Published in final edited form as:

Biochem Biophys Res Commun. 2011 October 7; 413(4): 588–593. doi:10.1016/j.bbrc.2011.09.006.

Pilot Analysis of the Plasma Metabolite Profiles Associated with Emphysematous *Chronic Obstructive Pulmonary Disease* Phenotype

Mikell Paige^a, Marie D. Burdick^b, Suhyon Kim^b, Junrui Xu^c, Jae K. Lee^d, and Y. Michael Shim^{b,e}

Mikell Paige: map65@georgetown.edu; Junrui Xu: jx5aj@virginia.edu; Jae K. Lee: jaeklee@virginia.edu; Y. Michael Shim: yss6n@virginia.edu

^a3970 Reservoir Road, NW, NRB EP15; Georgetown University, Drug Discovery Program, Lombardi Comprehensive Cancer Center, Washington, DC 20057, USA

^bPO BOX 800546, Division of Pulmonary & Critical Care, Department of Medicine, University of Virginia, Charlottesville, Va 22908-0546, USA

^cP.O. Box 400747, Departments of Systems & Information Engineering, University of Virginia, Charlottesville, VA 22904, USA

^dP.O.Box 800717, Division of Biostatistics, Departments of Public Health Sciences, University of Virginia School of Medicine, Charlottesville, VA 22908-0717, USA

Abstract

The current pilot study examined the hypothesis that cigarette smokers who developed an emphysematous phenotype of Chronic Obstructive Pulmonary Disease (COPD) were associated with distinctive patterns in their corresponding metabolomics profile as compared to those who did not. Peripheral blood plasma samples were collected from 38 subjects with different phenotypes of COPD. They were categorized into three groups: healthy non-smokers (n=16), smokers without emphysema (n=8), and smokers with emphysema (n=14). Ultra High Performance Liquid Chromatography/quadrupole–time-of-flight mass spectrometry techniques were used to identify a large number of metabolite markers (3,534). Unsupervised clustering analysis accurately separated the smokers with emphysema from others without emphysema and demonstrated potentials of this metabolomics data. Subsequently predictive models were created with a supervised learning set, and these predictive models were found to be highly accurate in identifying the subjects with the emphysematous phenotype of COPD with excellent sensitivity and specificity. Our methodology provides a preliminary model that differentiates an emphysematous COPD phenotype from other COPD phenotypes on the basis of the metabolomics profiles. These results also suggest that the metabolomics profiling could potentially guide the characterization of relevant metabolites that leads to an emphysematous COPD phenotype.

Keywords

Emphysema; Chronic obstructive pulmonary disease; Metabolite(s); Metabolomics; UPLS-QTOF-MS

^eCorresponding author. Phone:434-243-9212. Fax:434-924-9682.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Introduction

For the past 100 years, cigarette-induced Chronic Obstructive Pulmonary Disease (COPD) has become one of the most prevalent chronic adult diseases worldwide. It affects approximately 18 million people in the U.S. and many more around the world.[1,2] In 2005, more than 100,000 adults died of COPD which was ranked as the 4th most common etiology of death in the U.S., and more than six million in-patient hospital stays were attributed to a diagnosis of COPD with annual per patient expenditure of more than \$6,000.[1,2] While COPD appears to be a term designated for one disease, great variability of the disease phenotype is observed. For example, patients with chronic bronchitic phenotype suffer from irreversible airway obstruction with normal alveolar epithelium and intact pulmonary gas exchange while those with emphysematous phenotype primarily lose integrity of the alveolar epithelium and subsequently suffer from severe defect in pulmonary gas exchange in addition to the irreversible airway obstruction.[3,4,5,6] Only forty to seventy percent of cigarette smokers develop COPD even after substantial cigarette smoke exposure.[7,8] Of those who develop COPD, only about 25% develops emphysematous phenotype of COPD.[7,9] There is also a subgroup of emphysematous COPD patients who has a genetic trait called α 1-antitrypsin deficiency (AATD). Persons with this genetic disorder become substantially susceptible to cigarette smoke, which manifests as aggressive forms of emphysema at young ages.[10,11,12] Taken together, we can speculate that unique host characteristics like AATD, or still unknown etiologies, may either protect patients from or predispose patients to cigarette smoke-induced emphysema.

