Proteomic and Computational Analysis of Bronchoalveolar Proteins during the Course of the Acute Respiratory Distress Syndrome

Dong W. Chang^{1,2*}, Shinichi Hayashi^{1,2*}, Sina A. Gharib^{2,5,6*}, Tomas Vaisar³, S. Trevor King⁴, Mitsuhiro Tsuchiya⁴, John T. Ruzinski², David R. Park², Gustavo Matute-Bello^{1,2}, Mark M. Wurfel², Roger Bumgarner⁴, Jay W. Heinecke³, and Thomas R. Martin^{1,2}

¹Medical Research Service of the VA Puget Sound Healthcare System, Seattle, Washington; Divisions of ²Pulmonary and Critical Care Medicine and ³Endocrinology and Metabolism, Department of Medicine; and ⁴Department of Microbiology; ⁵Center for Lung Biology; and ⁶Fred Hutchinson Cancer Research Institute, University of Washington, Seattle, Washington

Rationale: Acute lung injury causes complex changes in protein expression in the lungs. Whereas most prior studies focused on single proteins, newer methods allowing the simultaneous study of many proteins could lead to a better understanding of pathogenesis and new targets for treatment.

Objectives: The purpose of this study was to examine the changes in protein expression in the bronchoalveolar lavage fluid (BALF) of patients during the course of the acute respiratory distress syndrome (ARDS).

Methods: Using two-dimensional difference gel electrophoresis (DIGE), the expression of proteins in the BALF from patients on Days 1 (n = 7), 3 (n = 8), and 7 (n = 5) of ARDS were compared with findings in normal volunteers (n = 9). The patterns of protein expression were analyzed using principal component analysis (PCA). Biological processes that were enriched in the BALF proteins of patients with ARDS were identified using Gene Ontology (GO) analysis. Protein networks that model the protein interactions in the BALF were generated using Ingenuity Pathway Analysis.

Measurements and Main Results: An average of 991 protein spots were detected using DIGE. Of these, 80 protein spots, representing 37 unique proteins in all of the fluids, were identified using mass spectrometry. PCA confirmed important differences between the proteins in the ARDS and normal samples. GO analysis showed that these differences are due to the enrichment of proteins involved in inflammation, infection, and injury. The protein network analysis showed that the protein interactions in ARDS are complex and redundant, and revealed unexpected central components in the protein networks.

Conclusions: Proteomics and protein network analysis reveals the complex nature of lung protein interactions in ARDS. The results provide new insights about protein networks in injured lungs, and identify novel mediators that are likely to be involved in the pathogenesis and progression of acute lung injury.

Keywords: acute respiratory distress syndrome; acute lung injury; proteomic analysis; bronchoalveolar lavage; 2D gel electrophoresis

The study of single pathways and one-dimensional protein-protein interactions in complex diseases, such as acute lung injury (ALI) and the acute respiratory distress syndrome (ARDS), places major limitations on understanding the pathophysiology

AT A GLANCE COMMENTARY

Scientific Knowledge on the Subject

Single inflammatory pathways do not completely account for the onset, perpetuation, or resolution of lung injury. New approaches accounting for the complexity of the inflammatory response and changes that occur during the course of acute lung injury are needed.

What This Study Adds to the Field

Proteomic and computational network analysis of proteins in injured lungs shows new relationships among proteins and identifies new groups of mediators that could be targets for novel treatments.

of these complex diseases. Previous studies suggest that numerous biological processes, including inflammation, apoptosis, and thrombosis, are involved in the pathogenesis of ALI and ARDS (1). Traditional research methods that explore single biological pathways cannot capture the complex interactions between these processes. In contrast, systems-based methodologies, such as proteomics, analyze global biological changes and provide an opportunity to examine the complexity that is inherent in human diseases, such as ALI and ARDS.

Proteomics methods have been applied to the study of lung injury by several groups of investigators. Bowler and coworkers used an electrophoresis-based proteomics method to show that patients with ARDS have differences in both the expression and post-translational modification of proteins in distal lung fluid as compared with healthy control subjects (2). Schnapp and colleagues used a shotgun proteomics approach to identify insulin-like growth factor-binding protein-3 (IGFBP-3) as a novel mediator of apoptotic pathways in acute lung injury (3). De Torre and coworkers identified markers of lung inflammation, such as S100 A8 and A9 proteins, in the BALF of subjects challenged with endobronchial endotoxin and patients with ARDS using SELDI-TOF and electrophoresis-based proteomics methods (4). These studies applied proteomics techniques to study proteins in ARDS lavage fluids at a single time, but they did not address the complex and dynamic changes that occur during the course of lung injury.

Thus, the purpose of this study was to use proteomic analysis to profile the changes in protein expression in the lungs at the onset and during the course of acute lung injury to examine the protein pathways that have important roles in its pathogenesis. We used a quantitative proteomics approach to profile proteins in the bronchoalveolar lavage (BAL) fluid (BALF) of patients with ARDS at Days 1, 3, and 7 after the onset of illness and compared the results with protein profiles in the BALF of healthy control

(Received in original form December 26, 2007; accepted in final form July 21, 2008)

*These authors contributed equally to this study.

Funding: NIH/NHLBI HL073996, NIH/NHLBI HL090298.

Correspondence and requests for reprints should be addressed to Thomas R. Martin, M.D., Pulmonary Research Laboratories, VA Puget Sound Health Care System, 1660 S. Columbian Way, 151L Seattle, WA 98108. E-mail: trmartin@u.washington.edu

This article has an online supplement, which is accessible from this issue's table of contents at www.atsjournals.org.

Am J Respir Crit Care Med Vol 178, pp 701-709, 2008

Originally Published in Press as DOI: 10.1164/rccm.200712-1895OC on July 24, 2008

Internet address: www.atsjournals.org

subjects. We then applied advanced methods in computational analysis to map complex protein interactions in the lung fluids and study how these interactions changed during the course of ARDS. This approach to protein network analysis identified novel mediators of acute lung injury, and showed that protein pathways were redundant and involved in multiple biological processes. These characteristics of the protein interactions in the lungs of patients with ARDS have important implications for the development of new molecular-based therapies. Some of the results of this study have been previously reported in the form of an abstract (5).

