Journal of proteome research

Article

pubs.acs.org/jpr

Feasibility of Early Detection of Cystic Fibrosis Acute Pulmonary Exacerbations by Exhaled Breath Condensate Metabolomics: A Pilot Study

Xiaoling Zang,[†] María Eugenia Monge,[†,‡] Nael A. McCarty,[§] Arlene A. Stecenko,[§] and Facundo M. Fernández*[,†]

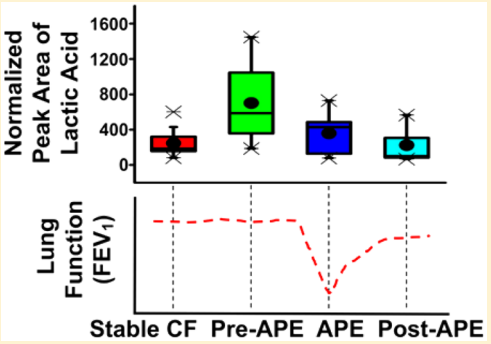
[†]School of Chemistry and Biochemistry, Georgia Institute of Technology, Atlanta, Georgia 30332, United States

[‡]Centro de Investigaciones en Bionanociencias (CIBION), Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Godoy Cruz 2390, C1425FQD, Ciudad de Buenos Aires, Argentina

[§]Emory+Children's Center for Cystic Fibrosis and Airways Disease Research and Department of Pediatrics, Emory University School of Medicine and Children's Healthcare of Atlanta, Atlanta, Georgia 30322, United States

Supporting Information

ABSTRACT: Progressive lung function decline and, ultimately, respiratory failure are the most common cause of death in patients with cystic fibrosis (CF). This decline is punctuated by acute pulmonary exacerbations (APEs), and in many cases, there is a failure to return to baseline lung function. Ultraperformance liquid chromatography quadrupole-time-of-flight mass spectrometry was used to profile metabolites in exhaled breath condensate (EBC) samples from 17 clinically stable CF patients, 9 CF patients with an APE severe enough to require hospitalization (termed APE), 5 CF patients during recovery from a severe APE (termed post-APE), and 4 CF patients who were clinically stable at the time of collection but in the subsequent 1−3 months developed a severe APE (termed pre-APE). A panel containing two metabolic discriminant features, 4-hydroxycyclohexylcarboxylic acid and pyroglutamic acid, differentiated the APE samples from the stable CF samples with 84.6% accuracy. Pre-APE samples were distinguished from stable CF samples by lactic acid and pyroglutamic acid with 90.5% accuracy and in general matched the APE signature when projected onto the APE vs stable CF model. Post-APE samples were on average more similar to stable CF samples in terms of their metabolomic signature. These results show the feasibility of detecting and predicting an oncoming APE or monitoring APE treatment using EBC metabolites.



KEYWORDS: *cystic fibrosis, acute pulmonary exacerbations, metabolomics, ultraperformance liquid chromatography mass spectrometry*

INTRODUCTION

Cystic fibrosis (CF) is a genetic disease caused by mutations in the gene encoding the cystic fibrosis transmembrane conductance regulator (CFTR) protein, leading to abnormal ion and water transport across epithelial cells.[1,2] Although multiple organs are affected by CF, over 90% of patients die from progressive pulmonary disease and subsequent respiratory failure.[3] CF lung disease is characterized by the triad of impaired mucociliary clearance, chronic polymicrobial bacterial infection, and neutrophil-dominated inflammation. This triad results in progressive decline in lung function that is punctuated by acute episodes of increased respiratory symptoms and often declines in lung function that can be marked. These episodes are termed acute pulmonary exacerbations (APEs). Therapies are intensified and hospitalization is often required in an attempt to restore lung function to baseline. For example, the CF Foundation Registry shows that 26% of CF patients younger than 18 years of age and 44% of those 18 years and older required hospitalization for an APE in 2014.[4] It is known that the frequency of APEs severe enough to require hospitalization adversely impacts quality of life, survival,[5] and associated health care costs.[6] Despite the clinical importance of APEs, there is still a general lack of knowledge regarding their pathophysiology,[7−10] resulting in nonuniform treatment decisions.[11]

Unlike asthma, the triggers for APEs in CF are still poorly defined. Viral infections, particularly RSV, rhinovirus, and influenza, are thought to be important initiating factors in CF APEs.[12,13] In addition, exposure to cigarette smoke or other pollutants[14] as well as nonadherence to daily maintenance therapy that has been shown to prevent APEs[15] may also be important. Controversy exists on the role of bacteria, with some evidence suggesting that an increased bacterial load of resident organisms is associated with APEs versus infection with new

Received: July 19, 2016
Published: November 8, 2016

DOI: 10.1021/acs.jproteome.6b00675
J. Proteome Res. XXXX, XXX, XXX−XXX

bacteria. Regardless of the trigger, a generally held notion is that intensification of bacterial infection and inflammation drives the clinical manifestations of APEs, so the mainstay of therapy is intensive antibiotic treatment and physically clearing the airways of debris.[16] However, there is a flaw with this approach, as 25% of CF patients with APEs severe enough to require hospitalization do not recover to their baseline lung function.[17] A discovery-based, untargeted approach is needed to identify pathways that are aberrant in CF with APEs compared to clinically stable CF in order to develop mechanistic hypotheses on their pathophysiology that can then be translated to improvements in early detection, preventative strategies, and new therapeutic approaches. The difficulty with targeted approaches can be seen in studies on inflammatory biomarkers. For example, a comparison of sputum from clinically stable CF patients and patients during exacerbations has suggested that there is a correlation between APEs and inflammatory mediators such as interleukin (IL)-1β, IL-8, and myeloperoxidase.[18−21] Other studies, however, failed to differentiate CF patients based on IL-8 alone, finding instead other potential protein biomarkers, such as soluble intercellular adhesion molecule-1, calprotectin, and calgranulin A and B.[22,23]

Metabolomics has become a valuable tool in studying biochemical processes,[24] with various studies having shown CF-induced alterations at the metabolome level in different biofluids. For example, the relative plasma linoleic acid and docosohexanoic acid concentration product, as determined by gas chromatography−mass spectrometry (GC−MS),[25] suggests abnormal essential fatty acid metabolism in CF patients. Investigation of bronchoalveolar lavage fluid (BALF) metabolite profiles in pediatric CF patients via nuclear magnetic resonance spectroscopy (NMR) has shown amino acids and lactate as markers of airway inflammation.[26] Increased concentrations of regulatory lipid mediators have been found in sputum from CF patients using liquid chromatography−tandem mass spectrometry (LC−MS/MS).[27] More recently, 92 differential metabolites discriminating between CF and non-CF subjects have been identified in serum with a discovery-based metabolomics approach, showing changes in lipid metabolism in addition to abnormalities in bile acid processing and decreased fatty acid β-oxidation.[28]

Due to its ease of collection, exhaled breath condensate (EBC) has been one of the preferred biofluids to study biochemical changes in the lung environment. EBC consists of aerosolized epithelial lining fluid containing volatile and nonvolatile compounds trapped and diluted by water vapor condensation.[29,30] EBC has distinct advantages over sputum to investigate the early phases of CF lung disease at a time where intervention could prevent organ damage, as EBC can be collected at all ages, whereas expectorated sputum can be collected only in older patients who have established disease. However, the disadvantage of EBC is that its components may be present in trace levels (nano- to micromolar concentration range), necessitating sensitive techniques for its analysis.

