Proteomic Studies in Acute Hypoxic Respiratory Failure

A THESIS
SUBMITTED TO THE FACULTY OF THE
UNIVERSITY OF MINNESOTA
BY

Maneesh Bhargava

IN PARTIAL FULFILMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

LYNDA ELLIS

August 2015

ACKNOWLEDGEMENTS

I would like to express my gratitude to Dr. Lynda Ellis for her supervision and guidance during my Ph.D. I am also fortunate to have Drs. David Ingbar and Chris Wendt as my mentors. Their insight and encouragement have been critical in all my academic endeavors. I am also thankful to Dr. Vipin Kumar and his group for their enthusiasm and eagerness to support this collaborative effort. It has been an honor to have extremely dedicated laboratory personnel, clinical staff and colleagues in the Division of Pulmonary, Allergy, Critical Care and Sleep Medicine. Their commitment to research was crucial in obtaining the biological samples for these studies.

I am very thankful to my family. My mother for her undaunted support, and my brother for purposeful prodding. I am extremely grateful to my wife, Mukta and my children, Divya and Rishi for their unfaltering love and affection. Finally, I want to thank all the patients who contributed to these studies

DEDICATION

This thesis is dedicated to my father whose memories and spirit of enterprise will always inspire me

ABSTRACT

Respiratory failure is a syndrome of impaired gas exchange resulting in abnormal oxygenation and carbon dioxide elimination. Lung damage seen in Acute Respiratory Distress Syndrome (ARDS) and Idiopathic Pneumonia Syndrome (IPS) cause acute respiratory failure and result in a high mortality and morbidity. Our objective is to gain novel insights into the pathways and biological processes that occur in response to diffuse lung injury by using comprehensive protein expression profiling in combination with bioinformatics tools. We characterized the protein expression in the Bronchoalveolar lavage fluid (BALF) from subjects with ARDS and also in hematopoietic stem cell transplantation (HSCT) recipients. For our studies, ARDS cases were grouped into survivors and non-survivors. The HSCT recipients were assigned to either infectious lung injury or IPS, i.e. non-infectious lung injury. The BALF samples were processed by desalting, concentration and removal of high abundance proteins. Enriched medium and low abundant protein fractions were trypsin digested and labeled with the iTRAQ reagent for mass spectrometry (MS). The complex mixture of iTRAQ labeled peptides was analyzed by 2D capillary LC-MS/MS on an Orbitrap Velos system in HCD mode for data-dependent peptide tandem MS. Protein identification employed a target decoy strategy using ProteinPilot. To determine the biologic relevance of the differentially expressed proteins we used Database for Visualization and Annotation for Integrated Discovery (DAVID) and Ingenuity Pathway Analysis (IPA). In the studies done on pooled BALF described in Chapter 3, we identified 792 proteins at a global FDR of ≤ 1%. The proteins that were more abundant in early phase survivors represented the GO groups involved in coagulation, fibrinolysis and wound healing, cation homeostasis and activation of the immune response. In contrast, non-survivors had evidence of carbohydrate catabolism,

collagen deposition and actin cytoskeleton reorganization. These proof of concept studies identified early differences in the BALF from ARDS survivors compared to non-survivors. As a follow-up, we characterized BALF from the individual subject with ARDS, 20 survivors and 16 non-survivors (Chapter 4). To accomplish this we performed six eight-plex iTRAQ LC-MS/MS experiments, and we identified 1122 unique proteins in the BALF. The proteins that had a differential expression between survivors and non-survivors represented three canonical pathways – acute phase response signaling, complement system activation, LXR/RXR activation- and four IPA Diseases and Functions- cellular movement, immune cell trafficking, hematological system development and inflammatory response. Similar to our prior studies, GO biological processes annotated to these proteins included programmed cell death, collagen metabolic processes, and acute inflammatory response. The sparse logistic regression model identified twenty proteins that predicted survival in ARDS. For the studies conducted in HSCT recipients (Chapter 5), we performed five eight-plex iTRAQ LC-MS/MS experiments and identified 1125 unique proteins. The proteins that had a differential expression between IPS and infectious lung injury enrich GO biological terms of immune response, leucocyte adhesion, coagulation, wound healing, cell migration, glycolysis, and apoptosis. In summary, the BALF protein expression profile identifies key differences in the biological processes in different subgroups of patients with diffuse lung injury. These differences position us to develop diagnostic and prognostic biomarkers and identify new targets for pharmacological therapy.

TABLE OF CONTENTS

List of Tables

List of Figures

LIST OF ABBREVIATIONS

ALI	Acute Lung Injury
ARDS	Acute respiratory distress syndrome
BALF	Bronchoalveolar lavage fluid
C3	Complement 3
CCSP	Club cell secretory protein
CF	Cystic Fibrosis
COPD	Chronic obstructive pulmonary disease
DAVID	Database for Annotation, Visualization, and Integrated Discovery
EF	Edema fluid
ELF	Epithelial Lining Fluid
ELISA	Enzyme-linked immunosorbent assay
FDR	False Discovery Rate
FiO2	Fraction of inspired oxygen
HCD	Higher energy collision induced dissociation
HDL	High density lipoproteins
HSCT	Hematopoietic stem cell transplant
IgA	Immunoglobulin A
IgG	Immunoglobulin G
IgM	Immunoglobulin M
IL	Interleukin
IPS	Idiopathic pneumonia syndrome
ITRAQ	isobaric tags for relative and absolute quantification
LC	Liquid chromatography
LDL	Low density lipoproteins

MALDI	Matrix-assisted laser desorption ionization
MAPK	Mitogen-activated protein kinase
MCX	mixed-mode cationic exchange
MGF	Mascot Generic Format
MMTS	Methyl methanethiosulfonate
MRM	Multiple reaction monitoring
MS	Mass spectrometer
MWCO	Molecular weight cut-off
PAGE	Polyacrylamide gel electrophoresis
PBSCT	Peripheral blood stem cell transplant
PSPEP	Proteomics System Performance Evaluation Pipeline
RAGE	Receptor for advanced glycation end products
SDS	Sodium dodecyl sulfate
SELDI	Surface enhanced laser desorption SELDI
SP-A	Surfactant protein A
SP-D	Surfactant protein D
SPE	Solid phase extraction (SPE)
SRM	Selected reaction monitoring
SILAC	*stable isotope labeling by amino acids in cell culture*
SWATH	Sequential Window Acquisition of all Theoretical fragment Ion
TAEB	Triethylammonium bicarbonate
TMT	Tandem mass tags
TNF	Tumor necrosis factor
UCB	Umbilical cord blood transplant
vWF	von Willibrand Factor

Chapter 1

Introduction to Thesis

Acute Respiratory Distress Syndrome (ARDS) is a devastating form of respiratory failure requiring treatment in the intensive care unit. Since its description, significant research has been performed to improve our knowledge about its pathogenesis, epidemiology and treatment options. However, the mortality from ARDS is approximately 40%. This high mortality is partly due to incomplete understanding of the lung repair mechanisms that are specifically lacking in the ARDS cases that do not recover from their illness. To address this gap in knowledge, we will characterize the protein expression in the distal lung space and identify the proteins and the biological process that are different in the ARDS survivors vs. non-survivors. These processes could be due to the differences in the compensatory mechanisms activated in response to lung injury. The proteins that enrich these biological processes are also expected to be robust biomarkers for prognosis. With advancements in mass spectrometers and bioinformatics tools, it is now possible to carry our comprehensive protein profiling with high throughput. For my thesis project, I will undertake a series of studies in the bronchoalveolar lavage fluid (BALF) from patients with ARDS and characterize the protein expression using high-resolution label based semi-quantitative proteomics. We will initially develop methods to obtain adequately mass spectrometer data and the bioinformatics pipeline for protein inference and quantification. This methods development will be carried out in pooled BALF that we have in hand. Next, we will characterize protein expression in BALF from prospectively enrolled individual patients with ARDS to identify the differences in the protein expression between ARDS survivors and non-survivors. These

studies will provide insights into the pathways and biological processes that differ in the two comparison groups and will also be biomarkers to predict outcomes in ARDS. We will also investigate the BALF in lung injury following hematopoietic stem cell transplant recipients to identify biological processes contributing to injury-repair in this setting. These studies will also provide markers for rapid diagnosis and early institution of treatment. Accomplishing these studies will provide experiential learning, create new knowledge and satisfy the requirements for my Ph.D.

Chapter 2

Application of Proteomics in

Acute Respiratory Distress Syndrome

Bhargava M, Higgins L, Wendt CH, Ingbar D. Application of Clinical Proteomics in Acute Respiratory Distress Syndrome. Clinical and Translational Med. 2014;3(34):111.

Biological systems function via intricate orchestrated cellular processes in which various cellular entities participate in a tightly regulated manner. Proteins are the 'work horse' of the cell and alterations of their behavior often are implicated in the development of diseases. Due to limitations in technology most of the initial biomedical research to determine the structure-function of the proteins was performed one molecule at a time. Since the completion of the human genome project there has been increasing interest to study the broader changes of proteins within a biological system, a field defined as Proteomics [1]. Prior reviews have focused on current techniques available at that time as applied to interstitial lung diseases [2, 3], lung cancer [4-6] and other lung diseases [7-9]. Some of these reviews have described the principles of electrophoresis, the gel-based methodologies and the basic principles of mass spectrometry (MS) [7]. With improvements in the MS platforms, the proteomics research has grown substantially from simply identifying proteins present in a clinical sample to the capability for absolute and relative quantification of proteins. With these advances, the field is now poised to identify candidate biomarkers and give insight into the biological mechanisms of disease. In this review, we highlight the principles and advances in proteomic platforms focusing on contemporary MS methodologies; discuss sample preparation challenges related to biofluids for pulmonary research and the application of current proteomic techniques in Acute Respiratory Distress Syndrome (ARDS).

PROTEOMICS METHODOLOGIES

Traditional proteome analysis began with 2-dimensional (2D) SDS-PAGE protein separation and differential analysis of gel spot patterns [10, 11]. Revolutionary ionization techniques- matrix-assisted laser desorption ionization (MALDI) [12] and electrospray ionization (ESI) [13] -have advanced all proteome pursuits starting in the mid-1990's [14-16].

Sample-specific details: Procurement of body fluid samples destined for proteomics projects must be controlled for protein loss, degradation, proteolysis and oxidative modifications [17-19]. Variability in sample handling should be minimized for quantitative analyses of protein expression levels to ensure conclusions are made based on biological variability not variability in sample handling. Wide dynamic ranges in protein abundances may limit or preclude detection limits for clinically interesting, low abundant proteins such as tissue leakage proteins and transcription factors [20, 21]. When protein dynamic range is wide (e.g., serum where protein abundance spans ten orders of magnitude), high abundant protein depletion with spin cartridges or columns is often necessary to maximize protein detection [20]. Assessment of the reproducibility of depletion products, when employed, is critical for both qualitative and quantitative projects [22].

Top-down analyses: 'Top-down' analyses of proteins by MS employ measurements on intact proteins [23, 24]. Two common technologies, MALDI and surface enhanced laser desorption (SELDI)- time of flight (TOF), provide protein profiles but do not provide protein identification. Thus, these have been utilized as screening methods for comparison of protein profiles from various sample types among populations of healthy and diseased patients for the pursuit of disease biomarker detection. Solid phase extraction (SPE) and chip-based techniques used for these top-down analyses are fast and efficient methods for intact protein purification, with the principal limitation that relatively small subsets of proteins are extracted and subsequently detected. SPE is employed for protein purification, desalting and concentration prior to MALDI-TOF MS detection. MALDI-TOF MS has been performed in both serum [25, 26] and BALF [27] for biomarker discovery. In a variation of MALDI-TOF MS, surface enhanced selective protein capture, an affinity-based chip method for protein extraction prior to SELDI-TOF

detection [28] has been used for biomarker discovery for subjects with pulmonary sarcoidosis.

Bottom-up analyses: in contrast to studying intact proteins, analysis of peptide mixtures obtained after proteolytic treatment of protein mixtures is called 'bottom-up' or 'shot-gun' proteomics [29, 30]. 'Bottom-up' proteomics studies are typically implemented for discovery-based experiments that provide protein identification and can also provide relative and absolute protein quantitative measurements with the appropriate experimental design. Two basic workflows for bottom-up proteomic studies are: 1) solution-based proteolytic digestion of protein extracts [31-34] such as done for studies in ARDS by others [34, 35] and our laboratory [36] 2) GeLC analysis, which entails one-dimensional (1D) SDS-PAGE separation of proteins, excision of consecutive gel regions and proteolytic digestion of proteins in each gel section [14, 37, 38]. The steps in a 'Bottom-up' proteomic workflow are shown in Figure 1 and include 1) Proteolytic digestion 2) Chromatographic peptide separation 3) Peptide Tandem MS 4) Database search for peptide identification and 5) Protein assembly.

Separation methods: Prior to MS protein identification and measurement peptide mixtures such as protein from excised gel band could be separated by I-D liquid chromatography (LC) [39]. 2D- LC is used for fractionation of complex peptide mixtures such as tissue or cellular proteins [29]. The first dimension typically separates peptides based on peptide pI or hydrophobicity in a high pH solvent. The second dimension separation is usually based on peptide hydrophobicity in a low pH solvent and is

performed ‘in-line’ with the MS-ESI interface between the column tip and MS orifice [40, 41]. In a less common approach, the second dimension LC eluent is directed onto a

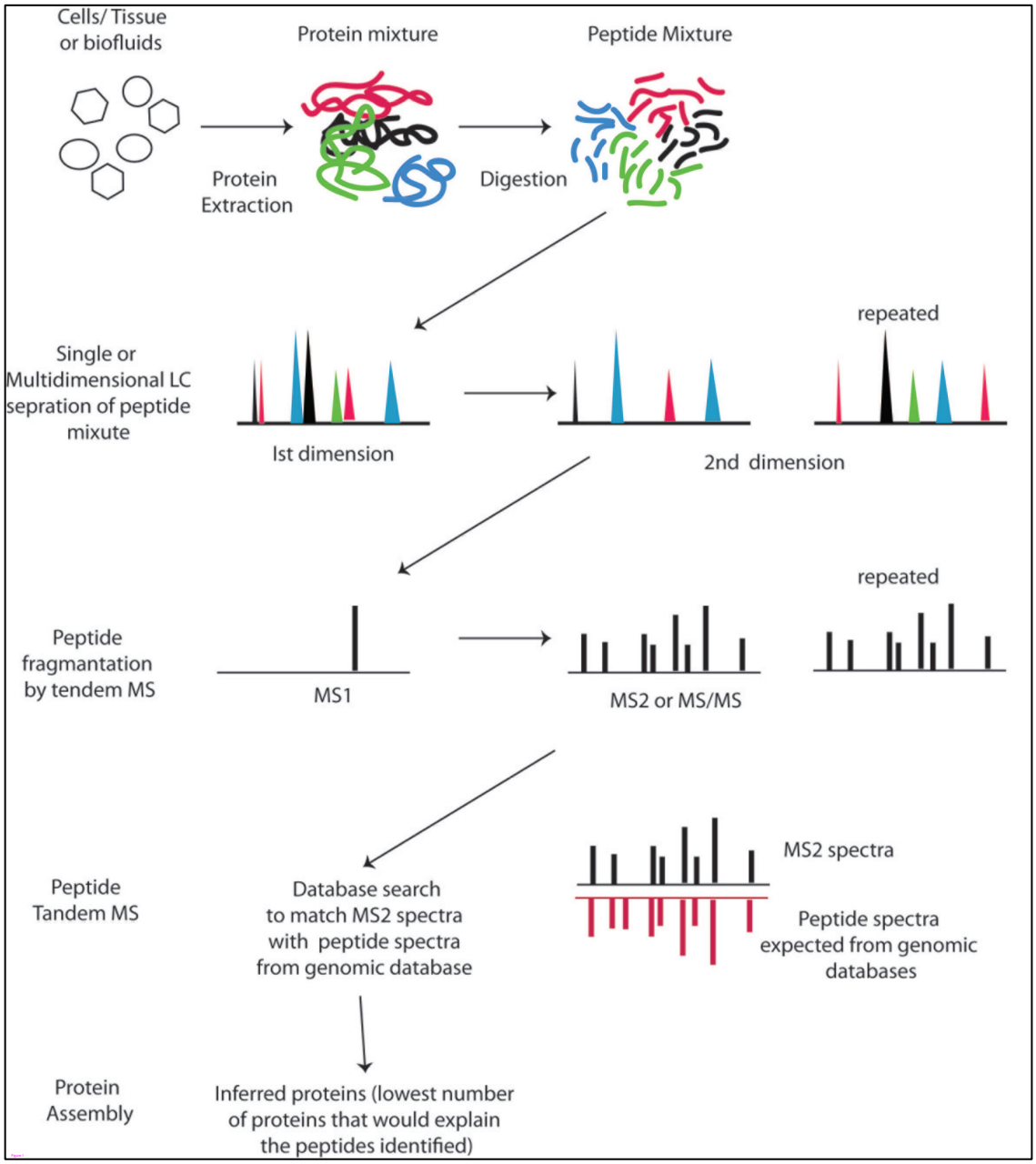

Figure 1: Workflow of ‘bottom-up’ or shotgun proteomics. *Protein extracts from cells, tissue or biofluids are prepared by mechanical (e.g., glass bead or homogenization) or chemical-based (precipitation, detergent solubilization) methods. Proteins are proteolytically digested into peptides, usually with trypsin, that are separated by 1D or 2D chromatographic separation. The final chromatographic step is performed in-line with the mass spectrometer. Two scan types are acquired: MS1 spectra contain intact peptide mass to charge (m/z) values; MS2 or tandem MS (MS/MS) spectra represent peptide fragment ion m/z values. Peptide MS1 and MS2 data are correlated with theoretical peptide m/z values with database search programs that use protein sequences as templates; parsimonious protein identifications with peptide matches are reported.*

metal plate or target for LC MALDI-TOF analysis [42, 43].

Peptide and Protein Identification: Peptide mass spectra generated by tandem MS are used for protein identification in bottom-up experiments. Program-specific algorithms compare theoretically derived peptide fragment pattern (generated *in silico*) to experimental peptide data [44-46]. Potential peptide database matches are ranked, scored and reported. Highest scoring peptides are used to generate a list of inferred proteins present in the complex mixture (protein assembly). Parsimonious protein assembly is used so the lowest number of inferred proteins would account for the detected peptides [47, 48]. Variations on database search algorithms provide a multitude of commercial and open source search programs for database searching, each of which has a unique peptide candidate scoring scheme and protein inference method. One or more peptide matches per protein are sufficient evidence for the detection of the protein in the sample [49]. False discovery rates of protein identification are available when the target protein database is reversed or scrambled and concatenated to the target database [50-52]. Public, species-specific protein data repositories that contain translated genomic sequences provide templates for the software programs (e.g., http://www.ncbi.nlm.nih.gov/protein and http://www.uniprot.org/).

Quantitative Proteomics: Methods for protein quantitation in clinical samples can provide either a relative or absolute quantitation. In the discovery phase of a project, relative protein quantitation is performed with the bottom-up, global approach from complex samples. Two discrete methods may be used for quantification: label-free [53] and differential isotopic labeling approaches [54, 55] (Figure 2). In both cases, equal amounts of protein extracts from multiple samples are processed by trypsin digestion and analyzed by LC-MS/MS.

Label-free quantitation: Peptide counts per protein [56, 57] or peptide peak area under the curve generated during chromatographic separation [58, 59] define label-free quantitation. Comparisons of peptide counts or peptide AUC across sample sets are performed with replicate measurements of each sample. Higher peptide count or AUC represents higher relative abundance when compared across samples (Figure 2, panel A). Label-free quantitation is challenging due to the inherent variability in the spectral level data and extensive post-processing required for minimizing this variability. This laborious approach has infrequently been used for studies in lung diseases.

Label based quantitation (SILAC, TMT, iTRAQ): The foundation of the peptide labeling approach is the incorporation of heavy isotopes into peptides or proteins by metabolic or chemical labeling.

- SILAC (stable isotope labeling by amino acids in cell culture) technique incorporates stable heavy isotopes into proteins via labeled amino acids added as a growth supplement during cell culture [60]. Cells are grown in similar media without heavy isotope labeled amino acids under different conditions that establish the comparative assay. Proteins from 'heavy' and 'light' labeled and are digested into peptides, mixed and analyzed by mass spectrometry. Mass spectrometric peak intensities for the 'heavy' or 'light' peptides are used for relative protein quantitation among the select sample types. Equal amounts of protein are used for each sample under comparison so that any differences in relative peptide/protein amount measured by mass spectrometry reflect differences between samples, not starting protein amounts. Thus, sample preparation must be optimized to ensure accurate and consistent protein quantitation of the starting samples.

- TMT and iTRAQ: Differential labeling of protein extracts from discrete samples can be multiplexed with the commercial TMT (tandem mass tags) and iTRAQ (isobaric tagging for absolute and relative quantitation) amine-specific chemical reagent tags [61-63] (Figure 2, panel B). Comparison of protein expression levels of 2 – 10 sample types is made with heavy isotope-labeled functional groups of

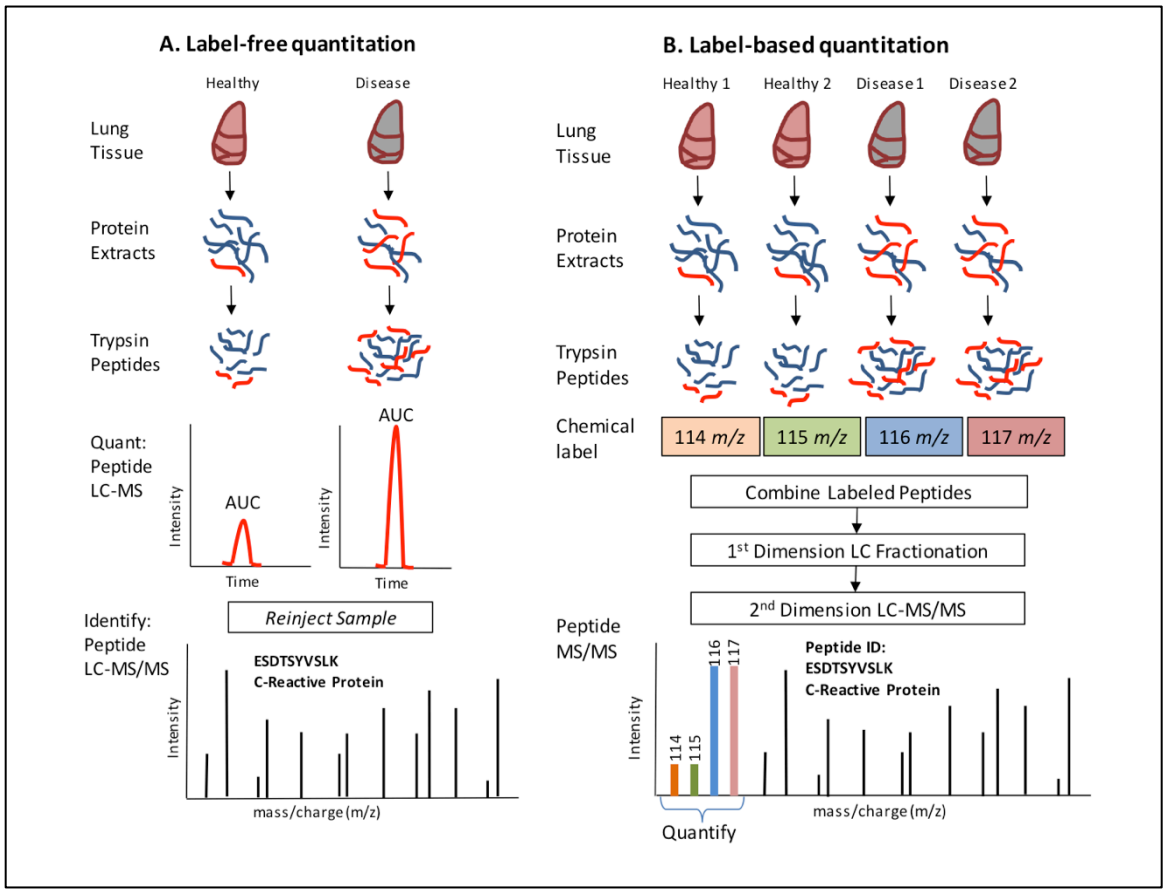

Figure 2: Principles of quantitative proteomics. *A) Label-free quantitation performed by peptide peak area under the curve. Proteins are extracted from tissue, proteolytically digested into peptides and analyzed by LC-MS. Analyte intensity versus retention time profiles are generated from which area under the curve (AUC) or summed peak intensities are calculated. Relative peptide amount in healthy versus disease sample is proportional to peak AUC or summed intensities. Targeted peptide identification is typically performed on a subsequent injection. B) Label-based quantitation with the iTRAQ® (isotope tagging for relative and absolute quantitation) 4plex workflow. Proteins from four individual samples are digested into peptides that are tagged with isobaric stable isotope labeled chemicals. Four chemical tags have 4 unique mass-to-charge (m/z) values that are produced during peptide tandem MS (MS/MS) and used for relative quantitation by relative peak intensity. Peptide fragment ions are used for peptide ID and protein inference.*

isobaric compounds that bind to peptide free amines [64]. These label-based methods allow for estimation of relative protein abundance [43].

Targeted Proteomics: Mass spectrometry can be employed as a targeted assay for the detection and precise quantitation of limited number of biomolecules identified from discovery-based experiments with selected reaction monitoring (SRM) MS [65, 66] or multiple reaction monitoring (MRM) MS assays. Protein detection or protein absolute quantification is achieved by selective measurement of peptides from proteolytic (e.g., tryptic) digestion of clinical samples on a specialized mass spectrometer, typically a triple quadrupole MS. The mass spectrometric acquisition method contains a list of the mass-to-charge values of the select peptides from the target protein(s) as well as the mass-to-charge values for one or more peptide fragment ions generated by tandem MS. The mass spectrometer acts as a selective mass-based detector for the chosen molecules; very low detection limits can be achieved, for instance, <10 fmol per molecule. MS measures peptides after separation by liquid chromatography. Chromatographic peptide peak integration is used for quantitation with the stable isotope dilution method using heavy isotope-labeled peptides as internal standards, which are spiked into the samples during work-up. The term MRM refers to an acquisition method for monitoring multiple peptide fragment ions per peptide as a measure of increasing specificity of detection for the select molecules. The sensitivity of SRM assays surpasses the sensitivity of data-dependent discovery based assays [67]. SRM methods provide a fast, cost-effective way to validate biomarker candidates or quantitative proteins from large sample sets. Targeted analyses require significant method development but provide a means for absolute quantitation of proteins with a low coefficient of variance [68-70].

