Published in final edited form as:

J Proteome Res. 2014 February 7; 13(2): 627–639. doi:10.1021/pr4007602.

A Large Scale, Ion Current based Proteomics Investigation of Bronchoalveolar Lavage Fluid in Chronic Obstructive Pulmonary Disease Patients

Chengjian Tu^{#1,2}, Manoj Jacob Mammen^{#3}, Jun Li^{1,2}, Xiaomeng Shen^{1,2}, Xiaosheng Jiang^{1,2}, Qiang Hu⁵, Jianmin Wang⁵, Sanjay Sethi^{*,3,4}, and Jun Qu^{*,1,2}

¹Department of Pharmaceutical Sciences, University at Buffalo, State University of New York, Buffalo, NY 14260 USA

² New York State Center of Excellence in Bioinformatics and Life Sciences, 701 Ellicott Street, Buffalo, NY 14203 USA

³University at Buffalo, SUNY

⁴WNY VA Healthcare System, NY 14203 USA

⁵Department of Biostatistics and Bioinformatics, Roswell Park Cancer Institute, Buffalo, NY 14203.

These authors contributed equally to this work.

Abstract

Proteomic analysis of bronchoalveolar lavage fluid (BALF) in chronic obstructive pulmonary disease (COPD) patients may provide new biomarkers and deeper understanding of the disease mechanisms but remains challenging. Here we describe an ion current based strategy for comparative analysis of BALF proteomes from patients with moderate and stable COPD vs. healthy controls. The strategy includes an efficient preparation procedure providing quantitative recovery and a nanoBLC/MS analysis with a long, heated column. Under optimized conditions, high efficiency and reproducibility were achieved for each step, enabling a “20Bplex” comparison of clinical subjects ($n=10/\text{group}$). Without depletion/fractionation, a total of 423 unique protein groups were quantified under stringent criteria with at least two quantifiable peptides. Seventy six proteins were determined as significantly altered in COPD, which represent a diversity of biological processes such as alcohol metabolic process, gluconeogenesis/glycolysis, inflammatory response, proteolysis, and oxidation reduction. Interestingly, altered alcohol metabolism responding to oxidant stress is a novel observation in COPD. The prominently elevated key enzymes involved in alcohol metabolism (e.g. ADH1B, ALDH2&ALDH3A1) may provide a reasonable explanation for a bewildering observation in COPD patients known for decades: the underestimation of the blood alcohol concentrations through breath tests. These discoveries could provide new insights for identifying novel biomarkers and pathological mediators in clinical studies.

*Corresponding Authors: Sanjay Sethi, M.D. VA WNY Healthcare System 3495 Bailey Avenue Buffalo, NY 14215 Phone: (716) 862B7875 Fax: (716) 862B6526 ssethi@buffalo.edu Jun Qu, Ph.D. The Department of Pharmaceutical Sciences 319 Kapoor Hall, University at Buffalo State University of New York Buffalo, NY 14260B1200 Phone: (716) 645B2844 x283 Fax: (716) 645B3693 junqu@buffalo.edu.

Keywords

Biomarker Discovery; Bronchoalveolar Lavage Fluid; Chronic Obstructive Pulmonary Disease; Peptide Extracted Ion Current

INTRODUCTION

Chronic obstructive pulmonary disease (COPD) is the third leading cause of death in the US and the fifth leading cause worldwide and its prevalence is expected to increase dramatically in the next decades (1B3). Despite its major impact on public health, the diagnosis and staging of COPD is still evolving, and the fundamental mechanisms involved in COPD onset and progression have not been fully elucidated, largely due to the complexity and heterogeneity of this disease. Therefore, the development of diseaseBspecific and sensitive biomarkers is highly desirable, as these will contribute substantially to the understanding disease pathogenesis and the development of novel diagnostic and therapeutic strategies.

Bronchoalveolar lavage (BAL) procedure permits the sampling of cellular and soluble materials from the distal airways and alveolar spaces. The proteins in bronchoalveolar lavage fluid (BALF) are either locally released by epithelial and inflammatory cells or from plasma exudation (4, 5). As a result, the expression profiling of BALF proteome in COPD patients not only provides a pool of valuable biomarker candidates, but also may obtain critical insights into pathogenesis and lung defense mechanisms (4, 5). However, despite the tremendous progresses of proteomic technologies in the recent years, quantitative characterization of BALF proteomics in COPD remains a daunting challenge. First, the variability among clinical subjects is high and the lavage procedure may introduce additional quantitative variations, e.g. the variable dilution effects during sampling (6). Therefore, the use of a relatively high number of clinical replicates (e.g. $n=10$ per group) to alleviate the effects of individual variability and to increase statistical power of results, and a reliable sample preparation procedure that provides high, quantitative recovery of BALF proteins are required.

Second, the BALF proteome is dominated by highBabundance proteins and encompasses a large proteinBconcentrationBdynamicBrange comparable to that of plasma (7, 8). Previous studies showed the top 6 most abundant plasmaBderived proteins in BALF may have accounted for >90% of the total protein mass (7). Owing to this challenge, recent works using two- dimensional gel electrophoresis-mass spectrometry (2BDE/MS) and surfaceBenhanced laser desorption/ionization timeBofBflight mass spectrometry (SELDIBTOFBMS) for differential analysis of BALF from COPD patients, though representing important progresses in the field, quantified a limited number of proteins (8, 9). Immunodepletion of the highBabundance proteins is widely practiced for the analysis of plasma (10, 11) and has been successfully applied in BALF (7). Nonetheless the technique may not be practical for investigating many individual BALF samples, because of the very limited amounts of BALF proteins from each subject. Moreover, though fractionation by techniques such as strong cation exchange may drastically improve the depth of analysis, it

is not feasible to perform multiple fractionations for each of the many BALF samples while maintaining a good quantitative accuracy and precision.

In this study, we addressed the above challenges by employing an extensive, reproducible and well-controlled proteomics strategy, which enables a confident quantitative comparison using many biological replicates and an in-depth characterization of BALF proteomes. The ion- currentBbased strategy includes i) an efficient sample preparation procedure that normalized the variable protein concentrations in BALF samples while providing high protein/peptide recovery; ii) a highBresolution and reproducible nano-LC separation (e.g. 7Bhr separation on a 75-cm-long column packed with small particles) that allows extensive analysis of BALF samples with a large biological replicates; iii) and an ion current area based label-free quantification method that provides high quantitative precision. The overall proteomics procedure was carefully optimized and controlled, and then applied in the profiling of BALF proteomes from patients with stable COPD (n=10) vs. healthy subjects (n=10), a “20Bplex” quantitative analysis. The significantlyBalterred proteins quantified with high confidence were subjected to functional annotation. Groups of proteins of high interest were further validated by immunoassays.

MATERIALS AND METHODS

Recruitment of Subjects

This study was performed on BALF obtained in the course of a study of innate lung defense in COPD. All procedures received approval of the Institutional Review Board, VAWNY Healthcare System, and were in strict accordance with institutional guidelines. After informed consent, volunteers were screened for inclusion into one of three groups: i) healthy nonBsmokers, ii) exBsmokers with COPD that had quit smoking for at least one year and iii) Current smokers with COPD. All participants underwent clinical assessment, routine spirometry and chest x-rays. Ex-smokers and healthy non-smokers had expired breath carbon monoxide of <0.02 ppm (Vitalograph Breath CO Monitor, Lenexa, KS).

All participants were at least 30 years of age. Inclusion criteria of COPD patients were: 1) At least 10 packByears of smoking history 2) FEV₁/FVC ratio below the 95% confidence limit of normal on spirometry; 3) chronic bronchitis by history and/or emphysema by chest x-ray and/or Computed Tomography (CT) scan; 4) absence of other lung disease, including asthma and bronchiectasis, based on clinical and radiological evaluation; 5) non-atopic by history; 6) no antibiotic or systemic steroid use for four weeks preceding enrollment. Exclusion criteria included a forced expiratory volume at one second (FEV₁) <35%, hypercapnia and co-morbid diseases that would render bronchoscopy unsafe. Healthy non-smokers had normal lung function and had no lung disease by clinical ad radiological evaluation.

For the purposes of this study, of the samples obtained from COPD patients, additional selection criteria were applied to obtain a phenotypically well-characterized group of patients (e.g. to avoid the complication of active smoking in COPD patients). We selected 10 subjects from group ii) who were all ex-smokers and had moderate COPD, as defined by

the GOLD criteria with an FEV₁% predicted of 50-80% (12). From the healthy controls, 10 samples were selected that were matched for age and sex.