On the basis of the above, we hypothesize that, regardless of the initiating trigger, CF exacerbations severe enough to require hospitalization will be associated with a specific metabolic signature in EBC that would allow potential pathways that are linked to the pathophysiology of APEs to be identified. We further hypothesize that this metabolic fingerprint could precede any symptoms or signs of an APE, that is, it will signal an impending exacerbation. Finally, we hypothesize that this chemical signature would return to the clinically stable signature following treatment for the APE. Therefore, in this study, we utilize a discovery-based metabolomics approach to analyze EBC samples collected from CF patients who are clinically stable compared to those with APE severe enough to require hospitalization, using ultraperformance liquid chromatography coupled to high-resolution mass spectrometry (UPLC−MS), in combination with supervised multivariate classification models. Once a discriminant metabolite profile was identified, we investigated its presence in the presymptomatic phase of an APE event and also its persistence following treatment for the APE.

MATERIALS AND METHODS

Chemicals

LC−MS grade methanol, purchased from J.T. Baker Avantor Performance Materials, Inc. (Center Valley, PA, USA), and ultrapure water with 18.2 MΩ·cm resistivity (Barnstead Nanopure UV ultrapure water system, USA) were used to prepare mobile phases and solutions. DL-Lactic acid lithium salt (∼99%) and myristoleic acid (≥99%) were purchased from MP Biomedicals, LLC (Solon, OH, USA), pyroglutamic acid (5-oxoproline) was from Anaspec, Inc. (San Jose, CA, USA), hydroxyacetone (96.4%) was from TCI America (Portland, OR, USA), 2-methylbutyric acid (98%), 3,3-dimethylglutaric acid (≥98%), and pimelic acid (≥98%) were from Alfa Aesar (Ward Hill, MA, USA), 4-methylvaleric acid (98.5%) was from Acros Organics (Morris, NJ, USA), 4-hydroxycyclohexanecarboxylic acid, D-lactaldehyde solution (1 M in H_2O), 3-hydroxybenzoic acid (99%), 4-hydroxybenzoic acid (99%), propionic acid (99%), isovaleric acid (99%), valeric acid (99%), adenosine (≥99%), *trans*-4-hydroxy-L-proline (≥99%), L-proline (≥99%), sucrose (≥99%), L-glutathione reduced (≥98.0%), D-tyrosine (99%), D-(+)-glucose monohydrate (≥99%), and D-(−)-fructose (≥99%) were from Sigma-Aldrich (St. Louis, MO, USA), and 8-isoprostane-d_4 (1050 mg/L), 5*S*,6*R*-lipoxin A_4 (100 mg/L), and 5*S*,6*S*-lipoxin A_4 (100 mg/L) were from Cayman Chemical Company (Ann Arbor, MI, USA).

Cohort Description

CF patients are usually seen in CF clinic every 3 months when stable and more frequently with exacerbations. EBC was collected during these regular clinic visits to the Emory +Children's CF Care Center in Atlanta, Georgia, after obtaining informed consent. The patient's clinical course was then followed over the subsequent months so that they could be grouped according to their APE status. Clinically stable CF was defined as CF subjects whose symptoms were at baseline, physical examination of the lungs was at baseline, lung function (FEV_1, i.e., forced expiratory volume in 1 s) was within 10% of the yearly baseline, and no new therapies (particularly antibiotics) were added at that clinic visit, plus the patient was seen at the next clinic visit 3 months later and was again classified as clinically stable. EBC was collected on 17 CF subjects meeting this definition of clinically stable (age range 14−39, mean (SD) age 28(7) years, 29.4% female). A severe APE was defined as an increase in respiratory symptoms (cough, sputum production) and/or changes in physical examination of the lungs (increase in crackles, decrease in airflow), at least a 10% decrease in FEV_1, and in the opinion of the clinician required hospitalization for treatment of the APE. EBC was collected in 9 such subjects with a severe APE at the time of hospitalization (age range 15−39, mean age 26(7) years, 55.6% female). In 5 subjects, EBC was collected 1−3 months after an APE event requiring hospital-

DOI: 10.1021/acs.jproteome.6b00675
J. Proteome Res. XXXX, XXX, XXX−XXX

ization, addressed as post-APE (age range 19−30, mean age 26(5) years, 40% female). Finally, EBC was collected in 4 subjects who were clinically stable as defined above but in the subsequent 1 to 3 months developed APEs severe enough to require hospitalization, addressed as pre-APE (age range 15−39, mean age 27(10) years, 50% female). Table S2 describes the details of the cohort. At the 0.05 level, the means of the age populations were not significantly different with the two-sample *t*-test for all possible pairs of sample classes. Among the 26 patients from whom all 35 EBC samples were collected, 6 patients had multiple samples collected at different stages of disease severity. In most cases, the samples were not drawn from the same APE episode, so they did not meet the criteria of paired samples. Due to the small number of samples available, we included all samples to maximize the sample size.

EBC Sample Collection and Preparation

EBC sample collection followed the guidelines approved by the Georgia Institute of Technology and the Emory University Institutional Review Boards (approval number IRB00000372). An R-Tube collector (Respiratory Research, Inc., Austin, TX, USA) was used to collect EBC samples, which were immediately frozen at −80 °C until processed. Prior to analysis, EBC samples were thawed and lyophilized at −40 °C and 100 mTorr for 24 h using a VirTis benchtop freeze-dryer (SP Industries, Stone Ridge, NY, USA). Sample residues were reconstituted in water/methanol (90:10 v/v) with a concentration factor of 20 and analyzed by UPLC−MS. Blank samples containing ultrapure water also went through the same sample preparation procedure. Prior to UPLC−MS, samples were randomly separated into two batches and analyzed on consecutive days together with solvent and sample preparation blanks. Quality control (QC) samples (5.50 μM L-glutathione (reduced), *trans*-4-hydroxy-L-proline, adenosine, D-(+)-glucose monohydrate, D-(−)-fructose, sucrose, 5*S*,6*R*-lipoxin A_4 and 5*S*,6*S*-lipoxin A_4, 8.85 μM D-tyrosine, 5.49 μM L-proline, and 5.51 μM 8-isoprostane-d_4 solution) were analyzed every 5 h to verify the stability of the retention times, peak shapes, and areas during the analysis. Chemical standards for metabolite identity validation were prepared in ultrapure water or methanol (or a mixture of those solvents), depending on their solubility.