SAMPLES FOR LUNG PROTEOMICS

Proteomic studies begin with protein extraction from a biological sample. Either tissue specimens and/or biological fluids can be used for proteomic investigations. Clinical-based samples, specifically body fluids, pose unique challenges for proteomics experiments due to the wide dynamic range of proteins typically present in most samples. Since MS is a concentration-dependent technique, the molecules of highest concentration in a sample are detected preferentially over lower abundant species. The presence of 'matrix' biomolecules such as mucins (e.g, large MW glycoproteins) and surfactants (e.g., phospho- lipoproteins) in pulmonary fluids complicate sample preparation since they must be removed during initial sample preparation steps. Sample cleanup and preparation methods must be developed and validated for specific applications. The initial step of protein extraction from either the cells or body fluids is the most critical for achieving successful and reproducible outcomes and is overall the most challenging step in a mass spectrometry-based proteomics experiment.

For lung diseases, including ARDS, it would be ideal to have lung tissue from an involved region for proteomic studies; however, lung biopsy specimens often are not available. Biological fluids that have been studied for extracellular proteins include plasma/serum. Using these fluids offers the benefit of repeated sampling but the lung-specific signal likely is diluted. Consequently other body fluids such as sputum [71], epithelial lining fluid (ELF) [72] lung edema fluid [73], exhaled breath condensate [74] and bronchoalveolar lavage fluid (BALF) have been investigated.

Sputum: Sputum consists of expectorated secretions from the respiratory tract. In a study, Nicholas et al. studied sputum proteins from one healthy smoker using 2-DE or

SDS-PAGE followed by Gel LC-MS/MS. By 2-DE over 600, features were present in the sputum. However only 61 proteins were identified when spots present in at least three replicate gels were excised and analyzed by MS/MS after in-situ trypsin digestion. Most of these proteins represented high abundance proteins previously reported in sputum, saliva, BAL and nasal lining fluid. In contrast, Gel LC-MS/MS provided extended coverage with identification of 191 human proteins, which also included low abundance proteins such as mucins, uteroglobin related protein, etc. The authors reported striking similarity between the proteome of the sputum and BAL [75] [20]. Gray et. al. [71] investigated sputum from healthy controls and subjects with an obstructive airways disease (asthma or COPD) and suppurative airway diseases (cystic fibrosis or bronchiectasis). These studies using top-down SELDI-TOF methodology identified approximately 50 (p-value <0.001) proteins peaks that differentiated healthy control subjects from patient's asthma or COPD and approximately 300 protein peaks (p-value <0.001) that differentiated healthy controls from subjects with bronchiectasis or CF. Calgranulin A, B, and C were more abundant in bronchiectasis and CF and not seen in COPD or asthma. In this study, club cell secretory protein (CCSP) was present in a lesser amount in both obstructive and suppurative lung diseases compared to healthy controls.

Bronchoalveolar lavage fluid (BALF): The epithelial lining fluid of the lung contains locally produced proteins that participate in a variety of different functions including defense mechanism, tissue remodeling, oxidant-antioxidant systems, inflammatory processes and cell growth. This fluid can be sampled directly by performing bronchoalveolar lavage. The proteins in BALF also may originate from diffusion from the serum; however comparison of serum and BALF proteomes demonstrates the presence of certain proteins at higher quantities in the BALF, suggesting alveolar and airway

epithelial cells specifically secrete some of these proteins [76]. Thus, BALF is particularly attractive to investigate in pulmonary diseases such as ARDS as it reflects the fluid most proximate to the site of injury.

Two-dimensional gel electrophoresis (2-DE) and LC-MS has been used for characterizing the protein expression in BALF [77-81]. Improvement in techniques for the separation of a complex protein mixture on a gel, such as isoelectric focusing (IEF), have been critical for development of 2-DE as a major tool for differential display proteomics [77] . One of the first studies mapping BALF proteins using 2-DE demonstrated mostly plasma proteins [81]. Subsequent studies using more sophisticated sample preparation technique have demonstrated a more comprehensive map of the BALF proteins [78, 79, 82] resulting in the creation of a database of BALF proteins [80, 83]. The 2-DE map created by characterizing both individual and pooled BALF from subjects with different lung conditions has resulted in visualization over 1200 silver stained spots and identification of 900 proteins that include intact proteins or protein subunits or fragments [83]. However, the major challenges in BALF proteomics are high salt and low protein content with wide dynamic range. Several of the sample preparation techniques used for 2-DE, such as desalting of the BALF, continue to be used for contemporary MS studies to address this issue. The removal of albumin [84] and other high abundance proteins, referred to as deep proteome profiling, has also improved identification of low abundance proteins [27, 85, 86] and is a useful strategy for LC-MS based proteomics. A recent report by Goodlet et. al. reviews studies applying shotgun proteomics to BALF [87]. Our laboratory has optimized BALF sample preparation for semi-quantitative protein expression studies using iTRAQ® LC-MS/MS for patients with ARDS. Initial studies using removal of six high abundant proteins (albumin, transferrin, IgG, IgA, haptoglobin and antitrypsin) resulted in identification of only 93 proteins at a

false discovery rate of 5% (abstract presented at ASPEN lung meeting). Optimization of sample preparation that included careful selection of spin columns for desalting and concentration of the BALF, depletion of 14 high abundance plasma proteins - albumin IgG, α1-antitrypsin, IgA, IgM, transferrin, haptoglobin, α2-macroglobulin, fibrinogen, complement C3, α1-Acid glycoprotein (orosomucoid), HDL (apolipoproteins A-I and A-II), LDL (mainly apolipoprotein B)- in combination with the use of high-resolution Orbitrap MS resulted in improved coverage with identification of approximately 724 proteins at 1% global FDR [34]. With improvement in the tools available, it is likely that challenges with BALF such as high dynamic range, protein loss during sample preparation, and variable states of dilution during sampling will be overcome, and a comprehensive database of BALF proteome will become available.

***Serum or plasma*:** Plasma and serum are attractive due to ease of collection thus permitting serial measurements. This could be extremely valuable in ARDS to understand the pathological changes that occur during the development and recovery stages of this disease when lung-specific biospecimens can be challenging to collect. Other advantages of identifying markers in serum or plasma include the ability to detect proteins with different tissue of origin such as the alveolar epithelial cells (SP-D, SP-A, RAGE), vascular endothelium (vWF), matrix metalloproteinase and mediators of inflammation [88]. However, barriers to successful plasma biomarkers include its high level of complexity of the proteome in addition to high abundance proteins limiting the systematic study of medium or low abundant proteins. Similar to BALF, immunodepletion of high abundance proteins has been used for plasma proteomics in ARDS [89, 90].

Other potential bio-fluids that could be investigated include urine, nasal lavage fluid, and pleural effusion fluid. However, currently there is limited evidence of the utility of these samples in the study of ARDS.

PROTEOMICS IN ARDS

ARDS is the acute respiratory failure with bilateral infiltrates due to permeability pulmonary edema resulting in hypoxia with a decrease in PaO_2 to FiO2 ratio in the absence of congestive heart failure [91-93]. ARDS continues to have a high mortality [94, 95]. American European Consensus Conference criterion used the term Acute lung injury (ALI) for a milder form of ARDS [93] but Berlin definition has suggested using mild ARDS instead of ALI [96]. Current knowledge is that ARDS is associated with an exuberant inflammatory response in the lung resulting in diffuse alveolar damage, surfactant dysfunction, epithelial and endothelial damage with loss of alveolar-capillary barrier and leakage of protein-rich edema fluid into the alveolus that results in impaired gas exchange. Following the exudative phase the lung attempts to repair itself by proliferation of type II alveolar epithelial cells which then differentiated into type I alveolar epithelial cells and ultimately leading to regeneration of the alveolar epithelium and clearing of edema fluid and cellular debris from the alveolus. Proteomics studies have been used to provide novel insight into the mechanisms underpinning the development of and recovery from ARDS and also to discover biomarkers of the disease (Table 1).

Initial attempts to study the proteome in ARDS were performed using gel-based platforms. First attempts at applying proteomics to ARDS were published by Bowler [73] where they studied plasma and edema fluid (EF) in 16 (age 55 ± 3 years) patients with ALI/ARDS (PF ratio 124 ± 15) and plasma and BALF in 12 normal non-smoking subjects

(age 25 ± 5 years). Studies performed using 2-DE demonstrated 300 distinct protein spots in healthy volunteers. In healthy controls, the protein profile was globally similar except that there was some variability in the intensity of protein spots. Multiple isoforms of some proteins such as SP-A, IgA, and IgM, were evident in the BALF. A few proteins were present only in the BALF and not in plasma, several proteins such as albumin, haptoglobulin, IgG, fibrinogen, apolipoprotein, clusterin-sulfated glycoprotein-2, transferrin, retinol binding protein, and transthyretin all had more intense staining in the plasma than BALF. In patients with ALI/ARDS the protein spot profile could be grouped into three patterns when compared to controls- 1) increased protein intensity, 2) decreased protein intensity or 3) modified expressions due to presence of post-translational modifications. The spots with increased relative intensity in EF of all ALI subject were of albumin, transferrin, IgG, and clusterin. In contrast, SP-A was seen in the BALF for all normal subjects but only one patient with ALI/ARDS. Similarly, alpha-1-anti trypsin was identified in all normal subjects but only half of ALI/ARDS patients. Haptoglobin and orosomucoid appeared to be have undergone post-translational modification in ALI/ARDS. The authors concluded that proteomics has a potential to study the air space in patient's with ALI/ARDS with the ability to identify post-translational modifications that would not be possible with other techniques.

In another study de Torre et al. [97] used top-down SELDI-TOF methodology and 2-DE with MALDI-TOF MS to identify BALF protein profile differences in ARDS compared to normal subjects. Study subjects included 11 cases within 72 hours of meeting the ARDS criterion and 33 healthy nonsmoking subjects challenged with either saline or endotoxin for induction of local lung inflammation followed by BAL in 6, 24 and 48 hours. Their studies revealed the presence of differentially expressed proteins in endotoxin challenged compared with saline challenged subjects. Three peaks at 14,18 and 28 kDa

were more prominent in the endotoxin challenged subjects. The inflammation persisted at 24 hours but decreased at 48 hours after the endotoxin challenge. The pattern from ARDS cases was similar to that seen at 6 hours after the endotoxin challenge with an increase in the 14 and 28 kDa peak intensity. Subsequent 2-DE combined with in-gel trypsin digestion with MALDI-TOF MS identified increased level of apolipoprotein A1, S100-A8, and A9 in subjects challenged with endotoxin and ARDS.

Other studies have used MS for characterizing global changes in BALF in patients with ARDS. In a study Chang et. al . [38] performed DIGE followed by MS-based proteomics in combination with *in silico* analysis to characterize serial changes in ARDS BALF at day 1 (n= 7), day 3 (n=8), and day 7 (n=5) and compared these to normal volunteers (n=9). Protein separation using DIGE showed an average of 991 protein spots in each group of patients. Of these 991 protein spots, 80 spots of interest were chosen for further study using tandem MALDI-TOF resulting in identification of 37 unique proteins that represented opsonins, antioxidants, basement membrane proteins, coagulation proteins and acute phase reactants. Twenty-two of these proteins were differentially expressed over time compared to controls. This type of study lends itself to functional analysis and Gene Ontology of these 22 proteins demonstrated processes involved in inflammation, response to microbials and response to stress/injury. An advantage of this approach is a sophisticated network analysis that revealed complex and redundant dynamic changes suggesting the complex nature of protein changes in ARDS. Several of the proteins that were previously known to be critical in ARDS such as TNF-alpha, IL-1beta, LBP, p38MAPK were central hubs in the identified networks in this study. Time course network analysis showed temporal dynamic changes. Compared to controls, on day one of the ARDS diagnosis there were increases in complement proteins, annexin A3, S100 protein, antiproteases, actin and extracellular matrix proteins in the BALF. In

contrast, surfactant protein-A, annexin A1, fibrinogen and fatty acid binding protein were decreased in ARDS compared to control. Differences between day one and day three of ARDS were less dramatic though complement C3 and preredoxin-2 showed a major difference. By day seven, there was evidence of regeneration of the lung epithelium, decreased cellular injury, cell turnover and resolution of lung injury.

Our laboratory has used label based quantitative 'bottom-up' proteomics (iTRAQ Orbitrap LC-MS/MS) and characterized protein expression from ARDS patients who had BALF collected either in early phase of ARDS (day 1-7 after intubation) or late phase (≥ 8 days post intubation) [36]. The goal of these studies was to identify differentially expressed proteins in early phase survivors when compared to early phase non-survivors and determine the biological processes that are lacking or over-expressed in the two groups with divergent outcomes. We identified 724 proteins (FDR ≤ 1%) of which 499 proteins had quantitative data available. The proteins that were overexpressed in early phase survivors represent six ontologies- three related to coagulation, fibrinolysis and wound healing, two related to iron and cation homeostasis and one related to immune system activation. In contrast, the early phase survivors had a signature of collagen deposition, carbohydrate catabolism, and actin cytoskeleton organization. Proteins that are differentially expressed in these biological processes could be potential biomarkers for prediction of outcomes in ARDS. In this study when early phase survivors were compared to late phase survivors, biological processes that were activated in late phase were cell migration and actin filament-based processes suggesting dynamic changes in the BALF occur in ARDS subjects who survive. The processes that get activated in late phase ARDS survivors could be potential targets to design novel therapeutics and be manipulated in early ARDS in patients predicted to have poor outcomes.

In a recent study, pooled plasma from patients with ARDS due to direct lung injury (n=6), indirect lung injury (n=5) and normal controls (n=15) were analyzed using semi-quantitative proteomics by iTRAQ with MALDI-TOF tandem MS [89]. Despite the depletion of albumin and IgG, the proteome coverage in that study was limited to identification of 2429 peptides with only 132 non-redundant inferred proteins. Of these 132 proteins only eleven proteins were differentially expressed in ARDS compared to controls, seven up-regulated and four down-regulated. The canonical pathways represented by these proteins were liver X receptor/retinoid X receptor (LXR/RXR) and farnesoid X receptor (FXR)/RXR activation, clathrin-mediated endocytosis signaling, atherosclerosis signaling, IL-12 signaling and production in macrophages, nitric oxide and reactive oxygen species production in macrophages, and complement system signaling. Due to the limited protein coverage and relatively small number of differentially expressed proteins, any protein pathway inference requires further investigation. This study highlights the ongoing challenges of plasma/serum proteomics due to a wide dynamic range and lack of deep proteome coverage in these biofluids.

In addition to BALF and plasma, exhaled breath condensate has been studied by SDS gel separation in combination with MALDI-TOF in patients with respiratory failure [98]. A high level of cytokeratin 2 and 10 was associated with increased peak inspiratory pressure; PEEP and ARDS score suggesting that cytokeratins correlated with mechanical stress. These studies are examples of how extended proteome coverage of lung biospecimens by different proteomics platforms and computational tools can lend new insights into the pathobiology of ARDS.

CONCLUSION AND FUTURE OF PROTEOMICS IN LUNG DISEASES

Significant strides have been made in several techniques that are available for large-scale studies of proteins in biological systems. Mass spectrometer based proteomics has evolved from the ability to identify proteins present in a complex mixture to its current state where both label-free and label-based methodologies can provide quantitative information regarding proteins with high precision. Label-based methodologies are currently used more widely, but one of the limitations of these techniques is co-isolation of more than one peptide for tandem MS, which would provide imprecise quantification. Label-free quantification with SRM and MRM requires prior information of the peptide behavior of the proteins of interest. Targeted proteomics with SRM or MRM is also dependent on sample processing prior to LC-MS and thus precludes measurement of low abundance proteins. Some of the newer techniques that implement unbiased data independent acquisition by mass spectrometry followed by targeted data extraction such as SWATH-MS (**S**equential **W**indowed data independent **A**cquisition of the **T**otal **H**igh-resolution **M**ass **S**pectra) [99] offers promise for high throughput precise quantification of a large number of proteins. Sophisticated bioinformatics algorithms are also being developed (inSeq) [100] which implement real-time assignment of the spectral matches allowing for improved accuracy of quantitation and also improved localization of post-translational modifications. Better understanding of post-translational modifications will allow more comprehensive mapping of networks and pathways implicated in certain diseases. In addition to advanced algorithms for protein inference, there is a major opportunity to understand the systems that are contributing to a disease state by integrating proteomics with other platforms such as next-generation sequencing and small molecule studies using metabolomics. This 'multi-dimensional data integration' would be key to developing targeted therapies for complex conditions like ARDS.

Tables for Chapter 2

Table 1: Studies in ARDS using proteomics platforms					
Year	**Proteomics Methodology**	**Sample Type**	**Number of subjects**	**Number of proteins identified**	**Reference**
2004	2DE-MALDI/TOF	Plasma and Edema fluid in ARDS and Plasma and BALF in controls	ALI/ARDS = 16, Controls= 12	300 distinct protein spots and 158 proteins identified.	Bowler [73]
2006	SELDI-TOF and 2DE+ MALDI TOF/TOF	BALF	ARDS= 11, Healthy nonsmoking controls = 33	Only differentially expressed proteins reported	de Torre [97]
2006	'Bottom-up' proteomics with LC-MS/MS	BALF	ARDS= 3	226, 291 and 659 proteins for the three patients studied	Schnapp [35]
2008	2DE-MALDI TOF/TOF	BALF	ARDS day 1= 7 ARDS Day 3= 8 ARDS day7= 5	991 protein spots seen. Only 80 protein spots analyzed by MS which represented 37 unique proteins	Chang, Martin [38]
2013	MALDI TOF/TOF	Pooled plasma	Direct injury = 6, Indirect injury= 5, healthy controls = 15	132 proteins	Chen [89]
2014	iTRAQ Orbitrap LC-MS/MS	Pooled BALF	Early phase ARDS survivors= 7, non-survivors = 8 & Late phase ARDS survivors= 7	724 proteins identified, 499 proteins quantified	Bhargava [36]

Chapter 3

Proteomic Profiles in Acute Respiratory Distress Syndrome Differentiates Survivors From Non-survivors

Bhargava M, Becker TL, Viken KJ, Jagtap PD, Dey S, Steinbach MS, Wu B, Kumar V, Bitterman PB, Ingbar DH, Wendt CH. Proteomic profiles in acute respiratory distress syndrome differentiates survivors from non-survivors. PLoS One. 2014;9(10):e109713. PubMed PMID: 25290099; PubMed Central PMCID: PMC4188744.

Acute Respiratory Distress Syndrome (ARDS) is characterized by the abrupt onset of tachypnea, hypoxia, and loss of lung compliance in response to infectious or inflammatory triggers [101]. Extensive research has improved our understanding of ARDS pathophysiology [102], epidemiology [94], treatment options [103, 104] and outcomes [94, 95], yet ARDS patients continue to have a high mortality rate. There is strong interest in identifying biomarkers to predict the development of ARDS in at-risk subjects [105-107], assist in diagnosis [108-111], and inform prognosis [109, 112-116]. Biomarkers enabling risk stratification would be not only useful in the clinical care setting, but also in clinical trials of new therapeutic interventions to phenotype clinical trial subjects and serve as surrogate endpoints.

Development of ARDS is associated with the activation of a large number of inflammatory mediators that damage the alveolar epithelium, endothelium, and basement membrane. Biomarkers based on the tissue of origin have been studied in both single center studies [107, 117, 118] and in NHLBI ARDS Network cohorts [109, 112, 117]. Most studies have focused on investigating an individual biomarker in blood, bronchoalveolar lavage fluid (BALF), or urine. Markers of inflammation such as interleukin-1β [119], interleukin 6 [104], and soluble TNF receptor I and II [120] are associated with poor prognosis in ARDS. Markers of endothelial damage including ICAM-1 [104, 117], Angiopoietin (Ang) [109], and Von Willibrand Factor (vWF) [121] correlate with higher mortality from ARDS. Poorer outcomes are also associated with higher plasma levels of SP-D (but not SP-A), a marker of type 2 alveolar epithelial cell damage [115], and receptor of advanced glycation end products (RAGE), a marker of type 1 alveolar epithelial cell damage. Several other molecules, such as those involved in coagulation [122], damage to the extracellular matrix [116], and oxidative stress (urine NO) [123] correlate with ARDS outcomes. A combination of biomarkers and clinical

predictors was found to be superior to clinical predictors or biomarkers alone for predicting mortality in ARDS [122]. However, the identification of a single biomarker or a combination of biomarkers that could be widely used has remained elusive [124] due to lack of correlation between the biochemical marker, pathophysiological variables, and clinical outcomes.

The primary aim of this study was to identify pathways of survival and stimulate new biomarker discovery by characterizing the BALF protein expression profile of ARDS survivors and non-survivors at different stages (early versus late) of disease progression. We analyzed medium and low abundant protein fractions in BALF samples by using contemporary high-resolution mass spectrometry (MS)-based proteomics techniques, along with quantitative labeling methodology. Our hypothesis was that patients who can survive ARDS would exhibit a distinct BALF protein profile during the early phase of mechanical ventilator support. Here, we show distinct differences in the BALF proteome between patients who survive ARDS from those who die. Moreover, the ontologies of differentially expressed proteins in late-phase survivors (cell migration and actin cytoskeleton organization) differ markedly from those in early-phase survivors, suggesting a critical role for these processes during lung repair. Enhancing these processes may provide new directions for therapy in ARDS.

METHODS

Study population

The University of Minnesota Institutional Review Board Human Subjects Committee approved this study. Patients were recruited from the University of Minnesota Medical Center. Informed consent for study participation was obtained from either the patient or the patient's legal representative. The early-phase ARDS BALF samples were available

from clinically indicated bronchoscopies with excess supernatant made available for these studies. The late-phase ARDS samples were excess supernatant BALF obtained from research bronchoscopies with consent from the patient or the surrogate. Bronchoalveolar lavage (100 mls normal saline) was performed using standard protocol in either the right middle lobe or left upper lobe (lingula).

For this study, patients were grouped based on the timing of the bronchoscopy – conducted in either the early phase of ARDS (Day 1-7) or the late phase (Day 8-35), referenced to the initiation of mechanical ventilation (designated Day 1) – and the outcome at the time of discharge (non-survivor or survivor). We thus studied patients in the early phase who were grouped into survivors or non-survivors and late-phase survivors. Late phase non-survivors were not included in this study as not enough BALF was available to perform the protein expression profile. The APACHE-II score was calculated to assess the severity of illness on the day of bronchoscopy for patients in the early phase of ARDS as previously described [125].

Sample preparation

BALF samples were processed as previously described [23] with some modifications. BALF containing equal amounts of protein from individual patients were pooled to collect a total of 4 mg protein for each group (early-phase survivors, early-phase non-survivors, and late-phase survivors). Pooled BALF was concentrated and desalted by centrifugation with an Amicon 3-MWCO spin filter (Millipore, catalog number UFC800396). To decrease the dynamic range, we enriched the medium and low abundance proteins by selectively immunodepleting the fourteen most abundant proteins in the concentrated samples on Seppro IgY 14 spin columns (Sigma-Aldrich, cat # SEP010). The Seppro IgY 14 spin columns deplete albumin, IgG, α1-antitrypsin, IgA,

IgM, transferrin, haptoglobin, α2-macroglobulin, fibrinogen, complement C3, α1-acid glycoprotein, apolipoprotein A-1, A-II and B. Per the manufacturer's instructions, each sample was mixed with the dilution buffer to a final volume of 500 µl, loaded onto the immunoaffinity depletion column, and incubated for 15 minutes at room temperature. To prevent saturation of the column, 250 µg of protein was depleted at a time. The unbound medium and low abundance proteins were collected in the flow through. Pooled samples from each representative group were required to have adequate protein concentrations since immunodepletion results in > 90% of the proteins being removed. An additional wash was performed with 0.5 ml of the dilution buffer. The depleted samples were then concentrated with an Amicon filter. A buffer exchange with 0.5 M triethylammonium bicarbonate (TEAB) was performed to remove TRIS, and the sample was concentrated with an Amicon filter. A Bradford protein assay was performed to quantify the enriched low abundant proteins.

iTRAQ labeling and 2D LC-Orbitrap MS

Enriched medium and low abundance proteins (50 µg from early-phase survivors and non-survivors, 25 µg from late phase survivors) were digested by trypsin and labeled with iTRAQ® reagent (AB Sciex, Foster City, CA) [27] for mass spectrometric analysis. The total peptide mixture was purified with an MCX Oasis cartridge (Waters, Milford, MA) before separation via two-dimensional liquid chromatography-mass spectrometry (2D LC-MS). LC and MS experimental details were previously reported [23]. Proteins were separated and concentrated offline in the 1st dimension into 15 peptide-containing fractions, collected in 2-minute intervals on a C18 Gemini column (Phenomenex, Torrance, CA) at pH 10, and in the 2nd dimension by a C18 reversed-phase capillary LC with a nano LC system (Eksigent, Dublin, CA). Data-dependent acquisition of the six most intense peaks per LC fraction was performed on an Orbitrap Velos system, with

HCD (higher energy collision induced dissociation) as the activation type for peptide tandem MS.

Database search for protein identification and quantification

Each of the 15 .RAW files generated from the Orbitrap Velos MS system was converted to mzML files by using msconvert, then converted to a ProteinPilot compatible Mascot Generic Format (MGF) with preselected iTRAQ reporter ions. The MGF files were searched against the Human UniProt database along with contaminant protein sequences (84,838 sequences in total; December 2012) using ProteinPilot version 4.5 and the following search parameters: Sample Type: iTRAQ 4-plex (peptide labeled); Cys-alkylation: MMTS; Instrument: Orbi MS, Orbi MS/MS; Run Quant; Use bias correction; Search focus on biological modifications and amino-acid substitutions; Thorough search and with a Detected Protein Threshold (Unused Protscore (Conf)): 10%. The ProteinPilot searches and subsequent generation of PSPEP (FDR) report and protein and peptide level summaries were generated within Galaxy-P [126]. Because MS data acquisition was performed on BALF samples after depletion of 14 high abundance plasma proteins, the high abundance proteins (or their fragment) were manually removed if they were present in the list of inferred proteins generated by ProteinPilot. Protein Summary with iTRAQ ratios (with early-phase survivors as the denominator for determining fold change) was processed through a workflow built within Galaxy-P so that it yielded UniProt accession numbers and gene names of differentially expressed proteins. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium [127] via the PRIDE partner repository with the dataset identifier PXD001095.

Statistical analysis

Differences in the clinical characteristics of the three participant groups were calculated by using ANOVA and, when appropriate, a post hoc Tukey test. For protein identification and quantification, multiple hypothesis correction was performed by controlling for false discovery rate (FDR), which measures the expected proportion of false positives among the statistically significant findings. The FDR cutoff was set at ≤1% (global) for protein identification in ProteinPilot. For quantification of protein abundance, each ratio (obtained by comparison of early-phase non-survivors to survivors or late-phase survivors to early- phase survivors) was compared to one; multiple hypothesis correction was performed by controlling the FDR set at ≤ 5% [128] and computing q-values with the mafdr routine in Matlab. Proteins with q-values less than 0.05 were retained for further analysis.