Recent technological advances in chromatography and mass spectroscopy have enabled extremely sensitive assays to simultaneously detect thousands of metabolites in human tissues. Subsequently, global profiling of metabolites (i.e. metabolomics) started to complement genomics and proteomics strategies to help us understand biological mechanisms of human diseases.[13] In this aspect, a metabolomics approach is an excellent tool to investigate the biology behind the heterogeneity of the COPD phenotypes. The current report describes a pilot analysis of peripheral blood plasma from healthy non-smokers, non-emphysematous smokers, and emphysematous smokers. The aim was to explore potential differences in their metabolite profiles, and then to create and test a model to predict the presence of an emphysematous COPD phenotype among cigarette smokers. Our results demonstrate that the metabolomics assessment can distinguish cigarette smokers with emphysema from those without emphysema or healthy non-smokers. The preliminary predictive model led us to a discovery of metabolites highly associated with the emphysematous COPD phenotype, and showed potentials for future investigation to help us understand the mechanisms involved in the pathogenesis of the emphysematous COPD.

Material and Methods

Subject Recruitment

All subjects were recruited at the University of Virginia and signed the informed consent, which was approved by the University of Virginia Institutional Review Board. All subjects were initially categorized into three groups on the basis of a complete pulmonary function test (PFT) consisting of spirometry for all subjects and carbon monoxide diffusion capacity for significant cigarette smokers. "Cigarette smokers" had significant smoking history defined as minimally 20 pack-year smoking (number of packs of cigarette smoked per day multiplied by the total number of years smoking). Emphysematous COPD subjects were defined by PFT findings with less than 0.70 ratio of the Forced Expiratory Volume in 1 second (FEV1) to Functional Vital Capacity (FVC) and less than 80% of the predicted carbon monoxide diffusion capacity (DLCO) by European Community for Coal and Steel

(ECCS) 1993 reference. Non-emphysematous smokers were defined as cigarette smokers with greater than 80% of the predicted DLCO even if the FEV1 to FVC ratio was less than 0.70. Healthy non-smokers were defined as having normal FEV1 to FVC ratio and normal percent predicted FVC (%FVC) as defined by the third National Health and Nutrition Examination Survey. This interpretation strategy for the PFT is based on the American Thoracic Society/European Respiratory Society Consensus Statement.[14]

Plasma Preparation

All subjects were fasted beginning the night prior to enrolling in our study. Last cigarette smoking was at least 8 hours prior to enrolling in our study—10 mL of peripheral blood was collected in a vacutainer (BD) with heparin sodium from subjects prior to the PFT. Blood was kept on ice and processed within 2 hours after collection. Initially the vacutainer was centrifuged for 10 minutes at 1,200 RPM with temperature maintained at 4 °C. The supernatant was collected. Plasma protein was precipitated by mixing 5 µL of plasma and 195 µL containing 4-nitrobenzoic acid and debrisoquine as internal standards [the stock solution was made as follows: 100 µL of a 1 mg/mL solution of 4-nitrobenzoic acid in MeOH and 10 µL of a 1 mg/mL solution of debrisoquine in H₂O was added to a solution of 10 mL of 2:1 ACN-H₂O]. The mixture was vortexed and then incubated on ice for 20 min. The mixture was centrifuged at 10,000 rpm for 10 min at 4 °C and the supernatant was transferred to auto-injection vials. The samples were frozen at -80 °C until analysis.

