lipid-predominant cluster* (C1) consisted of 1-palmitoyl-2-arachidonoyl-GPC (16:0/20:4n6), 1-stearoyl-2-arachidonoyl-GPC (18:0/20:4), arachidonate (20:4n6), uridine, 2-hydroxypalmitate, 2-hydroxystearate and lignoceroyl sphingomyelin (d18:1/24:0) (figure 4B). In PCA, these metabolites clustered closely (figure 4A). A member of C1, arachidonate (20:4n6), had the third highest mean decrease accuracy score and was therefore the third most important metabolite to class separation (figure 3).

Amino acid metabolites and two ascorbate and aldarate metabolites were identified in C2: arabinonate/xylonate, gulonate, n-acetylglutamine, methylsuccinate, n2-acetyllysine, lanthionine and threonate. This group of metabolites included n2-acetyllysine, the metabolite with the highest mean decrease accuracy score (figure 3) and clustered tightly in the PCA scores plot (figure 4A).

Amino acids and lipids were key co-contributors in C3, which consisted of proline, dimethylglycine, propionylcarnitine, isobutyrylcarnitine, azelate (nonanedioate), vanillylmandelate (VMA), prolylglycine and 4-methylcatechol sulfate (figure 4B). Interestingly, several members of C2 bore strong correlations with members of C3 (figure 4B).

In contrast to the prior clusters, sphingolipids were identified in C4, which consisted of sphingosine, sphinganine and sphingosine 1-phosphate (figure 4B). This group of metabolites was tightly grouped in the PCA scores plot (figure 4A). Furthermore, this cluster contained the metabolite with the second highest mean decrease accuracy score, sphingosine (figure 3).