Gene Ontology (GO) Enrichment Analysis

To gain insight into the biological significance of differentially expressed proteins, we used the Database for Annotation, Visualization, and Integrated Discovery (DAVID, http://david.abcc.ncifcrf.gov, search date 7/11/13) [129]. DAVID provides batch annotations for highlighting the most relevant GO term associated with a gene (or protein) list. Of the three GO terms annotated to a gene (molecular function, biological process, and cell compartment), we limited the biological process annotation to differentially expressed proteins. Functional annotation clustering analysis in DAVID was used to identify the combinations of genes, according to common biological function. DAVID generates an enrichment score for a group of genes indicating annotation term member associations in a given experiment. An enrichment score of 1.3 is equivalent to a non-log scale p-value of 0.05.

Individual Protein Quantification

Levels of selected proteins were measured by ELISA with commercially available kits (BlueGene Life Science Advance (MUC5AC), R and D Systems, (MMP9 and SP-D), APC Biomaterials LLC (club cell secretory protein), Abcam (Kiniongen, Antithrombin III, Ceruloplasmin, Plasminogen, Prothrombin), MyBioSource Inc (decay-accelerating factor, thioredoxin), AssyPro (Factor 12), Cloud-Clone Corp (Moesin), CusaBio (CD9), MBL International (S100A9) and CusaBio (Ezirin).

RESULTS

Characteristics of study participants

We analyzed BALF samples from 22 unique ARDS patients (Table 2 and Table 3): 7 in the early-phase survivor group (mean ARDS day of sample collection = 2.0 ± 1.15 days), 8 in the early phase non-survivor group (mean ARDS day of sample collection = 3.25 ± 2.19 days), and 7 in the late-phase survivor group (mean ARDS day of sample collection = 18.6 ± 13.3 days). Although the mean age of patients in the early-phase non-survivor group was higher than in the other two groups, this difference was not statistically significant (ANOVA p-value = 0.16). The three groups did not differ in the severity of gas exchange on the day of the bronchoscopy, APACHE-II score, BALF leukocyte count, or BALF neutrophil count. The average time from onset of ARDS to death in the early phase non-survivor group was 19.9 ± 14.5 days.

Proteins identified by peptide spectral matching and database searching

The ProteinPilot PSPEP FDR Summary reported 20,601 spectra matched to 10,355 distinct peptides at ≤ 1% global FDR for a total of 792 inferred proteins (Table S1, Protein Pilot PSPEP summary and protein identified at 1% FDR tab). High abundance proteins or their fragments (Table S1, High abundance / contaminants tab) that were

incompletely removed by the depletion column were removed manually from the protein list. Suspected contaminants or misidentified proteins such as trypsin, bovine albumin, and the reverse matches that occurred from use of the target decoy strategy for peptide identification were also manually removed. After exclusion of these proteins, the number of inferred proteins was 724 (Table S1, BALF proteome tab). These 724 proteins were used as the background for GO enrichment analysis (i.e. the "universe of identified BALF proteome"). Of these 724 proteins, quantitative spectral data were available on 499 to allow determination of the bias corrected relative abundance in the two comparison groups for this study (Table S1, BALF with quantification tab). All but three proteins had at least two peptides used for identification. Bias factors for the two comparison groups were 2.8 for the early phase non-survivors to survivors and 0.64 for early phase survivor to late phase survivor group. Bias factors were used for normalization of the protein quantification within ProteinPilot.

Proteins differentiating early-phase survivors and early-phase non-survivors

Controlling for an FDR of ≤ 5%, we identified 161 proteins that were differentially expressed in the BALF of early-phase survivors compared with early-phase non-survivors (Table S2, proteins with q-values ≤ 5%). Eighty-six of these proteins were more abundant in non-survivors (Table S2, high in non-survivors tab) and 75 were more abundant in survivors (Table S2, high in survivors tab). Gene Ontology enrichment analysis demonstrated significant differences in the biological processes represented by these differentially expressed proteins (Figure 3). The differentially expressed proteins represented six ontologies in survivors (Table 4): three involved in coagulation (fibrinolysis and coagulation and wound healing), two representing cellular ion homeostasis, and one involved in immune activation. In contrast, differentially expressed proteins mapped to three ontologies in non-survivors (Table 5, Table S2, GO non-

survivors tab). Non-survivors showed disruption of bioenergetics with evidence of carbohydrate catabolism and cellular damage as evidenced by disorganization of actin

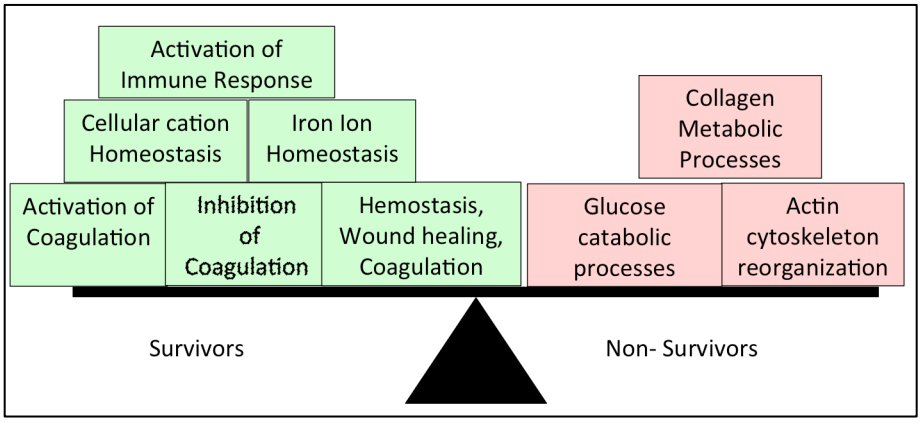

Figure 3: Biological processes represented by 165 proteins that are differentially expressed when early-phase non-survivors are compared to early-phase survivors. *GO enrichment analysis was performed using the universe of identified BALF proteins as a background. A Functional Annotation Clustering tool was used to group related biological processes. Annotation clusters with an enrichment score > 1.3 are shown. In the functional annotation-clustering tool, an enrichment score of 1.3 that corresponds to a non-log scale p-value of 0.05 was used as the cutoff for significance.*

filament based processes. Additionally, there was evidence of collagen biosynthesis in non-survivors early in ARDS.

Changes in the proteome among late-phase survivors

We identified 172 proteins (FDR at ≤ 5%) that demonstrated differential expression between early-phase and late-phase survivors of ARDS (Table S3, FDR q-values tab). Of these 172 proteins, 91 were more abundant in early-phase ARDS survivors (Table S3, high in early phase tab) and 81 were more abundant

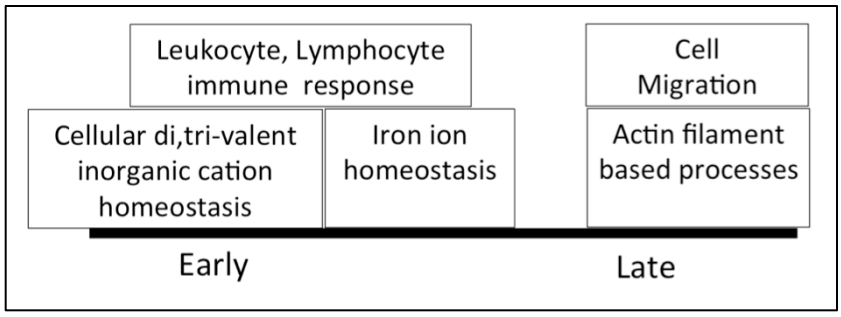

Figure 4: Biological processes represented by 175 proteins that are differentially expressed when late-phase survivors are compared to early-phase survivors.

in late-phase ARDS survivors (Table S3, high in late phase tab). Gene ontology enrichment analysis employing the Functional annotation clustering tool in DAVID identified three ontologies: Lymphocyte and leukocyte immune response, cellular cation homeostasis, and iron ion homeostasis (Table S3, GO early phase tab). In contrast, proteins that were more abundant in late-phase survivors represented two clusters of ontologies involved in lung repair: cell migration and actin cytoskeleton organization (Figure 4 and Table S3, GO late phase tab).

Changes in key proteins concentrations in individual samples

As we used pooled samples for our proteomic studies, we measured protein concentration from individual BALF samples by ELISA. Similar to the MS data, the level of club cell secretory protein was significantly higher in early phase non-survivors when compared to early phase survivors (2458 ± 1409 vs. 922 ± 534 ng/mL, p-value = 0.048, figure 5a). MMP 9 (93.51 ± 133.1 vs. 10 ± 11.87 ng/mL, p-value = 0.19, figure 5b) and Moesin (1.02 ± 0.52 vs. 2.63 ± 1.76 ng/ml, p-value 0.055, Figure 5c) demonstrated a non-significant increase in early phase non-survivors compared to survivors. Although MUC5A was higher in survivors, it did not reach statistical significance (10.74 ± 14.16 vs. 4.499 ±3.48, p-value = 0.29 whereas SP-D was not different

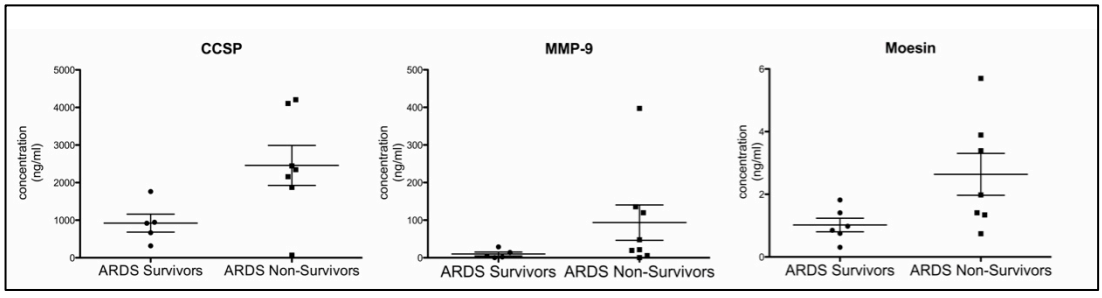

Figure 5: Protein levels of selected candidates. *ELISA was performed to quantify CCSP, Moesin and MMP9. Levels of these proteins were higher in early phase non-survivors in comparison to survivor (p-value < 0.05 t-test) for CCSP and < 0.1 for Moesin.*

To develop a panel of candidate proteins that could discriminate early phase survivors from non-survivors, we measured BALF levels of several key proteins that were higher in survivors and participated in biological processes listed in Table 4. BALF levels of Plasminogen, Factor 12, Antithrombin and Cerulopasmin were consistent with our iTRAQ® MS/MS findings (Figure 6). However, BALF levels of kininogen and Prothrombin (Factor 2) did not mirror the quantitative iTRAQ MS/MS data. We also measured levels of key proteins that participated in biological processes in early phase non-survivors listed in Table 5. S100A9 and Thioredoxin levels measured by ELISA mirrored the iTRAQ data (figure 7). Similar to the iTRAQ® MS data, Ezrin level measured by ELISA in pooled BA was higher in non-survivors compared to survivors (2.45 ng/ml vs. 0.737 ng/ml.) We were not able to detect decay accelerating factor and CD9 by ELISA in our BALF samples.

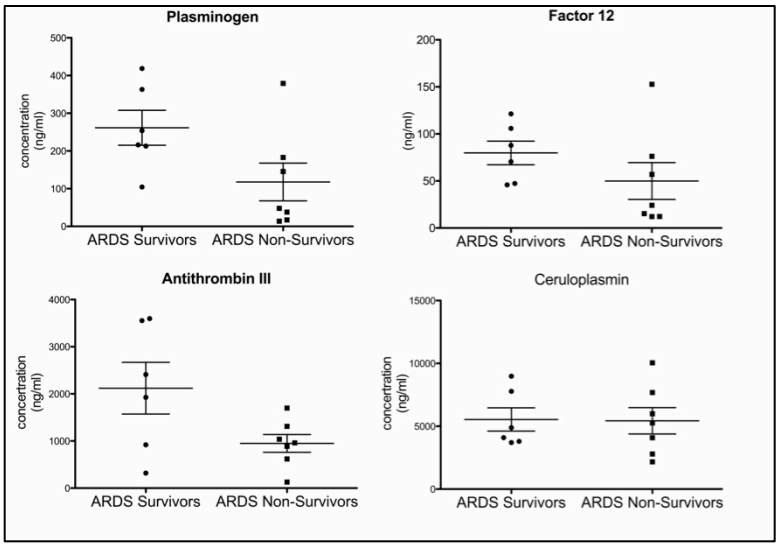

Figure 6 Protein levels of selected proteins that represent biological processes that are activated in early phase survivors (p-value plasminogen = 0.06, antithrombin III = 0.054, factor 12 = 0.2 and ceruloplasmin = 0.9).

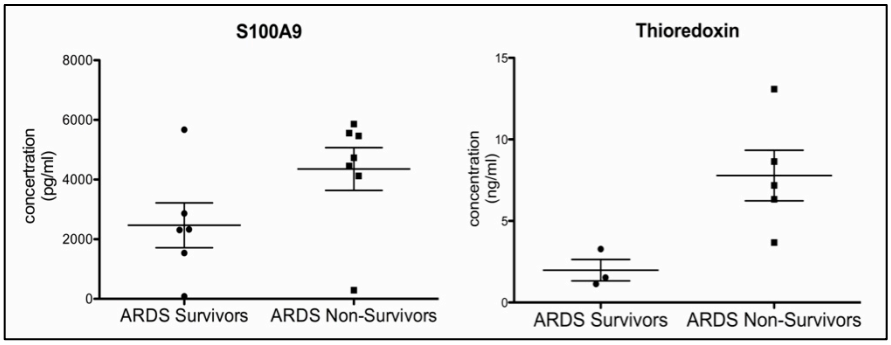

Figure 7: Protein levels of selected proteins that represented biological processes that are activated in early phase non-survivors. *Thioredoxin was identified in only 3 survivors. (p-value < 0.05 thioredoxin and < 0.1 for S-100)*

DISCUSSION

In this study, we achieved a deep coverage of the BALF proteome through the use of high-resolution mass spectrometry-based proteomics and an optimized sample preparation designed to enrich medium and low abundance protein fractions. This extent of proteome coverage has not been previously reported in human BALF from normal or diseased lungs [9, 34, 38, 130-132]. The semi-quantitative techniques used in this study reveal dynamic changes in the distal airspace of patients with ARDS. Differences in the BALF protein expression profile are seen early in the course of ARDS in patients who die compared with those who live. Further, GO enrichment analysis demonstrates highly informative, biologically coherent differences in the ontologies represented by the differentially expressed proteins when early-phase survivors are compared to late-phase survivors or early-phase non-survivors.

The BALF protein expression profile for patients early in ARDS was different comparing survivors to non-survivors. Patients who died had evidence of aberrant lung repair early on in the disease process, as evidenced by approximately a two-fold differential expression of type I, III and V collagen, a signature of activated fibroblasts, and newly synthesized collagen deposition. This is in line with previous reports that higher levels of collagen I and III in ARDS [133] reflect matrix remodeling locally in the lung. Increased levels of BALF type III procollagen also correlated with fatal outcomes in ARDS [134], In contrast, survivors early in the disease course demonstrated a more coordinated response that includes coagulation as the focal point, including plasminogen-mediated fibrinolysis. The balance of activation of coagulation and fibrinolysis is an important determination of the extent of fibrin deposition. In previous studies, tissue factor-mediated pro-coagulant [135], protein C mediated anticoagulant [136], and plasminogen mediated fibrinolysis pathways appeared to be important in ARDS [137]. Up-regulation

of plasminogen activator inhibitor in BALF suggests a shift from a profibrinolytic to an anti-fibrinolytic phenotype being associated with poorer outcomes [138]. In our study, a similar pattern of active collagen deposition suggesting an anti-fibrinolytic milieu even in the exudative phase in ARDS non-survivors is seen while ARDS survivors have a more prominent fibrinolytic milieu. This indicates a profound difference in response to alveolar injury in survivors compared to non-survivors.

Besides abnormal repair, non-survivors also have evidence of increased catabolism and cellular disruptions. In contrast, survivors demonstrate a coordinated activation of cation and iron homeostasis. Prior studies have shown the importance of iron in the development of ARDS [139, 140]. Cell and tissue damage resulting from inflammatory/ oxidative stress can ultimately be a consequence of disruption of normal iron metabolism. Patients with ARDS have increased concentrations of heme and non-heme iron that could lead to the generation of oxidative stress and resultant lung damage [141]. Polymorphisms in ferritin light chain and heme oxygenase have also been associated with increased susceptibility to ARDS [142]. In survivors, higher levels of several proteins involved in iron regulation such as ferritin heavy and light chain, hemopexin, and ceruloplasmin indicate better capacity to counteract the redox stress mediated by iron or other reactive oxygen species in the lungs.

In addition to giving insight into mechanisms of disease, differentially expressed proteins in the early phase of ARDS can be used to discern non-survivors from survivors for prognostication. The ideal biomarker would have biological significance related to lung injury and repair. Alternatively, a panel of proteins that represent the divergent biological processes in the two groups could be selected for testing in a separate cohort of well-phenotyped patients. As the proteomic platform that we used only provides relative

quantitation, complementary studies using multiplex ELISA or multiple reaction monitoring will be needed to measure absolute levels to select a limited number of proteins that could be further investigated. We used ELISA for measurement of protein levels for two main reasons. First, this provided validation that mass spectrometric measures of the protein amounts were accurate. The fold change for CCSP (sp|P11684|UTER_HUMAN, Table S2, high in the non-survivor tab, row 25) was 6.2 fold in the mass spectrometric studies. In line with these findings, the mean levels measured with ELISA demonstrated a > 4.5 fold higher abundance at the protein level. Similarly, other proteins produced in the lung, Moesin, MMP9 and MUC5A, also demonstrated a trend toward higher levels in non-survivors by ELISA and mass spectrometry studies, whereas surfactant D did not demonstrate a significant change when measured in individual BALF samples. These proteins could represent epithelial damage and be candidate proteins to test in a larger cohort of well-phenotyped subjects with ARDS.

CCSP is produced by small airway cells and has been implicated in regulating inflammatory responses in the lung. In patients with ventilator associated pneumonia, serum CCSP levels increased two days before the diagnosis of ARDS/ALI [143]. However, data regarding the utility of plasma CCSP levels is conflicting in small studies with one study demonstrating evidence in CCSP predicting mortality [144], while another study did not find any association of serum CCSP levels with mortality [145]. Our study suggests BALF CCSP levels alone or conjunction with other proteins could be a marker of epithelial damage and could predict mortality in ARDS.

ERM (ezirin-radixin-moesin) proteins co-localize in cell-matrix adhesion sites, filopodia, and membrane protrusions [146]. ERMs function by binding to and organizing the actin cytoskeleton [147] and in turn, stabilizing adherens junctions [148] and influencing cell

migration [149, 150]. In adult wild-type mice, moesin expression is limited to the alveolar epithelium of the distal lung. Moesin-deficient mice develop normally [151] demonstrate decreased moesin in the distal alveolar wall and have airspace enlargement. In response to bleomycin, moesin- deficient mice had lower survival [152] more inflammation, extensive alveolar destruction, hemorrhage and pulmonary edema, increased lung permeability, and a higher total BALF cell count. In moesin-deficient mice, fibrotic response to bleomycin was both earlier and more severe. This supports an involvement of moesin in injury-repair response in the lung.

Matrix metalloproteinase (MMP) are proteases that are involved in degradation of extracellular matrix. Type IV collagen is specific to the basement membrane and MMP-9 is a type IV collagenase. In ARDS, BALF MMP-9 levels were high compared to controls and correlated with the degree of collagen breakdown as determined by measuring collagen breakdown products (7S collagen) [153]. Early elevations of MMP-9 levels have also been found to be associated with prolonged duration of mechanical ventilation in pediatric ARDS patients [154]. In our study, although we did not compare the BALF MMP-9 level in ARDS with controls, higher BALF MMP-9 were seen in patients who died. Though speculative, this could be a marker of worse epithelial damage in non-survivors. We also measured MUC5A levels in the BALF by ELISA. MUC5A is a member of the mucins, large glycoproteins that form a protective biofilm covering the respiratory epithelial lining. MUC5AC is secreted mostly by the surface epithelial goblet cells [155]. MUC5AC transcript levels increase in airway epithelial cells upon the cyclic stretch, in mice with ventilator-induced lung injury and humans with ARDS [155]. Though the proteins levels of MUC5AC in our studies did not differentiate survivors and non-survivors, this may have resulted from our relatively small sample.

In this study we also compared the BALF protein expression profiles of early- and late-phase survivors. This comparison highlights the dynamic changes in the airspace milieu during repair. Chang and colleagues [19] characterized BALF in ARDS patients on day 1, 3, and 7; their results demonstrated striking differences between normal controls and ARDS patients on day 1, but less dramatic changes between days 1, 3 and 7. The changes seen in their studies reflected the alteration in the innate immune and oxidant pathways at day 3 and possibly lung regeneration at day 7. Similar to that study, we found that activation of the innate immune system and cation homeostasis were over-represented by proteins in early-phase survivors compared to late-phase survivors. However, in our late-phase survivors – whose samples were obtained 18 days after the onset of ARDS, much later than in previous studies – the ontologies were drastically different and included cell migration and actin cytoskeleton organization. Since all of these patients survived, these findings suggest a critical role of these processes in lung repair. The proteins that are represented in these ontologies could be potential targets to stimulate repair mechanisms as potential molecular targets for therapy in ARDS.

We acknowledge the limitations of our study. We were limited by the amount of available BALF. Therefore, the protein characterization was performed on pooled BALF samples. Pooled samples are subject to influence by a minority of outliers within the pool. We were limited to a single run for the mass-spectrometry. However, we demonstrate changes in protein levels using ELISA studies that were performed on individual subject samples. Another limitation is the study design, i.e. binary outcome, which can be influenced by some confounders not controlled in our study. Our approach was to enrich the medium and low abundance protein fractions by depleting high abundance proteins. We chose depletion over alternative method to avoid an unwieldy dynamic range with a subsequently limited depth of proteome coverage. Also, many of the abundant proteins

that were eliminated are found in the plasma, which can leak into the alveolar space during lung injury. Another limitation is our relatively small number of subjects. Despite the small sample size, our sample preparation optimization methods enabled us to identify some differentially expressed proteins successfully. Our findings provide a starting point for subsequent studies characterizing BALF in individual patients for biomarker identification in ARDS.

CONCLUSION

This study illustrates a framework whereby protein profiling can be used to identify panels of proteins that parallel the pathophysiological changes occurring in ARDS. We demonstrate dynamic changes in BALF protein expression during ARDS and also an early divergence in the protein expression profile in ARDS. Differences in absolute levels of the proteins that represent divergent biological processes in survivors and non-survivors will facilitate identification of prognostic biomarkers in ARDS.

Tables for Chapter 3

Table 2: Demographic and Clinical Characteristics of Subjects by Study Group				
Variable	**Early-phase ARDS survivors (N=7)**	**Early-phase ARDS non-survivors (N=8)**	**Late phase ARDS survivors (N=7)**	**p-value***
Age	42.29± 11.43	58.13±20.49	47.86± 10.07	0.16
Sex	M=5, F=2	M=6, F=2	M=5, F=2	
APACHE-II score	19.14 ± 7.4	19.75 ± 4.7		0.85
ARDS Day of BALF collection	2.0 ± 1.15	3.25 ± 2.19	18.6 ± 13.3	<0.001#
PF Ratio on day of bronchoscopy	143.7 ± 34.1	150 ± 71.1	161.5 ±83.7	0.137
BALF WBC count (cells/µl)	496 ± 342.6	364.4 ± 408.7	451.1 ± 471.5	0.75
BALF Neutrophils (%)	56.0 ±33.3	49.0 ± 38.1	40.6 ± 38.5	0.9327

One-way analysis of variance (ANOVA) with Tukey post-test
Statistically significant difference between late-phase survivors and early-phase survivors ($p < 0.05$) and between late-phase survivors and early-phase non-survivors ($p < 0.05$), but no difference between early-phase survivors and early-phase non-survivors.
PF ratio- PaO_2 to FiO_2 ratio

Table 3: Pulmonary history and clinical risk factors for ARDS in the study subjects			
Past Pulmonary History	**Early-phase ARDS survivors (n=7)**	**Early-phase ARDS Non-survivors (n=8)**	**Late phase ARDS survivors (n=7)**
None	3	6	1
VTE	1	0	0
Smoker (prior or current)	1	0	3
Lung infection	2	0	0
COPD	0	0	1
NSCLC	0	2	1
Prior ARDS	0	0	1
Risk factor for ARDS			
Disseminated candidiasis	1	0	0
Sepsis	4	2	0
Pneumonia, not specified	0	1	
Pneumonia, gram negative	0	1	2
Pneumonia, gram positive	0	2	0
Pneumonia, fungal	0	1	0
Pneumonia, viral	0	0	1
Pneumonia, aspiration	2	0	3
Pancreatitis	0	1	0
Unknown	0	0	1

Table 4: Early-phase ARDS Survivor Ontology Groups and Associated Proteins

GO Biological Process	Official Gene Symbol	Protein Name	Fold change*
Positive Regulation of blood coagulation	AHSG	Alpha-2-HS-glycoprotein	0.21
	APOH	Apolipoprotein H	0.21
	HRG	Histidine-rich glycoprotein	0.34
	PLG	Plasminogen	0.35
	F12	Coagulation factor XII	0.36
	F2	Coagulation factor II	0.57
	SERPINF2	Serpin peptidase inhibitor, member 2	0.38
Negative regulation of blood coagulation	AHSG	Alpha-2-HS-glycoprotein	0.21
	APOH	Apolipoprotein H	0.21
	KNG1	Kininogen 1	0.24
	PLG	Plasminogen	0.35
	F12	Coagulation factor XII	0.36
	APOE	Apolipoprotein E	0.44
	ANXA5	Annexin A5	0.45
	F2	Coagulation factor II	0.57
	ANXA2	Annexin A2	0.81
Regulation of body fluid levels	SERPINC1	Antithrombin III	0.16
	APOH	Apolipoprotein H	0.21
	KNG1	Kininogen 1	0.24
	PLG	Plasminogen	0.35
	F12	Coagulation factor XII	0.36
	ANXA5	Annexin A5	0.45
	F2	Coagulation factor II	0.57
	ANXA2	Annexin A2	0.81
Cellular cation homeostasis	KNG1	Kininogen 1	0.24
	HPX	Hemopexin	0.30
	SFTPD	Surfactant protein D	0.38
	APOE	Apolipoprotein E	0.44
	F2	Coagulation factor II	0.57

	FTL	Ferritin, light polypeptide	0.62
	FTH1	Ferritin, heavy polypeptide 1	0.62
	CD 55	Decay accelerating factor	0.78
	CP	Ceruloplasmin	0.84
Iron ion homeostasis	HPX	Hemopexin	0.30
	FTH1	Ferritin, heavy polypeptide 1	0.62
	FTL	Ferritin, light polypeptide	0.62
	CP	Ceruloplasmin	0.84
Positive regulation of immune response	C4BPA	Complement component 4 binding protein, alpha	0.09
	PLG	Plasminogen	0.35
	F12	Coagulation factor XII	0.36
	CFH	Complement factor H-related 2	0.45
	C1RL	Complement component 1r	0.56
	F2	Coagulation factor II	0.57
	CLU	Histone cluster 1	0.65
	C5	Complement component 5	0.66
	KRT1	Keratin 1	0.67
	CD55	Decay accelerating factor	0.78
	C8A	Complement component 8, alpha	0.79
	C6	Complement component 6	0.80
	APOH	Apolipoprotein H	0.21

*Fold change is relative to survivors, therefore a fold change < 1represents proteins more abundant in survivors

Table 5: Early-phase ARDS Non-Survivor Ontology Groups and Associated Proteins

GO Biological Process	Official Gene Symbol	Protein Name	Fold change
Actin filament-based process	TMSB4X	Thymosin-like 2	2.65
	EZR	Ezrin	2.15
	PFN1	Profilin 1	1.93
	VASP	Vasodilator-stimulated phosphoprotein	1.81
	CAP1	Adenylate cyclase-associated protein 1	1.58
	ARHGDIB	Rho GDP dissociation inhibitor (GDI) beta	1.52
	S100A9	S100 calcium binding protein A9	1.51
	FLNA	Filamin A, alpha	1.31
	MYH9	Non-muscle myosin, heavy chain 9	1.27
	STMN1	Stathmin 1	1.20
Glycolysis	GAPDHL6	Glyceraldehyde-3-phosphate dehydrogenase-like 6	2.01
	TXN	Thioredoxin	1.82
	PGK1	Phosphoglycerate kinase 1	1.73
	TPI1	Triosephosphate isomerase 1	1.63
	GPI	Glucose phosphate isomerase	1.63
	ENO1	Enolase 1, (alpha)	1.51
	PGAM1	Phosphoglycerate mutase 1	1.42
Collagen metabolic process	COL5A1	Type V collagen, alpha 1	2.33
	MUC5AC	Mucin 5AC	2.04
	COL3A1	Type III collagen alpha 1	1.99
	MMP9	Matrix metallopeptidase 9	1.84
	COL1A1	Type I collagen, alpha 1	1.75

Chapter 4

Protein Expression Profile in Acute Respiratory Distress Syndrome: Analyzing Individual Cases

Manuscript prepared for submission to American Journal of Respiratory and Critical Care Medicine

Acute Respiratory Distress Syndrome (ARDS) continues to have a high mortality of around 30–40% [95, 156, 157]. Some clinical studies have been undertaken to investigate prognostic biomarkers in ARDS. The markers studied arise from various tissue components of the lung such as type 2 alveolar epithelial cells [115, 158, 159], type 1 alveolar epithelial cells [110, 112, 160], lung matrix [108] and vascular endothelial cells [117, 121]. Additionally, proteins participating in coagulation/fibrinolysis [136, 138, 161], pro-inflammatory [119, 162-167] and anti-inflammatory [119, 120, 168] mediators have also been reported to have prognostic value. Several investigators have also focused on the predictive accuracy of a panel of markers [161, 169]. Combining clinical risk factors with eight biologic markers that included vFW, SP-D, TNFR I, IL-6, IL-8, ICAM-1, protein-C, and PAI-1 are associated with the outcome of ARDS [169]. However, the biomarkers studied had a fair degree of colinearity thus containing redundant information and adding little additional predictive value to clinical predictors alone. Search for orthogonal (uncorrelated) biomarkers either through unbiased discovery experiments or targeted examination of novel pathways (including those identified by genetic association studies) has been recommended to develop a multimarker predictor in ARDS [90, 170].