METHODS

Patient Population

Patients with ARDS, as defined by the American-European Consensus Conference, were enrolled at Harborview Medical Center (Seattle, WA), a tertiary, university-based hospital (6). The patients underwent fiberoptic bronchoscopy and BAL in either the right middle lobe or lingula on Days 1, 3, and 7, as described (7). The control BAL fluid samples were obtained from healthy, nonsmoking volunteers between the ages of 18 and 50. The experimental protocol was approved by the Institutional Review Board of the University of Washington. Informed consent was obtained from the patient or a legal representative.

Sample Preparation

The BALF was processed by centrifugation as described (7). For proteomic analysis, the BALF samples were concentrated by centrifugation to equivalent starting volumes (~500 μ l) using a 5-kD molecular weight filter (Amicon Ultra-15; Millipore, Billerica, MA). The samples were then spun through a 0.22- μ m filter to remove mucus and other insoluble products. The total protein concentrations were determined using the BCA Protein Assay (Pierce, Rockford, IL).

To improve the detection of low-abundance proteins in the proteomic analysis, all samples were depleted of six highly abundant serum proteins (albumin, transferrin, haptoglobin, antitrypsin, IgG, IgA) using a monoclonal IgG immunoaffinity HPLC column (Multiple Affinity Removal System; Agilent Technologies, Wilmington, DE). The BAL fluids were passed over the depletion column, which absorbed the high abundance proteins. The buffer of the column flow-through fraction was exchanged to 7 M urea, 2 M thiourea, 4% CHAPS, 10 mM Tris, pH 8.5 and concentrated to approximately 100 μ l using a 5-kD centrifugation filter (Amicon Ultra-15; Millipore). The final protein concentration was measured using the 2D-Quant Assay, which allows protein measurements in urea-based solutions (Amersham Biosciences, Piscataway, NJ). The specificity, reproducibility, and improvements in protein spot detection during proteomic analysis of BALF using this approach have been reported (8, 9).

Two-Dimensional Difference Gel Electrophoresis

Proteomic analysis of the BALF samples was performed using two-dimensional difference gel electrophoresis (DIGE; GE Healthcare, Piscataway, NJ). In DIGE, fluorescent dyes with unique absorbance and emission spectra (Cy3, Cy5) were used to differentially label a reference standard (Cy3) and each experimental sample (Cy5). The reference standard was generated by pooling small aliquots of each of the experimental and normal samples. Thus, the reference standard theoretically contained every protein spot that was detected in the experiment. The standard and experimental samples were labeled separately with Cy3 or Cy5 and run in a single two-dimensional gel in which proteins were separated first by their isoelectric points (first dimension) and then by molecular weights (second dimension). The gel was first scanned at a wavelength that was optimized to identify the proteins labeled with Cy3, then rescanned at a second wavelength that was optimized to identify proteins labeled with Cy5. The acquired images were then superimposed and the abundances of the protein spots in the experimental sample was expressed as a ratio between the spot of interest (labeled with Cy 5) and the corresponding spot in the reference standard (labeled with Cy3). Because the abundance of each protein spot was expressed as a ratio to a common reference standard, the influence of experimental factors that can confound the accurate measurement of protein spot abundance, such as differences in gel quality, was minimized (10). The sensitivity and reproducibility of the DIGE method for

detecting changes in protein abundance in complex biological samples have been reported (10).

For each subject, 75 μ g of reference standard and 75 μ g of BAL protein were labeled with Cy3 and Cy5, respectively. These samples were then applied onto a single rehydrated immobilized pH gradient (IPG) strip pH 4–7, 24 cm (GE Healthcare). Two-dimensional electrophoresis was performed as described in the online supplement, separating proteins by isoelectric point in the first dimension and molecular weight in the second dimension. To identify individual protein spots, the gels were scanned using the Typhoon 9400 Series Variable Imager (GE Healthcare) with excitation wavelengths of 532 nm for Cy3 and 580 nm for Cy5. This procedure was performed using the samples from each of the 20 patients with ARDS (Day 1, $n = 7$; Day 3, $n = 8$; Day 7, $n = 5$) and 9 control subjects, yielding a total of 29 separate gels.

Protein Spot Analysis

The protein spots were analyzed using the Decyder software program version 5.01 (GE Healthcare). The difference in-gel analysis (DIA) module of the Decyder program was used to measure the abundance of each of the protein spots in the experimental samples as a $\log_2$ ratio between the spot of interest in the experimental sample (Cy 5 wavelength) and the corresponding spot in the reference standard (Cy3 wavelength). Next, the biological variance analysis (BVA) module of the Decyder program was used to compare the expression of protein spots between the experimental groups. Spots that were detected in more than 50% of the gels were included in the analysis.

Identification of Protein Spots

For protein identification, a “picking gel” containing 500 μ g of protein from the pooled standard was generated using the same electrophoresis protocol as described above, and stained after the second dimension electrophoresis with SyproRuby (PerkinElmer, Waltham, MA). The spots on this picking gel were matched with the spots of interest from the analytical (CyDye) gels using the Decyder program. The protein spots of interest were excised from the picking gel using an automated spot picker (Ettan DIGE Spot-picker; Amersham) with a 2-mm picking head into a 96-well ZipPlateC18 (Millipore) for in-gel digestion and matrix-assisted laser desorption (MALDI) mass spectrometric analysis. Each gel spot containing a protein of interest was washed, dehydrated, and digested with trypsin in 50 μ l of 50 mM NH_4HCO_3 and 5 mM CaCl_2 . The peptides were eluted from the ZipTip of the plate in 80% acetonitrile containing 0.1% trifluoroacetic acid (TFA), spotted on the MALDI target (0.5 μ l), dried, and overlaid with 0.5 μ l of α -cyano-4-hydroxycinnamic acid (HCCA) matrix (5 mg/ml in 70% acetonitrile, 0.1% TFA) for MALDI-TOF/TOF mass spectrometry. The MALDI-TOF MS and MS/MS spectra were acquired on the ABI 4700 Proteomic analyzer MALDI tandem time-of-flight mass spectrometer with air as the collision gas (Applied Biosystems, Foster City, CA). The MS and MS/MS spectra were searched in combination against the human SwissProt protein database using the MASCOT search engine (v2.0). Only protein identifications with a confidence interval of greater than 95% based on MASCOT MOWSE were included in the analysis.