Bronchoscopy and BALF sample preparation

The research bronchoscopy was performed as has been described previously (13). The bronchoscope was wedged in a segment of the right middle lobe and four sequential aliquots of 50 ml of normal saline were introduced, with gentle suction to retrieve the lavage fluid after each aliquot. The returned fluid of the 2nd to 4th aliquot was pooled and further processed as the BALF. The BALF was kept on ice right away and processed expeditiously. Cells were recovered by centrifugation for 10 min at 500 g, and the supernatant was spiked with cocktails of protease and phosphatase inhibitors (Roche Applied Science, Indianapolis, IN), and then stored in 10-mL aliquots at -80 °C. The protein concentrations were measured using BCA Protein Assay (Pierce, Rockford, IL) before storage. In order to enrich the proteins and normalize protein concentrations, an acetone precipitation/on-pellet digestion procedure that provides reproducible protein/peptide recovery was employed (14, 15). Each sample was added with 1 volume of pre-chilled acetone with vortexing to form a cloudy suspension, and then stepwise adding another 5 volumes of cold acetone with continuous vortexing that resulted in compact sediments with small, uniform size. The mixture was then incubated overnight at -20 °C. After centrifugation at 16,000 g for 30 min at 4 °C, the supernatant was removed and the pellet was dissolved in a strong, detergent- containing buffer (50 mM Tris-formic acid (FA), 150 mM NaCl, 1% sodium deoxycholate, 2% Sodium Dodecyl Sulfate (SDS), 2% NP-40, pH 8.0) and adjusted to a proper final volume that gave all samples a uniform protein concentration (1 mg/mL). Proteins (100 µg) were reduced with Tris(2-carboxyl)phosphine (TCEP) (3 mM) for 10 min and then alkylated with 20 µM iodoacetamide (IAM) for 30 min in darkness. The extract is processed by a precipitation/on-pellet-digestion procedure (*Supplemental Methods*).

Precipitation/on-pellet-digestion

The mixture was precipitated by stepwise addition of 9 volumes of cold acetone with continuous vortexing and then incubated overnight at -20 °C. After centrifugation at 20,000 g for 20 min at 4 °C, the supernatant was removed and the pellet was allowed to air-dry. Two consecutive digestion steps were employed for the on-pellet- digestion. In step 1 (digestion-aided pellet dissolution), 300 µL Tris buffer (50 mM, pH 8.5) containing trypsin at an enzyme/substrate ratio of 1:30 (w/w) was added and incubated at 37 °C for 6 h with vortexing at 120 rpm in a New Brunswick shaking incubator (Edison, NJ); in step 2 (complete cleavage), another 300 µL trypsin solution was added at an enzyme/substrate ratio of 1:25 (w/w). The mixture was incubated at 37 °C overnight (12 h) to achieve a complete digestion.

Long gradient nano-RPLC/Mass Spectrometry

An extensive and reproducible separation was carried out by a unique LC/MS strategy. The Nano-RPLC system consisted of a Spark Endurance autosampler (Emmen, Holland) and an ultra-high pressure Eksigent (Dublin, CA) Nano-2D Ultra capillary/nano-LC system. To achieve a comprehensive separation of the complex peptide mixture, a nano-LC/nanospray

setup, which features low void volume and high chromatographic reproducibility (15), was employed. Mobile phase A and B were 0.1% formic acid in 2% acetonitrile and 0.1% formic acid in 88% acetonitrile, respectively. Samples containing 6 μ g of peptides were loaded onto a large-ID trap (300 μ m ID x 1 cm, packed with Zorbax 3- μ m C18 material) with 1% B at a flow rate of 10 μ L/min, and the trap was washed for 3 min before being brought in line with the nano-LC flow path. A series of nanoflow gradients (the flow rate was 250 nL/min) was used to back-flush the trapped samples onto the nano-LC column (75 μ m ID x 75 cm, packed with Pepmap 3- μ m C18 material) for separation. The nano-LC column was heated at 52 °C to greatly improve both chromatographic resolution and reproducibility. A 7Bhr shallow gradient was used to achieve sufficient peptide separation. The optimized gradient profile was as following: 3 to 8% B over 15 min; 8 to 24% B over 215 min; 24 to 38% B over 115 min; 38 to 63% B over 55 min; 63 to 97% B in 5 min, and finally isocratic at 97% B for 15 min.

An LTQ Orbitrap mass spectrometer (Thermo Fisher Scientific, San Jose, CA) was used for protein identification. The nanospray needle was washed and conditioned with dripping 50% methanol after every three runs to maintain reproducible ionization efficiency. The instrument was operating under data-dependent product ion mode. One scan cycle included an MS1 survey scan (m/z 310-2000) at a resolution of 60,000 followed by seven MS2 scans by CID activation mode, to fragment the top seven most abundant precursors in the survey scan. The target value for MS1 by Orbitrap was 8×10^6 , under which the Orbitrap was calibrated carefully for mass accuracy and FT transmission. The use of high target value on the Orbitrap enabled a highly sensitive detection without compromising the mass accuracy and resolution. The dynamic exclusion was enabled with the following settings: repeat count = 1; repeat duration = 30 s; exclusion list size = 500; exclusion duration = 40 s. The activation time was 30 ms with an isolation width was 3 Da for ITMS; the normalized activation energy was 35% and the activation q was 0.25. Ten samples each of BALF from patients with COPD and healthy controls were analyzed randomly.

Protein Identification and Ion-current-based Quantification

Individual raw files were searched against the UniProt human protein database (released on September, 2011) with a total of 20,261 protein entries using Sequest imbedded in Proteome Discoverer (version 1.2.0.208, Thermo-Scientific). The search parameters used were as follows: 15 ppm tolerance for precursor ion mass and 1.0 Da for fragment ion mass. Two missed cleavages were permitted for fully tryptic peptides. Carbamidomethylation of cysteines was set as a fixed modification and a variable modification of methionine oxidation was allowed during database search. The false discovery rate was estimated by a target-decoy search strategy (16), with a concatenated database containing both forward and reversed sequences. The thresholds determined for identification were: Δ CN 0.1 and Xcorr 1.85 for singly-charged (1+) peptides, 2.3 for 2+, 3.5 for 3+ and 3.8 for 4+... Ion current-based quantification was performed using the SIEVE v2.0.180, which includes components for chromatographic alignment, global intensity-based MS1 feature detection/extraction and protein identification/quantification assignment {Tu, 2013 #66}. The detailed procedure is shown in Supplemental Fig. 1. The raw files were imported and assigned into different experimental groups (e.g. control vs. COPD). The LC-MS runs were aligned with

an adaptive tiled algorithm- ChromAlign algorithm {Sadygov, 2006 #67}, and then quantitative frames containing the group of ion current AUC (area-under-the-curve) data from each peptide were defined by the threshold windows of $m/z < \pm 0.01$ amu and retention time $< \pm 1$ min for frame definition. Normalization of peak areas was achieved for individual samples based on the sum of total ion current (TIC) area in each run. In order to match the peptide ID represented by each frame, the identification files (.msf) produced by Proteome Discoverer (described above) were imported. Moreover, an eligible "Quantifiable protein" must have at least two distinct quantifiable peptides that met all above-mentioned identification and quantification criteria. Peptides shared among proteins are removed so that only peptides with unique protein assignments were considered. AUC (area-under-the-curve) of peptides were calculated and compared between the two experimental groups to assess relative expression ratios and then t-test and Fisher's Combined Probability test was performed. Missing AUC data, where a well-defined peak was not detected in a certain sample, was assigned as "0" by Sieve. In order to incorporate the ratios from peptide level to protein level, a variance-based model was employed, as shown in the following formula: Protein ratio = $\text{sum}(\text{peptide ratio}/\text{stdev}^2) / \text{sum}(1/\text{stdev}^2)$.

A pooled sample (equally-mixed healthy subjects and patients) was prepared and then analyzed 20 times to evaluate the quantitative precision and accuracy by the overall method. Additionally, to experimentally measure the extent of false-positive discovery of altered proteins and to estimate the proper cutoff thresholds, a "Sham" sample set, which is consisted 20 *individual* healthy control subjects that are free of COPD and without any on-going medical procedure, were analyzed and evaluated, following a procedure we described previously (17). Briefly, these samples were randomly assigned in either ShamBcontrol or Sham-experimental groups (N=10/group), and then the two groups were experimentally compared under the same procedure and cutoff thresholds as these employed for the COPD/healthy comparison. The extents of false-positive discovery were evaluated in the Sham samples.

Western Blot analyses

In order to provide independent verification of differentially expressed proteins in the 2 groups, Western blot analyses of selected proteins was performed. Protein samples (20 μ g/each) in lysis buffer (50 mM Tris-FA, 150 mM NaCl, 0.5% sodium deoxycholate, 2% SDS, 2% NPB40, pH8.0) and broad range markers (Santa Cruz, CA) were separated in small-sized 4-12% polyacrylamide gels by SDS-PAGE (Invitrogen). Proteins were electrophoretically transferred to nitrocellulose (NC) membranes. The membrane was then blocked for 1 h with Western blocking solution (Invitrogen) and sequentially incubated with a primary antibody followed by an appropriate secondary antibody conjugated with horseradish peroxidase (HRP) (Santa Cruz, CA). The positive immunoreactions were detected with xBray film by chemiluminescence using the ECL Western blotting kit (Pierce, Rockford, IL) and developed by a Kodak X-OMAT 2000A Processor. When reprobing a western blot, the stripping buffer from Fisher Scientific was used. The primary antibodies used in this study were as follows: rabbit polyclonal anti-Cathepsin D (1:500; Santa Cruz Biotechnology), mouse polyclonal anti- alcohol dehydrogenase 1B (ADH1B) (1:500; Novus Biologicals), mouse monoclonal anti- aldehyde dehydrogenase 2 (ALDH2) (1:200; Santa

Cruz Biotechnology), mouse monoclonal anti-aldehyde dehydrogenase 3 family, member A1 (ALDH3A1) (1:200; Santa Cruz Biotechnology), mouse monoclonal antiBfibrinogen β (FGB) (1:200; Santa Cruz Biotechnology), mouse monoclonal anti-galectin-3 (1:100; Santa Cruz Biotechnology), rabbit monoclonal anti- serum amyloid P antibody (SAP/APCS) (1:500; Abcam), goat polyclonal anti-transgelin-2 (1:200; Santa Cruz Biotechnology), and mouse monoclonal anti-albumin (ALB) (1:1000; Santa Cruz Biotechnology).