Ultraperformance Liquid Chromatography−Mass Spectrometry

Ultraperformance liquid chromatography−mass spectrometry (UPLC−MS) analyses were performed using a Waters Acquity UPLC H class system fitted with a Waters Acquity UPLC BEH C_{18} column (2.1 × 50 mm, 1.7 μm particle size, Waters Corporation, Milford, MA, USA) and coupled to a Xevo G2 QTOF mass spectrometer (Waters Corporation, Manchester, UK) with an electrospray ionization (ESI) source. The typical resolving power and mass accuracy of the Xevo G2 QTOF mass spectrometer were 25 000 fwhm and 1.8 ppm at m/z 554.2615, respectively. Gradient elution was employed in the chromatographic separation method using water (mobile phase A) and methanol (mobile phase B), with the following program: 0−1 min, 90−80% A; 1−3 min, 80−60% A; 3−5 min, 60−50% A; 5−10 min, 50−40% A; 10−15 min, 40−10% A; and 15−20 min, 10% A. The flow rate was constant at 0.3 mL min^{-1}. After each sample run, the column was re-equilibrated to the initial conditions in 6 min. The injection volume was 5 μL. The column and autosampler tray temperatures were set at 60 and 5 °C, respectively. The mass spectrometer was operated in negative ion mode with a probe capillary voltage of 2.0 kV and a sampling cone voltage of 12.0 V. The source and desolvation gas temperatures were set to 120 and 350 °C, respectively. The nitrogen gas desolvation flow rate was 650 L h^{-1}. The mass spectrometer was calibrated across the range of m/z 50−1500 using a 0.5 mM sodium formate solution prepared in 2-propanol/water (90:10 v/v). Data were drift corrected during acquisition using a leucine encephalin (m/z 554.2615) reference spray (LockSpray) infused at 4 μL min^{-1}. Data were acquired in the range of m/z 50−1500, and the scan time was set to 1 s. Technical duplicates were acquired in all cases, except for 3 samples with too little volume for replicates. For UPLC−MS/MS experiments, the product ion mass spectra were acquired with collision cell voltages between 7 and 35 V and sampling cone voltages of 12 or 30 V, depending on the analyte. Ultra-high-purity argon (≥99.999%) was used as the collision gas in UPLC−MS/MS experiments. Data acquisition and processing were carried out using MassLynx, version 4.1 (Waters Corp., Milford, MA, USA).

Data Analysis

Spectral features [retention time (t_R), m/z pairs] were extracted from UPLC−MS data using Progenesis QI, version 2.0 (Nonlinear Dynamics, Waters Corp.). The procedure included retention time alignment, peak picking, integration, and deconvolution to group together adducts derived from the same compound (Figure 1). Subsequently, m/z values of all extracted features were input into the Metlin database[31] to perform a broad search for chemical compound candidates with an error window of 20 ppm, and 16.7% features with no candidates in the database were removed from the list. The remaining features were normalized after blank subtraction. Furthermore, only features that were present in at least 50% of one group class were retained. These were subject to a more stringent search against the Human Metabolome Database (HMDB)[32] using the elemental formula of the compound candidates in Metlin, and only those that had candidates with endogenous human or microbial origins were retained. The remaining features were further confirmed by MS/MS experiments.

The feature matrix obtained after this procedure was utilized to build models for sample discrimination via orthogonal partial least-squares-discriminant analysis (oPLS-DA[33,34]) by comparing the sample classes pairwise (MATLAB, R2015a, The MathWorks, Natick, MA with PLS-Toolbox, version 8.0, Eigenvector Research, Inc., Manson, WA). Reverse interval PLS-DA (iPLS-DA) was applied to autoscaled feature abundances to find the optimum number of latent variables (LVs) and a feature panel that maximized classification accuracy. The iPLS-DA interval size was set to 1, and the maximum number of LVs was set to 6. Leave-one-out cross-validation (LOOCV) and contiguous block cross-validation (three data splits) were used for oPLS-DA model building.

Metabolite Identification Procedure

Spectral features with tentative candidates in HMDB were targeted for identification based on (i) the accurate mass and isotopic pattern, (ii) tandem MS experiments where the respective precursor ions were quadrupole selected, and (iii) further validation against chemical standards (when available). For those cases in which MS/MS spectra were not available in the Metlin database, fragmentation patterns were manually interpreted for metabolite annotation. Commercially available standards were analyzed under identical conditions as EBC samples to validate putative metabolite identities by chromato-

DOI: 10.1021/acs.jproteome.6b00675
J. Proteome Res. XXXX, XXX, XXX−XXX

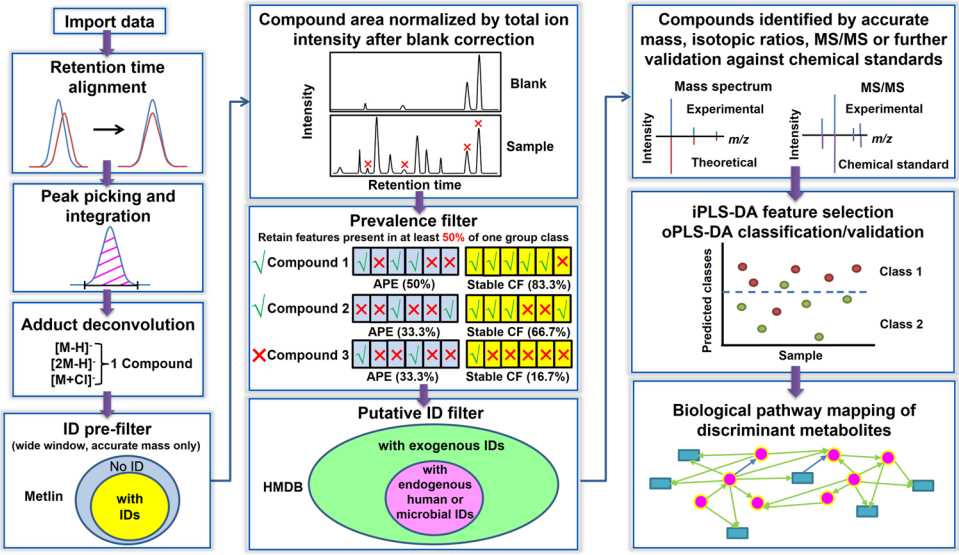


Figure 1. Data analysis workflow.

Table 1. Comparison of Discriminant Feature Panels

model/ panel #	classes compared	type of cross-validation	no. of features in initial set	no. of discriminant features in oPLS-DA model	discriminant feature codes	model accuracy	model specificity	model sensitivity
1	APE (9); stable CF (17)	3-block	9	2	397, 407	84.6	88.2	77.8
2	pre-APE (4); stable CF (17)	LOOCV	10	2	40, 407	90.5	94.1	75.0

graphic retention time matching and MS/MS fragmentation pattern matching.