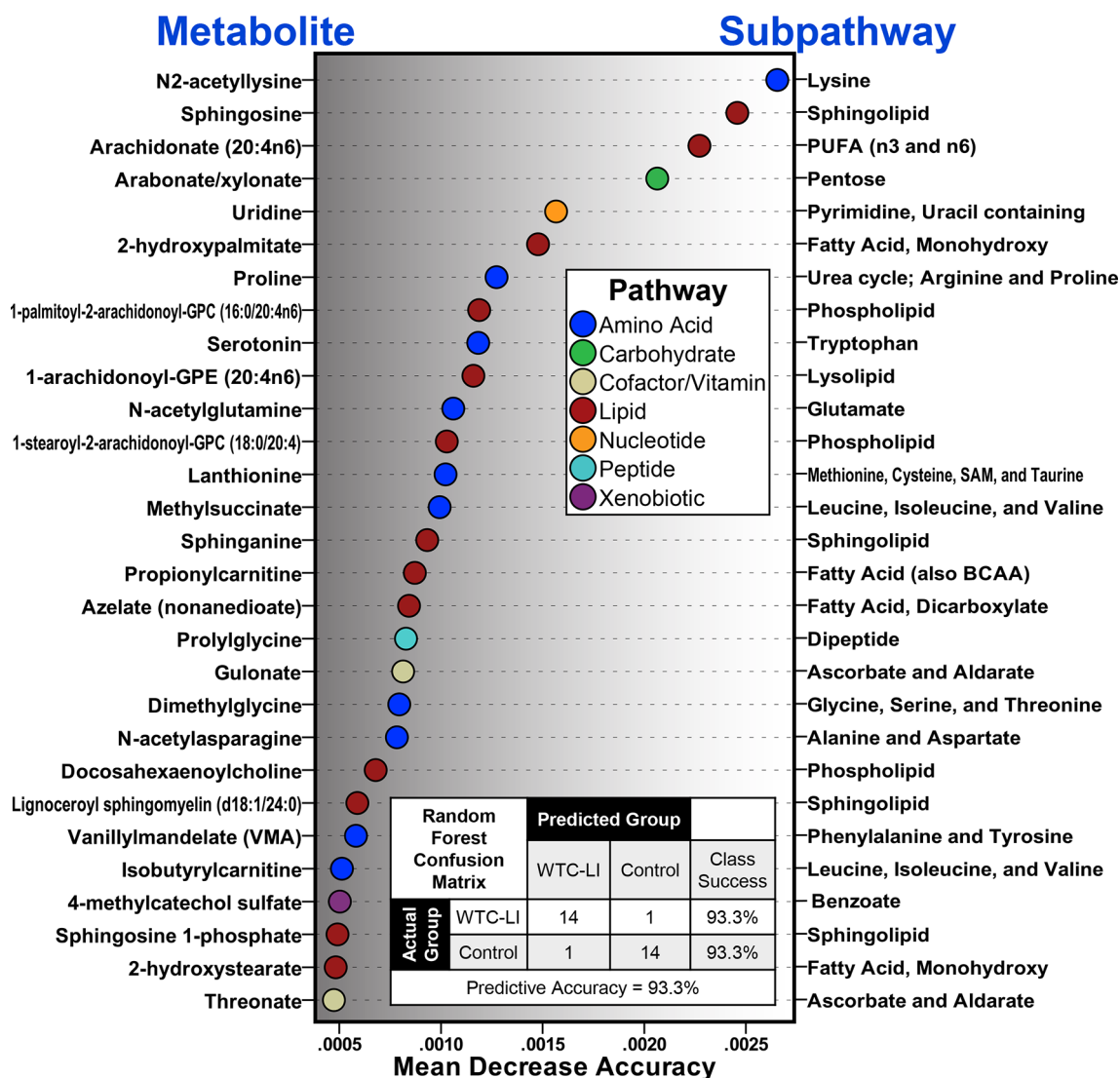


Figure 3 Random forests (RF) variable importance in projection. RF variable importance in projection is measured by mean decrease accuracy; the top 5% of metabolites important to class separation are shown. The confusion matrix shows classification accuracy of the refined profile. PUFA, polyunsaturated fatty acids.

Finally, a phospholipid, docosahexaenoylcholine, and an alanine and aspartate metabolite, n-acetylasparagine, were found in C5 (figure 4B). Members of C5 were strongly correlated with metabolites of other clusters, namely C1 (figure 4B).

DISCUSSION

The serum metabolome of the WTC-exposed firefighters soon after 11 September 2001 is metabolically rich and empowered by a well-described cohort that strengthens associations in the investigation of WTC-LI and systemic pathology; however, these associations are limited in interpretation as our study's samples are entirely collected within 200 days of WTC exposure. Although we can speculate on the metabolic environment prior to 11 September, inferences can only be made assuming that any individual component of MetSyn may not clinically change within such a short period of time. The state of health prior to

exposure can only speculatively, but not definitively, be considered by review of the early blood draws post exposure.

Prior literature closely links concurrent metabolic disease and lung function. In contrast, we have previously shown that these associations of MetSyn and future lung function loss occur in a cohort with normal lung function prior to exposure. Thus, we suggest that the metabolic environment soon after exposure can trigger an inflammatory cascade that contributes to persistent and irreversible lung damage.

Although the metabolome reflects the complex interaction between many different parent cells, our data suggest that the most biologically active WTC-LI-associated metabolites reflect lipids and amino acids. Several of these have been identified as associated with broader categories of OAD, and, in line with our hypothesis, these metabolites clustered in patterns representative of their established