Bioinformatics Analysis

Gene Ontology annotation was performed using the online tool High-Throughput GoMiner (<http://discover.nci.nih.gov/>) (18). Biological processes assigned by the software were manually examined, and re-grouped into respective categories. The prediction of secreted proteins was performed by using SignalP 4.1 Server (19). This version of SignalP was able to discriminate between signal peptides and transmembrane regions. The information on the tissue-specific expression was extracted from UniProt Knowledgebase. The theoretical molecular weight (MW) and isoelectric point (pI) were calculated by Compute pI/Mw tool (<http://web.expasy.org/>). We analyzed the differentially expressed protein list generated from the comparative profiling analysis using SIEVE package.

RESULTS

Demographics

Samples from 20 subjects with balanced demographic characteristics were carefully selected for the proteomics study and are shown in Table 1. Details of inclusion/exclusion criteria are shown in Experimental, and the strategy to achieve a well-controlled clinical study is discussed in the Discussion section.

The development and evaluation of the proteomics strategy

The label-free profiling approach is preferred in this study due to the need of investigating a relatively large number of biological replicates ($n=10$ clinical subjects per group). Furthermore, the ion-current-based method was chosen over spectral counting due to its potential for higher sensitivity of detecting altered proteins and superior accuracy when conducted in a well- controlled manner (20-22). Fig.1 illustrates the procedural schematics of the proteomics strategy, which consists the following steps: an efficient and quantitative procedure for normalization of protein concentrations in BALF, sample cleanup and digestion; an extensive, reproducible and high-resolution nano-LC separation followed by a sensitive MS detection; an ion-current-based quantification and the method evaluation. Each step was thoroughly optimized for high sensitivity, reproducibility and quantitative accuracy.

a) Normalization of protein concentrations and quantitative clean-up and digestion—The varied protein concentrations in procured BALF (in the range of 0.07-0.27 mg/mL) necessitate normalization. The precipitation procedure (*Method*) provided high and constant protein recoveries (83-90%) within the concentration range of the BALF samples, as shown in Supplemental Fig. 2. The pelleted proteins were then reconstituted with a strong extraction buffer (2% NP-40, 2% SDS and 1% sodium deoxycholate, etc) to a uniform

concentration. A quantitative precipitation/on-pellet-digestion procedure (14, 15, 23) was used to remove detergents and other non-protein components in the reconstituted samples, and then to perform complete digestion. High and reproducible peptide yields (mean=81.1%) with excellent inter-day reproducibility (RSD=7.3%) were achieved for the 20 samples (supplemental Fig. 3). Such a high level of recovery and reproducibility greatly contributed to a reliable ion-current-based quantification.

b) Ion-current-based strategy for the investigation of BALF proteomes—

To achieve an extensive and accurate quantification of the BALF proteome, a high-resolution and reproducible chromatographic strategy was developed, as described in the *Methods*. Separation was carried out on a long, heated nano-column (75-cm-long, 3- μ m and heated at 52°C) with a shallow, 7-hr gradient (*Method*). The peptide elution windows were > 345 min with an average peak width <30 s (FWHM, excluding peaks of some very abundant peptides from human serum albumin) and a peak capacity >580. A regular tip washing approach was employed to maintain stable ionization efficiency and thus constant MS response (*Method*). An Orbitrap analyzer with “overfilling” approach to enhance sensitivity without compromising MS accuracy and resolution (14, 24, 25), was used for sensitive detection.

The chromatographic conditions were optimized to achieve the following goals: i) the maximal peak capacity is obtained; ii) no perceivable loss of hydrophilic and hydrophobic peptides; iii) separation is reproducible among runs. Detailed conditions are shown in *Experimental*. To evaluate the analytical reproducibility, we performed 20 consecutive analyses of a pooled BALF sample. The assessment was carried out based on fifteen representative peptides that were randomly selected within each consecutive 20Bmin segment of the elution window (details not shown). The reproducibility for the twenty 7-h runs was excellent, as expressed by the low variations of 0.12-0.35% and 8-17% respectively for the retention times and area-under-the- curves (AUC) of peptide ion currents across the 15 peptides. Such an excellent reproducibility can be attributed to the efficient gradient mixing and dampened pump noise by the nano-LC/NSI setup, and the homogenous heating of the long column(17). We randomly split the 20 analyses into two groups ($n=10$ per group), and then analyzed the distributions of peptide and protein relative expression ratios (true value was 1) between the two groups. The results are shown in Fig. 2. For relative expression ratios of peptides, the majority (>96%) fell into the ratio range of 0.67-1.5 (left panel of Fig. 2), while for the relative expression ratios of proteins, nearly all (>99%) fell into the same range (right panel of Fig. 2). This result suggests the high reliability of the relative quantification. The information of the peptide/protein quantified by ion-current- based approach in the technical replicate dataset is shown in supplemental table I. We also employed spectral counting, a more traditional method for relative quantification in shotgun proteomic analyses (11, 26, 27), to perform the comparison between these two approaches. The ratios of proteins (true value was 1) determined by ion-current-based approach were more tightly centered around the theoretical value than by spectral counting (Supplemental Fig. 4), indicating better quantitative accuracy for relative protein quantification by ion-current-based approach. e Good quantitative consistencies were observed for the spectral count method among the 20 analyses and some representative data is shown in Supplemental Fig. 5. For example, the spectral counts for 1st analysis vs. the 2nd,

the 10th and the 20th analyses showed excellent agreement, with Pearson correlation coefficients of 0.9996, 0.9940, and 0.9926, respectively. These results agreed with the ion-current-based evaluation that highly reproducible analyses were achieved by the developed method.

Proteomics profiling in BALF of COPD and healthy control subjects

With the developed method, we performed an extensive contrast of the BALF proteomes of COPD patients ($n=10$) vs. healthy controls ($n=10$). A set of stringent criteria on both protein identification and quantification were employed to ensure confident discovery of altered proteins (*Methods*). Under the stringent criteria, totally 423 unique protein groups (supplemental Table II) were quantified with high confidence. The distributions of these proteins by molecular weight (MW), isoelectric point (pI), hydrophobicity and tissue expression are shown in Supplemental Fig. 6. As expected, these proteins covered a wide range of MW and pI (supplemental Fig. 5A and 5B). The cellular location information of 410 proteins was obtained by Gene Ontology analysis. The majority of proteins are related to extracellular region (182, 44.4%) and plasma membrane (119, 29.0%, supplemental Fig. 5C). As to the tissue expression distribution of the 423 proteins, proteins related to liver/plasma, lung, platelet and skeletal muscle present as the predominant components in BALF proteins (supplemental Fig. 5D). Secreted proteins was predicted using SignalP 4.1 Server (19) and 167 (39.5%) out of the 423 quantified proteins were designated as secreted proteins.

In order to determine the proper cutoff for confident discovery of altered proteins, we experimentally estimated the false-positive discovery of altered proteins. Procedures using Sham sample sets for false-discovery are described previously (17). In order to evaluate the effects of both biological and technical variations, 20 BALF samples from healthy projects were used. Under the optimal cutoff threshold (1.5-fold change and $p < 0.05$), only 1 false-positive were discovered (Supplemental Fig. 7). By comparison, 76 protein groups (out of a total of 423) showed significant changes in the COPD vs. healthy sample set.

Out of the 76 proteins significantly altered in the BALF of COPD patients vs. healthy subjects, 50 proteins were elevated and 26 were decreased in COPD. The list of the significantly elevated proteins is shown in Table II. Among the peptides assigned to elevated proteins, the majority (94.2%) have the relative expression ratio more than 1.0 as shown in Supplemental Fig. 8A. Similar results were obtained with the peptides assigned to decreased proteins (Supplemental Fig. 8B).

Functional annotation of the altered proteins in COPD group

We categorized the 76 altered proteins by their biological processes and cellular localizations using Gene Ontology analysis, where 72 and 76 proteins have known terms respectively in these two categories as shown in supplemental Table III and supplemental Table IV. In terms of biological processes, the altered proteins are involved in alcohol metabolic process (12 proteins), cell migration (11), inflammatory response (13), oxidation reduction process (9), platelet activation (10) and proteolysis (10), etc. (Fig. 3). Interestingly, altered proteins associated with alcohol metabolic process, oxidation reduction

and gluconeogenesis were all elevated in the BALF from COPD patients. A detailed discussion on these processes was shown below. As to the cellular locations, most of the differentially expressed proteins were located in extracellular region (38 proteins, 50%) and plasma membrane (25, 32.9%), as shown in supplemental Table IV. We also performed a thorough literature search which reveal that 23 out of the 76 altered proteins have been found in previous studies in lung tissue, serum or BALF, while the others are novel discoveries (details in Supplemental Table IV). The estimated abundances of these altered proteins are also shown in Supplemental Table IV.