RESULTS AND DISCUSSION

A total of 491 features were extracted from UPLC−MS negative ion mode data of the entire sample cohort by Progenesis software. Following Metlin filtering, 409 spectral features were kept (Figure 1). After deleting contaminant compounds such as known surfactants or plasticizers, background subtraction was applied to remove features in EBC samples that were also present in the sample blanks. If a feature had a maximum peak area in the blank runs that was one-third or more of the peak area of the same feature in EBC samples, then it was considered a contaminant[35] and its peak area in the corresponding EBC sample was set to 0. Otherwise, the maximum peak area in the blank samples was deducted from the feature peak areas in the EBC samples. Following background subtraction, features that had zero peak areas in 70% or more of the EBC samples from the studied class pairs were removed, resulting in 176 features that remained in the APE/stable CF class pair and 185 features in the pre-APE/stable CF class pair. This step was aimed at pruning out less significant groups of features. Of these, only features present in at least half of any group class were selected to increase the robustness of the final marker panel, leaving 144 features for the APE/stable CF classes and 159 features for the pre-APE/stable CF classes, when considered pairwise. As described in Figure 1, the feature data sets were further filtered to keep only those that also had tentative identities based on elemental formula searches in the HMDB database. Following this filtering, 20 features remained in the APE/stable CF class pair and 21 features in the pre-APE/stable CF class pair, and of these, only 9 and 10, respectively, could be confirmed by MS/MS experiments (Table S1) and were then subject to iPLS-DA feature selection process, as described in the next section. This rather stringent filtering approach was chosen to ensure that the features used for multivariate classification had a high certainty in terms of biochemical identity, therefore improving the chances of understanding their significance in the context of CF APE pathophysiology.

Classification Performance

Table 1 and Figure 2A describe results for the discrimination of 9 EBC samples collected from patients during an APE vs 17 EBC samples corresponding to stable CF patients. An optimum panel of 2 discriminant features (panel #1) was selected through the iterative iPLS-DA process. Following a three-block cross-validation approach, the corresponding oPLS-DA model yielded a classification accuracy of 84.6%, a sensitivity of 77.8%, and a specificity of 88.2%. One latent variable was used to build the oPLS-DA model that interpreted 44.7 and 36.3% variance from the *X* (feature peak areas) and *Y* (EBC class membership) blocks, respectively. Two EBC samples from the stable CF patient class and two samples from the APE class were systematically misclassified. Figure 2B,C shows box plots of peak areas for each discriminant feature in panel #1 and denotes fold changes obtained between the compared sample classes. The median instead of the mean peak area values were used to calculate fold changes to account for sample variability resulting from the relatively small sample size used in this class comparison. Interestingly, when the 4 pre-APE samples were used as an unknown sample set and input into this classification model, 3 out of 4 pre-APE samples were predicted as being similar to APE samples (Figure 2A), foreshadowing a metabolic fingerprint of APEs in the pre-APE EBC samples. Conversely, when the 5 post-APE samples were input into the APE vs stable CF PLS-DA model, 4 out of 5 post-APE samples were predicted as being like stable CF samples (Figure 2B), possibly suggesting that

DOI: 10.1021/acs.jproteome.6b00675
J. Proteome Res. XXXX, XXX, XXX−XXX

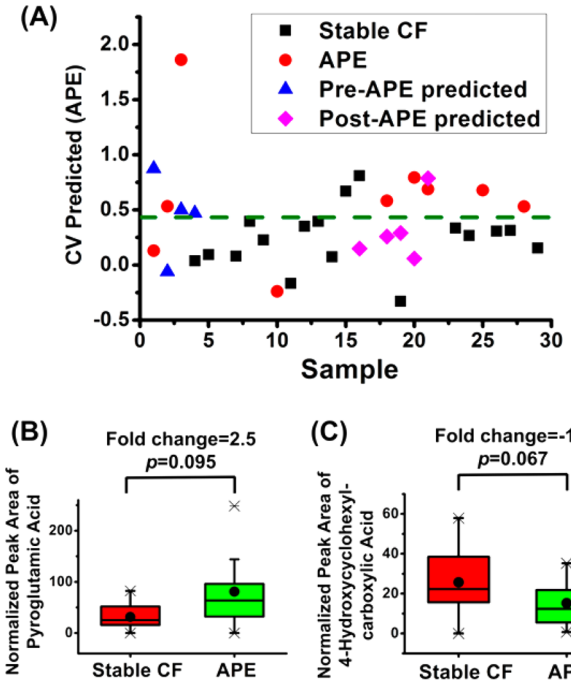


Figure 2. (A) oPLS-DA cross-validated classification plot and (B) box plots of peak areas of each discriminant feature in panel #1. (A) The *x*-axis represents sample number, and *y*-axis represents the cross-validated predicted scores of the oPLS-DA classification model. APE and stable CF samples are represented by red circles and black squares, respectively. The pre- and post-APE samples projected into the model are represented by blue triangles and magenta diamonds, respectively. The threshold for sample classification is represented by the green dashed line. Pre-APE samples are collected within 3 months before an APE event; APE samples are collected during an APE event; post-APE samples are collected within 3 months after an APE event; stable CF samples are collected for stable CF patients (3 months before or after collection there were no APE events). (B, C) Box plots for pyroglutamic acid and 4-hydroxycyclohexylcarboxylic acid, respectively, in EBC samples from stable CF and APE patient groups. Mean values are represented by a filled circle in the box; median values are represented by a line in the box; the edges of the box are 25th and 75th percentiles; the whisker extends to the most extreme values in data not including outliers, with a 99.3% coverage; outliers are represented by asterisks. The positive fold change is calculated as the ratio of median peak areas between APE and stable CF samples. The negative fold change is calculated as the negative ratio of median peak areas between stable CF and APE samples. *p* values are calculated from a Wilcoxon rank-sum test.

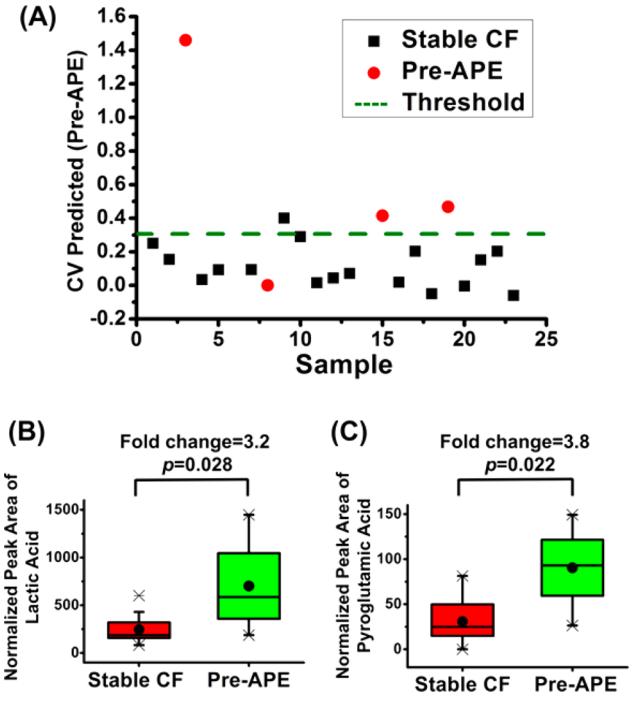


Figure 3. (A) oPLS-DA cross-validated classification plot and (B) box plots of peak areas of each discriminant feature in panel #2. (A) The *x*-axis represents sample number, and *y*-axis represents the cross-validated predicted scores by the oPLS-DA classification model. Pre-APE and stable CF samples are represented by red circles and black squares, respectively. The threshold for sample classification is represented by a green dashed line. (B, C) Box plots for lactic acid and pyroglutamic acid, respectively, in EBC samples from stable CF and pre-APE patient groups. Mean values are represented by a filled circle in the box; median values are represented by a line in the box; the edges of the box are the 25th and 75th percentiles; the whisker extends to the most extreme values in data not including outliers, with a 99.3% coverage; outliers are represented by asterisks. Fold changes are calculated as the ratio of median peak areas between pre-APE and stable CF samples. *p* values are calculated from a Wilcoxon rank-sum test.

following APE treatment the EBC metabolic profiles of the discriminant features of most post-APE patients resembled those of the stable CF patients.