The unbiased discovery by protein expression profiling technologies has been undertaken in ARDS to understand the disease biology and also to identify orthogonal candidate biomarker proteins [171, 172]. Studies using 2-dimensional electrophoresis showed increased edema fluid albumin, transferrin, IgG, and clusterin [73] and lower BALF levels of surfactant protein -A and α1-antitrypsin in ARDS subjects compared to controls. BALF from subjects challenged with endotoxin or ARDS showed similar protein expression profile with higher levels of apolipoprotein A1, S100 A-8 and A9 compared to healthy non-smoking controls [97]. A study characterizing serial changes at day 1, 3 and

7 of ARDS demonstrated differential expression over time of 22 proteins. These proteins were involved in inflammation, response to microbials and response to stress/injury. Compared to controls, at day 1 of ARDS, higher levels of complement proteins, annexin A3, S100, antiproteases, actin and extracellular matrix proteins were seen in BALF. In contrast, the BALF levels of surfactant proteins, annexin A1, fibrinogen, and fatty acid binding proteins were lower on day 1 of ARDS compared to controls. By day seven, there was the regeneration of lung epithelium, decreased cellular injury and resolution of lung injury [38]. These findings suggest that the dynamic changes that occur in the BALF fluid reflect the underlying pathophysiologic mechanisms involved in the resolution of lung damage in ARDS. We have performed BALF studies using high-resolution label based LC-MS/MS on pooled BALF and demonstrated early differences within one week of initiation of mechanical ventilation- in proteins expression profile in ARDS survivors compared to non-survivors [36].

The primary aim of this study is to extend the findings from our prior work where we demonstrated differences in pooled BALF from ARDS survivor and non-survivors. In this study, we characterize the protein expression profile in individual ARDS cases. We perform comprehensive label based quantitative protein profiling to identify the proteins and biological process that differentiate ARDS survivors from non-survivors.

METHODS

Study population

The University of Minnesota Institutional Review Board (IRB) Human Subjects Committee approved this study. Cases with ARDS at University of Minnesota Medical Center, who had a clinically indicated bronchoscopy and excess BALF available, were included in this study. All control subjects (except one) had respiratory failure but did not

meet AECC criteria for ARDS [93]. The study was designed prior to the publication of the Berlin definition. Only exclusions for the study included a history of HIV or viral hepatitis. All BALF samples were collected from bronchoscopies done in a standard protocol [36, 173]. The BALF samples were placed on ice for transportation and within 60 minutes of collection centrifuged at 500g at 4°C for 10 minutes. The cell-free supernatant was aliquoted into microcentrifuge tubes and stored at -80°C.

We divided the ARDS subjects into two groups based on the ultimate outcome at the time of hospital discharge- survivors and non-survivors. A global internal standard consisting of pooled BALF from 27 controls (pooled mastermix) was used to compare relative protein abundance across multiple isobaric tagging for relative and absolute quantification- two-dimensional liquid chromatography-tandem mass spectrometry (iTRAQ 2DLC-MS/MS) experiments. Differences in the clinical characteristics of the three participant groups were tested by Kruskal-Wallis One Way ANOVA with post hoc Dunn test or Chi-square for proportions when appropriate

Sample processing for protein profiling

Cell and the debris free supernatant were stored at -80°C and did not undergo any freeze-thaw cycles until sample processing. BALF containing at least 8 mg of protein (Bradford reagent, Bio-Rad cat#500-0006) was processed for LC-MS/MS from individual patient employing a protocol previously published with minor modifications [36]. Briefly, the BALF was concentrated and desalted using Amicon 3-MWCO filters, depleted of high abundance proteins with appropriate buffer exchanges for labeling with iTRAQ reagent

iTRAQ labeling and 2D LC-MS/MS

The enriched medium and low abundance protein fractions from BALF was digested with trypsin-gold (Promega cat#V5280), dried and suspended in 0.1- 0.2% formic acid (pH < 3) and MCX cation exchange (Oasis MCX Cartridge, Waters, Milford, MA, Cat no 186000254) was performed to remove SDS. Peptides were eluted, and adequacy of the trypsin digestion was confirmed by the analysis of 3µg of the tryptic-digested peptides with linear trap quadrupole MS (LTQ-MS). The remaining 37 µg of the peptide mixture was labeled with eight-plex iTRAQ reagent per manufacturer instructions (AB Sciex, Framingham, MA) as done previously [36, 43]. To compare the protein abundance across the different LC-MS/MS experiment, we used the same pooled mastermix as a global internal standard and as a reference for determination of relative protein abundance. In each LC-MS/MS experiment, two iTRAQ reporter ion channels were labeled with the mastermix to assure the accuracy of fold changes measurement while the remaining six channels contained study samples. To characterize thirty-six ARDS cases in the study, we performed six separate iTRAQ LC-MS/MS experiments. The labeling strategy for the 36 BALF samples studied is outlined in Table 6. To prevent reporter ion signal (channel) bias the pool- mastermix, ARDS survivors, and ARDS non-survivors samples were randomly placed in different iTRAQ reporter ion channels for each separate experiment.

iTRAQ-reagent labeled peptide mixture, was purified with an MCX Oasis cartridge before off-line peptide separation in the first dimension and if needed C18 stage tipping. Peptides were separated offline into 15 separate peptide-containing fractions collected in 2-minute intervals on a C18 Gemini column (Phenomenex, Torrance, CA) at pH 10. Peptide fractions were concentrated, purified by the stage tip procedure [182] with Empore SDB-RPS extraction disks [mixed mode strong cation exchange and reversed phase], 3M (St. Paul MN), and separated in the 2nd dimension by online C18 reversed-

phase capillary LC with a nano LC system (Eksigent, Dublin, CA). Data-dependent acquisition was performed on an Orbitrap Velos system with HCD (higher-energy collision induced dissociation) activation for peptide tandem MS. LC and MS experimental details were previously reported, with the exception that the activation time was 20 msec [36, 174].

Database search for protein identification and quantification

.RAW files obtained directly from the Orbitrap Velos XL Mass Spectrometer were imported into GalaxyP (https://usegalaxyp.org/ for public instance) for further processing (as described in z.umn.edu/ppingp). Within GalaxyP, all .RAW files were converted to mzml format using msconvert and then into a ProteinPilot compatible Mascot Generic Format (MGF) files with preselected iTRAQ reporter ions. The MGF files were searched against the target-decoy version of Human UniProt database along with contaminant protein sequences (88,304 sequences in total; Date Aug 1, 2014) using ProteinPilot version 4.5 and the following search parameters: Sample Type: iTRAQ 8-plex (peptide labeled); Cys-alkylation: MMTS; Instrument: Orbi MS, Orbi MS/MS; run quant; bias correction on; search focus on biological modifications and amino-acid substitutions; thorough search and with a detected protein threshold (Unused Protscore (Conf)): 10%. The ProteinPilot searches and subsequent generation of Proteomics System Performance Evaluation Pipeline Software (PSPEP) - FDR reports and protein and peptide level summaries were generated within Galaxy-P as previously described [126]. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium [127] via the PRIDE partner repository with the dataset identifier PXD002672.

The results of multiple iTRAQ LC-MS/MS experiments were aligned to compare protein levels using Protein Alignment Template vs. 2.00p (AB Sciex) [175]. The Protein Alignment Template provides the overlap of the proteins across multiple iTRAQ LC-MS/MS experiments by matching proteins in a 'reference master list' to the 'test list' of proteins identified in each individual iTRAQ LC-MS/MS run. For the alignment, we created a 'reference master list' by performing a database search using .RAW files from all six iTRAQ LC-MS/MS experiments. To ensure that the proteins in this list are high ID quality, a local FDR ≤ 5% was used as a threshold for the reference master list. As per the recommendation of the Protein Alignment Template, for the creation of feature table with quantitative values, the threshold of ≤ 5% global FDR was used for individual sets of the six-iTRAQ LC-MS/MS experiments. Protein Alignment Template resulted in aligning the ratios, p-values and error factors for the proteins across all six iTRAQ LC-MS/MS experiments by using accession numbers of isoforms within protein summary and UniProt database.

Statistical analysis

Identification of differentially expressed proteins between ARDS survivors and non-survivors: We performed inverse variance weighted ratio test to account for the peptide level variance in fold changes measured for each protein across multiple iTRAQ runs. Specifically, denote Xi as the log protein ratio for one group and Si its corresponding variance; denote Yj as the log protein ratio of another group and Vi its corresponding variance. Define the following Z-test statistic

$$Z = \frac{\sum_i X_i/S_i - \sum_j Y_j/V_j}{\sqrt{\sum_i 1/S_i + \sum_j 1/V_j}}$$

We can compare Z to standard normal distribution to compute significant p-values. Thus using a traditional α of 0.05 is not optimal. We controlled for multiple comparisons by FDR [176] corrected p-value ≤0.05 as done previously [36].

Identification of proteins that predict outcomes:

To identify the proteins that predict ARDS outcomes, we used leave-one-out cross-validation (LOO-CV) with logistic regression model for the survivorship status. Each time we leave out one observation (protein ratios from an ARDS case) as a testing sample, we select significant proteins and fit logistic regression models using all the other observations. The fitted logistic regression model estimates the surviving probability for the testing sample. This calculation was repeated for all proteins. This analysis was performed using functions built into MATLAB statistical computing environment (such as cvpartition and lassoglm).

Computational Analysis

To gain insight into the biological significance of differentially expressed proteins, we used the Database for Annotation, Visualization, and Integrated Discovery (DAVID, http://david.abcc.ncifcrf.gov) [129] as previously done [36, 43] using 'high stringency'. We focused on only GO biological processes and the molecular function. In addition, we also performed functional analysis using Ingenuity Pathway Analysis (IPA® QIAGEN, Redwood City www.quiagen.com/ingenuty Build 321510M, Version 21249400) and focused on the protein subsets represented in canonical pathways (signaling and metabolic) and also diseases and functions. This analysis was performed on proteins passing the FDR threshold of ≤ 5% for differential expression between ARDS survivors and non-survivors. We used the 'the universe of BALF proteins' identified in the six

iTRAQ LC-MS/MS runs as the background for the functional analysis. This was the same as the 'reference master list' used in the protein alignment template.

RESULTS

Characteristics of study subjects

We characterized BALF samples from 20 cases of ARDS who survived and 16 cases that died during that hospital stay. BALF from 27 subjects was used as controls. The controls were older than the ARDS survivors. No statistically significant difference in age was present between controls and ARDS non-survivors or ARDS survivors and non-survivors. No difference was present in the PO_2:FiO_2 ratio in ARDS survivors and non-survivors. Controls had a higher PO_2: FiO_2 ratio compared to the ARDS cases. In the three groups, there was no difference in the time from ARDS onset to BALF collection, the BALF leukocytes, neutrophil and lymphocyte counts (Table 7).

Protein identified in the BALF.

The Protein-Pilot PSPEP FDR summary showing the number of spectra, peptides and the proteins identified at 1% global FDR for the six-iTRAQ experiments is shown in Table 8. The number of proteins identified in each of the six iTRAQ LC- MS/MS run at a global FDR of ≤ 1%, were 850, 606, 1055, 865, 976 and 879 (Table S4). To identify the universe of the BALF protein, we also performed a database search using the .RAW files from all six LC-MS/MS experiments. This resulted in the identification of 1189 proteins at a local FDR of ≤ 5%. However, this included misidentified proteins, contaminants and proteins that were not completely removed by the IgY 14 depletion column. After manually removing these proteins, 1122 proteins were used for further analysis (Table S5).

Differentially expressed proteins in ARDS survivors vs. non-survivors

The variance-weighted t-test comparing the protein abundance between ARDS survivors and non-survivors identified 249 differentially expressed proteins (corrected p-value ≤ 0.05). These proteins are listed in Table S6 (supplemental data). Proteins with the greatest difference in abundance between ARDS survivors and non-survivors are shown in Table 9.

Biological relevance of the proteins that are differentially expressed

We performed disease and function analysis in IPA to identify the molecular/cellular

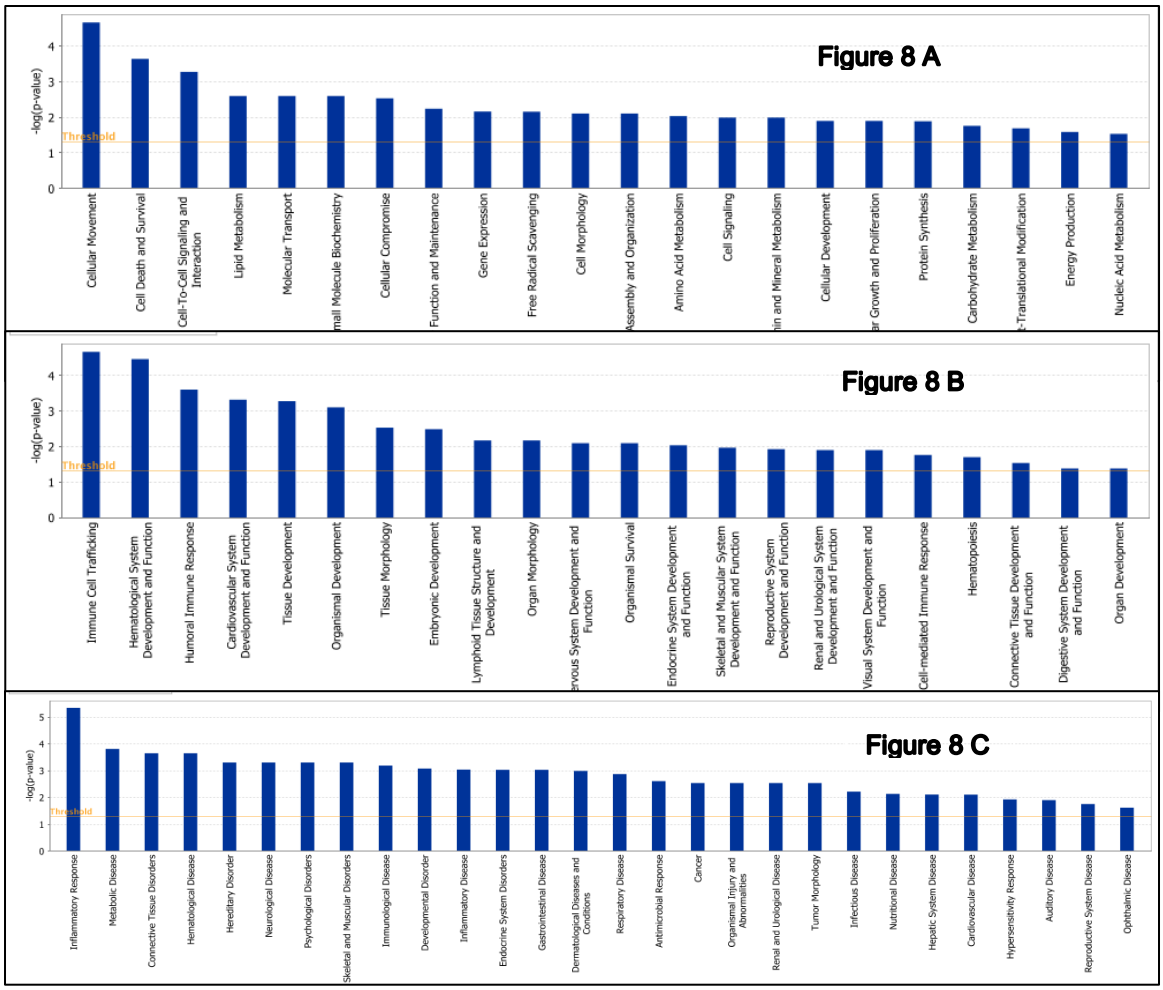

Figure 8: Functional analysis performed in IPA to determine the biological relevance of the 249 proteins differentially expressed between ARDS survivors and non-survivors*. The molecular and cellular function (Figure 8A), Physiological system development and function (Figure 8B) and the diseases and disorders (Figure 8C) significantly overrepresented by these proteins using a Fisher exact –log(p-value) > 1.3 are shown.*

functions, physiological system development and function and disease and disorders that are represented by the proteins that show differential expression in ARDS survivors and non-survivors (Figure 8). These proteins represented 22 molecular and cellular functions (Figure 8A), 22 physiological system development and functions (Figure 8B) and 28 diseases and disorders (Figure 8C) with a Fisher exact –log (p-value) of >1.3. The proteins represented in these processes are included in the supplemental table S7. On using a more stringent Benjamini and Hochberg corrected, –log (p-value) four processes remained significant. These were cellular movement (molecular/cellular function), immune cell trafficking and hematological system development and function (physiological system development and function) and inflammatory response (disease and disorder).

Complementary analysis performed using functional annotation clustering in DAVID identified two 'molecular function' clusters (Table 10) and three 'biological processes' clusters (Table 11) that passed the threshold of an enrichment score of >1.3 for statistical significance. Several other clusters involved in interesting and previously identified biological functions such as coagulation, fibrinolysis and cell motility were also identified but did not reach statistical significance. All the clusters identified are shown in Table S8. The annotation groups and the assigned proteins that passed the statistical threshold are shown in Table 10 and 11.

The canonical pathways are signaling and metabolic pathways that are represented by the gene or protein subsets. Three canonical pathways; Acute Phase Response Signaling, Complement System Activation, and LXR/RXR Activation were identified to be significantly enriched using IPA analysis (Figure 9). The proteins that were assigned to these pathways are shown in Table 12.

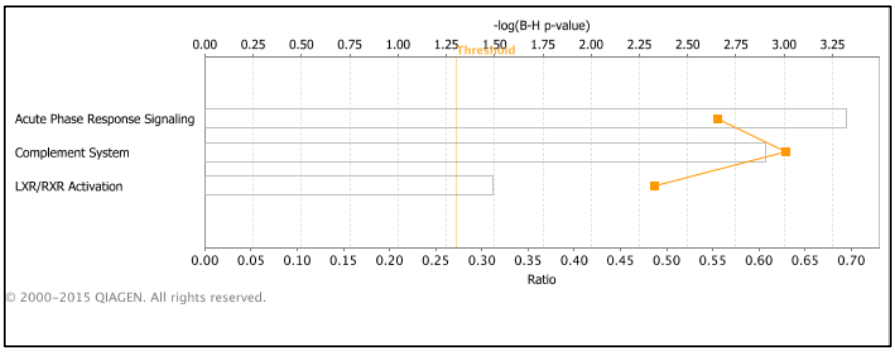

Figure 9: The canonical pathways that were enriched by differentially expressed proteins when survivors were compared to non-survivors*. The x axis displays the –log of the p-value calculated by the Fisher's exact right tailed test with the Bejamini and Hochberg correction. The orange point connected by thin line represents the ratio calculated by number of proteins in a given pathway that meet our cutoff criteria, divided by the total number of proteins that make up that pathway and that are in the BAL universe of proteins.*

Proteins that predict ARDS outcomes: The regression model identified 20 proteins that predict survivorship status. These proteins are shown in Table 13. These proteins either individually or as a panel could be tested as a biosignature to predict outcomes in ARDS.

DISCUSSION

In this study, we expand on our prior work where we used pooled BALF from ARDS survivors and non-survivors to demonstrate differences in the biological processes that are activated in response to lung injury. Here we characterize individual cases of ARDS and again see differences in acute inflammatory response and collagen metabolic processes between ARDS survivors and non-survivors. Additionally, cellular movement is also identified as an IPA 'disease and function' that differs between survivors and non-survivors. We also determined the GO molecular functions of the proteins with the difference in their levels between survivors and non-survivors. The molecular functions

assigned to a subset of these proteins include serine hydrolase activity and serine peptidase inhibitor activity. Our regression model identifies proteins that predict outcomes early in the course of ARDS. Additionally, we show the feasibility of performing high-resolution BALF protein expression profiling and statistical analysis in individual subjects with ARDS.

In our study, we found a cluster of proteins with GO molecular function annotation of serine hydrolase activity to be different between survivors and non-survivors of ARDS. Proteases are enzymes that are capable of cleaving and inactivating small peptides. Serine proteases are a family of enzymes that catalyze the hydrolysis of peptide bonds. In mammalian systems, serine proteases perform an important function in digestion, blood clotting and the complement system, and several activated clotting factors and complement factors have the protease activity. Activated neutrophils are an important source of proteolytic enzymes. We identified proteins that participate in clotting, complement cascade enzymes and mediators released by neutrophils in the cluster with serine hydrolase activity. Though these enzymes are involved in the innate immune response, in the lung, an unregulated release of neutrophilic mediators can paradoxically damage the lung epithelium and endothelium [177]. In addition to the release of neutrophil mediators, microbial proteins result in activation of complement cascade ultimately resulting in microbial killing but also the perpetuation of inflammation and direct tissue injury [178, 179]. Prior work has shown coagulation proteases to be important in ARDS and inflammatory disorders [180] likely by proteinase-activated receptors (PARS) mediated signaling [181]. Thus, the presence of proteases in the distal lung space contributes to tissue damage. Our data suggests the extracellular proteases could lead to changes in permeability resulting in gas exchange abnormalities seen in ARDS.

Several mechanisms that counteract the unregulated proteolytic activity also exist. In our study, the most significant cluster of proteins had GO term of 'serine – type endopeptidase inhibitor activity'. The proteins that were present in this cluster included SERPIN superfamily proteins, inter- α-trypsin Inhibitors (ITI), annexins and other proteins that are involved in inhibition of proteolytic activity of chymotrypsin, kallikrein, coagulation factors, cathepsins, complements and phospholipases. Serpins (serine peptidase inhibitors) make up the major source of peptidase inhibitors in the lung. Thirty-six human protein-coding serpin genes have been identified and are divided into clades according to their sequence homology [182]. Although most serpins inhibit serine proteases, some inhibit cysteine proteases, and some do not have protease inhibitory properties (such as thyroid/cortisol binding globulin). Several serpins have been implicated in lung diseases such as emphysema (SERPIN A1- α1-antirypsin), sepsis / ARDS (SERPIN C1-antitrombin III, SERPIN E1-plaminogen activator inhibitor), asthma (SERPIN E1) and idiopathic pulmonary fibrosis (SERPIN E1) [183]. We identified three serpins clade A members with inhibitory activity- serpin A3: α1 chymotrypsin, A4: kallistatin and A10: protein z dependent protease inhibitor- in our study. Other extracellular serpins that were identified included serpin C1 (antithrombin III), serpin D1 (heparin cofactor 2), and serpin G1 (plasma protease C1 inhibitor) in our studies. Serpin C1 and D1 inhibit coagulation factor II, IX, and X while serpin G1 acts in the coagulation cascade. We also identified serpin clade B members that typically are intracellular serine and cysteine peptidase inhibitors, in our studies, Serpin B1 (leukocyte elastase inhibitor), serpin B3 and serpin B6 were identified in the BALF. Though the mechanism by which intracellular serpins end up in the BALF is not clear but it is likely that they contribute to inhibition of proteolysis and protect lung damage due to exuberant proteolytic activity. Besides serpins, we also identified inter-α-trypsin inhibitors (ITI) in

the BALF. ITI is a family of proteins with a light chain with protease inhibitor activity and two heavy chains by which it interacts with hyaluronic acid [184]. The light chain also called as bikunin inhibits trypsin, chymotrypsin, plasmin, granulocyte elastase, and cathepsins. ITI levels in serum of septic patients are lower than healthy controls and are inversely related to mortality [185]. ITI inhibits the activation of both classical and alternate complement pathway and reduces lung injury in mice [186]. In our studies, the levels of all four isoforms of ITI were higher in survivors compared to non-survivors, suggesting a protective role of ITI in ARDS. These findings taken together the proteolytic vs. anti-proteolytic milieu in the distal lung spaces appears to be important in determining the extent of lung injury and could predict outcomes in ARDS.