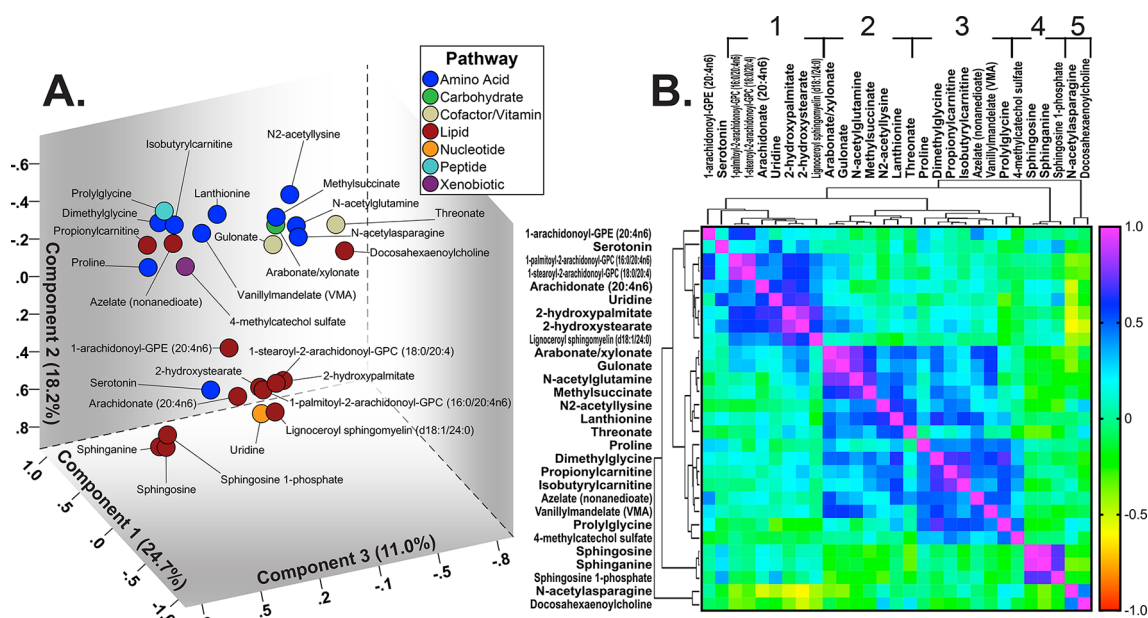


Figure 4 (A) Refined profile principal component analysis loading weights plot was used to derive insight into possible association of biomarkers. (B) Correlation heatmap. Correlation matrix of refined profile subjected to hierarchical clustering using Pearson correlation as a distance metric.

mechanisms; however, other clusters contain novel metabolites. Overall, these metabolite clusters represent biologically plausible signalling cascades.

Our exclusively sphingolipid cluster, C4, highlighted the importance of these established inflammatory and metabolic mediators. Sphingosine 1-phosphate is a pleiotropic inflammatory mediator involved in immune cell trafficking and asthmatic hyper-reactivity induction.³² Elevation of sphingosine, a product of sphingosine 1-phosphate degeneration, in cases of WTC-LI may indicate decreased bioavailability of sphingosine 1-phosphate, leading to compromised vascular integrity.³² We saw a decrease in sphingosine 1-phosphate levels in cases of WTC-LI. Of interest is our finding that cases of WTC-LI also had significantly higher HR and SBP.

In contrast to the homogeneity of C4, the metabolites in C1 were far more diverse, including lipids from a variety of subpathways. The importance of lignoceroyl sphingomyelin (d18:1/24:0) proposes that ceramide, a building block of sphingosine and a metabolite of LDL and very-low-density lipoprotein, may be involved in WTC-LI pathogenesis. Sphingomyelins are involved in ceramide synthesis and metabolism. Ceramide is involved in the synthesis and degradation of ceramide 1-phosphate. Ceramide 1-phosphate triggers release of arachidonate (20:4n6), the third most important metabolite to class separation in this study and a key mediator in the inflammatory cyclooxygenase (COX) pathway.³² Arachidonate (20:4n6), being the primary substrate in eicosanoid production, is involved in immune cell recruitment.^{33,34} Additionally, eicosanoids are pivotal in inflammatory pathogenesis. Moreover, changes in ceramide 1-phosphate correlate directly with changes in sphingosine and sphingosine 1-phosphate, suggesting that ceramide 1-phosphate is a precursor of sphingosine

1-phosphate.³² Together, these data elucidate interconnected pathways involved in early response to WTC-exposure that differentiate well cases of WTC-LI and controls based on the metabolome years pre diagnosis.

Chronic inflammation has been identified as a probable mediator of loss of lung function after exposure to WTC-PM. We know that metabolic dysfunction is not only a predictor of WTC-LI but also causes chronic inflammation. Azelate, a saturated FA in C1, was important to class separation. Saturated FAs activate TLR-4 in a mechanistic link between systemic inflammation and obesity, a key component of metabolic dysfunction.³² This activation of TLR-4 induces ceramide biosynthetic genes required for TLR-4-dependent insulin resistance.³² Glucose, a marker of insulin resistance, is a predictor of WTC-LI in the entire WTC-exposed firefighter cohort.