With the aim of investigating if it was possible to discriminate the 4 EBC samples collected from patients in a pre-APE state from the 17 EBC samples from stable CF patients, a new oPLS-DA model was developed using a LOOCV approach. This approach was chosen due to the small number of pre-APE samples. Table 1 shows that classification was indeed possible with an accuracy of 90.5%, a sensitivity of 75.0%, and a specificity of 94.1% using a two-feature discriminant metabolite panel (panel #2) selected by iPLS-DA (Figure 3A). One EBC sample from a pre-APE patient and one sample from a stable CF patient were misclassified with this model, which used 1 latent variable and interpreted 87.2 and 44.6% variance from the *X* and *Y* blocks, respectively. Figures 3B,C show box plots for each discriminant feature in panel #2, with the respective median fold changes obtained for the compared sample classes. Interestingly, feature #407 was common to panels #1 and #2, but feature #40 was selected only in panel #2, suggesting that biomarkers of the asymptomatic phase preceding an APE event may be somewhat different from those associated with biochemical processes occurring during an exacerbation. Overall, these results highlight the feasibility of early indication of an oncoming APE event using these small metabolite panels (Tables 1 and 2), a possibility that could have significant implications in terms of pre-emptive APE diagnostics, enabling detection and treatment before irreversible damage to lung function occurs.

The cross-validation of classification models prevented overfitting to some extent. In addition, permutation tests were also performed to further validate the models. A pairwise Wilcoxon signed rank test was chosen for the cross-validated residuals since the population could not be assumed to be normally distributed due to the small sample size. For the model classifying pre-APE/stable CF class pair, the probability that the unpermuted model was not significantly different from the permuted models was 0.024, indicating that the original model was significant and not overfitted at the 95% confidence level. For the model classifying the APE/stable CF class pair, the probability that the unpermuted and permuted models were

DOI: 10.1021/acs.jproteome.6b00675
J. Proteome Res. XXXX, XXX, XXX−XXX

Table 2. Chemical Identification of Discriminant Features in EBC

feature code	used in model/ panel	retention time (min)	experimental m/z	ion type	elemental formula	Δm (mDa)	tentative annotation	method for tentative annotation	metabolite ID validation (with standard)
40	2	0.48	89.0231	$[M-H]^-$	$C_3H_6O_3$	−0.8	lactic acid	(*a*), MS/MS (*b*)	t_R, MS/MS match
407	1, 2	0.48	128.0343	$[M-H]^-$	$C_5H_7NO_3$	−0.5	pyroglutamic acid (5-oxoproline)	(*a*), MS/MS (*b*)	t_R, MS/MS match
397	1	0.82	143.0701	$[M-H]^-$	$C_7H_{12}O_3$	−0.7	4-hydroxycyclohexyl-carboxylic acid	(*a*)	t_R, MS/MS match

[a]Accurate mass and isotopic pattern matched. [b]Metlin database matched.

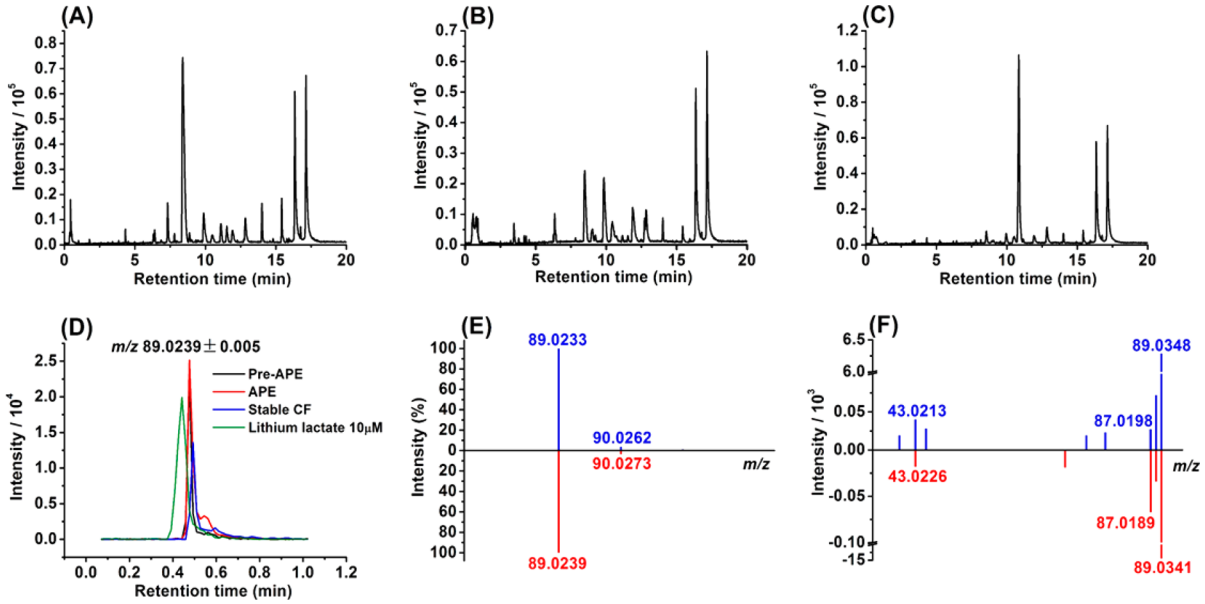


Figure 4. Base peak intensity (BPI) chromatograms obtained for EBC samples from the same patient at three different CF states: (A) pre-APE, (B) stable CF, and (C) during an APE event. (D) Extracted ion chromatogram for the discriminant feature with m/z 89.0239 ± 0.005 (lactic acid) generated from data in (A−C) and lithium lactate standard. (E) Experimental (top) and theoretical (bottom) mass spectra for the discriminant feature with m/z 89.0239 and t_R = 0.48 min. (F) MS/MS spectrum for m/z 89.0239 precursor ion using a collision cell voltage of 8 V and a sampling cone voltage of 30 V. The matching of MS/MS fragmentation between the experimental spectrum (top) and the chemical standard (bottom) is shown.