Cytokine Assays

The concentrations of IL-1 β and IL-6 were measured in the BALF from patients with ARDS using a cytometric bead-based immunoassay system (Luminex, Austin, TX) according to the manufacturer's protocols (R&D Systems, Minneapolis, MN). The concentration of TNF- α was measured using an ultra-sensitive ELISA kit according to the manufacturer's protocol (R&D Systems).

Data Analysis

Significance testing was performed using Student's t test, one-way ANOVA, and Pearson's chi-squared test, as appropriate. A P value ≤ 0.05 was used as the cutoff for statistical significance.

Differences in the abundance of protein spots in the ARDS BALF at Days 1, 3, and 7 compared with normal BALF were determined using the EDGE algorithm (11). Multiple hypothesis testing was addressed by false discovery analysis using a Q -value of ≤ 0.01 (11). When a protein was represented by more than one spot on the gel, the abundance values were averaged.

TABLE 1. CLINICAL CHARACTERISTICS OF PATIENTS WITH ACUTE RESPIRATORY DISTRESS SYNDROME

	ARDS Day 1 (N = 7)	ARDS Day 3 (N = 8)	ARDS Day 7 (N = 5)	P Value
Age, yr	42	46	42	0.95
Male/female	3/4	3/4*	1/3*	0.54
PaO ₂ /Fio ₂	166	154	234	0.38
APACHE II	23.4	24.3	23.4	0.32
Lung Injury Score	2.60	2.53	2.08	0.75
ICU Mortality (%)	28.6	25.0	0	0.43
Protein Concentration (μg/ml)	1151	544	264	0.15

Definition of abbreviations: APACHE II = Acute Physiology And Chronic Health Evaluation II; ARDS = acute respiratory distress syndrome; ICU = intensive care unit.
* Sex information missing in one patient.

Principal components analysis (PCA) was performed based on the covariance matrix of normalized protein abundance values for all subjects (12). PCA is an exploratory analytical tool that is used to (1) reduce the complexity of the dataset and (2) identify meaningful groups and associations in the dataset. PCA transforms a number of correlated variables (e.g., individual protein spot abundance levels in each experimental sample) into a smaller number of uncorrelated variables, called principal components. The first principal component accounts for as much of the variability in the data as possible, and each succeeding component accounts for successively decreasing amounts of the remaining variability. PCA was used in this study to cluster the experimental groups based on the expression of protein spots in the BALF. All protein spots that were present in more than 50% of the gels and identified by mass spectrometry were included in the PCA analysis.

Enriched functional categories within groups of proteins that were differentially expressed in the BALF of subjects with ARDS compared with healthy control subjects were determined using the DAVID program (Database for Annotation, Visualization, and Integrated Discovery) (13). A Benjamini-corrected $P \leq 0.05$ was used as the cutoff for significantly enriched biological processes.

Protein interaction networks reflecting temporal changes in the BALF protein expression during the development of ARDS were constructed using the Ingenuity Pathways Analysis software and database (IPA; Ingenuity Systems; Redwood City, CA) (14). This database has been manually curated from over 200,000 full-text, peer-reviewed scientific articles studying approximately 10,000 human, 8,000 mouse, and 5,000 rat genes. A molecular network of direct and indirect interactions among these mammalian orthologs has been developed

and served as the basis for creating smaller networks from the proteomics data. These subnetworks were built around proteins with the highest connectivity using an iterative algorithm and then merged together to create a larger overall network, called the interactome. Several protein hubs that had not been identified in the proteomics experiment but were critical in linking the detected proteins together were identified by the program and incorporated into the network. Details of the IPA program and database can be found at the manufacturer's website (http://www.ingenuity.com/products/pathways_analysis.html).

RESULTS

Patient Characteristics

The baseline clinical characteristics of the patients with ARDS are summarized in Table 1. There were no statistically significant differences in the clinical characteristics between the patients with ARDS at Days 1, 3, and 7 of disease. Sepsis was the clinical risk factor for ARDS in all of the patients. The trends in mortality, protein concentration, and P/F ratio in the Day 7 group suggest that the patients who remained alive and intubated on Day 7 may have had improvement in these clinical parameters as a result of resolving lung injury. The control group consisted of nine normal volunteers who did not smoke cigarettes and did not have underlying lung disease.

Identification of BALF Proteins Using DIGE

An average of 991 protein spots was detected in each gel across all experimental groups. Irrelevant spots were excluded based on an algorithm in the Decyder program designed to detect dust, defects in the gels, and other experimental artifacts. Of the protein spots, only those that were found in over 50% of the gels in every experimental group were considered for further analysis. Protein spots of interest were excised and analyzed by MALDI-TOF/TOF mass spectrometry for protein identification. A total of 80 protein spots were identified by tandem mass spectrometry, representing 37 unique proteins. The location of these 80 protein spots on a representative gel and their identifications are shown in Figure 1 and Table 2, respectively. Overall, the 37 proteins that were identified represented diverse protein families, including opsonins, antioxidants, basement membrane proteins, coagulation proteins, and serum acute-phase reactants, among others. Of the 37 proteins that were identified, 22 were differentially

Figure 1. Two-dimensional difference gel electrophoresis (DIGE) analysis of acute respiratory distress syndrome (ARDS) bronchoalveolar lavage fluid (BALF). Proteins in the BALF of patients with ARDS and control subjects were separated by isoelectric point and molecular weight using DIGE. An average of 991 protein spots was detected in each gel across all experimental groups. A total of 80 spots, representing 37 unique proteins, were identified using mass spectrometry. The locations of these spots are shown in a representative gel. The proteins identified in these spots are shown in Table 2. An illustration showing only the individual spots is included in the online supplement (Figure E1).