Similar to our study performed on pooled BALF from ARDS survivors and non-survivors; in this study we found differences in collagen metabolic processes. Here we found survivors had lower levels of MMP8, MMP9 and myeloblastin suggesting less collagen breakdown and also the lesser level of alpha type-1-collagen in ARDS survivors. Additionally there were higher levels of tenascin in ARDS survivors. High levels of tenascin are seen transiently in response to local injury [187] and down-regulation occurs when repair or scarring is accomplished [188]. Persistently increased levels of tenascin-C are associated with extracellular matrix deposition and fibrotic diseases such as pulmonary fibrosis [189]. These findings suggest an early difference in response to injury in ARDS survivors and non-survivors with increased matrix breakdown and collagen deposition in non-survivors. The role of tenascin in ARDS has not been explored, and its utility as a marker for persistent fibroproliferation also need characterization. In our previous study with pooled BALF coagulation / fibrinolysis also differed between survivors and non-survivors. Though we identified cluster with

annotation of fibrinolysis/wound healing in the current study, it did not meet statistical significance but appears to be important in lung injury/repair.

Similar to our studies in pooled BALF, in this study analysis of differentially expressed proteins identified 'IPA disease and function' of cellular movement to differ in survivors and non-survivors. ERM proteins (ezrin, radixin, and moesin) were again identified in the current study. Additionally, we identified CD9, a tetraspanin, also called as motility inducing protein 1 to be more abundant in ARDS non-survivors. We have shown CD9 to be expressed in primary alveolar epithelial cells, and higher levels of CD9 occur during recovery from in vivo hyperoxia-induced lung damage [43]. The mechanism by which higher levels of CD9 occurs in BALF in ARDS non-survivors are not known, but it could be due to either AT2 cell shedding or leakage from injured cells. This could be important as AT2 cell CD9 could contribute to their migration on a denuded basement membrane, an important factor in lung repair. If in non-survivors, high BALF CD9 correlates with low AT2 cell levels, this could provide a useful target to test and stimulate lung repair.

The study has limitations. We elected to deplete high abundance proteins to enrich medium and low abundance proteins. Depletion of high abundance proteins is an important issue in all proteomic studies, especially when protein quantification is performed. This methodology is important in data dependent MS acquisition where only selected MS1 precursor ions are selected for fragmentation. As high abundance proteins can preclude quantification of less abundance MS1 precursor ions, analyzing undepleted BALF could severely limit the findings of a study. The downside of depletion of higher abundance protein is that other proteins that bind to high abundance proteins might be co-depleted. It also increases the amount of sample handling needed before MS data acquisition. In this study, despite using high-resolution MS platforms, the

overlap of the proteins identified in the six iTRAQ runs was partial. We expect that with improvement in MS platforms, the depth and overlap in protein identified across multiple LC-MS/MS experiments will improve. More recent MS approaches, using data-independent acquisition (such as SWATH-MS)[99], also will directly address the limitation of MS platforms being employed for contemporary proteomic studies. We also acknowledge that serum proteins likely contaminated the BALF proteome to some extent in our study. Though this could be addressed by direct comparison of comprehensive protein expression changes in serum and BALF, BALF analysis is a critical first step that provides a framework for future studies of lung dysfunction in HSCT recipients.

Conclusions

The bronchoalveolar lavage fluid from patients with ARDS demonstrates early differences between survivors and non-survivors. These differences are likely due to the differences in the biological processes that are active in the ARDS patients who live compared to those who die from the disease. Similar to our prior studies where we characterized pooled BALF, characterization of individual ARDS cases showed that markers that suggest collagen deposition are different in ARDS survivors and non-survivors. Also, the differentially expressed proteins represented cell migration. The proteins that differ have molecular function annotation of serine hydrolases and serine peptidase inhibitors. A functional assay measuring the inhibitors of proteolysis rather than the use of protein's levels could provide prognostic information in ARDS. Future biomarker studies to determine the prognosis of ARDS should, therefore, investigate groups of proteins representing biological functions rather than proteins levels in the BALF.

Table 6: Labeling strategy for the iTRAQ experiments								
	113	114	115	116	117	118	119	121
iTRAQ-1	Control	Survivor 1	Survivor 2	Control	Survivor 3	Non-survivor 1	Non-survivor 2	Non-survivor 3
iTRAQ-2	Survivor 4	Control	Survivor 5	Survivor 6	Control	Non-survivor 4	Non-survivor 5	Non-survivor 6
iTRAQ-3	Survivor 7	Survivor 8	Control	Survivor 9	Non-survivor 7	Control	Non-survivor 8	Non-survivor 9
iTRAQ-4	Survivor 10	Survivor 11	Survivor 12	Control	Non-Survivor 10	Non-survivor 11	Control	Non-survivor 12
iTRAQ-5	Survivor 13	Survivor 14	Survivor 15	Non-survivor 13	Control	Non-survivor 14	Non-survivor 15	Control
iTRAQ-6	Control	Survivor 16	Survivor 17	Control	Survivor 18	Survivor 19	Non-survivor 16	Survivor 20

Table 7: Clinical characteristics of ARDS and control subjects. (Median & IQR)				
	Controls	ARDS Survivors	ARDS Non-survivors	P-value
Number	27	20	16	
Age (years)	66 (21)	42 (25)	59 (22)	0.004#
Gender (M/F)	12/15	12/8	7/9	0.5
ARDS day (days)	2.0 (2.0)	2.0 (3.5)	2.0 (7.0)	0.523
PF ratio	153 (196)	95 (82)	76 (107)	0.002#$
BAL Leukocytes (/µl)	342 (1983)	332 (537)	280 (641)	0.61
BAL Neutrophils (%)	60 (69)	66(34)	40 (56)	0.72
BAL Lymphocytes (%)	5.5 (12.25)	2.5 (5.3)	0.5 (3.8)	0.100

kruskal wallis ANOVA
# Post-Hoc Dunn test demonstrating significant difference in age between control and ARDS survivors. The differences in age between survivors and controls and non-survivors and survivors did not reach statistical significance.
$ post-hoc test demonstrating difference between control and ARDS non-survivors.

Table 8: PSPEP protein summary report for number spectra, peptides and proteins identified at ≤ 1% global FDR			
	Spectra	Peptides	Proteins
iTRAQ LC-MS/MS 1	37651	11623	850
iTRAQ LC-MS/MS 2	21183	7761	606
iTRAQ LC-MS/MS 3	25849	11865	1055
iTRAQ LC-MS/MS 4	24577	9061	865
iTRAQ LC-MS/MS 5	26623	11037	976
iTRAQ LC-MS/MS 6	26111	10389	879
Combined iTRAQ LC-MS/MS 1 to 6	160073	21148	1284

Table 9: ***Selected proteins with highest difference in abundance in BALF***					
Uniprot Accession Number	Protein Name	Mean variance weighted fold change		FDR (Comparing fold change in ARDS survivor vs. non-survivors)	Fold Change ARDS survivor: Non-survivor
		ARDS survivors: mastermix	ARDS non-survivor: marstermix		
Proteins which are high in survivors					
Q13510	Acid ceramidase	3.37	1.37	9.94E-04	2.45
P00740	Coagulation factor IX	1.36	0.55	1.40E-03	2.45
Q8IV08	Phospholipase D3	1.78	0.80	6.25E-08	2.22
Q8N257	Histone H2B type 3-B	1.26	0.58	1.53E-02	2.16
P30838	Aldehyde dehydrogenase, NADP-preferring	1.63	0.92	6.07E-30	1.78
S4R3Z2	Aldo-keto reductase family 1 member C3	1.63	0.94	3.07E-04	1.72
O60218	Aldo-keto reductase family 1 member B10	1.77	1.03	6.75E-05	1.72
P01011	Alpha-1-antichymotrypsin	1.70	1.02	4.78E-89	1.67
P02745	Complement C1q subcomponent subunit A	1.49	0.90	2.99E-05	1.65
P02748	Complement C9	1.04	0.63	8.59E-68	1.64
Proteins that are high in ARDS non-survivors					
G8JLH6	CD9 antigen (Fragment)	0.92	2.57	1.61E-06	0.36
P02774	Isoform 3 of Vitamin D-binding protein	0.33	0.92	1.38E-02	0.36
P05451	Lithostathine-1-alpha	0.86	2.18	7.16E-09	0.40
P02144	Myoglobin	1.06	2.55	8.15E-05	0.41
Q16378	Proline-rich protein 4	0.84	1.83	1.88E-03	0.46
P24158	Myeloblastin/Protinase 3	0.53	1.04	3.11E-04	0.51
P12724	Eosinophil cationic protein OS	0.91	1.55	3.06E-02	0.59
P15880	40S ribosomal protein S2	0.91	1.52	9.14E-05	0.60
P03973	Antileukoproteinase	0.94	1.54	3.08E-16	0.61
H7C2Z6	Grancalcin (Fragment)	0.79	1.29	3.12E-02	0.62

Table 10: GO Molecular Function identified using Functional Annotation Clustering algorithm in the Database for Visualization and Annotation for Integrated Discovery				
Molecular function	Proteins subsets assigned to the cluster		Fold Change Survivor: Non-survivor	Enrichment score
	Uniprot Accession Number	Protein Name		
Serine-type endopeptidase inhibitor activity	P01011	Alpha-1-antichymotrypsin	1.67	4.69
	P29622	Kallistatin	1.27	
	P08185	Corticosteroid-binding globulin[#]	1.64	
	P05543	Thyroxine-binding globulin[#]	1.61	
	P01019	Angiotensinogen [#]	1.15	
	G3V2W1	Protein Z-dependent protease inhibitor	1.51	
	P30740	Leukocyte elastase inhibitor	0.92	
	P29508	Serpin B3	1.08	
	P35237	Serpin B6	0.88	
	P01008	Antithrombin-III	1.12	
	P05546	Heparin cofactor 2	1.33	
	P36955	Pigment epithelium-derived factor	1.15	
	P05155	Plasma protease C1 inhibitor	0.88	
	P19827	Inter-alpha-trypsin inhibitor heavy chain H1	1.18	
	Q5T985	Inter-alpha-trypsin inhibitor heavy chain H2	1.22	
	Q06033	Inter-alpha-trypsin inhibitor heavy chain H3	1.45	
	Q14624	Inter-alpha-trypsin inhibitor heavy chain H4	1.09	
	P04083	Annexin A1	0.93	
	P12429	Annexin A3	0.75	
	P09525	Annexin A4	1.08	
	Q14508	WAP four-disulfide core domain protein 2	0.78	
	P02765	Alpha-2-HS-glycoprotein	1.42	
	P01031	Complement C5	1.48	
	P04196	Histidine-rich glycoprotein	1.11	
	P31025	Lipocalin-1	0.85	
	P27105	Erythrocyte band 7 integral membrane protein	0.83	
	P20742	Pregnancy zone protein	1.24	
	P07225	Vitamin K-dependent protein S	1.28	
	P11684	Club cell secretory protein	0.76	
	P03973	Antileukoproteinase	0.61	
	O43278	Kunitz-type protease inhibitor	1.15	
	P61981	14-3-3 protein gamma	0.85	
Serine hydrolase activity	P00740	Coagulation factor IX	2.45	1.39
	P00734	Prothrombin	1.52	
	P00748	Coagulation factor XII	1.34	
	P00747	Plasminogen	1.18	
	P03952	Plasma kallikrein	1.31	
	Q9NZP8	Complement C1r subcomponent-like protein	1.29	
	P06681	Complement C2	1.32	
	P00736	Complement C1r subcomponent	1.13	
	P22894	Neutrophil collagenase 1 (MMP 8)	0.84	
	P24158	Myeloblastin / Proteinase 3	0.51	
	Q9UHL4	Dipeptidyl peptidase 2	1.15	
	Q04756	Hepatocyte growth factor activator	1.25	
	P26927	Hepatocyte growth factor-like protein	1.29	

Table 11: GO Biological Processes identified using Functional Annotation Clustering				
Biological Process	Proteins subsets assigned to the cluster		Fold Change Survivor: Non-survivor	Enrichment score
	Uniprot Accession Number	Protein Name		
Acute Inflammatory response	P01011	Alpha-1-antichymotrypsin	1.67	3.2
	P02748	Complement component C9	1.64	
	P02745	Complement C1q subcomponent subunit A	1.65	
	P00734	Prothrombin	1.52	
	P01031	Complement C5	1.48	
	P21980	Protein-glutamine gamma-glutamyltransferase 2	1.45	
	P10909	Isoform 2 of Clusterin	1.43	
	P18428	Lipopolysaccharide-binding protein	1.42	
	P02765	Alpha-2-HS-glycoprotein	1.42	
	P07360	Complement component C8 γ chain	1.4	
	P02746	Complement C1q subcomponent subunit B	1.39	
	P13671	Complement component C6	1.38	
	P08603	Complement factor H	1.36	
	P00748	Coagulation factor XII	1.34	
	P06681	Complement C2	1.32	
	P03952	Plasma kallikrein	1.31	
	P07357	Complement component C8 α chain	1.31	
	Q9NZP8	Complement C1r subcomponent-like protein	1.29	
	P02743	Serum amyloid P-component	1.27	
	P04003	C4b-binding protein α chain	1.24	
	Q06830	Peroxiredoxin-1	1.15	
	O75636	Ficolin-3	1.18	
	P05362	Intercellular adhesion molecule 1	1.16	
	P01019	Angiotensinogen	1.15	
	P00736	Complement C1r subcomponent	1.12	
	P02790	Hemopexin	1,11	
	Q14624	Inter-alpha-trypsin inhibitor heavy chain H4	1.09	
	P27797	Calreticulin	0.91	
	P19320	Vascular cell adhesion protein 1	0.88	
	P08174	Isoform 5 of Complement decay-accelerating factor	0.87	
	P05155	Plasma protease C1 inhibitor	0.86	
	P06744	Glucose-6-phosphate isomerase	0.77	
	P27918	Properdin	0.73	
	P15090	Fatty acid-binding protein, adipocyte	0.62	
Programmed cell death	P07203	Glutathione peroxidase 1	1.64	1.97
	P02748	Complement component C9	1.64	
	P00734	Prothrombin	1.51	
	P01031	Complement C5	1.48	
	P21980	Protein-glutamine gamma-glutamyltransferase 2	1.45	

	P10909	Isoform 2 of Clusterin	1.43	
	P07360	Complement component C8 gamma chain	1.40	
	P13671	Complement component C6	1.38	
	O00115	Deoxyribonuclease-2-alpha	1.37	
	P07357	Complement component C8 alpha chain	1.31	
	P61626	Lysozyme C	1.18	
	P00747	Plasminogen	1.18	
	P08571	Monocyte differentiation antigen CD14	1.17	
	P14618	Pyruvate kinase PKM	1.15	
	P22314	Ubiquitin-like modifier-activating enzyme 1	1.11	
	P06396	Isoform 4 of Gelsolin	1.09	
	P62258	14-3-3 protein epsilon	1.09	
	P05107	Integrin beta-2	0.87	
	P62987	Ubiquitin-60S ribosomal protein L40	0.85	
	Q99828	Calcium and integrin-binding protein 1	0.77	
	P61604	10 kDa heat shock protein, mitochondrial	0.76	
	O43866	CD5 antigen-like	0.63	
Collagen metabolic processes	J3QSU6	Tenascin	1.28	1.53
	P12955	Xaa-Pro dipeptidase	1.16	
	P06727	Apolipoprotein A-IV	0.92	
	P14780	Matrix metalloproteinase-9	0.91	
	P22894	Neutrophil collagenase (MMP8)	0.84	
	P02452	Collagen alpha-1(I) chain	0.83	
	P24158	Myeloblastin / Proteinase 3	0.51	

Table 12: Canonical pathways enriched by the proteins that are differentially expressed between ARDS survivors and non-survivor			
Ingenuity Canonical Pathways	-log (B-H p-value)	Ratio	Proteins assigned to the pathway
Acute Phase Response Signaling	3.32E00	5.56E-01	SERPINF1, KLKB1, LBP, RBP4, C1R, PLG, C2, C9, ITIH2, F2, ITIH4, CP, HNRNPK, ITIH3, C5, SERPINA3, AGT, APCS, HRG, C4BPA, HPX, AHSG, SERPING1, SERPIND1, TTR
Complement System	2.9E00	6.3E-01	CFI, C1QB, C5, ITGB2, C6, C8G, C1R, C8B, CFH, C1QA, C8A, C2, CD55, C4BPA, C9, CR1, SERPING1
LXR/RXR Activation	1.49E00	4.87E-01	PON1, GC, ITIH4, APOA4, S100A8, MMP9, SERPINF1, CD14, LBP, RBP4, AGT, LYZ, VTN, CLU, C9, HPX, AHSG, A1BG, TTR

Table 13: Proteins that predict outcomes in ARDS			
Nucleophosmin	Properdin	AGR2	Cytochrome b5
Protein FAM49B	ATP-citrate synthase	Agrin Isoform 2	Adapter molecule crk
IGF binding protein 2	Nucleobindin-2	HSP 90-alpha	Mucin-1 subunit alpha
Isoform 2 Collagen alpha-1(IV) chain-isoform 2	LIM domain only protein 7	Erythrocyte band 7 membrane protein	Choline transporter-like protein- isoform 3
Ubiquitin carboxyl-terminal hydrolase	Poly (rC)-binding protein 1	Vascular non-inflammatory molecule 2	TGF –β induced protein ig-h3

Chapter 5

Protein Expression Profile in Lung Injury Following Hematopoietic Stem Cell Transplant

M Bhargava, KJ Viken, S Dey, MS Steinbach, B Wu, PD Jagtap, L Higgins, A Panoskaltsis-Mortari, DJ Weisdorf, V Kumar, M Arora, PB Bitterman, DH Ingbar, CH Wendt *(*Manuscript submitted to Biology of Blood and Marrow Transplant)

Hematopoietic stem cell transplant (HSCT) is a potentially curative treatment for otherwise fatal hematologic and lymphoid malignancies. However, pulmonary complications are the most common cause of death [190] and occur in up to 60-80% of allogeneic [190-193] and 25% of autologous HSCT recipients [194]. Reduced intensity preparatory regimens and use of alternate hematopoietic stem cell sources, such as peripheral blood stem cell transplant (PBSCT), have decreased the duration of neutropenia and lowered the incidence of infectious complications [195]. However, the early immune reconstitution and high levels of circulating cytokines associated with these strategies contribute to the development of inflammatory non-infectious pulmonary complications [196, 197]. Despite the introduction of many prophylactic strategies both infectious and non-infectious complications remain an important cause of transplant-related mortality and morbidity [191, 198].

Idiopathic Pneumonia Syndrome (IPS) is a non-infectious lung injury found in HSCT recipients. IPS is defined as “an idiopathic syndrome of pneumopathy after HSCT, with evidence of widespread alveolar injury and in which infectious etiologies and cardiac dysfunction, acute renal failure or iatrogenic fluid overload have been excluded” [199]. IPS occurs in up to 15% of allogeneic HSCT recipients after myeloablative conditioning [200-204]. Although it is less frequent with autologous HSCT [194, 205], the mortality of IPS remains high at approximately 80% [199, 200, 204, 206]. The median time to development of IPS has been reported to be 19 days after allogeneic (range 4-106 days) and 63 days after autologous HSCT (range 7-336), the period when the risk of infectious complications is also high. Escalated immunosuppression with high-dose steroids [207, 208] or TNF blockade [204, 209, 210] are current treatments for IPS. As these agents could worsen infectious lung injury, it is critical to differentiate infectious lung injury from IPS.

The goal of this study was to apply state-of-the-art protein expression profiling tools in cases of lung injury following HSCT or cellular therapy infusion to identify the proteins and biological processes that differentiate IPS from infectious lung injury in HSCT recipients. We performed comprehensive label-based quantitative protein profiling of BALF in patients undergoing HSCT. Our hypothesis was that subsets of proteins expressed in the BALF represent the biological processes responsible for lung injury and recovery. To identify these processes we performed Gene Ontology enrichment analysis and pathway analysis on proteins that showed differences in abundance in infectious lung injury and IPS.

METHODS

Study population

The University of Minnesota Institutional review board (IRB) Human Subjects Committee approved this study. Subjects were recruited from the Adult Blood and Marrow Transplant unit at the University of Minnesota Medical Center. Cases with a clinically indicated bronchoscopy within 180 days of HSCT were included. This period represents a higher risk period for lung injury and has been studied in a prior interventional study of IPS [207]. There was no exclusion based on underlying hematological disease, donor source, the degree of HLA match and conditioning regimen [207, 209]. Demographic and transplant characteristics of the study population are presented in Table 14. Amongst the study subjects, two underwent an autologous HSCT, twenty-three an allogeneic HSCT, and three patients were enrolled in a clinical trial of haploidentical natural-killer cell infusion for refractory acute myeloid leukemia. Twelve patients underwent myeloablative conditioning, and 16 underwent reduced intensity conditioning

prior to HSCT. Graft sources included bone marrow in 1, peripheral blood stem cells in 6, umbilical cord blood in 18 and natural killer cell in 3 patients.

All BALF samples were collected from clinically indicated bronchoscopies that were done according to a standard protocol [27, 36] with excess, cell-free supernatant used for these studies. The diagnosis of IPS or infectious lung injury was made using the criteria established by American Thoracic Society [199] (Table 15). All patients had evidence of widespread alveolar injury, abnormal respiratory physiology and no evidence of pulmonary edema or volume overload or acute kidney injury as the primary cause of respiratory dysfunction. A diagnosis of IPS required the absence of bacterial, viral, mycobacterial or fungal infection assessed by cultures, cytology or PCR-based studies. A diagnosis of infectious lung injury was made based on evidence of a pathogenic organism.

A global internal standard was used to compare relative protein abundance across multiple isobaric tagging for relative and absolute quantification- two-dimensional liquid chromatography-tandem mass spectrometry (iTRAQ 2DLC-MS/MS) experiments. This consisted of pooled BALF from 27 cases (mastermix) with respiratory failure without previous HSCT and who did not meet the criterion for acute respiratory distress syndrome (except in one case).

Sample preparation

Cell and debris free BALF supernatant were stored at -80°C within 60 minutes of collection. BALF samples containing at least 1.2 mg of proteins were processed for LC-MS/MS from individual patients employing a protocol previously published with minor modifications [36]. BALF was concentrated and desalted using Amicon 3-MWCO filters.

Hemoglobin depletion was performed with Hemoglobind (BioTech Support Group LLC, Monmouth Junction, NJ) per the manufacturer's instructions. Subsequent processing was similar to our prior study with immunoaffinity depletion of high abundance proteins to reduce the dynamic range and appropriate buffer exchanges for labeling with iTRAQ reagent [36].

iTRAQ labeling and 2D LC-MS/MS

The enriched BALF was digested with trypsin-gold (Promega cat#V5280), dried and suspended in 0.1- 0.2% formic acid (pH < 3) and MCX cation exchange (Oasis MCX Cartridge, Waters, Milford, MA, Cat no 186000254) was performed to remove SDS. Proteins were eluted and labeled with eight-plex iTRAQ reagent per the manufacturer (AB Sciex, Framingham, MA) instructions as described previously [36, 43]. To compare protein abundance across different LC-MS/MS experiment, we used the same pooled mastermix as a global internal standard. In each LC-MS/MS experiment, two iTRAQ reporter ion channels were labeled with the mastermix to assure the accuracy of fold-change measurements while the remaining six channels contained study samples. To characterize 30 samples in the study, we performed five separate iTRAQ LC-MS/MS experiments. The labeling strategy for the 30 BALF samples studied is outlined in Table S9. To prevent reporter ion signal (channel) bias the mastermix, IPS, and infectious lung injury samples were randomly placed in different iTRAQ reporter ion channels in each experiment.

Each iTRAQ-labeled peptide mixture was purified with an MCX Oasis cartridge before off-line peptide separation in the first dimension and if needed C18 stage tipping. Peptides were separated offline into 15 separate peptide-containing fractions collected in

2-minute intervals on a C18 Gemini column (Phenomenex, Torrance, CA) at pH 10. Peptide fractions were concentrated, purified by the Stage Tip procedure [179] with Empore SDB-RPS extraction disks [mixed mode strong cation exchange and reversed phase], 3M (St. Paul MN), and separated in the 2nd dimension by C18 reversed-phase capillary LC with a nano LC system (Eksigent, Dublin, CA) online with MS. Data-dependent acquisition was performed on an Orbitrap Velos system with HCD (higher-energy collision induced dissociation) activation for peptide tandem MS. LC and MS experimental details were previously reported, with the exception that the activation time was 20 msec [36, 174].

Database search for protein identification

.RAW files obtained directly from the Orbitrap Velos mass spectrometer were imported into GalaxyP (https://usegalaxyp.org/ for public instance) for further processing (as described in z.umn.edu/ppingp). Within GalaxyP, all .RAW files together (using a multifile format that has been recently changed to dataset collection) were converted to mzml format using msconvert and then into ProteinPilot compatible Mascot Generic Format (MGF) files with preselected iTRAQ reporter ions. The MGF files were searched against the target-decoy version of Human UniProt database along with contaminant protein sequences (84,838 target sequences in total; July 2014) using ProteinPilot version 4.5 and the following search parameters: Sample Type: iTRAQ 8-plex (peptide labeled); Cys-alkylation: MMTS; Instrument: Orbi MS, Orbi MS/MS; run quant; bias correction on; search focus on biological modifications and amino-acid substitutions; thorough search and with a detected protein threshold (Unused Protscore (Conf)): 10%. The ProteinPilot searches and subsequent generation of Proteomics System Performance Evaluation Pipeline Software (PSPEP) –FDR reports and protein and

peptide level summaries were generated within Galaxy-P as previously described [36]. All processing was conducted within the GalaxyP framework, which offers highly reproducible, robust, and easily shareable workflows [126, 211]. The mass spectrometry proteomics data have been deposited into the ProteomeXchange Consortium [127] via the PRIDE partner repository with the dataset PXD002437.

The results of multiple iTRAQ LC-MS/MS experiments were aligned to compare protein-level quantitative data using Protein Alignment Template vs. 2.00p (AB Sciex) [175]. The Protein Alignment Template provides the overlap of proteins across multiple iTRAQ LC-MS/MS experiments by matching proteins in a 'reference master list' to the 'test list' of proteins identified in each individual iTRAQ LC-MS/MS run. For this alignment, we created a reference master list by performing a database search using .RAW files from two iTRAQ LC-MS/MS experiments i.e. iTRAQ experiment 1 and 5. To ensure that the proteins in this list are of high ID quality, a local FDR ≤ 5% was used as a threshold for the reference protein list. As per the recommendation of the Protein Alignment Template, for creation of feature table with quantitative values, the threshold of ≤ 5% global FDR was used for 'test list' consisting of individual sets of the five-iTRAQ LC-MS/MS experiments. The Protein Alignment Template resulted in aligning of the ratios, p-values and error factors of the proteins across replicate experiments by using accession numbers of isoforms within protein summary and UniProt database.