Notably, several classes of FAs were revealed by this analysis. In C1, lignoceroyl sphingomyelin (d18:1/24:0) clustered with polyunsaturated fatty acids (PUFA)s including arachidonate (20:4n6), and FAs, monohydroxy, as well as phospholipids. By examination of figure 5, it becomes apparent that phospholipid and sphingolipid metabolism may be linked by FA metabolism. While little is known about the specific FAs and phospholipids identified in this analysis, we do know the inflammatory roles of the PUFA arachidonate (20:4n6). The metabolites correlated with arachidonate (20:4n6) may also be related to immune cell recruitment. Evidence of this relationship is the central cell membrane role of phospholipids and sphingolipids, that sphingosine 1-phosphate is a potent pro-inflammatory mediator and neutrophil chemotactant, and that elevated sphingolipids may be indicative of an increased, disproportionate response to perturbation. The information presented here warrants mechanistic research into

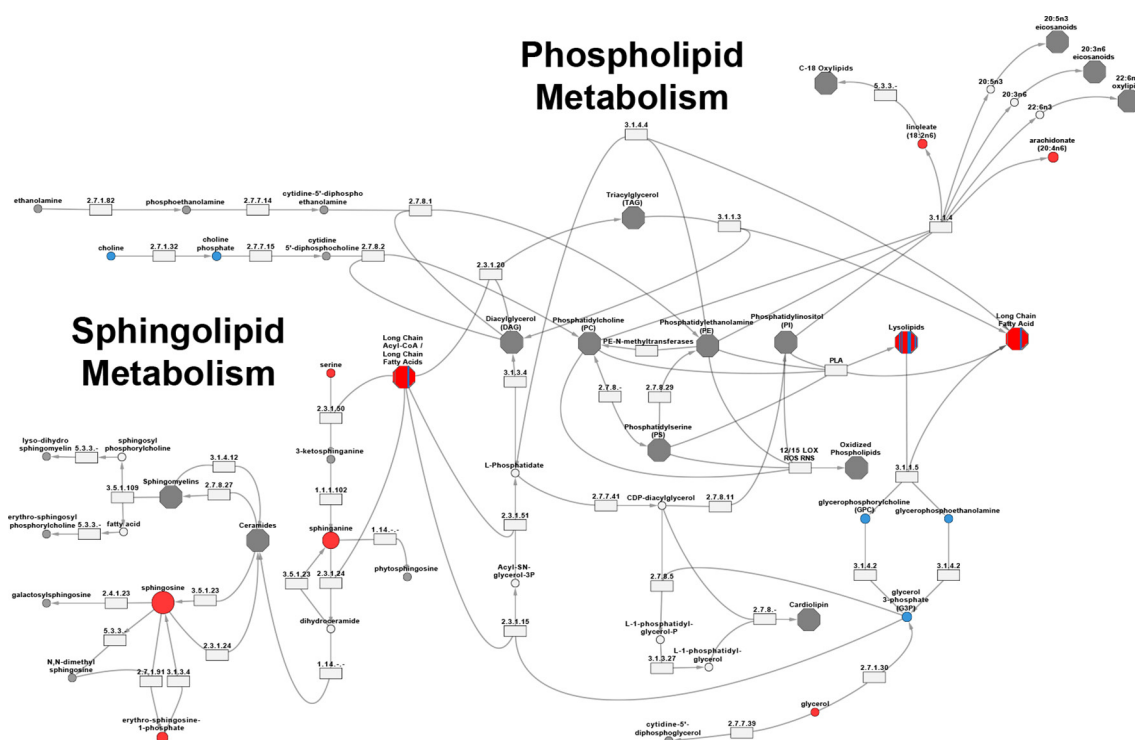


Figure 5 Pathway visualisation. Metabolic pathways of sphingolipids and phospholipids reveal pathways involving key metabolites, and that sphingolipid metabolism is linked with phospholipid metabolism by long-chain fatty acid metabolism. Node size correlates to fold change (red—up, blue—down, cases of World Trade Center Lung Injury/control).

this potential pathway as a therapeutic target to ameliorate derailed inflammatory responses.

As with previously discussed clusters, C2 contains some well-understood metabolites and some with unexplored roles. The most important metabolite to class separation is *n*-2-acetyllsine, a lysine metabolite; lysine is an essential amino acid involved in the formation of protein and acetylation of lysine residues plays a key role in epigenetic modulation of inflammatory responses, creating bromodomain docking sites.³⁵ In murine macrophages, pharmacological inhibition of a bromodomain attenuated nuclear factor- κ B-directed production of nitric oxide and interleukin-6.³⁵

Other essential amino acids presented in C2, namely several leucine, isoleucine and valine metabolites. These branched-chain amino acids (BCAAs) are hypothesised to be biologically active in respiratory illness due to their high anabolic capacity.³⁶ Attenuated blood BCAA concentrations have been observed in patients with chronic obstructive pulmonary disease (COPD).^{37–39} Dietary supplementation with BCAAs ameliorates COPD-related respiratory muscle weakness and weight loss, and low serum BCAA concentration has been identified in COPD cohorts.^{37 40 41} This is the first data suggesting low serum BCAA concentration may predispose to PM-related lung disease in our WTC FDNY cohort.