TABLE 2. PROTEINS IDENTIFIED IN BRONCHOALVEOLAR LAVAGE FLUID

Spot Number	Protein Identification	Q-Values
1	Actin, gamma 1 (ACTG1)*	4.16E-05
2, 3, 4, 5	α_1 -antitrypsin (SERPINA1)*	4.16E-05
6, 7	Annexin A1 (ANXA1)*	4.07E-03
8, 9	Annexin A3 (ANXA3)*	1.41E-04
10, 11, 12, 13	Annexin A5 (ANXA5)	4.42E-01
14	Apolipoprotein A1 (APOA1)*	4.16E-05
15	Heparan sulfate proteoglycan 2 (HSPG2)*	9.74E-03
16	β_2 microglobulin precursor (B2M)*	4.16E-05
17, 18, 19	Calgranulin A (S100A8)	3.54E-01
20, 21, 22	Calgranulin B (S100A9)*	4.16E-05
23, 24	Calgranulin C (S100A12)*	4.16E-05
25	Ceruloplasmin (CP)*	2.52E-03
26	Coactosin-like protein (COTL1)*	1.41E-04
27, 28, 29, 30	Complement C3 precursor (C3)*	4.58E-04
31	Complement C4 precursor (C4)*	4.16E-05
32	Complement C9 precursor (C9)*	9.15E-04
33, 34	Cystatin S (CST4)	5.85E-02
35	Cystatin S (CST4), cystatin SA (CST2)	5.85E-02
36, 37	Fatty acid-binding protein, adipocyte (FABP4)*	5.43E-04
38	Fatty acid-binding protein, epidermal (FABP5)	1.69E-02
39	Ferritin light chain (FTL)	2.95E-02
40	Fibrinogen alpha chain (FGA)*	3.92E-04
41, 42, 43, 44	Glutathione S-transferase P (GSTP1)	1.87E-01
45, 46, 47, 48	Hemoglobin (HBB)	4.37E-01
49, 50, 51	Hemopexin precursor (HPX)*	4.16E-05
52	Histidine-rich glycoprotein precursor (HRG)*	4.16E-05
53	Peptidyl-prolyl cis-trans isomerase A (PPIA)	1.13E-02
54	Peroxiredoxin 2 (PRDX2)	1.42E-01
55	Peroxiredoxin 5, mitochondrial (PRDX5)*	4.16E-05
56, 57, 58	Plasma retinol-binding protein (RBP4)	2.54E-01
59	Protein disulfide-isomerase A3 (PDIA3)*	4.16E-05
60, 61, 62, 63	Protein kinase C inhibitor protein-1 (HINT1)	2.95E-02
64, 65, 66, 67	Serum amyloid protein A (SAA1)	3.80E-01
68	Superoxide dismutase (SOD3)	1.51E-01
80	Surfactant protein-A (SFTP-A)*	2.63E-03
69, 70, 71, 72, 73, 74, 75, 76	Transthyretin (TTR)	2.95E-02
77	Transthyretin (TTR), calgranulin B (S100A9)*	4.16E-05
78	Triosephosphate isomerase (TPI1)	1.75E-01
79	Tropomyosin alpha 1/3/4, beta chain (TPM)	3.18E-01

* Proteins that were differentially expressed in ARDS BALF compared with controls ($Q \leq 0.01$).

expressed in the ARDS BALF over time as compared with controls (EDGE analysis, Q -value ≤ 0.01). These proteins are identified in Table 2 (*). The average abundance values of the protein spots in the normal and ARDS BALF samples, measured as a $\log_2$ ratio between the spot of interest in the experimental sample (Cy5 wavelength) and the corresponding spot in the reference standard (Cy3 wavelength), are shown in the online supplement (Table E1).

PCA

Principal component analysis (PCA) was used to reduce the complexity of the proteomics data and to examine global trends in protein expression in the lungs of patients with ARDS over the course of their disease. BALF samples were grouped based on the variance of their protein expression (Figure 2). The majority of the variance ($> 70\%$) in the protein expression between normal and ARDS BALF was captured by the first three principal components. PCA showed that the ARDS and normal samples cluster into distinct groups. Within the ARDS cluster, PCA did not discriminate among samples collected on Days 1, 3, and 7. These results show that the pathophysiologic events in ALI/ARDS lead to significant derangements in the lungs that are reflected in dramatic changes in the BALF proteins at the onset of

Figure 2. Principal components analysis (PCA). PCA was used to group the BALF samples based on their variances in protein expression. This analysis showed that the ARDS and normal BALF samples cluster into distinct groups in the proteome space. The PCA analysis did not discriminate between the ARDS BALF samples collected at Days 1, 3, and 7.

disease. By comparison, the PCA suggests that the changes that occur in the proteome of the lungs over the subsequent course of ALI/ARDS are more modest.

Functional Analysis

To better characterize the functional differences between proteins detected in the BALF of subjects with ARDS versus normal subjects, the 22 proteins whose expression changed during the progression of ARDS were analyzed using the Gene Ontology (GO) database. The GO analysis revealed that the ARDS BALF was significantly enriched in processes involved in inflammation, immunity, response to microbes, response to stress/injury, and enzyme inhibitor activity (Figure 3A). The proteins included in each of the functional categories are listed in the online supplement (Table E3). Many of the differentially expressed BALF proteins were common to several enriched categories, as shown in Figure 3B. GO analysis of the proteins that were differentially present in the ARDS BALF on Days 1, 3, and 7 compared with normal BALF showed that the top five biological processes that are enriched at each day of ARDS are similar (Table E2).