Validation of the differentially-expressed proteins of high interest

In this study, 8 altered proteins, cathepsin D, galectin 3, ADH1B, transgelin 2, fibronogen β , SAP, ALDH2, and ALDH3A1 were selected for further validation by Western blot analyses, owing to their potentially high clinical and functional significances (*cf.* the Discussion section). BALF samples from 5 *randomly-selected* COPD subjects and 5 *randomly-selected* control subjects were investigated. All of these eight proteins, which were found elevated in BALF samples from COPD patients in our proteomics data, also showed elevated levels with a p-value of less than 0.05 by Western blot analysis (Fig. 4). This confirmed the high reliability and quantitative accuracy of proteomics strategy. Moreover, the fact that the alterations of the 8 proteins were validated in a different set of individuals than these used for profiling, implied the general clinical applicability of these results. Because proteins that are used routinely as loading controls for cell and tissue samples (e.g. beta actin; glyceraldehyde 3-phosphate dehydrogenase and tubulin) are not applicable for BALF samples, we used the human serum albumin as the load control, whose level were found constant in BALF by both by Western blot (Fig. 4) and the ion- currentBbased profiling (Supplemental Table II).

DISCUSSION

The well-controlled clinical study design

Given the highly complex pathology of COPD, in order to attain a reliable proteomic profiling, we employed a group of 10 phenotypically wellBcharacterized COPD subjects. We excluded current smokers, as smoking is an inflammatory stimulus and induces acute changes in bronchial epithelial cells, which in turn could have significant impacts on the BALF proteome(8). All our participants were in the *stable* phase of their COPD, and free of exacerbation for at least 4 weeks. COPD exacerbations are episodes of increased respiratory symptoms related to infection-induced inflammation in the airway. Profound changes in the BALF proteome are likely to occur during such exacerbations; hence only stable patients were studied. Another source of clinical variability can be the severity of the disease process of COPD itself. The standard way to determine this severity is the FEV1 percent predicted, and on that basis, COPD severity is classified as mild, moderate, severe and very severe (40). We restricted our samples to those obtained from patients with moderate disease. The subgroups of patients not included in this study, i.e. current smokers, exacerbations, and other severities of COPD need to be examined in the future to provide more comprehensive insights into the pathogenesis of COPD.

Extensive and reproducible ion-current-based comparison of BALF proteomes in COPD vs healthy subjects

To comprehensively reveal the potential protein markers in BALF that are highly valuable for the diagnosis and mechanistic study of COPD, it is critical to achieve extensive, reproducible and well-controlled proteomic comparisons in COPD vs. healthy subjects. Given the expected high individual variability and the large protein concentration dynamic ranges, the ability to quantify with a relatively large number of biological replicates while achieving an inBdepth quantification in one experimental set is necessary.

A precipitation procedure providing quantitative protein recovery was implemented to normalize BALF protein concentrations and thus eliminating the effect of variable dilutions by the BAL procedure. To achieve high quantitative precision and accuracy, highly reproducible and efficient procedures for sample clean-up, digestion, chromatographic separation and MS detection were developed, optimized and evaluated. Under the optimized conditions, excellent precision and efficiency were achieved for each step, as demonstrated by the benchmark data. The high level of reproducibility allows for the reliable quantification of 20 subjects in one set. The used of a low-void-volume nano-LC/nanospray configuration provided significantly improved peak shapes, reduced tailing, and high run-to-run reproducibility(14, 15, 17); furthermore, the long, heated nano-column with a 7-hr gradient enabled efficient resolution of the BALF digests, permitting a relatively extensive analysis of the BALF proteome, especially for relatively lower abundance proteins(17). In this study, we found that using a 75-cm column with a 7-hr gradient resulted in ~2.5-fold more quantifiable proteins than with a 25-cm column and a 2.4-hr gradient. Furthermore, the extensive separation provides additional benefits for ion- current-based quantification, in that it significantly reduces isomeric interferences (e.g. peptide isomer or chemical noises), and thus markedly alleviating the problem of mismatching peptide peaks among runs.

Highly strict criteria for both identification and quantification were adopted to achieve a reliable discovery (*Methods*). Despite of the high complexity, wide protein concentration dynamic range and the stringent selection criteria, 423 unique protein groups were quantified with high confidence without immunodepletion or fractionation. Among these 76 protein groups were determined as altered proteins with low falseBdiscovery rate.

Physicochemical properties of human BALF proteins

As BALF is a largely under-characterized proteome, the study of the physicochemical properties of its components will provide valuable information. The comprehensive analysis in this study provided several interesting insights. Since the BALF samples were procured from the fluid lining the bronchoalveolar surface, the majority of the proteins found here are extracellular region and plasma membrane proteins. We found an average protein *pI* of 6.5 for the human BALF proteome, and the percentage of proteins with a *pI* value > 7.0 (72%) was noticeably higher than that in other human fluid proteomes such as those of plasma, urine and saliva (~55%B 65%) (28-32). Moreover, 290 proteins have MW > 60 kDa, accounting for 68.6% of all proteins. By comparison, the percentage of proteins with MW > 60 kDa is generally less than 50% in human plasma proteome (28-30), and is about 70% in human saliva and urinary proteomes (30, 31, 33-35). Additionally, the BALF, saliva

and urinary proteomes show very similar molecular weight distributions, which are different from that of plasma. This may be due to the renal clearance of plasma proteins > 60 kDa (36), a mechanism that is absent in the other fluid proteomes mentioned above.

Functional annotation and validation of altered proteins associated with pathological mechanisms in COPD

Functional annotation was performed for all the altered proteins. Selected groups of high interest are discussed in this section. Furthermore, with the emphasis of elevated proteins, which are more likely to serve as the candidates for circulating biomarkers, some key proteins were validated with Western blot analysis.

a) Proteins associated with lung inflammatory responses—Inflammatory response of lung to the inhalation of noxious particles such as cigarette smoking has been shown as a major contributory factor to the development of COPD (37). In this study, 13 altered proteins associated with the biological process of “inflammatory response” were observed. Among these, alpha-2-macroglobulin (A2M), fibrinogen (alpha, beta and gamma chain), galectinB3 and serum amyloid P (SAP) were elevated in COPD patients, and CD44, CD97, complement C1q, C2, C3, C4b and fibronectin were decreased (supplemental Table IV). Previous studies have demonstrated a lower surface expression of CD44 on alveolar macrophages (AM) in COPD (38), which may limit the ability of AM in COPD to clear apoptotic neutrophils due to the decreased hyaluronan-binding capacity by CD44, and thus result in perpetuation of inflammation (38, 39). Besides CD44, proteins C1q and surfactant protein A and D are also involved in the clearance of apoptotic cells (40). Consequently, the reduced C1q, surfactant protein A and D observed in COPD (supplemental Table IV) may also contribute to impaired AM clearance and sustained lung inflammation characteristic of this disease. Galectin-3, a beta-galactoside-binding lectin, is also involved in inflammation, as well as cell migration and cell-cell interactions (41, 42). Previous studies have showed increased expression of galectin-3 in small airway epithelial cells in COPD patients and in the sera from patients with lung cancer metastasis (42, 43). Inhibition of galectin-3 may serve as a potential therapeutic strategy for COPD and may help to inhibit progress to cancer and metastasis (44). We observed a significantly (2.4x) increased level of galectin-3 in COPD. This suggests galectin-3, could be used as a potential marker for COPD. The increase of galectin-3 was validated by western blot (Fig. 4).