Ultra Performance Liquid Chromatography–Quadrupole Time of Flight/Mass Spectroscopy(UPLC-qTOF/MS)

A known amount of exogenous leukotriene B₄ was added as a third quality control compound. Individual plasma samples were analyzed according to the method described below. An aliquot of the individual plasma sample was deposited in an auto-sampler vial, and 5 µL was separated on a 50 mm x 2.1 mm Acquity 1.7 µm C-18 reverse-phase column (Waters Corp) using an Acquity UPLC system (Waters Corp). The gradient mobile phase consisted by 0.1% formic acid (A) and acetonitrile containing 0.1% formic acid (B). The ten-minute UPLC method consisted of 0.5 minute of 100% solvent A followed by a gradient increase of solvent B gradient up to 100% for the remaining 9.5 minutes. The flow rate was set at 0.6 mL/min. The eluent was introduced by electrospray ionization into the mass spectrometer (Waters QTOF Premier) operating in negative mode. The capillary and sampling cone voltage were set to 3000 and 30 V, respectively. Source and desolvation temperature were set to 120 and 350 °C, respectively, and the cone and desolvation gas flow were set to 50.0 and 650.0 L/h, respectively. To maintain mass accuracy, sulfadimethoxine ([M-H]⁻ 309.0658) at a concentration of 500 pg/µL in 50% acetonitrile was used as a lock mass and injected at a rate of 0.08 µL/min. For MS scanning, data were acquired in centroid mode from 50 to 850 m/z.

Statistical Analysis

The mass chromatographic data were first inspected and analyzed by MassLynx and MarkerLynx softwares (Waters). The entire data of all subjects were examined for their overall consistency among all samples across all identified metabolites. First, we graphically examined whether their expression values were highly concordant and linearly correlated among all samples. We then evaluated overall Spearman correlation coefficients for each pair of the patient and control samples over all identified metabolites, and used the Fisher's Z-transformation normal test to assess their statistical significance. Prior to our statistical analysis, metabolomics data were statistically standardized, i.e. subtracted by mean and divided by standard deviation within each metabolite across all samples in order to

normalize our metabolomics data from different experimental setting and conditions. Using the normalized metabolomics data, we performed both unsupervised (hierarchical clustering) and supervised (multivariate classification) statistical learning techniques in order to explore the expression patterns of the metabolomics biomarkers on emphysema patients and controls and to identify and obtain highly robust metabolomics predictive model for stratifying emphysematous smokers from healthy controls or non-emphysematous smokers. Our unsupervised-learning hierarchical clustering analysis was conducted by 1) removing all metabolites with more than 1/3 zero values to avoid mathematical outlying artifacts (e.g. extremely small error estimates) and 2) applying a hierarchical clustering algorithm with average linkage and the Euclidean distance. This analysis was separately performed for the metabolites with retention times 0–10, 0–3, 3–7, and 7–10 mins. The clustering results were examined with clustering dendograms. Our supervised-learning statistical classification modeling was performed by using a Linear Discriminant Analysis (LDA) technique. This analysis was conducted with multiple analysis steps as follows: 1) all the metabolites with more than 1/3 zero values were removed. 2) Log transformation to the data set to balance the influence of highly and lowly expressed metabolites in prediction modeling was applied. If any data points were zero, we added a small value (half of the minimum of nonzero value of the entire data set) before the log transformation. 3) The whole set was divided into a training set [2/3 of balanced (disease and control) subjects] and a test set (1/3 of subjects), the latter completely kept aside from biomarker identification and predictive modeling for our evaluation and test of the final predictive models based on an independent set of subjects. 4) A nonparametric Wilcoxon rank sum test between emphysematous smokers and non-emphysematous smoker groups was performed with 100 times of bootstrapping to identify the top 12 robustly predictive metabolites that showed highly significant upper bounds for their bootstrapped p-values. 5) LDA prediction models were obtained by increasing the number of biomarkers from the most significant to the gradually less-significant metabolites (among the top 12 metabolites) on the training set. 6) The performance of the competing models on the test set was independently applied and evaluated. The final prediction performance was assessed with several statistical criteria, such as specificity, sensitivity, and the overall accuracy both for true positives and true negatives. 7) The best predictive model was selected with the highest overall accuracy.