Statistical analysis

Identification of differentially expressed proteins between ARDS survivors and non-survivors: To compare different proteins of two groups (ARDS survivor vs. non-survivors), w*e performed inverse weighted ratio (*accounting for the Error Factor) to account for the peptide level variance in fold changes measured for each protein across multiple iTRAQ runs. Specifically, denote Xi as the log protein ratio for one group and Si its corresponding variance; denote Yj as the log protein ratio of another group and Vi its corresponding variance. Define the following Z-test statistic

$$Z = \frac{\sum_i X_i/S_i - \sum_j Y_j/V_j}{\sqrt{\sum_i 1/S_i + \sum_j 1/V_j}}$$

We can compare Z to standard normal distribution to compute significant p-values. Thus using a traditional α of 0.05 is not optimal. We will control for multiple comparisons by FDR [176] corrected p-value ≤0.05 as done previously [36].

Computational Analysis

To gain insight into the biological significance of differentially expressed proteins, we used the Database for Annotation, Visualization, and Integrated Discovery (DAVID, http://david.abcc.ncifcrf.gov) [129] as previously reported [36, 43]. DAVID generates an enrichment score for a group of genes indicating annotation term member associations in a given experiment. We focused on clusters with an enrichment score ≥ 1.3 because it is equivalent to a non-log scale p-value of 0.05. In addition, to confirm the findings observed in DAVID we also performed functional analysis using Ingenuity Pathway Analysis (IPA® QIAGEN, Redwood City www.quiagen.com/ingenuty Build 321510M,

Version 21249400) and focused on the protein subsets represented in canonical pathways, upstream regulators and molecular and cellular functions. The significance of the association between the differentially expressed proteins and the canonical pathway was measured in 2 ways: 1) a ratio of the number of BALF proteins that mapped to the pathway divided by the total number of genes assigned to that canonical pathway as it provides information about the depth of association; 2) a Benjamini and Hochberg corrected p-value (obtained using the right-tailed Fisher Exact Test) determining the probability that the association between the BALF proteins and the canonical signaling pathway/biological function was explained by chance alone. All pathways with an FDR ≤ 0.05 are reported.

Identifying the universe of BALF proteins: The Protein Alignment Template uses a master reference list, which introduces a certain amount of arbitrariness into the result. Thus, to match corresponding proteins across runs, we also implemented our approach that seeks to overcome this limitation. This approach uses the ProteinPilot output files that characterize each protein in terms of the peptides that support its identification and, roughly speaking, matches proteins across runs when they share peptides. FDR filtering, as described previously, was performed prior to analysis to eliminate those proteins with less reliable identification. More specifically, our approach creates a graph, where each protein from any run is a node in the graph. Links are created between two nodes (proteins) for those nodes that share at least one peptide. We eliminated proteins that have links to protein within the same run or more than two proteins within another run. (This resulted in the elimination of a small number of proteins (~8). Each connected component of the graph was interpreted as representing a single protein and was used to align the ratios and P-values of the proteins across replicate experiments.

RESULTS

Characteristics of study participant

In this study, we characterized 12 BALF samples from 12 subjects with IPS and 18 BALF samples from 16 subjects with infectious lung injury collected between May 2008 and June 2013. The median time from transplantation to BALF sample collection was not different between IPS (median 33 days, range 15-138 days) and infectious lung injury (median 51 days, range 8-179 days). The indication and type of transplantation are shown in Table 14, and the clinical characteristics of these patients are shown in Table 16. Between the two groups, there was no difference in the age, time after HSCT when BALF samples were collected, the total BALF leukocyte count and the percent of BALF neutrophils, lymphocytes, and monocytes.

Protein identified by database searching

As most of the BALF samples in HSCT recipients appeared blood tinged, for the first iTRAQ LC-MS/MS experiment we tested whether removal of hemoglobin (in addition to high abundance proteins) would improve the depth of coverage. Hemoglobin removal (in addition to high abundance protein depletion) improved protein identification to 845 proteins at 1% global FDR compared to 496 proteins with high abundance protein depletion alone for one of the runs. Subsequent iTRAQ LC-MS/MS experiments were performed with hemoglobin depletion followed by high abundance protein depletion.

The ProteinPilot PSPEP FDR summary showing the number of spectra, peptides and the proteins identified at 1% global FDR for the five-iTRAQ experiments is shown in Table17. The total numbers of proteins identified in each of the five separate iTRAQ LC-MS/MS experiments were 845, 735, 532, 615 and 594 respectively. The ProteinPilot

summary report of the proteins identified is included in the Table S10. A total of 1125 unique proteins (Table S11) were identified in the five separate experiments performed of which 368 proteins were present in all five LC-MS/MS experiments.

Proteins differentially expressed in IPS

The 793 proteins identified in the reference list at ≤ 5% local FDR were aligned to the test proteins from individual iTRAQ LC-MS/MS experiment (Table S12, reference list aligned with tests tab). Of these 793 proteins, quantitative information was present to perform inverse variance weighted t-test on 558 proteins (Table S13, FDR BMT tab). Controlling for an FDR ≤ 5%, 132 proteins were differentially expressed when IPS was compared to infectious lung injury (Table S13 5%FDR tab). Proteins representing the greatest difference in abundance between IPS and infectious lung injury are shown in Table 18.

Biological processes represented by differentially expressed proteins.

To gain insight into biological processes, we performed Functional Annotation Clustering using DAVID on the differentially expressed proteins. Seventeen biological modules (functional annotation clusters with an enrichment score > 1.3) were identified (Table S14). Our analysis revealed proteins that mapped to processes of immune response and leukocyte adhesion (cluster 6,11). Other functional modules that were identified included those involved in blood coagulation, fibrinolysis and wound healing (cluster 1 and 2), cell migration (cluster 9,10), glycolysis (cluster 13) and apoptosis (cluster 8). There were three modules involved in lipid metabolism (cluster 3,12,15) and two involved in oxidative stress / cellular ion homeostasis (cluster 4, 5). We also identified

modules involved in multicellular organismal catabolic process (cluster 7), response to nutrient levels (cluster 14), protein oligomerization (cluster 16) and response to steroid hormone stimulus (cluster 17).

In addition, to understand the relative impact of changes in protein levels in the context of well-characterized pathways and to independently test the validity of the findings observed using DAVID, we performed IPA core-analysis, focused mainly on molecular and cellular functions, canonical pathways and the upstream regulators. Several of the GO terms that were over- represented by the differentially expressed proteins in DAVID were also identified by IPA core analysis. The top five molecular and cellular functions represented by the differentially expressed proteins included cell to cell signaling, cellular movement, cell death and survival, free radical scavenging and cellular growth and proliferation. Controlling for an FDR ≤ 5%, we identified 18 IPA canonical pathways that were represented by differentially expressed proteins (Table 19). These included processes identified by GO enrichment analysis including acute phase response signaling (APRS), complement system, coagulation system, intrinsic and extrinsic prothrombin activation, glycolysis, granulocyte adhesion and diapedesis (leukocyte migration) and glutathione redox reactions. Additionally pathways in the IPA knowledge base that were affected included clathrin-mediated endocytotic signaling, Rho GDI signaling, IL-12 signaling and production in macrophages.

Upstream regulator analysis in IPA demonstrated that lipopolysaccharide (LPS); nitrofurantoin, dexamethasone, beta-estradiol, and tretinoin were the top five upstream regulators of the differentially expressed proteins between IPS and infectious lung injury.

DISCUSSION

This is the first study that characterizes the BALF protein profile by mass spectrometry in HSCT recipients. In this study, we show deep coverage of the BALF proteome with identification of 1125 proteins. We also identified differentially expressed proteins that provide insights into the biological processes involved in the development of IPS in HSCT recipients. IPS continues to have a high mortality and rapid diagnosis and novel insights into the biology of the disease process will advance our goal of improving the outcomes of HSCT recipients with lung injury.

The five most significant canonical pathways identified in the BALF in our study were previously reported to be relevant to development of IPS when mass spectrometry studies were performed in blood comparing cases with IPS to control subjects with no lung damage following HSCT [212]. The pathways identified in the blood include acute phase reactants (APRS), complement system, coagulation system, LXR / RXR activation, and FXR/RXR activation. As BALF is the most proximate fluid to the site of injury, it is not surprising that BALF proteins are annotated to pathways that previously were implicated in the development of IPS. This also supports the appropriateness of BALF as a biofluid for discovery studies in lung injury as systemic changes occurring are identified both in the BALF and blood.

TNF- α plays an important role in the development of IPS and etanercept has been an effective agent in IPS [204, 208, 209]. However, the beneficial response to TNF blockade is not universal as a recent study reported that IPS cases already receiving corticosteroids did not have therapeutic benefit with etanercept [207]. In our study, although we did not identify TNF- α at the threshold for protein identification, there were

indirect indicators of TNF-α involvement in IPS after HSCT. In the upstream regulator analysis, LPS, a powerful stimulator of TNF-α was identified as the top regulator of the proteins that are differentially expressed in both IPS and infectious lung injury groups. We also identified two canonical pathways that are targets for TNF-α blockade. These pathways were Farsenoid X Receptor (FXR) / Retinoid X Receptor (RXR) activation and IL12 signaling and production in macrophages. FXR is a nuclear receptor and in the liver and intestinal epithelia it plays a crucial role in lipoprotein and bile acid metabolism [213]. Increasing literature supports the role of FXR in inflammation mediated by NF-κB in the liver [214] and intestinal epithelium [215]. In the lung, FXR recently was found to be expressed on the pulmonary endothelial cells [216, 217] and it protects from LPS induced lung injury, possibly via attenuation of P-selectin mediated neutrophil recruitment [216]. Interestingly in that study, FXR ligand attenuated TNF-α induced upregulation of P-selectin in a pure culture of endothelial cells, suggesting its role in TNF- α induced inflammation. This provides a potential mechanism by which selected cases with IPS respond to TNF-α inhibition. Our study identified L-selectin, but not P-selectin, to be more abundant in IPS. This will need further exploration.

In addition to FXR activation, the differentially expressed proteins also mapped to IL-12 signaling in macrophages. IL -12 belongs to a diverse family of cytokines that have both pro-inflammatory and anti-inflammatory roles. In macrophages, IL-12 favors the differentiation of Th1 cells stimulating the adaptive immune response. Also, IL-12 activates MAPK- mediated induction of tumor necrosis factor suggesting that it may be a link between innate and adaptive immune responses (IPA knowledgebase). Our study did not provide adequate information to conclude if this pathway is activated or inhibited. This also warrants further investigation.

Several canonical pathways were identified that could provide mechanistic insights into the injury repair process in the lungs of HSCT recipients. RhoGDI signaling was over-represented by the proteins differentially expressed in IPS compared to infectious lung injury. These differentially expressed proteins included the cytoplasmic proteins: beta-actin, ezrin, moesin and guanine nucleotide binding protein (all low in IPS) and the membrane proteins: CD44 and cadherin 1 (high in IPS). Rho GDI belongs to a family of small GTPases comprised of Cdc 42, Rac and Rho that regulate several critical processes including cell proliferation, apoptosis, cell differentiation and cell motility, which are important in lung injury repair cycle. Rho GDI signaling may be involved in the regulation of cell migration by regulation of ezrin-radixin-moesin (ERM) proteins. ERM proteins co-localize in cell-matrix adhesion sites, filopodia, and membrane protrusions [146]. ERMs function by binding to and organizing the actin cytoskeleton and in turn, stabilize adherens junctions that are involved in cell migration [147, 149, 150]. ERM protein activation requires c-terminal phosphorylation that involves the small GTP-binding Rho [218]. Upon activation, the n-terminus of the ERM protein interacts with trans-membrane proteins such as CD 44 subsequent stabilizing of the actin cytoskeleton or activating signaling molecules. These interactions during lung injury and repair will need better characterization but offer potential therapeutic targets

The study has several limitations. We elected to deplete high abundance proteins to enrich medium and low abundance proteins. This is an important issue in all proteomic studies, especially when protein quantification is performed. This methodology is important in data dependent MS acquisition where only selected MS1 precursor ions are selected for fragmentation. As high abundance protein can preclude quantification of less abundance MS1 precursor ions, analyzing undepleted BALF could severely limit the findings of a study. The downside of depletion of higher abundance protein is that other

proteins that bind to high abundance proteins might be co-depleted. It also increases the amount of sample handling needed before MS data acquisition. Another limitation is the small number of study subjects. Our study provides proof of concept for future work in this area. In this study, despite using high-resolution MS platforms, the overlap of the protein identified in the five-iTRAQ runs was partial. We expect that with improvement in MS platforms, the depth and overlap in proteins identified across multiple LC-MS/MS experiments will improve. More recent MS approaches, using data-independent acquisition (such as SWATH-MS), also will directly address the limitation of MS platforms being employed for contemporary proteomic studies. We also acknowledge that serum proteins likely contaminated the BALF proteome to some extent in our study. Though this could be addressed by direct comparison of comprehensive protein expression changes in serum and BALF, BALF analysis is a critical first step that provides a framework for future studies of lung dysfunction in HSCT recipients.

CONCLUSIONS

This is the first study to characterize the BALF proteome in subjects with lung injury following HSCT. We found major differences in the BALF proteome that distinguish IPS from infectious lung injury. These differences likely reflect the underlying mechanisms implicated in the development of lung damage and the repair response that is activated. Though somewhat speculative, we also identify possible mechanisms that could explain variable response to TNF blockade in IPS. Studies in a larger validation cohort of subjects will likely provide valuable insights into disease biology and also identify novel therapeutic targets for intervention.

Tables for Chapter 5

Table 14: Type of transplantation			
	IPS	Infection	p- value
Recipient age ((mean ± SD)	47 ± 15	49 ± 16	0.727
Gender (M/F) and %	3/9 (25 / 75%)	6/10 (37.5 / 62.5%)	0.48
Diagnosis Acute Leukemia (AML/ALL) Lymphoma (HD/NHL) MDS/ Myeloproliferative disorder (MDS,myelofibrosis, PV) Chronic Leukemia (CLL)	 7 2 2 1	 12 2 2 0	0.61
Type of transplant Autologous Allogeneic Haploidentical NK cell infusion	 1 10 1	 1 13 2	0.92
Donor Type Unrelated donor Autologous Haploidentical	 8 1 3	 13 1 2	0.66
Graft type DUCBT PBSCT Marrow Haploidentical NK cell infusion	 5 5 1 1	 13 1 0 2	0.067
HLA match Matched Mismatched Haploidentical	 3 5 3	 0 16 0	0.003
Conditioning Myeloablative Reduced intensity conditioning	 3 9	 9 7	0.13
Recipient / donor CMV status Positive / Negative Positive / Positive Negative / Negative Negative / Positive Positive (autologous transplant)	 3 3 4 1 1	 11 0 4 0 1	0.079

ALL- Acute lymphoblastic leukemia, AML Acute Myeloid Leukemia, MDS- Myelodysplastic Syndrome, CLL- Chronic lymphocytic leukemia, PV- Polycytehemia vera, DUCBT-Double umbilical cord blood transplant, PBSCT- peripheral blood stem cell transplant. chi-square test for all except age.

Table 15: IPS diagnostic criteria
1. **Presence of widespread alveolar injury**
a. CT of CXR evidence of bilateral of multilobar infiltrates
b. Abnormal respiratory physiology base on room air O_2 saturation of < 93% or need for supplemental oxygen to keep O_2 > 93
2. Absence of lower respiratory tract infection assessed by sputum or BAL fluid negative for pathogenic bacterial and nonbacterial organism[1]
a. Gram stain, fungal stain, AFB stain
b. Bacterial[2], fungal and viral (RSV CMV,adenovirus, parainfluenza, influenza A and B, rhinovirus).
c. Viral PCR studies for CMV, HSV, VZV, HHV-6
d. Pneumocystis jeroveci assay by cytology
3.Absence of cardiac dysfunction (echo, pro BNP), acute kidney injury and fluid overload

[1] *mixed oral flora and < rare candida or penicillium did not rule out IPS.*
[2] Any positive culture as quantitative bacterial cultures not always done.
CT computed tomography, CXR- chest X Ray, AFB- acid-fast bacilli, CMV- cytomegalo virus, PCR- polymerase chain reaction, HSV- herpes simplex virus, RSV- respiratory syncytial virus.

Table 16: Clinical characteristics of the study participants			
	IPS	**Infectious Lung Injury**	**P-value**
Number of BAL Samples	12	18	
Time from transplant to bronchoscopy (days)	60 ± 39	51 ± 50	0.60
S Creatinine (mg/dl) on day of bronchoscopy	1.2 ± 0.6	1.4 ± 0.8	0.47
BALF leucocytes (per µl)	314 ± 337	203 ± 173	0.24
BALF Neutrophils (%)	29± 35	20 ± 30	0.47
BALF lymphocytes (%)	8.7 ± 13	4.1 ± 6.9	0.23
BAL monocytes (%)	39 ± 43	59 ± 43	0.55

Table 17: PSPEP protein summary report for number of spectra, peptides and proteins identified at ≤ 1% global FDR			
	Spectra	**Peptides**	**Proteins**
iTRAQ1	21933	8635	845
iTRAQ2	24691	9466	735
iTRAQ3	21826	5842	532
iTRAQ 4	17893	7285	615
iTRAQ 5	16789	6515	594

Table 18: Selected proteins with highest difference in abundance in BALF of cases with IPS compared to infectious lung injury.					
Uniprot Accession Number	Protein Name	Mean variance weighted fold change IPS : mastermix	Mean Variance weighted Fold change Infection: marstermix	FDR (Comparing fold change in IPS vs infectious lung injury)	Fold Change IPS: Infection
Proteins that are high in IPS compared to infectious lung injury (ten proteins with highest differential expression)					
P24158	Myeloblastin	1.09	0.34	4.22E-04	3.23
P14151	L-selectin	2.60	0.89	6.79E-14	2.93
P07333	Macrophage colony-stimulating factor 1 receptor	2.71	0.93	1.58E-13	2.90
O95497	Pantetheinase	4.95	1.82	9.12E-06	2.73
Q15582	Transforming growth factor-beta-induced protein ig-h3	1.44	0.53	7.67E-03	2.72
Q9BTY2	Plasma alpha-L-fucosidase	4.00	1.53	9.33E-05	2.62
P13611	Isoform Vint of Versican core protein	2.64	1.03	4.55E-02	2.57
P12259	Coagulation factor V	7.75	3.51	1.97E-02	2.21
P13598	Intercellular adhesion molecule 2	2.55	1.24	1.44E-03	2.05
P07738	Bisphosphoglycerate mutase	2.10	1.02	7.45E-04	2.05
Proteins that are high in infectious lung injury compared to IPS (ten proteins with highest differential expression)					
P05109	Protein S100-A8	0.15	0.26	8.40E-06	0.57
P11684	Club Cell Secretory Protein	0.72	1.29	1.68E-03	0.56
Q96IY4	Carboxypeptidase B2	0.55	0.98	2.22E-03	0.56
Q8WUM4	Programmed cell death 6-interacting protein	1.11	2.00	5.99E-11	0.55
P15311	Ezrin	0.82	1.53	8.79E-36	0.54
E7EQB2	Kaliocin-1 (Fragment)	0.46	0.87	1.26E-33	0.53
P07108	Isoform 5 of Acyl-CoA-binding protein	1.00	1.97	2.05E-04	0.51
P02730	Band 3 anion transport protein	1.19	2.41	2.52E-02	0.49
P16104	Histone H2AX	0.21	0.56	1.82E-02	0.37
P08118	Beta-microseminoprotein	0.89	2.59	6.91E-09	0.34

Table 19: Ingenuity Pathway Analysis Canonical pathways represented by proteins that are differentially expressed between IPS and infectious lung injury.		
Canonical Pathway	**Corrected P value**	**Dataset proteins mapping to the pathway**
LXR/RXR Activation	1.02×10^{-11}	KNG1, APOE, APOA4, C3, APOH, A1BG, SERPINF2, PON1, LYZ, ITIH4, S100A8, GC, APOD, AGT
FXR/RXR Activation	3.99×10^{-9}	KNG1, APOE, PON1, APOA4, C3, APOH, ITIH4, GC, A1BG, SERPINF2, AGT, APOD
Acute Phase Response Signaling	5.37×10^{-9}	FTL, C3, APOH, CP, SERPINA3, SERPINF2, F2, C5, PLG, ITIH4, CFB, HRG, AGT
Complement System	3.24×10^{-6}	C3, CFB, CFI, C7, CFH, C5
Coagulation System	3.6×10^{-6}	KNG1, PLG, PROS1, F5, SERPINF2, F2
Atherosclerosis Signaling	4.77×10^{-5}	APOE, PON1, LYZ, VCAM1, APOA4, S100A8, PRDX6, APOD
Neuroprotective Role of THOP1 in Alzheimer's Disease	1.6×10^{-4}	KNG1, PLG, SERPINA3, ACE, AGT
Clathrin-mediated Endocytosis Signaling	7.2×10^{-4}	APOE, PON1, LYZ, APOA4, ACTB, S100A8, F2, APOD
Intrinsic Prothrombin Activation Pathway	7.9×10^{-4}	KNG1, PROS1, F5, F2
Extrinsic Prothrombin Activation Pathway	2.81×10^{-3}	PROS1, F5, F2
Production of Nitric Oxide and Reactive Oxygen Species in Macrophages	3.2×10^{-3}	APOE, PON1, LYZ, APOA4, S100A8, SIRPA, APOD
Agranulocyte Adhesion and Diapedesis	3.65×10^{-3}	SELL, VCAM1, ICAM2, EZR, ACTB, MSN, C5
Rapoport-Luebering Glycolytic Shunt	2.65×10^{-3}	PGAM1, BPGM
IL-12 Signaling and Production in Macrophages	3.65×10^{-3}	APOE, PON1, LYZ, APOA4, S100A8, APOD
Glycolysis	7.8 x 10 -3	TPI1, PGAM1, BPGM
RhoGDI Signaling	$1.\ 19 \times 10^{-2}$	CDH1, EZR, ACTB, CD44, GNB2, MSN
Granulocyte Adhesion and Diapedesis	1.2×10^{-2}	SELL ,VCAM1, CAM2, EZR, MSN, C5
Leukocyte Extravasation Signaling	2.1×10^{-2}	VCAM1, TIMP1, EZR, ACTB, CD44, MSN

A1BG: alpha1-B glycoprotein (P04217), ACE angiotensin 1 converting enzyme (P12821), ACTB: beta actin (P60709), AGT: angiotensinogen (P01019), APOA4 apolipoprotein A-IV (P06727), APOD: apoliporotein D (C9JF17), APOE: apolipoprotein E (P02649), APOH: apoliporotein H (P02749), BPGM: 2,3 bisphosphoglycerate (P07738), C3: complement component 3 (P01024), C5: complement component 5 (P01024), CFB: complement factor B (B4E124), CP: ceruloplasmin (P00450), CS: complement component S (P01031), CD44: CD 44 molecule (P60709), CDH1: cadherin 1 (P12830), EZR: ezrin (P15311)F2: coagulation factor 2, F5: coagulation factor 5 (P12259), (P00374), FTL: ferretin light polypeptide (P02792), GC: group- specific component (vitamin D binding protein, P02774), GNP2 : guanine- nucleotide binding protein (P62879), HRG: histidine rich glycoprotein (P04196), ITHH4: inter-alpha – trypsin inhibitor heavy chain (B72KHJ8), KNG1: ICAM 2: intercellular adhesion molecule 2 (P13598), kininogen 1 (P01042), LYZ: lysozyme (P61626), MSN: moesin (P26038), PON1: paroxanase 1 (P27169), PGAM1: phosphoglycero mutase 1 (P18669(, PLG: plasminogen (P00747), PRDX6: preredoxin 6 (P30041), PROS1: protein S (P07225), S100A8: S 100 calcium binding protein A8 (P05109), SERPIN A3: serpin peptidate inhibitor, clade A (P01011), SERPINF2: serpin peptidase inhibitor clade F (P08697), SELL: L- selectin (P14151), SIRPA: signal regulatory protein alpha (P78324), VCAM1: vascular cell adhesion molecule 1,

Chapter 6

Conclusion of Thesis

Proteomics, a large-scale study of proteins, is rapidly evolving due to advances in the tools for mass spectrometry and downstream bioinformatics analysis for protein assembly and biological relevance of the identified proteins. In Chapter 1, basic principles of protein identification and quantification and their application in Acute Respiratory Distress Syndrome (ARDS) are outlined. We then apply a label-based quantitative proteomics methodology in ARDS cases. These studies were initially performed on pooled bronchoalveolar lavage fluid (BALF) from ARDS survivors in early and late phase of the disease and early phase ARDS non-survivors (Chapter 2). Existing biospecimens and bioinformatics tools were used for these studies where we found early differences in protein expression and ARDS. Gene Ontology enrichment analysis demonstrated a coordinated response to lung damage in survivors while non-survivors had evidence of durable scar formation and catabolism. As a follow-up of these studies, we then characterized BALF in individual cases of ARDS, 16 non-survivors, and 20 survivors in six eight-plex iTRAQ LC-MS/MS experiments. In addition to the bioinformatics tools used for studies on pooled BALF, we used Protein Alignment Template to resolve the protein identification ambiguity that is introduced due to parsimonious protein assembly for identification of protein groups. This results in lack of consistently identifying the proteins with same accession number across different iTRAQ LC-MS/MS runs. Additionally, due to differences in the number of unique peptide contribution to the measurement of relative abundance, an absolute ratio of a certain protein cannot be compared across different iTRAQ LC-MS/MS experiment despite using a global internal standard for calibration across different iTRAQ runs. To account for this peptide level variability in protein quantification, we performed an error factor weighted t-test to compare the mean protein abundance of proteins in ARDS survivors and non-survivors. The differentially expressed proteins revealed differences in the mechanisms involved in lung injury and repair in ARDS survivors. Moreover, we also

identified differences in serine proteases and protease inhibitors to be different in survivors and non-survivors suggesting the overall proteolytic activity in the distal lung to be an important factor for prognosis in ARDS. Thus a functional assay rather than a protein level could be more relevant for prognostication. In addition to characterizing the BALF proteome in ARDS, we also studied the changes in the BALF that occur in lung injury following hematopoietic stem cell transplantation. Specifically, we compared the change in lung injury due to infection to non-infectious lung damage called IPS to determine if the pathways and biological processes that are activated in infection and IPS are different. Similar to studies in ARDS, eight-plex iTRAQ LC-MS/MS was performed on 30 HSCT recipients. We identified 132 proteins to be differentially expressed in IPS when compared to infectious lung injury. These proteins are involved in two canonical pathways that are specifically modulated by TNF; FXR / RXR activation and IL2 signaling and production in macrophages; therefore, selective targeting could be an important strategy in the treatment of IPS. In this thesis, in addition to demonstrating the feasibility of comprehensive BALF protein profiling we have also identified key differences in the mechanism responsible for lung injury and repair in two separate conditions where diffuse lung damage is present. In doing so we have established a pipeline for analysis and interpretation of ITRAQ LC-MS/MS data from multiple runs gained new insights that would ultimately identify both biomarkers and target for therapy for lung repair.