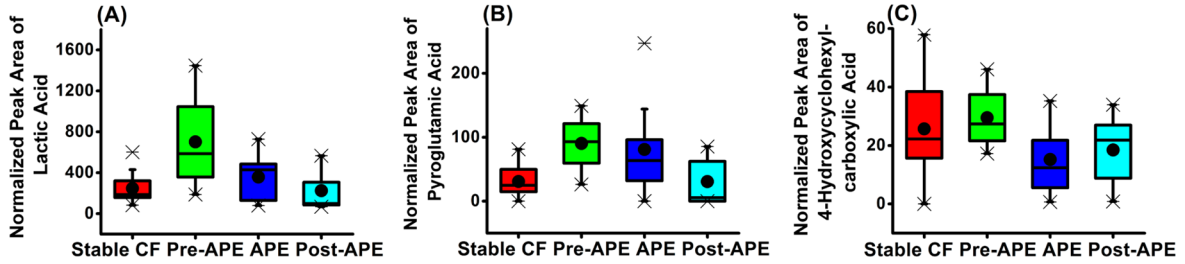


Figure 5. Box plots for three discriminant metabolites, (A) lactic acid, (B) pyroglutamic acid, and (C) 4-hydroxycyclohexylcarboxylic acid, in different subgroups of the sample cohort. Mean values are represented by a filled circle in the box; median values are represented by a line in the box; the edges of the box are the 25th and 75th percentiles; the whisker extends to the most extreme values in data not including outliers, with a 99.3% coverage; outliers are represented by asterisks.

indistinguishable was 0.087, indicating that the model was significant at least at the 90% confidence level.

Identification of Discriminant Metabolites and Their Biological Roles

Figure 4 describes the procedure used for unambiguous chemical identification of the discriminant features used in the oPLS-DA panels, using feature #40 as an example. Metabolic fingerprints from the same patient at three different CF states, pre-APE, stable, and during an APE event, are illustrated with the corresponding base peak intensity chromatograms (BPI) displayed in Figure 4, panels A−C, respectively. Extracted ion chromatograms (EICs) (Figure 4D) and corresponding mass spectra (Figure 4E, top) were obtained for features selected by iPLS-DA. According to the experimental monoisotopic mass of each feature, a series of possible candidates were generated from database searches in Metlin followed by HMDB and selected after matching of the observed and theoretical isotopic patterns (Figure 4E). Next, fragmentation patterns obtained from MS/MS experiments were compared to MS/MS spectra in Metlin, if available, or interpreted manually. Finally, the tentatively identified metabolites were confirmed by matching retention

DOI: 10.1021/acs.jproteome.6b00675
J. Proteome Res. XXXX, XXX, XXX−XXX

times and fragmentation patterns with chemical standards, whenever possible (Figure 4D,F).

Lactic acid (feature #40), which was selected by iPLS-DA for the model comparing pre-APE patient samples with stable CF ones, had a significant median fold increase of 3.2 from stable CF to pre-APE patient samples ($p < 0.05$, Figure 3B). Interestingly, a previous study using NMR reported increased levels of lactate in BALF from CF patients with high inflammation compared to those with low inflammation,[26] confirming that increased inflammation both prior to and during an APE event could be also detected in EBC by LC−MS. Increased lactate production has also been reported in patients with respiratory distress syndrome, finding it proportional to lung injury severity.[36] Lactic acid levels in the studied cohort possibly reflect the status of different stages preceding and following an APE event (Figure 5A). The higher levels of lactic acid in the pre-APE and APE patients compared to stable CF patients could possibly result from the increasingly hypoxic environment in CF lungs due to poorly cleared thick mucus developing on epithelial surfaces,[37] known to lead to an increased lactate conversion from pyruvate in anaerobic glycolysis. Lactate is also a glucose precursor in gluconeogenesis, and elevated gluconeogenesis has been found in CF-related diabetes (CFRD), possibly contributing to abnormal glucose tolerance in CF.[38] The decreasing trend of lactic acid from the pre-APE to the APE group in the studied cohort might be understood by considering that APE patients were treated with intravenous antibiotic therapy, so their inflammatory phenotype might be different from pre-APE patients, who had not yet been aggressively treated.

Interestingly, feature #407, identified as pyroglutamic acid (5-oxoproline), was present in both marker panels, suggesting that its relative alterations may reflect processes occurring both during APE as well as during the 3 month time window preceding the APE episode. Pyroglutamic acid had median fold increases from stable CF to APE and to pre-APE samples of 2.5 and 3.8, respectively (Figures 2B and 3C). Interestingly, in a recent serum metabolomics study of 31 CF vs 31 non-CF children reported by Joseloff et al., pyroglutamic acid was also identified as an important metabolite responsible for discrimination between CF and non-CF subjects.[28] This metabolite is a known intermediate in the gamma glutamyl cycle, a pathway for the biosynthesis and degradation of glutathione, and is thus related to redox imbalance. CF mutations cause a primary dysfunction in the glutathione system, leading to a systematic glutathione deficiency in the respiratory epithelial lining fluid, which is aggravated by oxidative burden.[39,40] Interestingly, decreased levels of glutathione have also been detected during exacerbations in EBC of children with asthma, hinting at some common mechanisms.[41] Figure 5B illustrates the relative concentrations of pyroglutamic acid in the different sample classes.

Feature #397 was selected by iPLS-DA for the model classifying APE from stable CF EBC samples, with a median fold decrease of 1.8 from stable CF to APE samples (Figure 2C). This feature was identified by both UPLC−MS and MS/MS and validated using a standard to be 4-hydroxycyclohexylcarboxylic acid, which is a relatively rare organic acid involved in gut microbial mammalian cometabolism, and it is a metabolite typically found in urine.[42,43] However, it has never before been reported in EBC to our knowledge. Interestingly, this type of metabolic gut−lung crosstalk has also been found to be associated with inflammatory bowel disease, in which the pulmonary inflammation is reported to accompany the main inflammatory processes in the bowel.[44] It is yet unclear, however, if these inflammatory processes are manifested through similar alterations in the respective lung and bowel metabolomes, as suggested by this finding. Figure 5C illustrates the relative concentrations of 4-hydroxycyclohexylcarboxylic acid in the different sample classes.

CONCLUSIONS

In our study, an untargeted UPLC−MS metabolomics method coupled to multivariate statistical analysis allowed identification of EBC metabolites related to APE events in CF patients. Orthogonal PLS-DA multivariate classification yielded acceptable accuracies (84.6%, 90.5%), sensitivities (77.8%, 75.0%), and specificities (88.2%, 94.1%) in distinguishing the 9 APE or 4 pre-APE EBC samples from the 17 stable CF samples, respectively. The discriminant metabolites included lactic acid, pyroglutamic acid, and 4-hydroxycyclohexylcarboxylic acid. Specifically, lactic acid was identified as a key biomarker for predicting an oncoming APE event. Since the limitation of this study is the relatively small sample size, a larger patient cohort will be needed to fully validate the present findings. Despite this limitation, these results show promise for new avenues that will allow for the detection of APEs and even the prediction of an oncoming APE event using EBC metabolites.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jproteome.6b00675.