Network Analysis

To examine the relationships between these functional groups of proteins in greater detail, an *in silico* protein network was created based on known protein interactions derived from the biomedical literature. The proteins detected in the proteomic analysis are shown in a red to green color scale corresponding to their relative abundance as compared with the pooled standard used in the proteomic analysis (Figure 4). The connectivity among these identified proteins was enhanced by systematically incorporating proteins or other biological compounds, such as hormones, that have known interactions with the identified proteins but were not detected in the proteomic analysis. These additional proteins and mediators are shown in yellow. To validate their inclusion into the network, several of the proteins that were added by the IPA program were measured in the ARDS BALF using sensitive immunoassays (Figure 5). The concentrations of TNF- α and IL-1 β were low, but measurable, in these dilute BALF samples. The concentration of IL-6 was higher at the onset of ALI, but progressively decreased over the course of disease. These findings are concordant with our previously published work that showed that TNF- α , IL-1 β , IL-6, and LPS-binding protein (LBP), among other proteins, were detectable on Days 1, 3, and 7 of ARDS (12–15). All of these proteins were nearly undetectable in normal BALF.

Figure 3. Functional analysis of ARDS BALF proteins. (A) Gene ontology (GO) analysis of proteins in ARDS BALF, organized by functional categories of proteins that are enriched in ARDS BALF. (B) Examples of proteins that are included in more than one category in the GO analysis.

The computational network analysis in this study was an attempt to capture the complex and dynamic nature of the changes in the lungs of patients with ALI/ARDS. The structure of the network that was generated from the proteomic analysis highlighted the complex interactions among its various nodes and identified several proteins as central hubs of connectivity. There is increasing evidence that the functional stability of biological networks is critically dependent on such hubs (15). In the ARDS protein network, many of the central hubs were proteins that have been implicated previously in the pathogenesis of acute lung injury, including TNF- α , IL-1 β , LBP, and p38 MAPK (12–16). Other highly connected nodes, however, were proteins and biological compounds whose roles have not been well studied in ALI, including retinoic acid, β -estradiol, annexin A1, the S100 proteins, and the EGF receptor (EGFR).

The temporally dynamic characteristic of ARDS was revealed by the time course network analysis. Significant differences were seen in the expression of many members of the network between normal controls and Day 1 of ARDS (Figure 4). Complement proteins, antiproteases, annexin A3, S100 proteins (S100A9, S100A12), actin, and extracellular matrix proteins (basement membrane-specific heparan sulfate) were increased in the BALF on the first day of ARDS compared with normal BALF. Conversely, several proteins, including surfactant protein-A, annexin A1, fibrinogen, and fatty acid-binding protein, were decreased on Day 1 of ARDS. The changes in the expression of these proteins in the ARDS BALF likely reflected perturbations in innate immune, oxidative, and apoptotic pathways, as well as cell and extracellular matrix breakdown during the acute phase of lung injury.

In contrast to the significant changes in protein expression between the normal and Day 1 ARDS BALF, the differences

between Days 1 and 3 of ARDS were less dramatic (Figure 4). Nevertheless, many proteins changed in abundance between these times. Some proteins, such as complement C3 and peroxiredoxin 2, showed major differences, while others, including superoxide dismutase, complement C4, histidine-rich glycoprotein, and protein disulfide isomerase A3, had modest, but detectable, changes. These changes likely reflect changes in innate immune and oxidant pathways in the early stages of lung injury. The expression profile of the network at Days 3 versus 7 of ARDS also revealed several proteins that changed significantly in expression (Figure 4). These included annexin A3 (decreased), surfactant protein-A (increased), and actin (decreased), among others. These changes likely reflect regeneration of the lung epithelium, decreased cellular injury and turnover, and the resolution of lung injury. However, the magnitudes of the changes in most proteins in the network between Days 3 and 7 of ARDS were modest. The changes in protein expression on Days 1, 3, and 7 of ARDS compared with normal control subjects are summarized in Figure 6.

DISCUSSION

In this study we used a quantitative electrophoresis-based proteomics method (DIGE) to profile the changes in the expression of 37 proteins in the BALF of healthy subjects and patients on Days 1, 3, and 7 after the onset of ARDS. Bioinformatics and computational analysis showed important differences in the BALF proteins between normal subjects and patients with ARDS that reflect selective enrichment of specific biological processes, such as immunity, defense responses, and inflammatory responses to pathogens and injury. These findings are consistent with the current paradigm of the mechanisms leading

Figure 4. Network analysis of ARDS BALF. *In silico* protein networks were created using the proteins identified in the proteomic analysis and published information about protein-protein interactions. The proteins are shown in a red to green color scale, corresponding to their relative abundance in the ARDS samples compared with the pooled standards used in the DIGE analysis (red = higher; green = lower). To enhance the connectivity of the network, proteins and other biological compounds that were not identified in the proteomic analysis, but had known interactions with the identified proteins, were added using the Ingenuity program (yellow). Several nodes, including TNF- α , IL-1 β , LPS-binding protein, p38 MAPK, β -estradiol, retinoic acid, and the S100 proteins, were highly connected to other proteins in the network. In addition, many proteins showed temporal changes in expression over the course of ARDS. Only 36 out of the 37 identified proteins are shown in the network because fatty acid-binding protein (epidermal, FABP5) was not connected to any of the other proteins and, therefore, not included.

to the development of lung injury, and suggest that critical events in the progression of ARDS can be captured by proteomic analysis of the BALF.