BIBLIOGRAPHY

1. James P: **Protein identification in the post-genome era: the rapid rise of proteomics.** *Q Rev Biophys* 1997, **30:**279-331.
2. Rottoli P, Bargagli E, Landi C, Magi B: **Proteomic analysis in interstitial lung diseases: a review.** *Current opinion in pulmonary medicine* 2009, **15:**470-478.
3. Bons JA, Drent M, Bouwman FG, Mariman EC, van Dieijen-Visser MP, Wodzig WK: **Potential biomarkers for diagnosis of sarcoidosis using proteomics in serum.** *Respiratory medicine* 2007, **101:**1687-1695.
4. Lehtio J, De Petris L: **Lung cancer proteomics, clinical and technological considerations.** *Journal of proteomics* 2010, **73:**1851-1863.
5. Cho WC: **Proteome profiling for the identification of lung cancer signatures.** *Expert review of proteomics* 2011, **8:**689-692.
6. Ostroff RM, Bigbee WL, Franklin W, Gold L, Mehan M, Miller YE, Pass HI, Rom WN, Siegfried JM, Stewart A, et al: **Unlocking biomarker discovery: large scale application of aptamer proteomic technology for early detection of lung cancer.** *PLoS One* 2010, **5:**e15003.
7. Hirsch J, Hansen KC, Burlingame AL, Matthay MA: **Proteomics: current techniques and potential applications to lung disease.** *Am J Physiol Lung Cell Mol Physiol* 2004, **287:**L1-23.
8. Miller I, Eberini I, Gianazza E: **Proteomics of lung physiopathology.** *Proteomics* 2008, **8:**5053-5073.
9. Magi B, Bini L, Perari MG, Fossi A, Sanchez JC, Hochstrasser D, Paesano S, Raggiaschi R, Santucci A, Pallini V, Rottoli P: **Bronchoalveolar lavage fluid protein composition in patients with sarcoidosis and idiopathic pulmonary fibrosis: a two-dimensional electrophoretic study.** *Electrophoresis* 2002, **23:**3434-3444.
10. O'Farrell PH: **High resolution two-dimensional electrophoresis of proteins.** *J Biol Chem* 1975, **250:**4007-4021.
11. Anderson L, Anderson NG: **High resolution two-dimensional electrophoresis of human plasma proteins.** *Proc Natl Acad Sci U S A* 1977, **74:**5421-5425.
12. Tanaka K, Waki, H., Ido, Y., Akita, S., Yoshida, Y., Yoshida, T. : **Protein and Polymer Analyses up to m/z 100 000 by Laser Ionization Time-of-Flight Mass Spectrometry.** *Rapid Commun Mass Spectrom* 1988, **02:**151-153.
13. Fenn JB, Mann M, Meng CK, Wong SF, Whitehouse CM: **Electrospray ionization for mass spectrometry of large biomolecules.** *Science* 1989, **246:**64-71.
14. Shevchenko A, Wilm M, Vorm O, Mann M: **Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels.** *Anal Chem* 1996, **68:**850-858.
15. Vestling MM, Fenselau C: **Polyvinylidene difluoride (PVDF): an interface for gel electrophoresis and matrix-assisted laser desorption/ionization mass spectrometry.** *Biochem Soc Trans* 1994, **22:**547-551.
16. Wilm M, Mann M: **Analytical properties of the nanoelectrospray ion source.** *Anal Chem* 1996, **68:**1-8.
17. West-Nielsen M, Hogdall EV, Marchiori E, Hogdall CK, Schou C, Heegaard NH: **Sample handling for mass spectrometric proteomic investigations of human sera.** *Anal Chem* 2005, **77:**5114-5123.
18. Hubel A, Spindler R, Skubitz AP: **Storage of human biospecimens: selection of the optimal storage temperature.** *Biopreservation and biobanking* 2014, **12:**165-175.

19. Landi MT, Caporaso N: **Sample collection, processing and storage.** *IARC scientific publications* 1997**:**223-236.
20. Prieto DA, Johann DJ, Jr., Wei BR, Ye X, Chan KC, Nissley DV, Simpson RM, Citrin DE, Mackall CL, Linehan WM, Blonder J: **Mass spectrometry in cancer biomarker research: a case for immunodepletion of abundant blood-derived proteins from clinical tissue specimens.** *Biomarkers in medicine* 2014, **8:**269-286.
21. Corthals GL, Wasinger VC, Hochstrasser DF, Sanchez JC: **The dynamic range of protein expression: a challenge for proteomic research.** *Electrophoresis* 2000, **21:**1104-1115.
22. Gundry RL, White MY, Nogee J, Tchernyshyov I, Van Eyk JE: **Assessment of albumin removal from an immunoaffinity spin column: critical implications for proteomic examination of the albuminome and albumin-depleted samples.** *Proteomics* 2009, **9:**2021-2028.
23. Han X, Jin M, Breuker K, McLafferty FW: **Extending top-down mass spectrometry to proteins with masses greater than 200 kilodaltons.** *Science* 2006, **314:**109-112.
24. Thomas CE, Kelleher NL, Mizzen CA: **Mass spectrometric characterization of human histone H3: a bird's eye view.** *J Proteome Res* 2006, **5:**240-247.
25. Nelsestuen GL, Harvey SB, Zhang Y, Kasthuri RS, Sinaiko AR, Ely EW, Bernard GR, Homoncik M, Jilma B: **Top-down proteomic analysis by MALDI-TOF profiling: Concentration-independent biomarkers.** *Proteomics Clinical applications* 2008, **2:**158-166.
26. Kasthuri RS, Wroblewski M, Jilma B, Key NS, Nelsestuen GL: **Potential biomarkers of an exaggerated response to endotoxemia.** *Biomarkers : biochemical indicators of exposure, response, and susceptibility to chemicals* 2007, **12:**287-302.
27. Zhang Y, Wroblewski M, Hertz MI, Wendt CH, Cervenka TM, Nelsestuen GL: **Analysis of chronic lung transplant rejection by MALDI-TOF profiles of bronchoalveolar lavage fluid.** *Proteomics* 2006, **6:**1001-1010.
28. Jr GW, Cazares LH, Leung SM, Nasim S, Adam BL, Yip TT, Schellhammer PF, Gong L, Vlahou A: **Proteinchip(R) surface enhanced laser desorption/ionization (SELDI) mass spectrometry: a novel protein biochip technology for detection of prostate cancer biomarkers in complex protein mixtures.** *Prostate cancer and prostatic diseases* 1999, **2:**264-276.
29. Washburn MP, Wolters D, Yates JR, 3rd: **Large-scale analysis of the yeast proteome by multidimensional protein identification technology.** *Nature biotechnology* 2001, **19:**242-247.
30. Yates JR, 3rd: **Mass spectrometry and the age of the proteome.** *Journal of mass spectrometry : JMS* 1998, **33:**1-19.
31. Gundry RL, White MY, Murray CI, Kane LA, Fu Q, Stanley BA, Van Eyk JE: **Preparation of proteins and peptides for mass spectrometry analysis in a bottom-up proteomics workflow.** *Current protocols in molecular biology / edited by Frederick M Ausubel [et al]* 2009, **Chapter 10:**Unit10 25.
32. Kim SC, Chen Y, Mirza S, Xu Y, Lee J, Liu P, Zhao Y: **A clean, more efficient method for in-solution digestion of protein mixtures without detergent or urea.** *J Proteome Res* 2006, **5:**3446-3452.
33. Wisniewski JR, Zougman A, Nagaraj N, Mann M: **Universal sample preparation method for proteome analysis.** *Nat Methods* 2009, **6:**359-362.

34. Gharib SA, Nguyen E, Altemeier WA, Shaffer SA, Doneanu CE, Goodlett DR, Schnapp LM: **Of mice and men: comparative proteomics of bronchoalveolar fluid.** *Eur Respir J* 2010, **35:**1388-1395.
35. 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.
36. Bhargava M, Becker TL, Viken KJ, Jagtap PD, Dey S, Steinbach MS, Wu B, Kumar V, Bitterman PB, Ingbar DH, Wendt CH: **Proteomic profiles in acute respiratory distress syndrome differentiates survivors from non-survivors.** *PLoS One* 2014, **9:**e109713.
37. Piersma SR, Warmoes MO, de Wit M, de Reus I, Knol JC, Jimenez CR: **Whole gel processing procedure for GeLC-MS/MS based proteomics.** *Proteome science* 2013, **11:**17.
38. Chang DW, Hayashi S, Gharib SA, Vaisar T, King ST, Tsuchiya M, Ruzinski JT, Park DR, Matute-Bello G, Wurfel MM, et al: **Proteomic and computational analysis of bronchoalveolar proteins during the course of the acute respiratory distress syndrome.** *American journal of respiratory and critical care medicine* 2008, **178:**701-709.
39. Gygi SP, Rochon Y, Franza BR, Aebersold R: **Correlation between protein and mRNA abundance in yeast.** *Molecular and cellular biology* 1999, **19:**1720-1730.
40. Yang F, Shen Y, Camp DG, 2nd, Smith RD: **High-pH reversed-phase chromatography with fraction concatenation for 2D proteomic analysis.** *Expert review of proteomics* 2012, **9:**129-134.
41. Wang Y, Yang F, Gritsenko MA, Wang Y, Clauss T, Liu T, Shen Y, Monroe ME, Lopez-Ferrer D, Reno T, et al: **Reversed-phase chromatography with multiple fraction concatenation strategy for proteome profiling of human MCF10A cells.** *Proteomics* 2011, **11:**2019-2026.
42. Mueller DR, Voshol H, Waldt A, Wiedmann B, Van Oostrum J: **LC-MALDI MS and MS/MS--an efficient tool in proteome analysis.** *Sub-cellular biochemistry* 2007, **43:**355-380.
43. Bhargava M, Dey S, Becker T, Steinbach M, Wu B, Lee SM, Higgins L, Kumar V, Bitterman PB, Ingbar DH, Wendt CH: **Protein expression profile of rat type two alveolar epithelial cells during hyperoxic stress and recovery.** *Am J Physiol Lung Cell Mol Physiol* 2013, **305:**L604-614.
44. Eng JK, McCormack AL, Yates JR: **An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database.** *Journal of the American Society for Mass Spectrometry* 1994, **5:**976-989.
45. Perkins DN, Pappin DJ, Creasy DM, Cottrell JS: **Probability-based protein identification by searching sequence databases using mass spectrometry data.** *Electrophoresis* 1999, **20:**3551-3567.
46. Taylor JA, Johnson RS: **Sequence database searches via de novo peptide sequencing by tandem mass spectrometry.** *Rapid Commun Mass Spectrom* 1997, **11:**1067-1075.
47. Zhang B, Chambers MC, Tabb DL: **Proteomic parsimony through bipartite graph analysis improves accuracy and transparency.** *J Proteome Res* 2007, **6:**3549-3557.
48. Nesvizhskii AI, Aebersold R: **Interpretation of shotgun proteomic data: the protein inference problem.** *Mol Cell Proteomics* 2005, **4:**1419-1440.
49. Gupta N, Pevzner PA: **False discovery rates of protein identifications: a strike against the two-peptide rule.** *J Proteome Res* 2009, **8:**4173-4181.

50. 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.
51. Kall L, Storey JD, MacCoss MJ, Noble WS: **Assigning significance to peptides identified by tandem mass spectrometry using decoy databases.** *J Proteome Res* 2008, **7:**29-34.
52. Shen C, Sheng Q, Dai J, Li Y, Zeng R, Tang H: **On the estimation of false positives in peptide identifications using decoy search strategy.** *Proteomics* 2009, **9:**194-204.
53. Chelius D, Bondarenko PV: **Quantitative profiling of proteins in complex mixtures using liquid chromatography and mass spectrometry.** *J Proteome Res* 2002, **1:**317-323.
54. Barr JR, Maggio VL, Patterson DG, Jr., Cooper GR, Henderson LO, Turner WE, Smith SJ, Hannon WH, Needham LL, Sampson EJ: **Isotope dilution--mass spectrometric quantification of specific proteins: model application with apolipoprotein A-I.** *Clinical chemistry* 1996, **42:**1676-1682.
55. Gygi SP, Rist B, Gerber SA, Turecek F, Gelb MH, Aebersold R: **Quantitative analysis of complex protein mixtures using isotope-coded affinity tags.** *Nature biotechnology* 1999, **17:**994-999.
56. 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.
57. Zhou W, Liotta LA, Petricoin EF: **The spectra count label-free quantitation in cancer proteomics.** *Cancer genomics & proteomics* 2012, **9:**135-142.
58. Tabb DL, MacCoss MJ, Wu CC, Anderson SD, Yates JR, 3rd: **Similarity among tandem mass spectra from proteomic experiments: detection, significance, and utility.** *Anal Chem* 2003, **75:**2470-2477.
59. Xie F, Liu T, Qian WJ, Petyuk VA, Smith RD: **Liquid chromatography-mass spectrometry-based quantitative proteomics.** *J Biol Chem* 2011, **286:**25443-25449.
60. Ong SE, Blagoev B, Kratchmarova I, Kristensen DB, Steen H, Pandey A, Mann M: **Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics.** *Mol Cell Proteomics* 2002, **1:**376-386.
61. Ross PL, Huang YN, Marchese JN, Williamson B, Parker K, Hattan S, Khainovski N, Pillai S, Dey S, Daniels S, et al: **Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents.** *Mol Cell Proteomics* 2004, **3:**1154-1169.
62. Pierce A, Unwin RD, Evans CA, Griffiths S, Carney L, Zhang L, Jaworska E, Lee CF, Blinco D, Okoniewski MJ, et al: **Eight-channel iTRAQ enables comparison of the activity of six leukemogenic tyrosine kinases.** *Mol Cell Proteomics* 2008, **7:**853-863.
63. Dayon L, Hainard A, Licker V, Turck N, Kuhn K, Hochstrasser DF, Burkhard PR, Sanchez JC: **Relative quantification of proteins in human cerebrospinal fluids by MS/MS using 6-plex isobaric tags.** *Anal Chem* 2008, **80:**2921-2931.
64. Guo Y, Singleton PA, Rowshan A, Gucek M, Cole RN, Graham DR, Van Eyk JE, Garcia JG: **Quantitative proteomics analysis of human endothelial cell membrane rafts: evidence of MARCKS and MRP regulation in the sphingosine 1-phosphate-induced barrier enhancement.** *Mol Cell Proteomics* 2007, **6:**689-696.

65. Kuhn E, Wu J, Karl J, Liao H, Zolg W, Guild B: **Quantification of C-reactive protein in the serum of patients with rheumatoid arthritis using multiple reaction monitoring mass spectrometry and 13C-labeled peptide standards.** *Proteomics* 2004, **4:**1175-1186.
66. Tai SS, Bunk DM, White Et, Welch MJ: **Development and evaluation of a reference measurement procedure for the determination of total 3,3',5-triiodothyronine in human serum using isotope-dilution liquid chromatography-tandem mass spectrometry.** *Anal Chem* 2004, **76:**5092-5096.
67. JW H: **A new linear ion trap mass spectrometer.** *Rapid Commun Mass Spectrom* 2002, **16:**512-526.
68. Yocum AK, Chinnaiyan AM: **Current affairs in quantitative targeted proteomics: multiple reaction monitoring-mass spectrometry.** *Brief Funct Genomic Proteomic* 2009, **8:**145-157.
69. Yocum AK, Khan AP, Zhao R, Chinnaiyan AM: **Development of selected reaction monitoring-MS methodology to measure peptide biomarkers in prostate cancer.** *Proteomics* 2010, **10:**3506-3514.
70. Bereman MS, MacLean B, Tomazela DM, Liebler DC, MacCoss MJ: **The development of selected reaction monitoring methods for targeted proteomics via empirical refinement.** *Proteomics* 2012, **12:**1134-1141.
71. Gray RD, MacGregor G, Noble D, Imrie M, Dewar M, Boyd AC, Innes JA, Porteous DJ, Greening AP: **Sputum proteomics in inflammatory and suppurative respiratory diseases.** *American journal of respiratory and critical care medicine* 2008, **178:**444-452.
72. Kipnis E, Hansen K, Sawa T, Moriyama K, Zurawel A, Ishizaka A, Wiener-Kronish J: **Proteomic analysis of undiluted lung epithelial lining fluid.** *Chest* 2008, **134:**338-345.
73. 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.** *American journal of physiology Lung cellular and molecular physiology* 2004, **286:**L1095-1104.
74. Griese M, Noss J, von Bredow C: **Protein pattern of exhaled breath condensate and saliva.** *Proteomics* 2002, **2:**690-696.
75. Nicholas B, Skipp P, Mould R, Rennard S, Davies DE, O'Connor CD, Djukanovic R: **Shotgun proteomic analysis of human-induced sputum.** *Proteomics* 2006, **6:**4390-4401.
76. Wattiez R, Falmagne P: **Proteomics of bronchoalveolar lavage fluid.** *Journal of chromatography B, Analytical technologies in the biomedical and life sciences* 2005, **815:**169-178.
77. Lenz AG, Meyer B, Weber H, Maier K: **Two-dimensional electrophoresis of dog bronchoalveolar lavage fluid proteins.** *Electrophoresis* 1990, **11:**510-513.
78. Lindahl M, Stahlbom B, Tagesson C: **Two-dimensional gel electrophoresis of nasal and bronchoalveolar lavage fluids after occupational exposure.** *Electrophoresis* 1995, **16:**1199-1204.
79. Wattiez R, Hermans C, Bernard A, Lesur O, Falmagne P: **Human bronchoalveolar lavage fluid: two-dimensional gel electrophoresis, amino acid microsequencing and identification of major proteins.** *Electrophoresis* 1999, **20:**1634-1645.
80. Wattiez R, Hermans C, Cruyt C, Bernard A, Falmagne P: **Human bronchoalveolar lavage fluid protein two-dimensional database: study of interstitial lung diseases.** *Electrophoresis* 2000, **21:**2703-2712.

81. Bell DY, Hook GE: **Pulmonary alveolar proteinosis: analysis of airway and alveolar proteins.** *The American review of respiratory disease* 1979, **119:**979-990.
82. Sabounchi-Schutt F, Astrom J, Eklund A, Grunewald J, Bjellqvist B: **Detection and identification of human bronchoalveolar lavage proteins using narrow-range immobilized pH gradient DryStrip and the paper bridge sample application method.** *Electrophoresis* 2001, **22:**1851-1860.
83. Noel-Georis I, Bernard A, Falmagne P, Wattiez R: **Database of bronchoalveolar lavage fluid proteins.** *Journal of chromatography B, Analytical technologies in the biomedical and life sciences* 2002, **771:**221-236.
84. 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.
85. Au JS, Cho WC, Yip TT, Yip C, Zhu H, Leung WW, Tsui PY, Kwok DL, Kwan SS, Cheng WW, et al: **Deep proteome profiling of sera from never-smoked lung cancer patients.** *Biomedicine & pharmacotherapy = Biomedecine & pharmacotherapie* 2007, **61:**570-577.
86. Li SQ, Qi HW, Wu CG, Zhang XJ, Yang SG, Zhao X, Wu Z, Wang Y, Que HP, Liu SJ: **Comparative proteomic study of acute pulmonary embolism in a rat model.** *Proteomics* 2007, **7:**2287-2299.
87. Nguyen EV, Gharib SA, Schnapp LM, Goodlett DR: **Shotgun MS proteomic analysis of bronchoalveolar lavage fluid in normal subjects.** *Proteomics Clinical applications* 2014.
88. Levitt JE, Gould MK, Ware LB, Matthay MA: **The pathogenetic and prognostic value of biologic markers in acute lung injury.** *Journal of intensive care medicine* 2009, **24:**151-167.
89. Chen X, Shan Q, Jiang L, Zhu B, Xi X: **Quantitative proteomic analysis by iTRAQ for identification of candidate biomarkers in plasma from acute respiratory distress syndrome patients.** *Biochem Biophys Res Commun* 2013, **441:**1-6.
90. Gerszten RE, Accurso F, Bernard GR, Caprioli RM, Klee EW, Klee GG, Kullo I, Laguna TA, Roth FP, Sabatine M, et al: **Challenges in translating plasma proteomics from bench to bedside: update from the NHLBI Clinical Proteomics Programs.** *Am J Physiol Lung Cell Mol Physiol* 2008, **295:**L16-22.
91. Ranieri VM, Rubenfeld GD, Thompson BT, Ferguson ND, Caldwell E, Fan E, Camporota L, Slutsky AS: **Acute respiratory distress syndrome: the Berlin Definition.** *JAMA : the journal of the American Medical Association* 2012, **307:**2526-2533.
92. Ferguson ND, Fan E, Camporota L, Antonelli M, Anzueto A, Beale R, Brochard L, Brower R, Esteban A, Gattinoni L, et al: **The Berlin definition of ARDS: an expanded rationale, justification, and supplementary material.** *Intensive care medicine* 2012, **38:**1573-1582.
93. 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.
94. Rubenfeld GD, Caldwell E, Peabody E, Weaver J, Martin DP, Neff M, Stern EJ, Hudson LD: **Incidence and outcomes of acute lung injury.** *N Engl J Med* 2005, **353:**1685-1693.
95. Villar J, Blanco J, Anon JM, Santos-Bouza A, Blanch L, Ambros A, Gandia F, Carriedo D, Mosteiro F, Basaldua S, et al: **The ALIEN study: incidence and**

outcome of acute respiratory distress syndrome in the era of lung protective ventilation. *Intensive Care Med* 2011, **37:**1932-1941.
96. Force ADT, Ranieri VM, Rubenfeld GD, Thompson BT, Ferguson ND, Caldwell E, Fan E, Camporota L, Slutsky AS: **Acute respiratory distress syndrome: the Berlin Definition.** *JAMA* 2012, **307:**2526-2533.
97. de Torre C, Ying SX, Munson PJ, Meduri GU, Suffredini AF: **Proteomic analysis of inflammatory biomarkers in bronchoalveolar lavage.** *Proteomics* 2006, **6:**3949-3957.
98. Gessner C, Dihazi H, Brettschneider S, Hammerschmidt S, Kuhn H, Eschrich K, Keller T, Engelmann L, Sack U, Wirtz H: **Presence of cytokeratins in exhaled breath condensate of mechanical ventilated patients.** *Respiratory medicine* 2008, **102:**299-306.
99. Gillet LC, Navarro P, Tate S, Rost H, Selevsek N, Reiter L, Bonner R, Aebersold R: **Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis.** *Mol Cell Proteomics* 2012, **11:**O111 016717.
100. Bailey DJ, Rose CM, McAlister GC, Brumbaugh J, Yu P, Wenger CD, Westphall MS, Thomson JA, Coon JJ: **Instant spectral assignment for advanced decision tree-driven mass spectrometry.** *Proc Natl Acad Sci U S A* 2012, **109:**8411-8416.
101. Ashbaugh DG, Bigelow DB, Petty TL, Levine BE: **Acute respiratory distress in adults.** *Lancet* 1967, **2:**319-323.
102. Suratt BT, Parsons PE: **Mechanisms of acute lung injury/acute respiratory distress syndrome.** *Clin Chest Med* 2006, **27:**579-589; abstract viii.
103. Guerin C, Reignier J, Richard JC, Beuret P, Gacouin A, Boulain T, Mercier E, Badet M, Mercat A, Baudin O, et al: **Prone Positioning in Severe Acute Respiratory Distress Syndrome.** *The New England journal of medicine* 2013.
104. Panoskaltsis-Mortari A, Ingbar DH, Jung P, Haddad IY, Bitterman PB, Wangensteen OD, Farrell CL, Lacey DL, Blazar BR: **KGF pretreatment decreases B7 and granzyme B expression and hastens repair in lungs of mice after allogeneic BMT.** *Am J Physiol Lung Cell Mol Physiol* 2000, **278:**L988-999.
105. Cohen MJ, Brohi K, Calfee CS, Rahn P, Chesebro BB, Christiaans SC, Carles M, Howard M, Pittet JF: **Early release of high mobility group box nuclear protein 1 after severe trauma in humans: role of injury severity and tissue hypoperfusion.** *Critical care* 2009, **13:**R174.
106. Gallagher DC, Parikh SM, Balonov K, Miller A, Gautam S, Talmor D, Sukhatme VP: **Circulating angiopoietin 2 correlates with mortality in a surgical population with acute lung injury/adult respiratory distress syndrome.** *Shock* 2008, **29:**656-661.
107. Bersten AD, Hunt T, Nicholas TE, Doyle IR: **Elevated plasma surfactant protein-B predicts development of acute respiratory distress syndrome in patients with acute respiratory failure.** *American journal of respiratory and critical care medicine* 2001, **164:**648-652.
108. Katayama M, Ishizaka A, Sakamoto M, Fujishima S, Sekiguchi K, Asano K, Betsuyaku T, Kotani T, Ware LB, Matthay MA, Hashimoto S: **Laminin gamma2 fragments are increased in the circulation of patients with early phase acute lung injury.** *Intensive Care Med* 2010, **36:**479-486.
109. Calfee CS, Gallagher D, Abbott J, Thompson BT, Matthay MA: **Plasma angiopoietin-2 in clinical acute lung injury: prognostic and pathogenetic significance.** *Critical care medicine* 2012, **40:**1731-1737.