Results

Patient Baseline Characteristics

Patients' baseline clinical characteristics are described in the Table 1. There were a total of 38 subjects, 23 females and 15 males. The overall baseline characteristics of these patients were similar. The healthy control subjects did not have any history of cigarette smoke exposure. Non-emphysematous smokers and emphysematous smokers had a comparable degree of lifetime exposure to cigarette smoke. Emphysematous smokers were slightly older than the non-emphysematous smokers or healthy control subjects. As expected, emphysematous smokers had significantly lower diffusion capacity for carbon monoxide. In summary, these inclusion and exclusion criteria were able to effectively separate COPD subjects with an emphysematous phenotype from those without an emphysematous phenotype.

Metabolite Identification

Initially individual plasma samples from each group were pooled together into three groups and analyzed by UPLC to broadly examine for potential differences in their metabolites. Reproducibility of the chromatographic elution was confirmed by comparing the retention time of the spiked leukotriene B₄ in addition to standard internal controls (Online material). The retention times for the leukotriene B₄ metabolite were between 4.11 and 4.12 minutes

varying only by 0.01 minute. Inspection of the UPLC chromatogram demonstrated potentials in distinguishing the emphysematous smokers from non-emphysematous smokers or healthy non-smoking controls (Online Material). Individual plasma samples were then analyzed of which a total of 3534 metabolites were identified. Internal consistency of the metabolite detection by the UPLC-QTOF/MS was investigated by plotting pairwise scatter plots of all metabolites. Emphysematous, non-emphysematous, and healthy control samples were highly reproducible within each experiment (38 sets of data) with a fixed condition (Online material). With our described UPLC-QTOF/MS protocol, highly concordant metabolite data could be obtained within these 38 patient set (metabolomics-wide correlation coefficients: mean 0.914, 95% CI [0.796, 0.966], high concordance (Fisher's Z-transformation normal test p-value < 0.0E-09, Online material).

Unsupervised Clustering

In order to explore the potential of distinguishing emphysematous COPD patients from non-emphysematous patients, we then subjected our metabolomics data through unsupervised clustering analysis by a blinded operator. Clustering based on metabolites within the full spectrum of the retention time (0–10 minutes) demonstrated potentials to separate emphysematous smokers from others with reasonably high accuracy (Figure 1A). 9 out of 14 emphysematous smokers (64.3%) clustered together (Figure 1A, inside the black box). Only 2 out of 16 healthy control subjects (12.5%) and 2 out of 8 non-emphysematous smokers (25%) clustered with the emphysematous smokers. Based on the initial survey of the UPLC chromatograms (online material), we anticipated that the accuracy of the clustering may improve by categorizing the metabolites according to the retention time. Same clustering analysis was conducted after grouping the metabolites into three blocks of retention times (0–3 minutes, 3–7 minutes, and 7–10 minutes). No significant clustering was noted in the metabolites with the retention times 0–3 minutes and 7–10 minutes (Figure 1B & 1D). However, significant clustering was noted in the metabolites with retention times between 3–7 minutes (Figure 1C). 14 out of 14 emphysematous smokers (100%) clustered together (Figure 1C, inside the black box). Only 4 out of 16 healthy control subjects (25%) and 3 out of 8 non-emphysematous smokers (37.5%) clustered with the emphysematous smokers. *Details of the Top 50 metabolites (retention time and m/z) from the retention times 3 to 7 minutes are shown in the online material.*