The up-regulation of the mRNA of signal transducers and activators of transcription 3 (Stat3) has been previously described in lung tissues in COPD (45, 46). We found significant up-regulation of five downstream proteins, A2M, SAP, fibrinogen alpha chain (FGA), fibrinogen beta chain (FGB) and fibrinogen gamma chain (FGG) in BALF (Table II). Western blot analysis confirmed the significant increase of these proteins (Fig. 4) in BALF of COPD patients.

b) Key proteins associated with gluconeogenesis/glycolysis—Alterations in energy expenditure and macronutrient metabolism may play a role in weight loss in patients with COPD (47-49). Recently, *in vivo* evidence suggested that whole-body glucose production, clearance, and oxidation, and rate of glycolysis are elevated in COPD (47). We

observed that gluconeogenesis/glycolysis associated proteins such as alpha enolase, gamma enolase, fructose biphosphate aldolase C, fructose 1,6 biphosphatase 1 and malate dehydrogenase were elevated in BALF from COPD (supplemental Table III). BALF gamma enolase elevation in COPD has been reported previously (50). The extent to which this observed elevation of gluconeogenesis/glycolysis in the airway lumen accounts for the elevated whole- body glucose production, clearance and rate of glycolysis in COPD, and contribute to the weight loss in COPD needs to be determined.

c) Up-regulated alcohol metabolism in response to the oxidative stress in

COPD—The association of alcohol metabolism with COPD has not been reported, despite alcohol abuse has been known for centuries to impair lung defense against inflammation (51). Ingested alcohol diffuses from the bronchial circulation through the airway epithelium where it vaporizes into the conducting airways of the lung (52). As a result, the concentration of alcohol exhaled through the lung is routinely used to estimate blood alcohol concentration (BAC) due to the excellent linear relationship between the two (53). Interestingly, for decades, many investigators have observed significant underestimation of BAC thru measuring the exhaled alcohol concentrations in the COPD patients (54-56). However, the molecular mechanism of this phenomenon is still unknown. In this study, 12 proteins related to alcohol metabolic process were observed to be significantly upBregulated (Table II). Using Western blot, the significant up- regulation of several key enzymes, ADH1B, ALDH2, and ALDH3A1, were further validated in the BALF from COPD patients (Fig.4). ADH1B and ALDH2 are key enzymes in alcohol metabolism, oxidizing the majority of alcohol to acetaldehyde and then acetaldehyde to acetate, respectively (57), as illustrated in Fig. 5. We speculate that the highly expressed ADH1B and ALDH2 in BALF in COPD may play a key role in the underestimation of the BAC thru breath test in COPD patients.

We speculate the cause of elevated alcohol metabolism is the response to the oxidative stress, which has critical implications in the pathogenesis of COPD (58, 59). ALDH2 and ALDH3A1 were recognized as oxidative stress response proteins that protect against oxidative damage (60, 61). Besides transferring acetaldehyde to acetate, ALDH2 and ALDH3A1 also detoxify other aromatic and aliphatic aldehydes such as 4-hydroxy-2-nonenal (4-HNE) to 4-hydroxy-2-enoic acid (4-HNA), which are usually generated during oxidative stress as a result of lipid peroxidation (61, 62), as illustrated in Fig. 5. ALDH2 also plays a role in the biotransformation of nitroglycerin (GTN), by its reductase activity, to 1,2-glyceryl dinitrate (GDN) for the production of nitric oxide (NO), a prominent antioxidant (63). Besides ADH1B, ALDH2, and ALDH3A1, the upBregulation of six other proteins involved in oxidant reduction were discovered in the BALF from COPD patients (Table II). Among them, the increased malate dehydrogenase 1 (MDH1, producing oxaloacetate) and isocitrate dehydrogenase 1 (IDH1, producing 2- oxoglutarate) may result in more acidic compounds such as acetate, 4-HNA and oxaloacetate, to be produced in the BALF in COPD, which would result in acidification of exhaled breath condensate (EBC) in COPD patients (Fig. 5). Consistent with our observation, it has been demonstrated that the average EBC pH in COPD was reduced and this phenomenon was referred as “acidopnea” (64, 65), to which the detailed mechanisms have not been elucidated before this study. Here the observed

increase of alcohol metabolism and oxidation reduction in BALFs from COPD patients, afforded an explanation of this phenomenon.

CONCLUSIONS

We conducted an extensive profiling of the BALF proteomes in COPD patients vs. healthy controls with 20 subjects, using a well characterized but phenotypically similar group of patients with COPD and well matched healthy controls. As the BALF is a unique fluid proteome that is constituted by proteins permeated/exuded from tissues cells and the circulation, it hosts a wealth of potential biomarkers that are valuable for mechanism studies and diagnosis of COPD. This study represents several methodological advances in sample preparation, chromatographic separation and ion-current-based quantification to enable an extensive and reproducible quantification of BALF proteomes. In spite of the tremendous protein concentration dynamic ranges in BALF, we achieved the high-quality quantification of 423 unique proteins, among which 76 proteins displayed significantly altered expression. These altered proteins clearly reflect the known complex pathophysiology of COPD in many relevant biological processes, and the novel observation of up-regulated alcohol metabolism not only revealed a new mechanism for response to oxidative stress in COPD, but also revealed the mechanisms for bewildering, decades-long clinical observations of the underestimation of BAC in COPD patients. Finally, the ion current based quantitative strategy can be adapted to a wide range of clinical investigations.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by NIH grants U54HD071594, DA027528 and HL103411, industrial funds thru the Center of Protein Therapeutics Consortium, and an AHA award 12SDG9450036 to Jun Qu. This work was also supported by NIH grant HL082561 to Sanjay Sethi and University of Buffalo start up funds awarded to Manoj Mammen.

Abbreviations

AUC	area under curve
BALF	bronchoalveolar lavage fluid
COPD	chronic obstructive pulmonary disease
GO	gene ontology
IPA	ingenuity pathway analysis
LTQ Orbitrap	a linear ion trap combined with an orbitrap analyzer mass spectrometer

References

1. Chapman KR, Mannino DM, Soriano JB, Vermeire PA, Buist AS, Thun MJ, Connell C, Jemal A, Lee TA, Miravittles M, Aldington S, Beasley R. Epidemiology and costs of chronic obstructive pulmonary disease. *Eur Respir J*. 2006; 27:188–207. [PubMed: 16387952]

2. Sullivan SD, Ramsey SD, Lee TA. The economic burden of COPD. *Chest*. 2000; 117:5S–9S. [PubMed: 10673466]
3. Halpin DM, Miravittles M. Chronic obstructive pulmonary disease: the disease and its burden to society. *Proc Am Thorac Soc*. 2006; 3:619–623. [PubMed: 16963544]
4. Plymoth A, Lofdahl CG, Ekberg-Jansson A, Dahlback M, Lindberg H, Fehniger TE, Marko-Varga G. Human bronchoalveolar lavage: biofluid analysis with special emphasis on sample preparation. *Proteomics*. 2003; 3:962–972. [PubMed: 12833521]
5. Magi B, Bargagli E, Bini L, Rottoli P. Proteome analysis of bronchoalveolar lavage in lung diseases. *Proteomics*. 2006; 6:6354–6369. [PubMed: 17133372]
6. Casado B, Iadarola P, Luisetti M, Kussmann M. Proteomics-based diagnosis of chronic obstructive pulmonary disease: the hunt for new markers. *Expert Rev Proteomics*. 2008; 5:693–704. [PubMed: 18937559]
7. Wu J, Kobayashi M, Sousa EA, Liu W, Cai J, Goldman SJ, Dorner AJ, Projan SJ, Kavuru MS, Qiu Y, Thomassen MJ. Differential proteomic analysis of bronchoalveolar lavage fluid in asthmatics following segmental antigen challenge. *Mol Cell Proteomics*. 2005; 4:1251–1264. [PubMed: 15951573]
8. Plymoth A, Yang Z, Lofdahl CG, Ekberg-Jansson A, Dahlback M, Fehniger TE, Marko-Varga G, Hancock WS. Rapid proteome analysis of bronchoalveolar lavage samples of lifelong smokers and never-smokers by micro-scale liquid chromatography and mass spectrometry. *Clin Chem*. 2006; 52:671–679. [PubMed: 16497942]
9. Merkel D, Rist W, Seither P, Weith A, Lenter MC. Proteomic study of human bronchoalveolar lavage fluids from smokers with chronic obstructive pulmonary disease by combining surface-enhanced laser desorption/ionization-mass spectrometry profiling with mass spectrometric protein identification. *Proteomics*. 2005; 5:2972–2980. [PubMed: 16075419]
10. Liu T, Qian WJ, Mottaz HM, Gritsenko MA, Norbeck AD, Moore RJ, Purvine SO, Camp DG 2nd, Smith RD. Evaluation of multiprotein immunoaffinity subtraction for plasma proteomics and candidate biomarker discovery using mass spectrometry. *Mol Cell Proteomics*. 2006; 5:2167–2174. [PubMed: 16854842]
11. Tu C, Rudnick PA, Martinez MY, Cheek KL, Stein SE, Slebos RJ, Liebler DC. Depletion of abundant plasma proteins and limitations of plasma proteomics. *J Proteome Res*. 2010; 9:4982–4991. [PubMed: 20677825]
12. Vestbo J, Hurd SS, Agusti AG, Jones PW, Vogelmeier C, Anzueto A, Barnes PJ, Fabbri LM, Martinez FJ, Nishimura M, Stockley RA, Sin DD, Rodriguez-Roisin R. Global Strategy for the Diagnosis, Management and Prevention of Chronic Obstructive Pulmonary Disease, GOLD Executive Summary. *Am J Respir Crit Care Med*. 2012
13. Berenson CS, Wrona CT, Grove LJ, Maloney J, Garlipp MA, Wallace PK, Stewart CC, Sethi S. Impaired alveolar macrophage response to *Haemophilus* antigens in chronic obstructive lung disease. *Am J Respir Crit Care Med*. 2006; 174:31–40. [PubMed: 16574934]
14. Duan X, Young R, Straubinger RM, Page B, Cao J, Wang H, Yu H, Canty JM, Qu J. A straightforward and highly efficient precipitation/on-pellet digestion procedure coupled with a long gradient nano-LC separation and Orbitrap mass spectrometry for label-free expression profiling of the swine heart mitochondrial proteome. *J Proteome Res*. 2009; 8:2838–2850. [PubMed: 19290621]
15. Tu C, Li J, Young R, Page BJ, Engler F, Halfon MS, Canty JM Jr. Qu J. Combinatorial peptide ligand library treatment followed by a dual-enzyme, dual-activation approach on a nanoflow liquid chromatography/orbitrap/electron transfer dissociation system for comprehensive analysis of swine plasma proteome. *Anal Chem*. 2011; 83:4802–4813. [PubMed: 21491903]
16. Elias JE, Haas W, Faherty BK, Gygi SP. Comparative evaluation of mass spectrometry platforms used in large-scale proteomics investigations. *Nat Methods*. 2005; 2:667–675. [PubMed: 16118637]
17. Tu C, Li J, Bu Y, Hangauer D, Qu J. An ion-current-based, comprehensive and reproducible proteomic strategy for comparative characterization of the cellular responses to novel anti-cancer agents in a prostate cell model. *J Proteomics*. 2012; 77:187–201. [PubMed: 22982362]