One well-known metabolite found in C3, VMA, is indicative of epinephrine and norepinephrine release (both known stress hormones). Epinephrine is responsible for increased release of FAs and is a potential explanation of raised levels of FAs, such as propionylcarnitine and azelate

(nonanedioate) (C3), in cases of WTC-LI. The pleiotropic inflammatory transcription factor c-Jun N-terminal kinases (JNK) is activated by exposure to free FAs and is involved in the signalling cascade of the receptor for advanced glycation end products (RAGE).⁴² We are currently investigating the roles of RAGE and JNK in WTC-LI.^{23 43 44}

As hypothesised, this analysis can identify biologically plausible signalling pathways relevant to the development of WTC-LI. While established mediators such as sphingolipids were reviewed, several potentially novel mediators, including FAs, lysine, BCAAs (in their predispositional nature), glutamine and stress hormone metabolites, have been elucidated. Some of our findings, especially the COX mediators, sphingolipids and BCAAs, mimic similar findings in COPD and pollutant-exposed cohorts. The similarities observed between the metabolome of this select group of WTC-exposed firefighters and that of other cohorts with inflammatory disease support generalisability to OAD cohorts, and support validity of findings in this select group of WTC-exposed firefighters (additional discussion of metabolites identified in the clustering can be found in the supplement).

Our study has several limitations. The metabolome provides only a contemporaneous snapshot of small organic compounds present in serum. As such, our ability to single out only WTC-LI-pathogenesis-related metabolites is limited; however, we attempted to mitigate the prevalence of spurious discoveries through implementation of machine learning algorithms. Given the limited sample size of the current study, confounders have been relatively well controlled by selection from a homogeneous subject

pool adhering to numerous criteria. Assuringly, the groups significantly differ in few metabolic biomarkers, only SBP, HR and BMI; other metabolic biomarkers and biomarkers of MetSyn are not significantly different between groups. Effects due to baseline BMI differences were mitigated by case definition based on percent predicted values. Additional unknown effects due to differences in exposure and duration of exposure are also minimised by the absence of significant differences in these variables between cases of WTC-LI and controls in subjects with metabolome assessed. Because distant history of smoking can predispose to non-resolving inflammation, our primary distinction is between 'ever' and 'never' smokers. Only never-smokers were included in this analysis; the metabolomic profile of ever-smoking WTC-exposed subjects remains unknown. Additionally, further profiling, including genomic, of this cohort must be performed to understand WTC-LI susceptibility.

It is important to note that any machine learning model is specific to its training data. However, the bootstrap-aggregating procedure followed by RF helps avoid overfitting that may typically be produced by a high variable-to-observation ratio. Simply put, the potential for false discovery is decreased by the high numbers of decision trees grown. Another aspect of this analysis is its strength in identifying intrinsic data patterns. While these patterns do not establish causal relationships, we assess model validity by out-of-bag classification procedures and support findings with pertinent literature. Due to the lack of an independent validation cohort, our findings have not been validated yet, but we acknowledge that this type of external validation would significantly strengthen the replicability and generalisability of our findings and methods.

Findings from this study have identified several metabolic pathways that may contribute to the pathogenesis of WTC-LI and similar lung injury-related conditions. The clusters differentiating WTC-exposed subjects with and without lung injury were primarily composed of lipid and protein metabolites, and therefore diet may be an important modifiable determinant of WTC-LI risk.

Dietary FAs may be able to alter the lipidomic signature that characterised the WTC-LI subjects. Specifically, a diet that is low in saturated fat intake and has a low omega-6-to-omega-3 FA ratio may help to correct high ceramide and the imbalance in phospholipid-derived long-chain PUFA metabolites (eg, arachidonate, docosahexaenoylcholine), which could have downstream beneficial effects on inflammatory and insulin signalling pathways.^{45–47}