Predictive Modeling

Once the unsupervised clustering demonstrated a potential for these metabolomics data, we searched for the most highly and robustly predictive 12 biomarkers (two sample t-test p-values between 0.001 and 0.015). Gradually increasing the number of biomarkers from the top 2 to the top 12, we constructed and evaluated 11 predictive models. The whole data set was divided as follows: 2/3 learning test set and 1/3 independent test set. We used a multivariate supervised learning strategy with LDA and evaluated the performance of each of the candidate predictive models, e.g. from the top 2 biomarker model to the top 12 biomarker model, with rigorous n-fold cross-validation 100 times. Comparison of each metabolite used in this model is shown with their respective retention times in Figure 2. The predictive model was then tested by assessing how accurately this model can identify the phenotypes of smokers by a blinded operator. We found that our metabolomics-based predictive models could achieve 76.4% – 88.9% overall accuracy (Table 2). The model based on the top 7 biomarkers was the most sensitive with 96.5% sensitivity. The model based on the top 6 biomarkers was the most specific with 86.5% specificity. The model based on the top 7 biomarkers was most accurate with 88.9% overall accuracy.

Discussion

Very little information is available with regards to the metabolite biomarkers of COPD, and even less information is available with regards to the metabolite biomarkers associated with the emphysematous COPD phenotype. Given the paucity of past studies, we expected that our study would be an excellent pilot investigation to provide important information and future direction in identifying an appropriate model(s) and methodology(s) to narrow metabolite candidates relevant to pulmonary emphysema.

Peripheral blood plasma was the biological matrix we used for our investigation because it can be easily collected by a minimally invasive method. Our study is unique because of the close attention paid to separate pulmonary phenotypes among smokers. The plasma specimens were obtained from three groups of individuals separated by the presence or absence of significant cigarette smoking history and by the presence or absence of emphysematous COPD. The MS negative mode of the UPLC-QTOF method identified 3534 metabolites as potential markers. Unsupervised clustering analysis was able to separate the emphysematous patients into a distinctive cluster, and the clustering became more accurate with better than 80% sensitivity and as high as 100% specificity by using the retention time as a primary filter for the metabolite screening. We then used the same data set to build a predictive model by a supervised analyst. Our data demonstrated that there are a number of metabolite combinations with which the presence of emphysematous COPD can be predicted with accuracy close to 90%. These preliminary results provided us with an operational lead as to how metabolomics can be applied to identify patients with emphysematous phenotypes due to cigarette smoke exposure.

Our study also had a few shortcomings that are worth mentioning. First, our case subjects were cigarette smokers who already developed emphysema, and the observed patterns of the metabolites in these subjects may be different as compared to the subjects in the developing stages of the disease. Although our strategy was necessary as a pilot study, this is a typical pitfall of a cross-sectional, case-controlled study design. A larger prospective study is necessary to answer causal interaction between the metabolites identified in this study and the emphysematous phenotypes. Second, our subjects were recruited from a single center.

Emphysema is a complex disease caused by complicated interactions among genes, behaviors, and environments. Therefore, we would need to determine if our models would still stand true and validated in more heterogeneous populations. Third, our subjects also had other co-morbid diseases including hypertension, atherosclerosis, diabetes, and hyperlipidemia. Prevalence of these diseases was less than 10% except hypertension. However, presence of hypertension had minimal effect on the unsupervised clustering, so we believe that these co-morbidities did not introduce any significant confounding effects on our results. Fourth, the sample size of our pilot study was small. Therefore, we kept the analytical chemist and bio-statisticians blinded throughout the sample assays and data analysis. A blinded analytical chemist generated the metabolites data, and a blinded bio-statistician conducted data analysis. We believe that our careful approach is an extremely important aspect of quality control in order to minimize the confounding effects from small sample size. We caution that even though this pilot study provided a promising potential for metabolomic profiling to study the emphysema COPD phenotype, performance of the biomarkers and their respective predictive models need to be confirmed again. We are presently pursuing a prospective study with a larger sample size to test the predictive models proposed in this communication.