Chemical identification of features in EBC with tentative identities in HMDB (Table S1) and cohort information (Table S2) (PDF)

AUTHOR INFORMATION

Corresponding Author

*E-mail: facundo.fernandez@chemistry.gatech.edu. Phone: 404 385 4432. Fax: 404 385 6447.

ORCID

Facundo M. Fernández: 0000-0002-0302-2534

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

Pilot funding was provided by the Emory+Children's Pediatric Center, supported by Emory University and Children's Healthcare of Atlanta. This study used EBC samples made available by the CF Biospecimen Registry, a part of the Emory+Children's Center for CF and Airways Disease Research. This project was also supported by philanthropic donations to the research program of the Emory+Children's CF Center of Excellence and by a pilot grant within the CF@LANTA Research Development Program as funded by the Cystic Fibrosis Foundation (MCCART15R0). M.E.M. is currently a Research Staff member from CONICET (Consejo Nacional de Investigaciones Científicas y Técnicas, Argentina).

REFERENCES

(1) Kreindler, J. L. Cystic fibrosis: Exploiting its genetic basis in the hunt for new therapies. *Pharmacol. Ther.* **2010**, *125*, 219−229.

DOI: 10.1021/acs.jproteome.6b00675
J. Proteome Res. XXXX, XXX, XXX−XXX

(2) Davis, P. B.; Drumm, M.; Konstan, M. W. Cystic fibrosis. *Am. J. Respir. Crit. Care Med.* **1996**, *154*, 1229−1256.
(3) Ramsey, B. W. Management of pulmonary disease in patients with cystic fibrosis. *N. Engl. J. Med.* **1996**, *335*, 179−188.
(4) *CF Patient Registry*. www.cff.org/Our-Research/CF-Patient-Registry.
(5) Britto, M. T.; Kotagal, U. R.; Hornung, R. W.; Atherton, H. D.; Tsevat, J.; Wilmott, R. W. Impact of recent pulmonary exacerbations on quality of life in patients with cystic fibrosis. *Chest* **2002**, *121*, 64−72.
(6) Lieu, T. A.; Ray, G. T.; Farmer, G.; Shay, G. F. The cost of medical care for patients with cystic fibrosis in a health maintenance organization. *Pediatrics* **1999**, *103*, e72.
(7) Rosenfeld, M.; Gibson, R. L.; McNamara, S.; Emerson, J.; Burns, J. L.; Castile, R.; Hiatt, P.; McCoy, K.; Wilson, C. B.; Inglis, A.; Smith, A.; Martin, T. R.; Ramsey, B. W. Early pulmonary infection, inflammation, and clinical outcomes in infants with cystic fibrosis. *Pediatr. Pulmonol.* **2001**, *32*, 356−366.
(8) Fuchs, H. J.; Borowitz, D. S.; Christiansen, D. H.; Morris, E. M.; Nash, M. L.; Ramsey, B. W.; Rosenstein, B. J.; Smith, A. L.; Wohl, M. E. Effect of aerosolized recombinant human DNase on exacerbations of respiratory symptoms and on pulmonary function in patients with cystic fibrosis. The Pulmozyme Study Group. *N. Engl. J. Med.* **1994**, *331*, 637−642.
(9) Saiman, L.; Marshall, B. C.; Mayer-Hamblett, N.; Burns, J. L.; Quittner, A. L.; Cibene, D. A.; Coquillette, S.; Fieberg, A. Y.; Accurso, F. J.; Campbell, P. W., III Macrolide Study, G. Azithromycin in patients with cystic fibrosis chronically infected with Pseudomonas aeruginosa: a randomized controlled trial. *JAMA* **2003**, *290*, 1749−1756.
(10) Ramsey, B. W.; Boat, T. F. Outcome measures for clinical trials in cystic fibrosis. Summary of a Cystic Fibrosis Foundation consensus conference. *J. Pediatr.* **1994**, *124*, 177−192.
(11) Goss, C. H.; Burns, J. L. Exacerbations in cystic fibrosis. 1: Epidemiology and pathogenesis. *Thorax* **2007**, *62*, 360−367.
(12) Wark, P. A.; Tooze, M.; Cheese, L.; Whitehead, B.; Gibson, P. G.; Wark, K. F.; McDonald, V. M. Viral infections trigger exacerbations of cystic fibrosis in adults and children. *Eur. Respir. J.* **2012**, *40*, 510−512.
(13) van Ewijk, B. E.; van der Zalm, M. M.; Wolfs, T. F.; van der Ent, C. K. Viral respiratory infections in cystic fibrosis. *J. Cystic Fibrosis* **2005**, *4*, 31−36.
(14) Goss, C. H.; Newsom, S. A.; Schildcrout, J. S.; Sheppard, L.; Kaufman, J. D. Effect of ambient air pollution on pulmonary exacerbations and lung function in cystic fibrosis. *Am. J. Respir. Crit. Care Med.* **2004**, *169*, 816−821.
(15) Smyth, A.; Elborn, J. S. Exacerbations in cystic fibrosis: 3−Management. *Thorax* **2008**, *63*, 180−184.
(16) Stenbit, A. E.; Flume, P. A. Pulmonary exacerbations in cystic fibrosis. *Curr. Opin. Pulm. Med.* **2011**, *17*, 442−447.
(17) Sanders, D. B.; Bittner, R. C. L.; Rosenfeld, M.; Hoffman, L. R.; Redding, G. J.; Goss, C. H. Failure to Recover to Baseline Pulmonary Function after Cystic Fibrosis Pulmonary Exacerbation. *Am. J. Respir. Crit. Care Med.* **2010**, *182*, 627−632.
(18) Osika, E.; Cavaillon, J. M.; Chadelat, K.; Boule, M.; Fitting, C.; Tournier, G.; Clement, A. Distinct sputum cytokine profiles in cystic fibrosis and other chronic inflammatory airway disease. *Eur. Respir. J.* **1999**, *14*, 335−338.
(19) Ordonez, C. L.; Henig, N. R.; Mayer-Hamblett, N.; Accurso, F. J.; Burns, J. L.; Chmiel, J. F.; Daines, C. L.; Gibson, R. L.; McNamara, S.; Retsch-Bogart, G. Z.; Zeitlin, P. L.; Aitken, M. L. Inflammatory and microbiologic markers in induced sputum after intravenous antibiotics in cystic fibrosis. *Am. J. Respir. Crit. Care Med.* **2003**, *168*, 1471−1475.
(20) Francoeur, C.; Denis, M. Nitric oxide and interleukin-8 as inflammatory components of cystic fibrosis. *Inflammation* **1995**, *19*, 587−598.