Diet may also help address the altered amino acid profile seen in WTC-LI subjects. In patients with advanced lung disease (COPD), BCAA supplements have been found to increase BCAA concentrations and improve health outcomes,^{38, 48} however, in COPD, inadequate intake of energy, protein and BCAAs likely contributes importantly to the low BCAA concentrations. Although there is no data yet available on the dietary intakes of WTC-LI subjects, based on their sociodemographic background and prevalence of adiposity-based chronic disease and cardiometabolic risk factors, we assume that they tend to consume

adequate energy and follow a typical Western diet, which, especially for males, tends to be high in protein.⁴⁹

Importantly, it remains unclear to what extent, if any, diet contributes to the metabolomic signature of WTC-LI subjects, and what is the optimal diet for preventing, managing and treating WTC-LI and similar lung injuries. Additional research, including randomised clinical trials, is needed to examine the impact of diet and other modifiable lifestyle behaviours on WTC-LI progression.

Author affiliations

¹Department of Medicine, Division of Pulmonary, Critical Care and Sleep Medicine, New York University School of Medicine, New York, USA

²Department of Population Health, Division of Health and Behavior, New York University School of Medicine, New York, USA

³Department of Environmental Medicine, New York University School of Medicine, New York, USA

⁴Bureau of Health Services and Office of Medical Affairs, Fire Department of New York, Brooklyn, New York, USA

⁵Department of Population Health, Division of Biostatistics, New York University School of Medicine, New York, USA

⁶Department of Medicine, Pulmonary Medicine Division, Montefiore Medical Center and Albert Einstein College of Medicine, Brooklyn, New York, USA

Acknowledgements The authors thank the FDNY first responders for their selflessness and sacrifice in the line of duty.

Contributors All authors made substantial contributions to the study. All authors participated in study conception and design, data analysis and interpretation, and manuscript preparation. Primary investigator: AN; study design: AN, SK and GC; statistical analysis: GC, SK, ML and AN; data interpretation: GC, SK, SHH, EJC, RL, DES, ML, DJP and AN. All authors participated in writing and revision of the report and approval of the final version.

Funding NHLBI R01HL119326, CDC/NIOSH U01-OH011300 and Contract # 200-2011-3978. This work was also partially funded by the NYU-HHC CTSI supported by grant UL1TR000038 from the National Center for Advancing Translational Sciences of the NIH and by the Saperstein Scholars Fund.

Disclaimer The funding agencies did not participate in the study design; collection, analysis and interpretation of data; in the writing of the report; and in the decision to submit the article for publication.

Competing interests None declared.

Patient consent Not required.

Ethics approval All subjects, at the time of enrolment, consented to analysis of their information and samples for research according to Institutional Review Board approved protocols at Montefiore Medical Center (#07-09-320) and New York University (#16-01412).

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement AN is the primary investigator and guarantor of the content of this manuscript, including the data and analysis. Sharing of human data is governed by the World Trade Center (WTC) Clinical Center of Excellence program maintained by the Fire Department of New York (FDNY). All investigators will need to enter into a data use agreement with the FDNY WTC Clinical Center of Excellence. Additional information about this database may be obtained through DJP. He can be reached by email at prezand@fdny.nyc.gov.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

REFERENCES

1. Naveed B, Weiden MD, Kwon S, *et al*. Metabolic syndrome biomarkers predict lung function impairment: a nested case-control study. *Am J Respir Crit Care Med* 2012;185:392–9.