18. Zeeberg BR, Qin H, Narasimhan S, Sunshine M, Cao H, Kane DW, Reimers M, Stephens RM, Bryant D, Burt SK, Elnekave E, Hari DM, Wynn TA, Cunningham-Rundles C, Stewart DM, Nelson D, Weinstein JN. High-Throughput GoMiner, an 'industrial-strength' integrative gene ontology tool for interpretation of multiple-microarray experiments, with application to studies of Common Variable Immune Deficiency (CVID). *BMC Bioinformatics*. 2005; 6:168. [PubMed: 15998470]
19. Petersen TN, Brunak S, von Heijne G, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods*. 2011; 8:785–786. [PubMed: 21959131]
20. Nilsson T, Mann M, Aebersold R, Yates JR 3rd, Bairoch A, Bergeron JJ. Mass spectrometry in high-throughput proteomics: ready for the big time. *Nat Methods*. 2010; 7:681–685. [PubMed: 20805795]
21. Oda Y, Huang K, Cross FR, Cowburn D, Chait BT. Accurate quantitation of protein expression and site-specific phosphorylation. *Proc Natl Acad Sci U S A*. 1999; 96:6591–6596. [PubMed: 10359756]
22. Ryu S, Gallis B, Goo YA, Shaffer SA, Radulovic D, Goodlett DR. Comparison of a label-free quantitative proteomic method based on peptide ion current area to the isotope coded affinity tag method. *Cancer Inform*. 2008; 6:243–255. [PubMed: 19259412]
23. Duan X, Abuqayyas L, Dai L, Balthasar JP, Qu J. High-throughput method development for sensitive, accurate, and reproducible quantification of therapeutic monoclonal antibodies in tissues using orthogonal array optimization and nano liquid chromatography/selected reaction monitoring mass spectrometry. *Anal Chem*. 2012; 84:4373–4382. [PubMed: 22519810]
24. Qu J, Lesse AJ, Brauer AL, Cao J, Gill SR, Murphy TF. Proteomic expression profiling of *Haemophilus influenzae* grown in pooled human sputum from adults with chronic obstructive pulmonary disease reveal antioxidant and stress responses. *BMC Microbiol*. 2010; 10:162. [PubMed: 20515494]
25. Wang H, Straubinger RM, Aletta JM, Cao J, Duan X, Yu H, Qu J. Accurate localization and relative quantification of arginine methylation using nanoflow liquid chromatography coupled to electron transfer dissociation and orbitrap mass spectrometry. *J Am Soc Mass Spectrom*. 2009; 20:507–519. [PubMed: 19110445]
26. Liu H, Sadygov RG, Yates JR 3rd. A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal Chem*. 2004; 76:4193–4201. [PubMed: 15253663]
27. Zybailov B, Coleman MK, Florens L, Washburn MP. Correlation of relative abundance ratios derived from peptide ion chromatograms and spectrum counting for quantitative proteomic analysis using stable isotope labeling. *Anal Chem*. 2005; 77:6218–6224. [PubMed: 16194081]
28. Jin WH, Dai J, Li SJ, Xia QC, Zou HF, Zeng R. Human plasma proteome analysis by multidimensional chromatography prefractionation and linear ion trap mass spectrometry identification. *J Proteome Res*. 2005; 4:613–619. [PubMed: 15822942]
29. Tu CJ, Dai J, Li SJ, Sheng QH, Deng WJ, Xia QC, Zeng R. High-sensitivity analysis of human plasma proteome by immobilized isoelectric focusing fractionation coupled to mass spectrometry identification. *J Proteome Res*. 2005; 4:1265–1273. [PubMed: 16083276]
30. Yan W, Apweiler R, Balgley BM, Boontheung P, Bundy JL, Cargile BJ, Cole S, Fang X, Gonzalez-Begne M, Griffin TJ, Hagen F, Hu S, Wolinsky LE, Lee CS, Malamud D, Melvin JE, Menon R, Mueller M, Qiao R, Rhodus NL, Sevinsky JR, States D, Stephenson JL, Than S, Yates JR, Yu W, Xie H, Xie Y, Omenn GS, Loo JA, Wong DT. Systematic comparison of the human saliva and plasma proteomes. *Proteomics Clin Appl*. 2009; 3:116–134. [PubMed: 19898684]
31. Sun W, Li F, Wu S, Wang X, Zheng D, Wang J, Gao Y. Human urine proteome analysis by three separation approaches. *Proteomics*. 2005; 5:4994–5001. [PubMed: 16281181]
32. Denny P, Hagen FK, Hardt M, Liao L, Yan W, Arellanno M, Bassilian S, Bedi GS, Boontheung P, Cociorva D, Delahunty CM, Denny T, Dunsmore J, Faull KF, Gilligan J, Gonzalez-Begne M, Halgand F, Hall SC, Han X, Henson B, Hewel J, Hu S, Jeffrey S, Jiang J, Loo JA, Ogorzalek Loo RR, Malamud D, Melvin JE, Miroshnichenko O, Navazesh M, Niles R, Park SK, Prakobphol A, Ramachandran P, Richert M, Robinson S, Sondej M, Souda P, Sullivan MA, Takashima J, Than S, Wang J, Whitelegge JP, Witkowska HE, Wolinsky L, Xie Y, Xu T, Yu W, Ytterberg J, Wong DT, Yates JR 3rd, Fisher SJ. The proteomes of human parotid and submandibular/sublingual gland

- salivas collected as the ductal secretions. *J Proteome Res.* 2008; 7:1994–2006. [PubMed: 18361515]
33. Jia L, Zhang L, Shao C, Song E, Sun W, Li M, Gao Y. An attempt to understand kidney's protein handling function by comparing plasma and urine proteomes. *PLoS One.* 2009; 4:e5146. [PubMed: 19381340]
 34. Hu S, Li Y, Wang J, Xie Y, Tjon K, Wolinsky L, Loo RR, Loo JA, Wong DT. Human saliva proteome and transcriptome. *J Dent Res.* 2006; 85:1129–1133. [PubMed: 17122167]
 35. Adachi J, Kumar C, Zhang Y, Olsen JV, Mann M. The human urinary proteome contains more than 1500 proteins, including a large proportion of membrane proteins. *Genome Biol.* 2006; 7:R80. [PubMed: 16948836]
 36. Schenk S, Schoenhals GJ, de Souza G, Mann M. A high confidence, manually validated human blood plasma protein reference set. *BMC Med Genomics.* 2008; 1:41. [PubMed: 18793429]
 37. Pauwels RA, Buist AS, Calverley PM, Jenkins CR, Hurd SS. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease. NHLBI/WHO Global Initiative for Chronic Obstructive Lung Disease (GOLD) Workshop summary. *Am J Respir Crit Care Med.* 2001; 163:1256–1276. [PubMed: 11316667]
 38. Pons AR, Noguera A, Blanquer D, Sauleda J, Pons J, Agusti AG. Phenotypic characterisation of alveolar macrophages and peripheral blood monocytes in COPD. *Eur Respir J.* 2005; 25:647–652. [PubMed: 15802338]
 39. Teder P, Vandivier RW, Jiang D, Liang J, Cohn L, Pure E, Henson PM, Noble PW. Resolution of lung inflammation by CD44. *Science.* 2002; 296:155–158. [PubMed: 11935029]
 40. Vandivier RW, Ogden CA, Fadok VA, Hoffmann PR, Brown KK, Botto M, Walport MJ, Fisher JH, Henson PM, Greene KE. Role of surfactant proteins A, D, and C1q in the clearance of apoptotic cells in vivo and in vitro: calreticulin and CD91 as a common collectin receptor complex. *J Immunol.* 2002; 169:3978–3986. [PubMed: 12244199]
 41. Saravanan C, Liu FT, Gipson IK, Panjwani N. Galectin-3 promotes lamellipodia formation in epithelial cells by interacting with complex N-glycans on alpha3beta1 integrin. *J Cell Sci.* 2009; 122:3684–3693. [PubMed: 19755493]
 42. Yao H, Rahman I. Current concepts on the role of inflammation in COPD and lung cancer. *Curr Opin Pharmacol.* 2009; 9:375–383. [PubMed: 19615942]
 43. Iurisci I, Tinari N, Natoli C, Angelucci D, Cianchetti E, Iacobelli S. Concentrations of galectin-3 in the sera of normal controls and cancer patients. *Clin Cancer Res.* 2000; 6:1389–1393. [PubMed: 10778968]
 44. Thijssen VL, Poirier F, Baum LG, Griffioen AW. Galectins in the tumor endothelium: opportunities for combined cancer therapy. *Blood.* 2007; 110:2819–2827. [PubMed: 17591944]
 45. Qu P, Roberts J, Li Y, Albrecht M, Cummings OW, Eble JN, Du H, Yan C. Stat3 downstream genes serve as biomarkers in human lung carcinomas and chronic obstructive pulmonary disease. *Lung Cancer.* 2009; 63:341–347. [PubMed: 18614255]
 46. Grubek-Jaworska H, Paplinska M, Hermanowicz-Salamon J, Bialek-Gosk K, Dabrowska M, Grabczak E, Domagala-Kulawik J, Stepień J, Chazan R. IL-6 and IL-13 in Induced Sputum of COPD and Asthma Patients: Correlation with Respiratory Tests. *Respiration.* 2012
 47. Kao CC, Hsu JW, Bandi V, Hanania NA, Kheradmand F, Jahoor F. Glucose and pyruvate metabolism in severe chronic obstructive pulmonary disease. *J Appl Physiol.* 2012; 112:42–47. [PubMed: 22016370]
 48. Creutzberg EC, Schols AM, Bothmer-Quaedvlieg FC, Wouters EF. Prevalence of an elevated resting energy expenditure in patients with chronic obstructive pulmonary disease in relation to body composition and lung function. *Eur J Clin Nutr.* 1998; 52:396–401. [PubMed: 9683390]
 49. Engelen MP, Deutz NE, Wouters EF, Schols AM. Enhanced levels of whole-body protein turnover in patients with chronic obstructive pulmonary disease. *Am J Respir Crit Care Med.* 2000; 162:1488–1492. [PubMed: 11029366]
 50. Karnak D, Beder S, Kayacan O, Ibis E, Oflaz G. Neuron-specific enolase and lung cancer. *Am J Clin Oncol.* 2005; 28:586–590. [PubMed: 16317269]
 51. Sisson JH. Alcohol and airways function in health and disease. *Alcohol.* 2007; 41:293–307. [PubMed: 17764883]