There are also several important points in our study which merit further discussion. First, our study is one of the very first unique studies demonstrating that the metabolomics

profiling of the human plasma has the potential to identify emphysematous COPD phenotypes. Recently, the ECLIPSE study has been published to describe the COPD phenotypes.[15,16,17] In these studies, computer tomography (CT) of the chest was used to identify subjects with emphysematous COPD. However, similar to other studies in the past, phenotypic distinction between the emphysematous COPD and non-emphysematous cigarette smokers was not a primary focus of these studies. Therefore, we believe that our study provides an important new dimension to understanding the pathogenesis of emphysema.

Second, healthy control subjects who never smoked can be considered as human subjects with undetermined future due to the lack of exposure to the inciting agent, i.e. cigarette smoke. Population based epidemiologic studies demonstrated that about 40 to 70% of cigarette smokers develop COPD and 25% of the COPD patients develop emphysema.[7,9] Based on these observations, we can apply the results of our clustering data and rationally speculate that the 25% of our healthy control subjects who clustered together with the emphysematous COPD subjects might have developed emphysema if they were to smoke a significant amount of cigarette. This is a tantalizing glimpse to the potentials of our metabolomics data which needs to be explored in a future prospective study.

Third, because COPD has been set aside as a disease of self-inflicted damages (i.e. cigarette smoking), efforts to understand the mechanisms of emphysema pathogenesis have been dismal at best. This is also an extremely difficult task, especially in case of diseases like emphysema, which depends on complex interactions of multiple genes and environment. This has led to paucity of novel therapeutics to date. As demonstrated in the Figure 2, we believe that the metabolites identified for our predictive models are excellent leads to pursue mechanistic studies of emphysema pathogenesis in the future. Identifications of these metabolites and their associated pathways would lead to possibilities of future targets for diagnostic and therapeutic interventions.

We conclude that simultaneous analytical measurement of multiple metabolites in human plasma results in strong potentials with which the emphysematous phenotypes of COPD can be distinguished. The accuracy of our predictive model was excellent, and the model could be applied for a number of future applications including early diagnosis and screening. This can be accomplished by further refinement of the proposed metabolomics predictive models in a prospective cohort study. This pilot study will also pave the way toward exciting new tools for future novel target discovery in the arenas of emphysema therapeutics.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was funded in part by NIH/NHLBI K08HL091127 (Y. Michael Shim), Flight Medical Research Institute (Y. Michael Shim), American Lung Association/Virginia Thoracic Society (Y. Michael Shim). The authors thank the Proteomics & Metabolomics Shared Resource at Georgetown University - Lombardi Comprehensive Cancer Center. Shared Resource is partially supported by NIH/NCI grant P30-CA051008. We also thank Dr. Iggy Khan (Waters Corp.) for MassLynx and MarkerLynx software support.

References

1. MMWR, Deaths from COPD - US-2000-2005. 2008.
<http://www.cdc.gov/mmwr/preview/mmwrhtml/mm5745a4.htm>
2. Miller JD, Foster T, Boulanger L, et al. Direct costs of COPD in the U.S.: an analysis of Medical Expenditure Panel Survey (MEPS) data. COPD. 2005; 2:311–318. [PubMed: 17146996]