We examined temporal changes in the BALF proteome during the progression of ARDS by generating protein networks that mapped detailed protein interactions. Our approach has several key advantages. By incorporating proteins and biological compounds that were not identified from our proteomics experiments, but are known to be highly connected with other members of the interactome in the lungs, the network analysis correctly predicted a number of critical mediators in the pathogenesis of

ARDS. These proteins included TNF- α , IL-1 β , IL-6, LBP, and p38 MAPK. Previous studies showed that TNF- α , IL-1 β , IL-6, and LBP are present in the BALF of patients with ARDS and contribute to the pro-inflammatory environment (16–20). In addition, p38 MAPK has also been shown to regulate key physiologic responses, including neutrophil recruitment and clearance of edema fluid, during the development of lung injury (20–23). The proteins that were added into the network were predominantly cytokines, intracellular signaling proteins, and transcription factors that exist in extremely low concentrations in the BALF. These proteins were not identified in the proteomic

Figure 5. Measurement of proteins in BALF that were identified by the *in silico* analysis. TNF- α , IL-1 β , and IL-6 were measured in the ARDS BAL fluids using sensitive bead-based immunoassays. Undetectable values were assigned the value for the lower limit of detection in each assay.

analysis because they were below the threshold of detection for the DIGE method. This is supported by the fact that we were able to detect several of the proteins that were added to the network, including TNF- α , IL-1 β , and IL-6, in the ARDS BALF samples using sensitive immunoassays (Figure 5).

The unprompted addition of these proteins with previously identified roles in ARDS to the *in silico* network has two

important implications. First, it supports the interpretation that the protein networks generated in this study accurately model the proteome of the injured lung. Second, it suggests that the nodes in the network that have not been well described in the pathogenesis of ARDS may also represent novel mediators of disease.

β -Estradiol is an example of one such novel mediator. In the network, β -estradiol had multiple predicted interactions with proteins such as TNF- α , p38 MAPK, IL-6, IL-1 β , annexin proteins, and serum acute phase reactants, among others. Although the role of sex-related hormones in the pathogenesis of ARDS is unknown, clinical studies suggest that women have better outcomes in sepsis and ARDS (24, 26). These clinical observations are supported by animal studies showing that estrogen suppresses lung inflammation by modulating pro-inflammatory cytokines and vascular cell adhesion molecules (24, 25, 27–29). Although the precise mechanism by which sex hormones influence the systemic inflammatory response is unknown, the network analysis in this study suggests that β -estradiol may dampen the lung inflammation of patients with ARDS by modulating the effects of pro-inflammatory cytokines, such as TNF- α , IL-1 β , and IL-6.

In addition to identifying novel mediators of disease, the network analysis also revealed important global characteristics of the protein-to-protein relationships that occur in lungs of patients with ARDS. The most striking feature was the complex topology of the protein interactions. Nested in this complexity were two features that have important implications for the development of molecular strategies to treat patients with ARDS. First, many of the protein pathways were redundant. For example, the p38 MAPK pathway has been implicated in the development of the exaggerated inflammation that is seen in ARDS (20). The network analysis showed that p38 MAPK may trigger part of its biological effect by increasing the expression of TNF- α , among other proteins. However, the network also revealed that the activity of TNF- α is influenced by other proinflammatory proteins, such as IL-1 β , IL-6, and the S100 proteins. This redundancy in augmenting TNF- α activity is an example of how targeted blockade of one protein or pathway (such as p38 MAPK) is unlikely to completely abrogate the pathogenic signals (such as exaggerated inflammation) in a complex disease such as ARDS.

In addition to redundancy, another feature of the protein interactome was that many protein nodes were involved in multiple biological processes. Using TNF- α as an example, the network analysis showed that in addition to proinflammatory proteins, TNF- α interacted with bacterial opsonins (complement proteins, SFTPA), mediators of apoptosis (annexin A5, annexin A1), and proteins involved in lung repair (EGFR) (30–32). This suggests that TNF- α modulates pathways involved in innate immunity and repair in the lungs, in addition to inflammation. Thus, the indiscriminant blockade of TNF- α to reduce lung

Figure 6. Protein expression in the lungs during the course of ARDS. The proteins detected in the ARDS BALF were clustered based on similarities in expression patterns over the course of disease. Changes in the expression of the proteins that were identified in the proteomic analysis are shown in a red to green color scale corresponding to the relative abundance compared with the pooled standard (red = higher; green = lower).

inflammation has the potential to disrupt important homeostatic mechanisms, such as host defense and repair, and ultimately lead to unexpected adverse consequences.

The complex characteristics of the protein network in the lungs of patients with ARDS illustrate the challenges associated with identifying effective therapeutic targets in human diseases. The redundancy of protein pathways suggests that single protein interventions may not effectively disrupt pathogenic mechanisms. Furthermore, the overlapping functions of single proteins in numerous biological processes highlight the difficulty in generating focused therapeutic strategies that do not disrupt other important homeostatic pathways. In light of these challenges, network analysis offers a way to examine these biologic redundancies and overlapping functions and identify rational targets for molecular-based interventions that maximize clinical efficacy and minimize adverse effects. This may be achieved by targeting nodes within the network that are not highly connected themselves but critically influence densely connected hubs in the interactome.

Based on this premise, the protein network in this study predicts that the S100 proteins (S100A8 and S100A9) might be appropriate candidates for targeted intervention in ARDS because these proteins interact with several key modulators of ARDS, including TNF- α , IL-6, and p38 MAPK, but are not highly connected to other noninflammatory proteins in the interactome. The S100A8 and S100A9 proteins belong to a family of calcium-binding proteins that are abundant in the cytoplasm of neutrophils and monocytes, and are important mediators of inflammation in sepsis (33). Thus, the blockade of S100 proteins may result in an effective antiinflammatory effect by simultaneously modulating several proinflammatory pathways without perturbing other important homeostatic pathways in the lungs. This hypothesis is supported by recent work from Vogl and colleagues, who showed that S100A8 and S100A9 (also known as Mrp 8 and Mrp 14, respectively) mediate TNF- α -induced inflammatory responses in murine models of sepsis (33). Furthermore, mice that lack S100A8 and S100A9 were protected from septic shock after the administration of endotoxin or *Escherichia coli*-induced abdominal sepsis.