52. George SC, Hlastala MP, Souders JE, Babb AL. Gas exchange in the airways. *J Aerosol Med.* 1996; 9:25–33. [PubMed: 10172721]
53. Hlastala MP. The alcohol breath test--a review. *J Appl Physiol.* 1998; 84:401–408. [PubMed: 9475843]
54. Wilson A, Sitar DS, Molloy WD, McCarthy D. Effect of age and chronic obstructive pulmonary disease on the Breathalyzer estimation of blood alcohol level. *Alcohol Clin Exp Res.* 1987; 11:440–443. [PubMed: 3314557]
55. Honeybourne D, Moore AJ, Butterfield AK, Azzan L. A study to investigate the ability of subjects with chronic lung diseases to provide evidential breath samples using the Lion Intoxilyzer 6000 UK breath alcohol testing device. *Respir Med.* 2000; 94:684–688. [PubMed: 10926340]
56. Russell JC, Jones RL. Breath ethyl alcohol concentration and analysis in the presence of chronic obstructive pulmonary disease. *Clin Biochem.* 1983; 16:182–187. [PubMed: 6851082]
57. Kayaalti Z, Soylemezoglu T. Distribution of ADH1B, ALDH2, CYP2E1 *6, and CYP2E1 *7B genotypes in Turkish population. *Alcohol.* 2010; 44:415–423. [PubMed: 20598484]
58. Boots AW, Haenen GR, Bast A. Oxidant metabolism in chronic obstructive pulmonary disease. *Eur Respir J Suppl.* 2003; 46:14s–27s. [PubMed: 14621103]
59. Kinnula VL. Focus on antioxidant enzymes and antioxidant strategies in smoking related airway diseases. *Thorax.* 2005; 60:693–700. [PubMed: 16061713]
60. Pappa A, Brown D, Koutalos Y, DeGregori J, White C, Vasiliou V. Human aldehyde dehydrogenase 3A1 inhibits proliferation and promotes survival of human corneal epithelial cells. *J Biol Chem.* 2005; 280:27998–28006. [PubMed: 15905174]
61. Pappa A, Chen C, Koutalos Y, Townsend AJ, Vasiliou V. Aldh3a1 protects human corneal epithelial cells from ultraviolet- and 4-hydroxy-2-nonenal-induced oxidative damage. *Free Radic Biol Med.* 2003; 34:1178–1189. [PubMed: 12706498]
62. Vasiliou V, Pappa A, Petersen DR. Role of aldehyde dehydrogenases in endogenous and xenobiotic metabolism. *Chem Biol Interact.* 2000; 129:1–19. [PubMed: 11154732]
63. Chen Z, Foster MW, Zhang J, Mao L, Rockman HA, Kawamoto T, Kitagawa K, Nakayama KI, Hess DT, Stamler JS. An essential role for mitochondrial aldehyde dehydrogenase in nitroglycerin bioactivation. *Proc Natl Acad Sci U S A.* 2005; 102:12159–12164. [PubMed: 16103363]
64. Borrill Z, Starkey C, Vestbo J, Singh D. Reproducibility of exhaled breath condensate pH in chronic obstructive pulmonary disease. *Eur Respir J.* 2005; 25:269–274. [PubMed: 15684290]
65. Effros RM, Casaburi R, Su J, Dunning M, Torday J, Biller J, Shaker R. The effects of volatile salivary acids and bases on exhaled breath condensate pH. *Am J Respir Crit Care Med.* 2006; 173:386–392. [PubMed: 16284109]

Fig.1.

The scheme of the overall proteomics strategy for the multiplexed comparative analysis of BALFs derived from COPD patients vs. healthy controls ($n=10$ per group). A reproducible, sensitive and extensive ion-current-based quantification procedure that is capable of analyzing a relatively high number of clinical replicates, was developed and optimized. False-positive discovery was evaluated with profiling 20 healthy subjects following a previously described strategy(17).

Fig. 2.

Evaluation of the analytical accuracy and precision for quantification of BALF proteome by performing 20 analyses of a pooled BALF sample. The 20 runs were randomly split into two groups, and the density plots of peptide and protein ratios between the two groups (true value was 0.0 on log scale) were calculated by the ion-current-based method;

Fig.3. Biological process classification of the differentially expressed proteins based upon Gene Ontology annotation. The arrows↑ and ↓ represent the elevated and decreased in COPD, respectively.

Fig. 4.

Western blot analysis of altered proteins in BALFs from five ex-smokers with COPD and five non-smoking healthy controls. Experimental molecular masses of the major band are shown. Human serum albumin, an abundant protein that exhibits constant relative levels in BALF total proteins, was used as a loading control. A) Western-blot images; B) Analysis of the intensity of bands on the Western-blot image. ADH1B = alcohol dehydrogenase 1B, ALDH2 = Aldehyde dehydrogenase 2, ALDH3A1 = Aldehyde dehydrogenase 3 member A1.

Fig. 5.

Illustration of the metabolic pathways of alcohol metabolism and oxidation reduction that may account for the underestimated BAC and acidopnea in COPD. The enzymes of alcohol dehydrogenase 1B (ADH1B), aldehyde dehydrogenase 2 (ALDH2) and 3A1 (ALDH3A1), malate dehydrogenase 1 (MDH1) and isocitrate dehydrogenase 1 (IDH1) were increased in BALF from patients with COPD (as shown in red), causing a series of metabolic cascades shown here. NOx, mono-nitrogen oxides; GTN, nitroglycerin; ROS, reactive oxygen species.

Table I

Clinical parameters of non-smoking healthy subjects and ex-smokers with stable COPD.

Variable	Control subjects	COPD subjects (GOLD stage 2)	<i>p</i> -value
Age (years)	63.4 ± 11.73	67.8 ± 8.5	NS
Sex (M/F)	6/4	7/3	NS
Race			NS
Caucasian	8	10	
African-American	2	0	
BMI (kg/m ²)	28.45±4.18	32.00±9.71	NS
Smoking Pack years	NA	56.6 ± 17.2	P<0.001
Years patient quit smoking	NA	12.9 ± 4.42	
FEV1 (% predicted)	96.29 ± 14.83	65.86 ± 8.1	P<0.001

BMI: body mass index. FEV1: forced expiratory volume in 1 second. M: male, F: female.

NS: Not significant.

The significantly elevated proteins in BALF of COPD patients (n=10) versus healthy controls (n=10) by the well-controlled ion-current-based approach.