3. Pauwels RA, Buist AS, Ma P, et al. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: National Heart, Lung, and Blood Institute and World Health Organization Global Initiative for Chronic Obstructive Lung Disease (GOLD): executive summary. *Respir Care*. 2001; 46:798–825. [PubMed: 11463370]
4. Sluiter HJ, Koeter GH, de Monchy JG, et al. The Dutch hypothesis (chronic non-specific lung disease) revisited. *Eur Respir J*. 1991; 4:479–489. [PubMed: 1855577]
5. Cosio MG, Guerassimov A. Chronic obstructive pulmonary disease. Inflammation of small airways and lung parenchyma. *Am J Respir Crit Care Med*. 1999; 160:S21–25. [PubMed: 10556164]
6. Izquierdo JL, Almonacid C, Parra T, Perez J. Systemic and lung inflammation in 2 phenotypes of chronic obstructive pulmonary disease. *Arch Bronconeumol*. 2006; 42:332–337. [PubMed: 16945263]
7. Raftery C, Girodet PO. Epidemiology of COPD, *European respiratory review* : an official journal of the European Respiratory Society. 2009; 18:213–221.
8. Postma DS, Siafakas NM. Epidemiology of Chronic Obstructive Pulmonary Disease. *Eur Respir Mon*. 1998:41–73.
9. Halbert RJ, Natoli JL, Gano A, et al. Global burden of COPD: systematic review and meta-analysis. *The European respiratory journal : official journal of the European Society for Clinical Respiratory Physiology*. 2006; 28:523–532. [PubMed: 16611654]
10. Koczulla R, Bittkowski N, Andress J, et al. The German registry of individuals with alpha-1-antitrypsin deficiency--a source for research on patient care. *Pneumologie*. 2008; 62:655–658. [PubMed: 18855308]
11. Stoller JK, Fromer L, Brantly M, et al. Primary care diagnosis of alpha-1 antitrypsin deficiency: issues and opportunities. *Cleve Clin J Med*. 2007; 74:869–874. [PubMed: 18183837]
12. Mulgrew AT, Taggart CC, McElvaney NG. Alpha-1-antitrypsin deficiency: current concepts. *Lung*. 2007; 185:191–201. [PubMed: 17562108]
13. Nicholson JK, Wilson ID, Lindon JC. Pharmacometabonomics as an effector for personalized medicine. *Pharmacogenomics*. 2011; 12:103–111. [PubMed: 21174625]
14. Pellegrino R, Viegi G, Brusasco V, et al. Interpretative strategies for lung function tests. *The European respiratory journal : official journal of the European Society for Clinical Respiratory Physiology*. 2005; 26:948–968. [PubMed: 16264058]
15. Hurst JR, Vestbo J, Anzueto A, et al. Susceptibility to exacerbation in chronic obstructive pulmonary disease. *The New England journal of medicine*. 2010; 363:1128–1138. [PubMed: 20843247]
16. Agusti A, Calverley PM, Celli B, et al. Characterisation of COPD heterogeneity in the ECLIPSE cohort. *Respiratory research*. 2010; 11:122. [PubMed: 20831787]
17. Pillai SG, Kong X, Edwards LD, et al. Loci identified by genome-wide association studies influence different disease-related phenotypes in chronic obstructive pulmonary disease. *American journal of respiratory and critical care medicine*. 2010; 182:1498–1505. [PubMed: 20656943]

Highlights

- We recruited cigarette smokers with and without emphysema.
- We examined differences in plasma metabolite profiles of these patients.
- Metabolomics predictive model identified emphysema subjects with 90% accuracy.
- Metabolite profiling identified 12 potential targets for future studies of emphysema.
- This pilot study will lead to important prospective study of metabolomics in emphysema.

Figure 1.

Unsupervised clustering analysis of all metabolites detected by the QTOF-MS(-) mode in 38 subjects' plasma by a blinded operator. Panel A: Clustering analysis of the metabolites from the entire span of the retention time. The box with black lines indicates potentials of clustering EM subjects from No-EM. Panel B: Clustering analysis of the metabolites from the retention time 0 to 3 minutes. Panel C: Clustering analysis of the metabolites from the retention time 3 to 7 minutes. The box with black lines indicates potentials of clustering EM subjects from No-EM. Panel D: Clustering analysis of the metabolites from the retention time 7 to 10 minutes. Vertical distance is proportional to the relative degrees of clustering. White bar = healthy non-smoking control individual. Grey bar (No-EM) = non-emphysematous cigarette smoking individual. Black bar (EM) = emphysematous cigarette smoking individual.

Figure 2.

Comparative levels of the top 12 metabolites used for the predictive models. All metabolites from the 12 panels were compared between the non-emphysematous smoker group (No-EM) and emphysematous smoker group (EM). m/z = MS negative mode molecular weight. RT = retention time. P value < 0.05 was considered statistically significant. Horizontal and vertical bars are medians $\pm$ interquartile ranges.