2. Friedman SM, Maslow CB, Reibman J, *et al.* Case-control study of lung function in World Trade Center Health Registry area residents and workers. *Am J Respir Crit Care Med* 2011;184:582–9.
3. Weiden MD, Kwon S, Caraher E, *et al.* Biomarkers of World Trade Center Particulate Matter Exposure: Physiology of Distal Airway and Blood Biomarkers that Predict FEV₁ Decline. *Semin Respir Crit Care Med* 2015;36:323–33.
4. Banauch GI, Hall C, Weiden M, *et al.* Pulmonary function after exposure to the World Trade Center collapse in the New York City Fire Department. *Am J Respir Crit Care Med* 2006;174:312–9.
5. Prezant DJ, Weiden M, Banauch GI, *et al.* Cough and bronchial responsiveness in firefighters at the World Trade Center site. *N Engl J Med* 2002;347:806–15.
6. Aldrich TK, Gustave J, Hall CB, *et al.* Lung function in rescue workers at the World Trade Center after 7 years. *N Engl J Med* 2010;362:1263–72.
7. Weiden MD, Naveed B, Kwon S, *et al.* Cardiovascular biomarkers predict susceptibility to lung injury in World Trade Center dust-exposed firefighters. *Eur Respir J* 2013;41:1023–30.
8. Nolan A, Naveed B, Comfort AL, *et al.* Inflammatory biomarkers predict airflow obstruction after exposure to World Trade Center dust. *Chest* 2012;142:412–8.
9. Webber MP, Glaser MS, Weakley J, *et al.* Physician-diagnosed respiratory conditions and mental health symptoms 7–9 years following the World Trade Center disaster. *Am J Ind Med* 2011;54:661–71.
10. Haider SH, Kwon S, Lam R, *et al.* Predictive Biomarkers of Gastroesophageal Reflux Disease and Barrett's Esophagus in World Trade Center Exposed Firefighters: a 15 Year Longitudinal Study. *Sci Rep* 2018;8:3106.
11. Schenck EJ, Echevarria GC, Girvin FG, *et al.* Enlarged pulmonary artery is predicted by vascular injury biomarkers and is associated with WTC-Lung Injury in exposed fire fighters: a case-control study. *BMJ Open* 2014;4:e005575.
12. Grundy SM, Cleeman JI, Daniels SR, *et al.* Diagnosis and management of the metabolic syndrome: an American Heart Association/National Heart, Lung, and Blood Institute Scientific Statement. *Circulation* 2005;112:2735–52.
13. Aguilar M, Bhuket T, Torres S, *et al.* Prevalence of the metabolic syndrome in the United States, 2003–2012. *JAMA* 2015;313:1973–4.
14. Tiengo A, Fadini GP, Avogaro A. The metabolic syndrome, diabetes and lung dysfunction. *Diabetes Metab* 2008;34:447–54.
15. Wei Y, Wang Z, Chang CY, *et al.* Global metabolomic profiling reveals an association of metal fume exposure and plasma unsaturated fatty acids. *PLoS One* 2013;8:e77413.
16. Dudka I, Kossowska B, Senhadri H, *et al.* Metabonomic analysis of serum of workers occupationally exposed to arsenic, cadmium and lead for biomarker research: a preliminary study. *Environ Int* 2014;68:71–81.
17. McGee JK, Chen LC, Cohen MD, *et al.* Chemical analysis of World Trade Center fine particulate matter for use in toxicologic assessment. *Environ Health Perspect* 2003;111:972–80.
18. Rom WN, Reibman J, Rogers L, *et al.* Emerging exposures and respiratory health: World Trade Center dust. *Proc Am Thorac Soc* 2010;7:142–5.
19. Oldiges M, Lütz S, Pflug S, *et al.* Metabolomics: current state and evolving methodologies and tools. *Appl Microbiol Biotechnol* 2007;76:495–511.
20. Bino RJ, Hall RD, Fiehn O, *et al.* Potential of metabolomics as a functional genomics tool. *Trends Plant Sci* 2004;9:418–25.
21. de Raad M, Fischer CR, Northen TR. High-throughput platforms for metabolomics. *Curr Opin Chem Biol* 2016;30:7–13.
22. Breiman L. Random Forests. *Mach Learn* 2001;45:5–32.
23. Caraher EJ, Kwon S, Haider SH, *et al.* Receptor for advanced glycation end-products and World Trade Center particulate induced lung function loss: A case-cohort study and murine model of acute particulate exposure. *PLoS One* 2017;12:e0184331.
24. Banauch GI, Dhala A, Prezant DJ. Pulmonary disease in rescue workers at the World Trade Center site. *Curr Opin Pulm Med* 2005;11:160–8.
25. Weiden MD, Ferrier N, Nolan A, *et al.* Obstructive airways disease with air trapping among firefighters exposed to World Trade Center dust. *Chest* 2010;137:566–74.