Table II

Accession Number	Gene Name	Protein name	No. quantified peptides	Ratio (COPD / Normal)	Standard Deviation	p-value	GO Annotation: Cellular Location
P60709	ACTB	Actin_cytoplasmic 1	9	1.56	0.2	< 0.01	Cytoplasm
P07108	DBI	Acyl_CoA_binding protein	3	1.75	1.92	< 0.01	Cytoplasm
P00325	ADH1B	Alcohol dehydrogenase 1B	2	12.86	13.23	< 0.01	Cytoplasm
P30838	ALDH3A1	Aldehyde dehydrogenase family 3 member A1	9	3.33	0.89	< 0.01	Cytoplasm
P05091	ALDH2	Aldehyde dehydrogenase_mitochondrial	12	2.11	1.4	0.04	Cytoplasm
P42330	AKR1C3	Aldo_keto reductase family 1 member C3	2	4.11	1.55	0.02	Cytoplasm
P01023	A2M	Alpha_2_macrolobulin	41	2.31	0.21	< 0.01	Extracellular
P06733	ENO1	Alpha_enolase	9	1.61	0.5	< 0.01	Plasma membrane
P12429	ANXA3	Annexin A3	7	3.34	0.46	< 0.01	Plasma membrane
P08758	ANXA5	Annexin A5	7	2.1	0.63	< 0.01	Cytoplasm
P02647	APOA1	Apolipoprotein A_I	18	2.13	0.28	< 0.01	Extracellular; Plasma membrane
P04114	APOB	Apolipoprotein B_100	22	3.56	0.63	< 0.01	Extracellular; Plasma membrane
P02656	APOC3	Apolipoprotein C_III	2	4.24	3.13	< 0.01	Extracellular
P53004	BLVRA	Biliverdin reductase A	2	2.66	3.48	0.01	Cytoplasm
Q6NUK1	SLC25A24	Calcium_binding mitochondrial carrier protein SCaMC_1	2	3.33	6.76	< 0.01	Cytoplasm
P07339	CTSD	Cathepsin D	10	2.57	0.85	0.05	Extracellular
Q9UBR2	CTSZ	Cathepsin Z	4	7.92	6.9	< 0.01	Extracellular
P16070	CD44	CD44 antigen	5	0.56	0.22	< 0.01	Plasma membrane
P48960	CD97	CD97 antigen	2	0.65	0.67	0.03	Extracellular; Plasma membrane
P23528	CFL1	Cofilin_1	5	1.56	0.48	< 0.01	Cytoplasm
P02747	CIQC	Complement C1q subcomponent subunit C	3	0.32	0.25	< 0.01	Extracellular
P06681	C2	Complement C2	15	0.65	0.14	< 0.01	Extracellular
P01024	C3	Complement C3	74	0.63	0.05	< 0.01	Extracellular
P0C0L5	C4B	Complement C4_B	47	0.65	0.08	< 0.01	Extracellular
P27487	DPP4	Dipeptidyl peptidase 4	2	0.43	0.38	< 0.01	Extracellular; Plasma membrane
P02671	FGA	Fibrinogen alpha chain	16	2.78	0.52	< 0.01	Extracellular; Plasma membrane
P02675	FBG	Fibrinogen beta chain	21	3.17	0.4	< 0.01	Extracellular; Plasma membrane

Accession Number	Gene Name	Protein name	No. quantified peptides	Ratio (COPD / Normal)	Standard Deviation	p-value	GO Annotation: Cellular Location
P02679	FGG	Fibrinogen gamma chain	16	2.74	0.91	< 0.01	Extracellular; Plasma membrane
P02751	FN1	Fibronectin	41	0.39	0.07	< 0.01	Extracellular
P21333	FLNA	Filamin_A	25	7.12	0.47	< 0.01	Extracellular; Plasma membrane
Q13642	FHL1	Four and a half LIM domains protein 1	2	5.53	6.49	< 0.01	Plasma membrane
P09467	FBP1	Fructose_1_6_bisphosphatase 1	8	1.6	0.99	0.02	Cytoplasm
P09972	ALDOC	Fructose_bisphosphate aldolase C	3	1.91	1.31	0.05	Cytoplasm
P17931	LGALS3	Galectin_3	3	2.52	2.22	< 0.01	Extracellular; Plasma membrane
Q08380	LGALS3BP	Galectin_3_binding protein	10	0.57	0.1	< 0.01	Extracellular
P09104	ENO2	Gamma_enolase	2	1.7	1.36	< 0.01	Plasma membrane
P20142	PGC	Gastricsin	2	0.19	0.16	< 0.01	Extracellular
P46439	GSTM5	Glutathione S_transferase Mu 5	2	11.26	14.64	< 0.01	Cytoplasm
P00738	HP	Haptoglobin	23	3.22	0.26	< 0.01	Extracellular
P00739	HPR	Haptoglobin_related protein	2	0.63	0.34	< 0.01	Extracellular
P01859	IGHG2	Ig gamma_2 chain C region	3	0.58	0.22	< 0.01	Extracellular
P06331	HV209	Ig heavy chain V_II region ARH_77	2	0.64	0.39	< 0.01	Extracellular
P06314	KV404	Ig kappa chain V_IV region B17	2	0.6	0.3	< 0.01	Extracellular
P04220	MUCB	Ig mu heavy chain disease protein	7	4.36	0.52	< 0.01	Membrane fraction
P05362	ICAM1	Intercellular adhesion molecule 1	12	0.6	0.14	< 0.01	Extracellular; Plasma membrane
075874	IDH1	Isocitrate dehydrogenase [NADP] cytoplasmic	4	1.88	1.03	0.04	Cytoplasm
P13645	KRT10	Keratin_type I cytoskeletal 10	12	2.47	0.46	< 0.01	Cytoskeleton
P35527	KRT9	Keratin_type I cytoskeletal 9	4	1.71	1.13	0.04	Cytoplasm
P04264	KRT1	Keratin_type II cytoskeletal 1	12	1.65	0.46	< 0.01	Plasma membrane; Cytoplasm
Q5XKE5	KRT79	Keratin_type II cytoskeletal 79	2	3.43	2.88	< 0.01	Cytoskeleton
P51884	LUM	Lumican	5	1.93	1.01	< 0.01	Extracellular
Q9UEW3	MARCO	Macrophage receptor MARCO	3	0.36	0.22	< 0.01	Plasma membrane
P40925	MDH1	Malate dehydrogenase_ cytoplasmic	7	1.68	0.69	0.02	Cytoplasm
Q13421	MSLN	Mesothelin	7	1.88	1.43	0.03	Extracellular; Plasma membrane
P15941	MUC1	Mucin_1	3	0.64	0.44	< 0.01	Extracellular; Plasma membrane
P35749	MYH11	Myosin_11	2	4.01	4.03	< 0.01	Cytoplasm
096009	NAPSA	Napsin_A	4	0.39	0.27	< 0.01	Extracellular

Accession Number	Gene Name	Protein name	No. quantified peptides	Ratio (COPD / Normal)	Standard Deviation	p-value	GO Annotation: Cellular Location
P62937	PPIA	Peptidyl prolyl cis-trans isomerase A	3	1.92	1.34	< 0.01	Extracellular
P30041	PRDX6	Peroxiredoxin_6	3	1.79	1.09	0.01	Cytoplasm
P30086	PEBP1	Phosphatidylethanolamine-binding protein 1	4	1.9	0.95	< 0.01	Extracellular
Q15149	PLEC	Plectin	6	2.31	0.34	< 0.01	Plasma membrane
Q7Z7M9	GALNT5	Polypeptide N-acetylgalactosaminyltransferase 5	2	0.64	0.57	0.01	Cytoplasm
P41222	PTGDS	Prostaglandin-H2 Disomerase	2	2.03	1.46	< 0.01	Extracellular
Q81WL2	SFTPA1	Pulmonary surfactant-associated protein A1	13	0.28	0.03	< 0.01	Extracellular
P35247	SFTPD	Pulmonary surfactant-associated protein D	8	0.41	0.11	< 0.01	Extracellular
Q8NFJ5	GPRC5A	Retinoic acid-induced protein 3	3	0.64	0.44	0.02	Plasma membrane
P07998	RNASE1	Ribonuclease pancreatic	5	0.63	0.19	< 0.01	Extracellular
P02743	APCS	Serum amyloid P-component	3	8.23	0.95	< 0.01	Extracellular
075368	SH3BGRL	SH3 domainbinding glutamic acid-rich-like protein	3	1.89	1.35	< 0.01	Cytoplasm
095436	SLC34A2	Sodiumdependent phosphate transport protein 2B	3	0.61	0.14	< 0.01	Plasma membrane
000560	SDCBP	Syntenin_1	4	0.66	0.39	0.02	Plasma membrane
P37837	TALD01	Transaldolase	3	2.09	0.51	0.03	Cytoplasm
Q15582	TGFB1	Transforming growth factor_beta-induced protein ig_h3	7	0.58	0.18	< 0.01	Extracellular
P37802	TAGLN2	Transgelin_2	7	1.74	0.97	< 0.01	Plasma membrane
P61088	UBE2N	Ubiquitin-conjugating enzyme E2 N	2	1.65	0.85	< 0.01	Cytoplasm
P08670	VIM	Vimentin	26	3.12	0.55	0.03	Plasma membrane; Cytoplasm