Table 1

Subject baseline characteristics.

	CTL (n=16)	No-Em (n=8)	EM (n=14)	ANOVA	No-EM vs EM
Age	53 ± 7.4	56 ± 5.9	61 ± 9.5	0.012	n.s.
Gender	M(8), F(8)	M(2), F(6)	M(5), F(9)	-	-
Height (inches)	68 ± 3.6	67 ± 3.6	66 ± 3.6	0.693	-
Weight (LBS)	165 ± 35	175 ± 35	189 ± 47	0.466	-
BMI	25.7 ± 6.1	27.2 ± 3.8	30.0 ± 6.6	0.146	-
Race	C(14) AA(2)	C(7) A(1)	C(14)	-	-
Hypertension	1/16	2/8	6/14	-	-
Hyperlipidemia	0/16	0/8	3/14	-	-
Atherosclerosis	0/16	0/8	2/14	-	-
Cirrhosis	0/16	0/8	0/14	-	-
Renal Failure	0/16	0/8	0/14	-	-
Diabetes	0/16	0/8	2/14	-	-
Cigarette (pk-yr)	0	28 ± 6	32 ± 11	<0.001	n.s.
FEV1/FVC	0.81 ± 0.06	0.82 ± 0.04	0.57 ± 0.12	<0.001	<0.001
%FEV1	97 ± 11	101 ± 7	72 ± 30	<0.001	<0.001
%FVC	99 ± 12	102 ± 7	91 ± 23	0.021	n.s.
%DLCO	-	94 ± 5	61 ± 13	-	<0.001

All values are mean ± standard deviation. CTL = healthy subjects without smoking. No-EM = Subjects with significant cigarette smoking history (greater than 20 pack-year) but without emphysema by PFT. EM = Subjects with significant cigarette smoking history (greater than 20 pack-year) and with emphysema by PFT. M = male, F = female. Cigarette = number of cigarette pack per day x number of years smoked. FEV1 = Forced Expiratory Volume in 1 second. FVC = Functional Vital Capacity. %FEV1 = Percentage predicted FEV1 by the third National Health And Nutrition Survey. %FVC = Percentage predicted FVC by the third National Health And Nutrition Survey. %DLCO = Percentage predicted by the European Community for Coal and Steel (ECCS) 1993 reference. n.s. = p value not significant (greater than 0.05).

Table 2

Accuracy of the predictive models.

	Specificity (95%CI)	Sensitivity(95%CI)	Accuracy(95%CI)
2 Biomarkers	0.718(0.683–0.754)	0.984(0.975–0.992)	0.853(0.835–0.871)
3 Biomarkers	0.667(0.633–0.700)	0.976(0.966–0.986)	0.809(0.791–0.828)
4 Biomarkers	0.857(0.826–0.887)	0.898(0.875–0.920)	0.871(0.855–0.888)
5 Biomarkers	0.833(0.807–0.860)	0.901(0.880–0.922)	0.872(0.858–0.886)
6 Biomarkers	0.865(0.838–0.892)	0.916(0.898–0.934)	0.886(0.869–0.902)
7 Biomarkers	0.795(0.765–0.825)	0.965(0.952–0.978)	0.889(0.873–0.906)
8 Biomarkers	0.813(0.781–0.845)	0.929(0.910–0.948)	0.871(0.851–0.890)
9 Biomarkers	0.775(0.739–0.811)	0.925(0.905–0.945)	0.855(0.835–0.875)
10 Biomarkers	0.727(0.687–0.766)	0.921(0.902–0.940)	0.815(0.791–0.839)
11 Biomarkers	0.685(0.645–0.725)	0.849(0.822–0.876)	0.764(0.740–0.789)
12 Biomarkers	0.712(0.671–0.752)	0.836(0.806–0.867)	0.766(0.738–0.793)

All values are mean (lower and upper boundaries of the 95% confidence interval)