26. Tsukiji J, Cho SJ, Echevarria GC, *et al.* Lysophosphatidic acid and apolipoprotein A1 predict increased risk of developing World Trade Center-lung injury: a nested case-control study. *Biomarkers* 2014;19:159–65.
27. Boudonck KJ, Rose DJ, Karoly ED, *et al.* Metabolomics for early detection of drug-induced kidney injury: review of the current status. *Bioanalysis* 2009;1:1645–63.
28. Boudonck KJ, Mitchell MW, Nemet L, *et al.* Discovery of metabolomics biomarkers for early detection of nephrotoxicity. *Toxicol Pathol* 2009;37:280–92.
29. Nielsen KA, Tattersall DB, Jones PR, *et al.* Metabolite formation in dhurrin biosynthesis. *Phytochemistry* 2008;69:88–98.
30. Ryals J, Lawton K, Stevens D, *et al.* Metabolon, Inc. *Pharmacogenomics* 2007;8:863–6.
31. Hackstadt AJ, Hess AM. Filtering for increased power for microarray data analysis. *BMC Bioinformatics* 2009;10:11.
32. Maceyka M, Spiegel S. Sphingolipid metabolites in inflammatory disease. *Nature* 2014;510:58–67.
33. Schwartz J. Role of polyunsaturated fatty acids in lung disease. *Am J Clin Nutr* 2000;71:393s–6.
34. Calder PC. n-3 polyunsaturated fatty acids, inflammation, and inflammatory diseases. *Am J Clin Nutr* 2006;83:1505S–19.
35. Filippakopoulos P, Knapp S. Targeting bromodomains: epigenetic readers of lysine acetylation. *Nat Rev Drug Discov* 2014;13:337–56.
36. Engelen MP, Rutten EP, De Castro CL, *et al.* Supplementation of soy protein with branched-chain amino acids alters protein metabolism in healthy elderly and even more in patients with chronic obstructive pulmonary disease. *Am J Clin Nutr* 2007;85:431–9.
37. Wang L, Tang Y, Liu S, *et al.* Metabonomic profiling of serum and urine by (1)H NMR-based spectroscopy discriminates patients with chronic obstructive pulmonary disease and healthy individuals. *PLoS One* 2013;8:e65675.
38. Kutsuzawa T, Shioya S, Kurita D, *et al.* Plasma branched-chain amino acid levels and muscle energy metabolism in patients with chronic obstructive pulmonary disease. *Clin Nutr* 2009;28:203–8.
39. Yoneda T, Yoshikawa M, Fu A, *et al.* Plasma levels of amino acids and hypermetabolism in patients with chronic obstructive pulmonary disease. *Nutrition* 2001;17:95–9.
40. Jameson JS, Rogers J, Misiewicz JJ, *et al.* Oral or intravenous erythromycin has no effect on human distal colonic motility. *Aliment Pharmacol Ther* 1992;6:589–95.
41. Rogers RM, Donahoe M, Costantino J. Physiologic effects of oral supplemental feeding in malnourished patients with chronic obstructive pulmonary disease. A randomized control study. *Am Rev Respir Dis* 1992;146:1511–7.
42. Hotamisligil GS. Inflammation and metabolic disorders. *Nature* 2006;444:860–7.
43. Kwon S, Haider SH, Carher EJ, *et al.* RAGE Contributes to Particulate-Induced Lung Function Loss and Hyperreactivity: Mitigating the Persistent Effects of a Single Intense Particulate Exposure, in D96. *Damaging Effects Of Inhaled Particulates* 2016:A7517.
44. Haider SH, Zhang L, Crowley G, *et al.* Persistence of World Trade Center Particulate Induced Hyperresponsiveness and the Role of RAGE. *Am J Respir Crit Care Med* 2017;195:A4647.
45. Holland WL, Bikman BT, Wang LP, *et al.* Lipid-induced insulin resistance mediated by the proinflammatory receptor TLR4 requires saturated fatty acid-induced ceramide biosynthesis in mice. *J Clin Invest* 2011;121:1858–70.
46. Lands WE, Libelt B, Morris A, *et al.* Maintenance of lower proportions of (n - 6) eicosanoid precursors in phospholipids of human plasma in response to added dietary (n - 3) fatty acids. *Biochim Biophys Acta* 1992;1180:147–62.
47. Fekete K, Marosvölgyi T, Jakobik V, *et al.* Methods of assessment of n-3 long-chain polyunsaturated fatty acid status in humans: a systematic review. *Am J Clin Nutr* 2009;89:2070S–84.
48. Chuang YC, Shaw HM, Chen CC, *et al.* Short-term glutamine supplementation decreases lung inflammation and the receptor for advanced glycation end-products expression in direct acute lung injury in mice. *BMC Pulm Med* 2014;14:115.
49. Mishra S, Goldman JD, Sahyoun NR, *et al.* Dietary protein intake by meal types in adults aged 51 years and over: WWEIA, NHANES 2011–2012. *Journal of Food Composition and Analysis* 2017;64(Part 1):93–6.