This study has several potential limitations. First, only a subset of the lung proteome was examined using DIGE proteomic analysis. As a result, a comprehensive molecular model of ARDS could not be generated, because doing so using a limited subset of the overall lung proteome could lead to over-interpretation of the data and erroneous conclusions. Instead, we focused the computational analysis on identifying novel molecular connections associated with the identified proteins that could have a role in the pathophysiology of lung injury. In doing so, the computational analysis confirmed several previous findings in the BALF of patients with ARDS and suggested new pathway interactions. Future studies using alternative methods, such as mass spectrometry-based proteomic analysis (shotgun proteomics), will permit the analysis of a larger set of the BALF proteome and the development of more comprehensive molecular models of ALI and ARDS. In addition, new methods need to be developed that will provide precise measurements of temporal changes in global protein networks. Although the number of BALF samples analyzed in each experimental group was modest, two important features of this study suggest that the findings are robust. First, we used a proteomics approach (DIGE) that incorporates a pooled standard and minimizes the experimental variability associated with gel-to-gel comparisons. Thus, the DIGE method increased the likelihood that the changes in protein expression between the experimental groups were true biological differences. Second, we only analyzed protein spots that were detected in more than 50% of the samples in every experimental group. Thus, the proteome

that we examined in this study is likely to be consistently found in the lungs of patients with ARDS. Because the control group used in this study included healthy volunteers that were not mechanically ventilated, it is possible that some of the differences in the proteomic analysis may be attributable to critical illness or mechanical ventilation, rather than lung injury. Previous studies of mechanically ventilated patients that did not have lung injury showed that the lungs of these patients do not contain increased amounts of total protein, neutrophils, or cytokines (34). This supports the interpretation that the changes seen in the proteomic analysis are likely due to the development of ARDS. Another limitation of this study is the dependence of the network analysis on our current knowledge of protein interactions. Thus, it is subject to any inaccuracies and shortcomings in the biomedical literature. Furthermore, there is likely to be an inherent bias in the identification of densely connected nodes, since many of these hubs have been previously well studied, and there is an abundance of information on their interactions. Therefore, predictions based on the protein network generated in this study will require further validation of biological relevance.

Despite these limitations, the computational proteomics approach used in this study confirms previous findings about the role of individual proteins in ARDS and provides insights about novel proteins and mediators that are likely to be involved in the progression of lung injury. In addition, the analysis shows the dynamic nature of differential protein expression in the lungs of patients with ARDS, the complexity of the protein-protein interactions, and the utility of network analysis in identifying potential targets for molecular-based interventions. Future studies that examine the specific changes in these networks over time are likely to provide important insight about the key biological processes that interact to produce lung injury.

Conflict of Interest Statement: None of the authors has a financial relationship with a commercial entity that has an interest in the subject of this manuscript.

References

1. Ware LB, Matthay MA. The acute respiratory distress syndrome. *N Engl J Med* 2000;342:1334-1349.
2. Bowler RP, Duda B, Chan ED, Enghild JJ, Ware LB, Matthay MA, Duncan MW. Proteomic analysis of pulmonary edema fluid and plasma in patients with acute lung injury. *Am J Physiol Lung Cell Mol Physiol* 2004;286:L1095-L1104.
3. Schnapp LM, Donohoe S, Chen J, Sunde DA, Kelly PM, Ruzinski J, Martin T, Goodlett DR. Mining the acute respiratory distress syndrome proteome: identification of the insulin-like growth factor (IGF)/IGF-binding protein-3 pathway in acute lung injury. *Am J Pathol* 2006;169:86-95.
4. De Torre C, Ying SX, Munson PJ, Meduri GU, Suffredini AF. Proteomic analysis of inflammatory biomarkers in bronchoalveolar lavage. *Proteomics* 2006;6:3949-3957.
5. Hayashi S, Chang DW, Wurfel MM, Tsuchiya M, Ruzinski J, Bumgarner RE, Heinecke JW, Park DR, Martin TM. Proteomic analysis of bronchoalveolar lavage fluid from patients with acute respiratory distress syndrome using high-abundant protein depletion and two-dimensional difference gel electrophoresis. *Proc Am Thorac Soc* 2006;3:A571.
6. Bernard GR, Artigas A, Brigham KL, Carlet J, Falke K, Hudson L, Lamy M, Legall JR, Morris A, Spragg R. The American-European Consensus Conference on ARDS: definitions, mechanisms, relevant outcomes, and clinical trial coordination. *Am J Respir Crit Care Med* 1994;149:818-824.
7. Steinberg K, Milberg JA, Martin TR, Maunder R, Cockrill B, Hudson L. Evolution of bronchoalveolar cell populations in the adult respiratory distress syndrome. *Am J Respir Crit Care Med* 1994;150:113-122.
8. Bjorhall K, Miliotis T, Davidson P. Comparison of different depletion strategies for improved resolution in proteomic analysis of human serum samples. *Proteomics* 2005;5:307-317.
9. Chang DW, Colucci G, Vaisar T, King T, Hayashi S, Matute-Bello G, Bumgarner R, Heinecke J, Martin TR, Domenighetti GM. Two non-