(21) Sloane, A. J.; Lindner, R. A.; Prasad, S. S.; Sebastian, L. T.; Pedersen, S. K.; Robinson, M.; Bye, P. T.; Nielson, D. W.; Harry, J. L. Proteomic analysis of sputum from adults and children with cystic fibrosis and from control subjects. *Am. J. Respir. Crit. Care Med.* **2005**, *172*, 1416−1426.
(22) Wolter, J. M.; Rodwell, R. L.; Bowler, S. D.; McCormack, J. G. Cytokines and inflammatory mediators do not indicate acute infection in cystic fibrosis. *Clin. Diagn. Lab. Immunol.* **1999**, *6*, 260−265.
(23) Salva, P. S.; Doyle, N. A.; Graham, L.; Eigen, H.; Doerschuk, C. M. TNF-alpha, IL-8, soluble ICAM-1, and neutrophils in sputum of cystic fibrosis patients. *Pediatr. Pulmonol.* **1996**, *21*, 11−19.
(24) Patti, G. J.; Yanes, O.; Siuzdak, G. Innovation: Metabolomics: the apogee of the omics trilogy. *Nat. Rev. Mol. Cell Biol.* **2012**, *13*, 263−269.
(25) Batal, I.; Ericsoussi, M. B.; Cluette-Brown, J. E.; O'Sullivan, B. P.; Freedman, S. D.; Savaille, J. E.; Laposata, M. Potential utility of plasma fatty acid analysis in the diagnosis of cystic fibrosis. *Clin. Chem.* **2007**, *53*, 78−84.
(26) Wolak, J. E.; Esther, C. R., Jr.; O'Connell, T. M. Metabolomic analysis of bronchoalveolar lavage fluid from cystic fibrosis patients. *Biomarkers* **2009**, *14*, 55−60.
(27) Yang, J.; Eiserich, J. P.; Cross, C. E.; Morrissey, B. M.; Hammock, B. D. Metabolomic profiling of regulatory lipid mediators in sputum from adult cystic fibrosis patients. *Free Radical Biol. Med.* **2012**, *53*, 160−171.
(28) Joseloff, E.; Sha, W.; Bell, S. C.; Wetmore, D. R.; Lawton, K. A.; Milburn, M. V.; Ryals, J. A.; Guo, L.; Muhlebach, M. S. Serum metabolomics indicate altered cellular energy metabolism in children with cystic fibrosis. *Pediatr. Pulmonol.* **2014**, *49*, 463−472.
(29) Mutlu, G. M.; Garey, K. W.; Robbins, R. A.; Danziger, L. H.; Rubinstein, I. Collection and analysis of exhaled breath condensate in humans. *Am. J. Respir. Crit. Care Med.* **2001**, *164*, 731−737.
(30) Hunt, J. Exhaled breath condensate: An evolving tool for noninvasive evaluation of lung disease. *J. Allergy Clin. Immunol.* **2002**, *110*, 28−34.
(31) Smith, C. A.; O'Maille, G.; Want, E. J.; Qin, C.; Trauger, S. A.; Brandon, T. R.; Custodio, D. E.; Abagyan, R.; Siuzdak, G. METLIN: a metabolite mass spectral database. *Ther. Drug Monit.* **2005**, *27*, 747−751.
(32) Wishart, D. S.; Jewison, T.; Guo, A. C.; Wilson, M.; Knox, C.; Liu, Y.; Djoumbou, Y.; Mandal, R.; Aziat, F.; Dong, E.; et al. HMDB 3.0—the human metabolome database in 2013. *Nucleic Acids Res.* **2013**, *41*, D801−D807.
(33) Worley, B.; Powers, R. Multivariate analysis in metabolomics. *Curr. Metabolomics* **2013**, *1*, 92−107.
(34) Boccard, J.; Rutledge, D. N. A consensus orthogonal partial least squares discriminant analysis (OPLS-DA) strategy for multiblock Omics data fusion. *Anal. Chim. Acta* **2013**, *769*, 30−39.
(35) Kirwan, J. A.; Weber, R. J.; Broadhurst, D. I.; Viant, M. R. Direct infusion mass spectrometry metabolomics dataset: a benchmark for data processing and quality control. *Sci. Data* **2014**, *1*, 140012.
(36) De Backer, D.; Creteur, J.; Zhang, H.; Norrenberg, M.; Vincent, J. L. Lactate production by the lungs in acute lung injury. *Am. J. Respir. Crit. Care Med.* **1997**, *156*, 1099−1104.
(37) Worlitzsch, D.; Tarran, R.; Ulrich, M.; Schwab, U.; Cekici, A.; Meyer, K. C.; Birrer, P.; Bellon, G.; Berger, J.; Weiss, T.; Botzenhart, K.; Yankaskas, J. R.; Randell, S.; Boucher, R. C.; Doring, G. Effects of reduced mucus oxygen concentration in airway Pseudomonas infections of cystic fibrosis patients. *J. Clin. Invest.* **2002**, *109*, 317−325.
(38) Hardin, D. S.; Ahn, C.; Rice, J.; Rice, M.; Rosenblatt, R. Elevated gluconeogenesis and lack of suppression by insulin contribute to cystic fibrosis-related diabetes. *J. Invest. Med.* **2008**, *56*, 567−573.
(39) Roum, J. H.; Buhl, R.; McElvaney, N. G.; Borok, Z.; Crystal, R. G. Systemic deficiency of glutathione in cystic fibrosis. *J. Appl. Physiol.* **1993**, *75*, 2419−2424.
(40) Hudson, V. M. New insights into the pathogenesis of cystic fibrosis: pivotal role of glutathione system dysfunction and implications for therapy. *Treat. Respir. Med.* **2004**, *3*, 353−363.
(41) Corradi, M.; Folesani, G.; Andreoli, R.; Manini, P.; Bodini, A.; Piacentini, G.; Carraro, S.; Zanconato, S.; Baraldi, E. Aldehydes and glutathione in exhaled breath condensate of children with asthma exacerbation. *Am. J. Respir. Crit. Care Med.* **2003**, *167*, 395−399.
(42) Sewell, A. C.; Bohles, H. J. 4-Hydroxycyclohexanecarboxylic acid: a rare compound in urinary organic acid analysis. *Clin. Chem.* **1991**, *37*, 1301−1302.

DOI: 10.1021/acs.jproteome.6b00675
J. Proteome Res. XXXX, XXX, XXX−XXX

(43) Zheng, X.; Xie, G.; Zhao, A.; Zhao, L.; Yao, C.; Chiu, N. H.; Zhou, Z.; Bao, Y.; Jia, W.; Nicholson, J. K.; Jia, W. The footprints of gut microbial-mammalian co-metabolism. *J. Proteome Res.* **2011**, *10*, 5512−5522.
(44) Wang, H.; Liu, J. S.; Peng, S. H.; Deng, X. Y.; Zhu, D. M.; Javidiparsijani, S.; Wang, G. R.; Li, D. Q.; Li, L. X.; Wang, Y. C.; Luo, J. M. Gut-lung crosstalk in pulmonary involvement with inflammatory bowel diseases. *World J. Gastroenterol.* **2013**, *19*, 6794−6804.

DOI: 10.1021/acs.jproteome.6b00675
J. Proteome Res. XXXX, XXX, XXX−XXX