- bronchoscopic methods of sampling the lungs of patients with the acute respiratory distress syndrome (ARDS). *Clin Proteomics* 2008; 29:334–341.
10. Alban A, David SO, Bjorkesten L, Andersson C, Sloge E, Lewis S, Currie I. A novel experimental design for comparative two-dimensional gel analysis: two-dimensional difference gel electrophoresis incorporating a pooled internal standard. *Proteomics* 2003;3:36–44.
 11. Leek JT, Monsen E, Dabney AR, Storey JD. EDGE: extraction and analysis of differential gene expression. *Bioinformatics* 2006;22:507–508.
 12. Saeed AI, Sharov V, White J, Li J, Liang W, Bhagabati N, Braisted J, Klapa M, Currier T, Thiagarajan M, *et al.* TM4: a free, open-source system for microarray data management and analysis. *Biotechniques* 2003;34:374–378.
 13. Dennis G Jr, Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, Lempicki RA. DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol* 2003;4:3.
 14. Calvano SE, Xiao W, Richards DR, Felciano RM, Baker HV, Cho RJ, Chen RO, Brownstein BH, Cobb JP, Tschoeke SK, *et al.* A network-based analysis of systemic inflammation in humans. *Nature* 2005;437: 1032–1037.
 15. Luscombe NM, Babu MM, Yu H, Snyder M, Teichmann SA, Gerstein M. Genomic analysis of regulatory network dynamics reveals large topological changes. *Nature* 2004;431:308–312.
 16. Pugin J, Ricou B, Steinberg KP, Suter PM, Martin TR. Proinflammatory activity in bronchoalveolar lavage fluids from patients with ARDS, a prominent role for interleukin-1. *Am J Respir Crit Care Med* 1996; 153:1850–1856.
 17. Martin TR, Mathison JC, Tobias PS, Leturcq DJ, Moriarty AM, Maunder RJ, Ulevitch RJ. Lipopolysaccharide binding protein enhances the responsiveness of alveolar macrophages to bacterial lipopolysaccharide. Implications for cytokine production in normal and injured lungs. *J Clin Invest* 1992;90:2209–2219.
 18. Park WY, Goodman RB, Steinberg KP, Ruzinski JT, Radella F II, Park DR, Pugin J, Skerrett SJ, Hudson LD, Martin TR. Cytokine balance in the lungs of patients with acute respiratory distress syndrome. *Am J Respir Crit Care Med* 2001;164:1896–1903.
 19. Goodman RB, Strieter RM, Martin DP, Steinberg KP, Milberg JA, Maunder RJ, Kunkel SL, Walz A, Hudson LD, Martin TR. Inflammatory cytokines in patients with persistence of the acute respiratory distress syndrome. *Am J Respir Crit Care Med* 1996;154:602–611.
 20. Schnyder-Candrian S, Quesniaux VF, Di Padova F, Maillet I, Noulain N, Couillin I, Moser R, Erard F, Vargaftig BB, Ryffel B, *et al.* Dual effects of p38 MAPK on TNF-dependent bronchoconstriction and TNF-independent neutrophil recruitment in lipopolysaccharide-induced acute respiratory distress syndrome. *J Immunol* 2005;175:262–269.
 21. Becker PM, Verin AD, Booth MA, Liu F, Birukova A, Garcia JG. Differential regulation of diverse physiological responses to VEGF in pulmonary endothelial cells. *Am J Physiol Lung Cell Mol Physiol* 2001;281:L1500–L1511.
 22. Lee JW, Fang X, Dolganov G, Fremont RD, Bastarache JA, Ware LB, Matthay MA. Acute lung injury edema fluid decreases net fluid transport across human alveolar epithelial type II cells. *J Biol Chem* 2007; 282:24109–24119.
 23. Roux J, Kawakatsu H, Gartland B, Pespeni M, Sheppard D, Matthay MA, Canessa CM, Pittet JF. Interleukin-1beta decreases expression of the epithelial sodium channel alpha-subunit in alveolar epithelial cells via a p38 MAPK-dependent signaling pathway. *J Biol Chem* 2005;280:18579–18589.
 24. Gee AC, Sawai RS, Differding J, Muller P, Underwood S, Schreiber MA. The influence of sex hormones on coagulation and inflammation in the trauma patient. *Shock* (In press).
 25. Majetschak M, Christensen B, Obertacke U, Waydhas C, Schindler AE, Nast-Kolb D, Schade FU. Sex differences in posttraumatic cytokine release of endotoxin-stimulated whole blood: relationship to the development of severe sepsis. *J Trauma* 2000;48:832–839. (discussion 839–840).
 26. Schroder J, Kahlke V, Staubach KH, Zabel P, Stuber F. Gender differences in human sepsis. *Arch Surg* 1998;133:1200–1205.
 27. Schroder J, Kahlke V, Book M, Stuber F. Gender differences in sepsis: genetically determined? *Shock* 2000;14:307–310.
 28. Angele MK, Knoferl MW, Schwacha MG, Ayala A, Cioffi WG, Bland KI, Chaudry IH. Sex steroids regulate pro- and anti-inflammatory cytokine release by macrophages after trauma-hemorrhage. *Am J Physiol* 1999;277:C35–C42.
 29. Angele MK, Knoferl MW, Ayala A, Bland KI, Chaudry IH. Testosterone and estrogen differently effect Th1 and Th2 cytokine release following trauma-haemorrhage. *Cytokine* 2001;16:22–30.
 30. Bolger MS, Ross DS, Jiang H, Frank MM, Ghio AJ, Schwartz DA, Wright JR. Complement levels and activity in the normal and LPS-injured lung. *Am J Physiol Lung Cell Mol Physiol* 2007;292:L748–L759.
 31. Sorensen GL, Husby S, Holmskov U. Surfactant protein A and surfactant protein D variation in pulmonary disease. *Immunobiology* 2007;212: 381–416.
 32. Greene KE, Wright JR, Steinberg KP, Ruzinski JT, Caldwell E, Wong WB, Hull W, Whitsett JA, Akino T, Kuroki Y, *et al.* Serial changes in surfactant-associated proteins in lung and serum before and after onset of ARDS. *Am J Respir Crit Care Med* 1999;160:1843–1850.
 33. Vogl T, Tenbrock K, Ludwig S, Leukert N, Ehrhardt C, van Zoelen MA, Nacken W, Foell D, van der Poll T, Sorg C, *et al.* Mrp8 and Mrp14 are endogenous activators of Toll-like receptor 4, promoting lethal, endotoxin-induced shock. *Nat Med* 2007;13:1042–1049.
 34. Weiland JE, Davis WB, Holter JF, Mohammed JR, Dorinsky PM, Gadek JE. Lung neutrophils in the adult respiratory distress syndrome: clinical and pathophysiologic significance. *Am Rev Respir Dis* 1986;133:218–225.