110. Fremont RD, Koyama T, Calfee CS, Wu W, Dossett LA, Bossert FR, Mitchell D, Wickersham N, Bernard GR, Matthay MA, et al: **Acute lung injury in patients with traumatic injuries: utility of a panel of biomarkers for diagnosis and pathogenesis.** *The Journal of trauma* 2010, **68:**1121-1127.
111. Ware LB, Fremont RD, Bastarache JA, Calfee CS, Matthay MA: **Determining the aetiology of pulmonary oedema by the oedema fluid-to-plasma protein ratio.** *The European respiratory journal* 2010, **35:**331-337.
112. Calfee CS, Ware LB, Eisner MD, Parsons PE, Thompson BT, Wickersham N, Matthay MA, Network NA: **Plasma receptor for advanced glycation end products and clinical outcomes in acute lung injury.** *Thorax* 2008, **63:**1083-1089.
113. Calfee CS, Ware LB, Glidden DV, Eisner MD, Parsons PE, Thompson BT, Matthay MA: **Use of risk reclassification with multiple biomarkers improves mortality prediction in acute lung injury.** *Critical care medicine* 2011, **39:**711-717.
114. Suratt BT, Eisner MD, Calfee CS, Allard JB, Whittaker LA, Engelken DT, Petty JM, Trimarchi T, Gauthier L, Parsons PE: **Plasma granulocyte colony-stimulating factor levels correlate with clinical outcomes in patients with acute lung injury.** *Critical care medicine* 2009, **37:**1322-1328.
115. Eisner MD, Parsons P, Matthay MA, Ware L, Greene K, Acute Respiratory Distress Syndrome N: **Plasma surfactant protein levels and clinical outcomes in patients with acute lung injury.** *Thorax* 2003, **58:**983-988.
116. McClintock DE, Starcher B, Eisner MD, Thompson BT, Hayden DL, Church GD, Matthay MA: **Higher urine desmosine levels are associated with mortality in patients with acute lung injury.** *American journal of physiology Lung cellular and molecular physiology* 2006, **291:**L566-571.
117. Calfee CS, Eisner MD, Parsons PE, Thompson BT, Conner ER, Jr., Matthay MA, Ware LB: **Soluble intercellular adhesion molecule-1 and clinical outcomes in patients with acute lung injury.** *Intensive care medicine* 2009, **35:**248-257.
118. Cheng IW, Ware LB, Greene KE, Nuckton TJ, Eisner MD, Matthay MA: **Prognostic value of surfactant proteins A and D in patients with acute lung injury.** *Crit Care Med* 2003, **31:**20-27.
119. Parsons PE, Eisner MD, Thompson BT, Matthay MA, Ancukiewicz M, Bernard GR, Wheeler AP: **Lower tidal volume ventilation and plasma cytokine markers of inflammation in patients with acute lung injury.** *Critical care medicine* 2005, **33:**1-6; discussion 230-232.
120. Parsons PE, Matthay MA, Ware LB, Eisner MD: **Elevated plasma levels of soluble TNF receptors are associated with morbidity and mortality in patients with acute lung injury.** *American journal of physiology Lung cellular and molecular physiology* 2005, **288:**L426-431.
121. Ware LB, Eisner MD, Thompson BT, Parsons PE, Matthay MA: **Significance of von Willebrand factor in septic and nonseptic patients with acute lung injury.** *American journal of respiratory and critical care medicine* 2004, **170:**766-772.
122. Ware LB, Matthay MA, Parsons PE, Thompson BT, Januzzi JL, Eisner MD: **Pathogenetic and prognostic significance of altered coagulation and fibrinolysis in acute lung injury/acute respiratory distress syndrome.** *Critical care medicine* 2007, **35:**1821-1828.
123. McClintock DE, Ware LB, Eisner MD, Wickersham N, Thompson BT, Matthay MA: **Higher urine nitric oxide is associated with improved outcomes in**

patients with acute lung injury. *American journal of respiratory and critical care medicine* 2007, **175:**256-262.
124. Cribbs SK, Martin GS: **Biomarkers in acute lung injury: are we making progress?** *Critical care medicine* 2008, **36:**2457-2459.
125. Knaus WA, Draper EA, Wagner DP, Zimmerman JE: **APACHE II: a severity of disease classification system.** *Crit Care Med* 1985, **13:**818-829.
126. Timothy Griffin JC, James Johnson, Ebbing de Jong, Getiria Onsongo and Pratik Jagtap: **Galaxy-P: Transforming MS-based proteomic informatics via innovative workflow development, dissemination, standardization and transparency.** In *Book Galaxy-P: Transforming MS-based proteomic informatics via innovative workflow development, dissemination, standardization and transparency* (Editor ed.^eds.). City; 2013.
127. Vizcaino JA, Deutsch EW, Wang R, Csordas A, Reisinger F, Rios D, Dianes JA, Sun Z, Farrah T, Bandeira N, et al: **ProteomeXchange provides globally coordinated proteomics data submission and dissemination.** *Nature biotechnology* 2014, **32:**223-226.
128. Storey JD: **A direct approach to false discovery rates.** *J R Statist Soc* 2002, **64:**479-498.
129. Huang da W, Sherman BT, Lempicki RA: **Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources.** *Nat Protoc* 2009, **4:**44-57.
130. Kosanam H, Sato M, Batruch I, Smith C, Keshavjee S, Liu M, Diamandis EP: **Differential proteomic analysis of bronchoalveolar lavage fluid from lung transplant patients with and without chronic graft dysfunction.** *Clin Biochem* 2012, **45:**223-230.
131. 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.
132. 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.
133. Farjanel J, Hartmann DJ, Guidet B, Luquel L, Offenstadt G: **Four markers of collagen metabolism as possible indicators of disease in the adult respiratory distress syndrome.** *The American review of respiratory disease* 1993, **147:**1091-1099.
134. Clark JG, Milberg JA, Steinberg KP, Hudson LD: **Type III procollagen peptide in the adult respiratory distress syndrome. Association of increased peptide levels in bronchoalveolar lavage fluid with increased risk for death.** *Ann Intern Med* 1995, **122:**17-23.
135. Idell S, Koenig KB, Fair DS, Martin TR, McLarty J, Maunder RJ: **Serial abnormalities of fibrin turnover in evolving adult respiratory distress syndrome.** *The American journal of physiology* 1991, **261:**L240-248.
136. Ware LB, Fang X, Matthay MA: **Protein C and thrombomodulin in human acute lung injury.** *Am J Physiol Lung Cell Mol Physiol* 2003, **285:**L514-521.
137. Ware LB, Bastarache JA, Wang L: **Coagulation and fibrinolysis in human acute lung injury--new therapeutic targets?** *Keio J Med* 2005, **54:**142-149.
138. Prabhakaran P, Ware LB, White KE, Cross MT, Matthay MA, Olman MA: **Elevated levels of plasminogen activator inhibitor-1 in pulmonary edema**

fluid are associated with mortality in acute lung injury. *Am J Physiol Lung Cell Mol Physiol* 2003, **285:**L20-28.
139. Quinlan GJ, Chen Y, Evans TW, Gutteridge JM: **Iron signalling regulated directly and through oxygen: implications for sepsis and the acute respiratory distress syndrome.** *Clin Sci (Lond)* 2001, **100:**169-182.
140. Connelly KG, Moss M, Parsons PE, Moore EE, Moore FA, Giclas PC, Seligman PA, Repine JE: **Serum ferritin as a predictor of the acute respiratory distress syndrome.** *American journal of respiratory and critical care medicine* 1997, **155:**21-25.
141. Ghio AJ, Carter JD, Richards JH, Richer LD, Grissom CK, Elstad MR: **Iron and iron-related proteins in the lower respiratory tract of patients with acute respiratory distress syndrome.** *Critical care medicine* 2003, **31:**395-400.
142. Lagan AL, Quinlan GJ, Mumby S, Melley DD, Goldstraw P, Bellingan GJ, Hill MR, Briggs D, Pantelidis P, du Bois RM, et al: **Variation in iron homeostasis genes between patients with ARDS and healthy control subjects.** *Chest* 2008, **133:**1302-1311.
143. Determann RM, Millo JL, Waddy S, Lutter R, Garrard CS, Schultz MJ: **Plasma CC16 levels are associated with development of ALI/ARDS in patients with ventilator-associated pneumonia: a retrospective observational study.** *BMC pulmonary medicine* 2009, **9:**49.
144. Lesur O, Langevin S, Berthiaume Y, Legare M, Skrobik Y, Bellemare JF, Levy B, Fortier Y, Lauzier F, Bravo G, et al: **Outcome value of Clara cell protein in serum of patients with acute respiratory distress syndrome.** *Intensive Care Med* 2006, **32:**1167-1174.
145. Kropski JA, Fremont RD, Calfee CS, Ware LB: **Clara cell protein (CC16), a marker of lung epithelial injury, is decreased in plasma and pulmonary edema fluid from patients with acute lung injury.** *Chest* 2009, **135:**1440-1447.
146. Berryman M, Franck Z, Bretscher A: **Ezrin is concentrated in the apical microvilli of a wide variety of epithelial cells whereas moesin is found primarily in endothelial cells.** *J Cell Sci* 1993, **105 (Pt 4):**1025-1043.
147. Niggli V, Rossy J: **Ezrin/radixin/moesin: versatile controllers of signaling molecules and of the cortical cytoskeleton.** *Int J Biochem Cell Biol* 2008, **40:**344-349.
148. Mangeat P, Roy C, Martin M: **ERM proteins in cell adhesion and membrane dynamics.** *Trends Cell Biol* 1999, **9:**187-192.
149. Arpin M, Chirivino D, Naba A, Zwaenepoel I: **Emerging role for ERM proteins in cell adhesion and migration.** *Cell Adh Migr* 2011, **5:**199-206.
150. Okayama T, Kikuchi S, Ochiai T, Ikoma H, Kubota T, Ichikawa D, Fujiwara H, Okamoto K, Sakakura C, Sonoyama T, et al: **Attenuated response to liver injury in moesin-deficient mice: impaired stellate cell migration and decreased fibrosis.** *Biochim Biophys Acta* 2008, **1782:**542-548.
151. Doi Y, Itoh M, Yonemura S, Ishihara S, Takano H, Noda T, Tsukita S: **Normal development of mice and unimpaired cell adhesion/cell motility/actin-based cytoskeleton without compensatory up-regulation of ezrin or radixin in moesin gene knockout.** *J Biol Chem* 1999, **274:**2315-2321.
152. Hashimoto S, Amaya F, Matsuyama H, Ueno H, Kikuchi S, Tanaka M, Watanabe Y, Ebina M, Ishizaka A, Tsukita S: **Dysregulation of lung injury and repair in moesin-deficient mice treated with intratracheal bleomycin.** *Am J Physiol Lung Cell Mol Physiol* 2008, **295:**L566-574.

153. Torii K, Iida K, Miyazaki Y, Saga S, Kondoh Y, Taniguchi H, Taki F, Takagi K, Matsuyama M, Suzuki R: **Higher concentrations of matrix metalloproteinases in bronchoalveolar lavage fluid of patients with adult respiratory distress syndrome.** *Am J Respir Crit Care Med* 1997, **155:**43-46.
154. Kong MY, Li Y, Oster R, Gaggar A, Clancy JP: **Early elevation of matrix metalloproteinase-8 and -9 in pediatric ARDS is associated with an increased risk of prolonged mechanical ventilation.** *PLoS One* 2011, **6:**e22596.
155. Koeppen M, McNamee EN, Brodsky KS, Aherne CM, Faigle M, Downey GP, Colgan SP, Evans CM, Schwartz DA, Eltzschig HK: **Detrimental role of the airway mucin Muc5ac during ventilator-induced lung injury.** *Mucosal immunology* 2013, **6:**762-775.
156. Esteban A, Frutos-Vivar F, Muriel A, Ferguson ND, Penuelas O, Abraira V, Raymondos K, Rios F, Nin N, Apezteguia C, et al: **Evolution of mortality over time in patients receiving mechanical ventilation.** *Am J Respir Crit Care Med* 2013, **188:**220-230.
157. Wang CY, Calfee CS, Paul DW, Janz DR, May AK, Zhuo H, Bernard GR, Matthay MA, Ware LB, Kangelaris KN: **One-year mortality and predictors of death among hospital survivors of acute respiratory distress syndrome.** *Intensive Care Med* 2014, **40:**388-396.
158. Greene KE, Ye S, Mason RJ, Parsons PE: **Serum surfactant protein-A levels predict development of ARDS in at-risk patients.** *Chest* 1999, **116:**90S-91S.
159. Ishizaka A, Matsuda T, Albertine KH, Koh H, Tasaka S, Hasegawa N, Kohno N, Kotani T, Morisaki H, Takeda J, et al: **Elevation of KL-6, a lung epithelial cell marker, in plasma and epithelial lining fluid in acute respiratory distress syndrome.** *Am J Physiol Lung Cell Mol Physiol* 2004, **286:**L1088-1094.
160. Uchida T, Shirasawa M, Ware LB, Kojima K, Hata Y, Makita K, Mednick G, Matthay ZA, Matthay MA: **Receptor for advanced glycation end-products is a marker of type I cell injury in acute lung injury.** *Am J Respir Crit Care Med* 2006, **173:**1008-1015.
161. McClintock D, Zhuo H, Wickersham N, Matthay MA, Ware LB: **Biomarkers of inflammation, coagulation and fibrinolysis predict mortality in acute lung injury.** *Crit Care* 2008, **12:**R41.
162. Park WY, Goodman RB, Steinberg KP, Ruzinski JT, Radella F, 2nd, 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.
163. Suter PM, Suter S, Girardin E, Roux-Lombard P, Grau GE, Dayer JM: **High bronchoalveolar levels of tumor necrosis factor and its inhibitors, interleukin-1, interferon, and elastase, in patients with adult respiratory distress syndrome after trauma, shock, or sepsis.** *Am Rev Respir Dis* 1992, **145:**1016-1022.
164. Meduri GU, Headley S, Kohler G, Stentz F, Tolley E, Umberger R, Leeper K: **Persistent elevation of inflammatory cytokines predicts a poor outcome in ARDS. Plasma IL-1 beta and IL-6 levels are consistent and efficient predictors of outcome over time.** *Chest* 1995, **107:**1062-1073.
165. Donnelly SC, Strieter RM, Kunkel SL, Walz A, Robertson CR, Carter DC, Grant IS, Pollok AJ, Haslett C: **Interleukin-8 and development of adult respiratory distress syndrome in at-risk patient groups.** *Lancet* 1993, **341:**643-647.

166. Takala A, Jousela I, Takkunen O, Kautiainen H, Jansson SE, Orpana A, Karonen SL, Repo H: **A prospective study of inflammation markers in patients at risk of indirect acute lung injury.** *Shock* 2002, **17:**252-257.
167. Bajwa EK, Khan UA, Januzzi JL, Gong MN, Thompson BT, Christiani DC: **Plasma C-reactive protein levels are associated with improved outcome in ARDS.** *Chest* 2009, **136:**471-480.
168. Calfee CS, Eisner MD, Ware LB, Thompson BT, Parsons PE, Wheeler AP, Korpak A, Matthay MA, Acute Respiratory Distress Syndrome Network NHL, Blood I: **Trauma-associated lung injury differs clinically and biologically from acute lung injury due to other clinical disorders.** *Crit Care Med* 2007, **35:**2243-2250.
169. Ware LB, Koyama T, Billheimer DD, Wu W, Bernard GR, Thompson BT, Brower RG, Standiford TJ, Martin TR, Matthay MA, Network NACT: **Prognostic and pathogenetic value of combining clinical and biochemical indices in patients with acute lung injury.** *Chest* 2010, **137:**288-296.
170. Pepe MS, Thompson ML: **Combining diagnostic test results to increase accuracy.** *Biostatistics* 2000, **1:**123-140.
171. Bhargava M, Wendt CH: **Biomarkers in acute lung injury.** *Translational research : the journal of laboratory and clinical medicine* 2012, **159:**205-217.
172. Bhargava M, Higgins LA, Wendt CH, Ingbar DH: **Applicaion of Clinical Proteomics in Acute Respiratory Distress Syndrome.** *Clincal and Translational Medicine* 2014, **3**.
173. Reilly C, Cervenka T, Hertz MI, Becker T, Wendt CH: **Human neutrophil peptide in lung chronic allograft dysfunction.** *Biomarkers : biochemical indicators of exposure, response, and susceptibility to chemicals* 2011, **16:**663-669.
174. Lin-Moshier Y, Sebastian PJ, Higgins L, Sampson ND, Hewitt JE, Marchant JS: **Re-evaluation of the role of calcium homeostasis endoplasmic reticulum protein (CHERP) in cellular calcium signaling.** *J Biol Chem* 2013, **288:**355-367.
175. Champion MM, Williams EA, Kennedy GM, Champion PA: **Direct detection of bacterial protein secretion using whole colony proteomics.** *Mol Cell Proteomics* 2012, **11:**596-604.
176. Y B: **Discovering the False Discovery Rate.** *J R Statist Soc B* 2010, **72:**405-416.
177. Moraes TJ, Zurawska JH, Downey GP: **Neutrophil granule contents in the pathogenesis of lung injury.** *Current opinion in hematology* 2006, **13:**21-27.
178. Ward PA: **The dark side of C5a in sepsis.** *Nature reviews Immunology* 2004, **4:**133-142.
179. Sun S, Zhao G, Liu C, Wu X, Guo Y, Yu H, Song H, Du L, Jiang S, Guo R, et al: **Inhibition of complement activation alleviates acute lung injury induced by highly pathogenic avian influenza H5N1 virus infection.** *Am J Respir Cell Mol Biol* 2013, **49:**221-230.
180. Coughlin SR: **Thrombin signalling and protease-activated receptors.** *Nature* 2000, **407:**258-264.
181. Jose RJ, Williams AE, Chambers RC: **Proteinase-activated receptors in fibroproliferative lung disease.** *Thorax* 2014, **69:**190-192.
182. Heit C, Jackson BC, McAndrews M, Wright MW, Thompson DC, Silverman GA, Nebert DW, Vasiliou V: **Update of the human and mouse SERPIN gene superfamily.** *Human genomics* 2013, **7:**22.

183. Askew DJ, Silverman GA: **Intracellular and extracellular serpins modulate lung disease.** *Journal of perinatology : official journal of the California Perinatal Association* 2008, **28 Suppl 3:**S127-135.
184. Zhuo L, Kimata K: **Structure and function of inter-alpha-trypsin inhibitor heavy chains.** *Connective tissue research* 2008, **49:**311-320.
185. Lim YP, Bendelja K, Opal SM, Siryaporn E, Hixson DC, Palardy JE: **Correlation between mortality and the levels of inter-alpha inhibitors in the plasma of patients with severe sepsis.** *J Infect Dis* 2003, **188:**919-926.
186. Garantziotis S, Hollingsworth JW, Ghanayem RB, Timberlake S, Zhuo L, Kimata K, Schwartz DA: **Inter-alpha-trypsin inhibitor attenuates complement activation and complement-induced lung injury.** *J Immunol* 2007, **179:**4187-4192.
187. Kaarteenaho-Wiik R, Lakari E, Soini Y, Pollanen R, Kinnula VL, Paakko P: **Tenascin expression and distribution in pleural inflammatory and fibrotic diseases.** *The journal of histochemistry and cytochemistry : official journal of the Histochemistry Society* 2000, **48:**1257-1268.
188. Fassler R, Sasaki T, Timpl R, Chu ML, Werner S: **Differential regulation of fibulin, tenascin-C, and nidogen expression during wound healing of normal and glucocorticoid-treated mice.** *Exp Cell Res* 1996, **222:**111-116.
189. Kaarteenaho-Wiik R, Mertaniemi P, Sajanti E, Soini Y, Paakko P: **Tenascin is increased in epithelial lining fluid in fibrotic lung disorders.** *Lung* 1998, **176:**371-380.
190. Sharma S, Nadrous HF, Peters SG, Tefferi A, Litzow MR, Aubry MC, Afessa B: **Pulmonary complications in adult blood and marrow transplant recipients: autopsy findings.** *Chest* 2005, **128:**1385-1392.
191. Chi AK, Soubani AO, White AC, Miller KB: **An update on pulmonary complications of hematopoietic stem cell transplantation.** *Chest* 2013, **144:**1913-1922.
192. Cordonnier C, Bernaudin JF, Bierling P, Huet Y, Vernant JP: **Pulmonary complications occurring after allogeneic bone marrow transplantation. A study of 130 consecutive transplanted patients.** *Cancer* 1986, **58:**1047-1054.
193. Afessa B: **Pretransplant pulmonary evaluation of the blood and marrow transplant recipient.** *Chest* 2005, **128:**8-10.
194. Afessa B, Abdulai RM, Kremers WK, Hogan WJ, Litzow MR, Peters SG: **Risk factors and outcome of pulmonary complications after autologous hematopoietic stem cell transplant.** *Chest* 2012, **141:**442-450.
195. Nusair S, Breuer R, Shapira MY, Berkman N, Or R: **Low incidence of pulmonary complications following nonmyeloablative stem cell transplantation.** *Eur Respir J* 2004, **23:**440-445.
196. Weisdorf DJ: **Diffuse alveolar hemorrhage: an evolving problem?** *Leukemia* 2003, **17:**1049-1050.
197. Spitzer TR: **Engraftment syndrome following hematopoietic stem cell transplantation.** *Bone Marrow Transplant* 2001, **27:**893-898.
198. Diab KJ, Yu Z, Wood KL, Shmalo JA, Sheski FD, Farber MO, Wilkes DS, Nelson RP, Jr.: **Comparison of pulmonary complications after nonmyeloablative and conventional allogeneic hematopoietic cell transplant.** *Biol Blood Marrow Transplant* 2012, **18:**1827-1834.
199. Panoskaltsis-Mortari A, Griese M, Madtes DK, Belperio JA, Haddad IY, Folz RJ, Cooke KR: **An official American Thoracic Society research statement: noninfectious lung injury after hematopoietic stem cell transplantation:**

idiopathic pneumonia syndrome. *American journal of respiratory and critical care medicine* 2011, **183:**1262-1279.
200. Fukuda T, Hackman RC, Guthrie KA, Sandmaier BM, Boeckh M, Maris MB, Maloney DG, Deeg HJ, Martin PJ, Storb RF, Madtes DK: **Risks and outcomes of idiopathic pneumonia syndrome after nonmyeloablative and conventional conditioning regimens for allogeneic hematopoietic stem cell transplantation.** *Blood* 2003, **102:**2777-2785.
201. Keates-Baleeiro J, Moore P, Koyama T, Manes B, Calder C, Frangoul H: **Incidence and outcome of idiopathic pneumonia syndrome in pediatric stem cell transplant recipients.** *Bone Marrow Transplant* 2006, **38:**285-289.
202. Huisman C, van der Straaten HM, Canninga-van Dijk MR, Fijnheer R, Verdonck LF: **Pulmonary complications after T-cell-depleted allogeneic stem cell transplantation: low incidence and strong association with acute graft-versus-host disease.** *Bone Marrow Transplant* 2006, **38:**561-566.
203. Patriarca F, Skert C, Bonifazi F, Sperotto A, Fili C, Stanzani M, Zaja F, Cerno M, Geromin A, Bandini G, et al: **Effect on survival of the development of late-onset non-infectious pulmonary complications after stem cell transplantation.** *Haematologica* 2006, **91:**1268-1272.
204. Yanik G, Hellerstedt B, Custer J, Hutchinson R, Kwon D, Ferrara JL, Uberti J, Cooke KR: **Etanercept (Enbrel) administration for idiopathic pneumonia syndrome after allogeneic hematopoietic stem cell transplantation.** *Biol Blood Marrow Transplant* 2002, **8:**395-400.
205. Wong R, Rondon G, Saliba RM, Shannon VR, Giralt SA, Champlin RE, Ueno NT: **Idiopathic pneumonia syndrome after high-dose chemotherapy and autologous hematopoietic stem cell transplantation for high-risk breast cancer.** *Bone Marrow Transplant* 2003, **31:**1157-1163.
206. Crawford SW, Fisher L: **Predictive value of pulmonary function tests before marrow transplantation.** *Chest* 1992, **101:**1257-1264.
207. Yanik GA, Horowitz MM, Weisdorf DJ, Logan BR, Ho VT, Soiffer RJ, Carter SL, Wu J, Wingard JR, Difronzo NL, et al: **Randomized, double-blind, placebo-controlled trial of soluble tumor necrosis factor receptor: enbrel (etanercept) for the treatment of idiopathic pneumonia syndrome after allogeneic stem cell transplantation: blood and marrow transplant clinical trials network protocol.** *Biol Blood Marrow Transplant* 2014, **20:**858-864.
208. Tizon R, Frey N, Heitjan DF, Tan KS, Goldstein SC, Hexner EO, Loren A, Luger SM, Reshef R, Tsai D, et al: **High-dose corticosteroids with or without etanercept for the treatment of idiopathic pneumonia syndrome after allo-SCT.** *Bone Marrow Transplant* 2012, **47:**1332-1337.
209. Yanik GA, Grupp SA, Pulsipher MA, Levine JE, Schultz KR, Wall DA, Langholz B, Dvorak CC, Alangaden K, Goyal RK, et al: **TNF receptor inhibitor therapy for the treatment of children with idiopathic pneumonia syndrome (IPS). A joint Pediatric Blood and Marrow Transplant Consortium (PBMTC) and Children's Oncology Group (COG) study (ASCT0521).** *Biol Blood Marrow Transplant* 2014.
210. Yanik GA, Ho VT, Levine JE, White ES, Braun T, Antin JH, Whitfield J, Custer J, Jones D, Ferrara JL, Cooke KR: **The impact of soluble tumor necrosis factor receptor etanercept on the treatment of idiopathic pneumonia syndrome after allogeneic hematopoietic stem cell transplantation.** *Blood* 2008, **112:**3073-3081.
211. Jagtap P, Bandhakavi S, Higgins L, McGowan T, Sa R, Stone MD, Chilton J, Arriaga EA, Seymour SL, Griffin TJ: **Workflow for analysis of high mass**

accuracy salivary data set using MaxQuant and ProteinPilot search algorithm. *Proteomics* 2012, **12:**1726-1730.

212. Schlatzer DM, Dazard JE, Ewing RM, Ilchenko S, Tomcheko SE, Eid S, Ho V, Yanik G, Chance MR, Cooke KR: **Human biomarker discovery and predictive models for disease progression for idiopathic pneumonia syndrome following allogeneic stem cell transplantation.** *Mol Cell Proteomics* 2012, **11:**M111 015479.
213. Shaik FB, Prasad DV, Narala VR: **Role of farnesoid X receptor in inflammation and resolution.** *Inflammation research : official journal of the European Histamine Research Society [et al]* 2014.
214. Wang YD, Chen WD, Wang M, Yu D, Forman BM, Huang W: **Farnesoid X receptor antagonizes nuclear factor kappaB in hepatic inflammatory response.** *Hepatology* 2008, **48:**1632-1643.
215. Vavassori P, Mencarelli A, Renga B, Distrutti E, Fiorucci S: **The bile acid receptor FXR is a modulator of intestinal innate immunity.** *J Immunol* 2009, **183:**6251-6261.
216. Zhang L, Li T, Yu D, Forman BM, Huang W: **FXR protects lung from lipopolysaccharide-induced acute injury.** *Molecular endocrinology* 2012, **26:**27-36.
217. He F, Li J, Mu Y, Kuruba R, Ma Z, Wilson A, Alber S, Jiang Y, Stevens T, Watkins S, et al: **Downregulation of endothelin-1 by farnesoid X receptor in vascular endothelial cells.** *Circulation research* 2006, **98:**192-199.
218. Kahsai AW, Zhu S, Fenteany G: **G protein-coupled receptor kinase 2 activates radixin, regulating membrane protrusion and motility in epithelial cells.** *Biochim Biophys Acta* 2009.

Supplementary Tables

Chapter 3

TableS1: BALF proteins identified using iTRAQ MS/MS. doi:10.1371/journal.pone.0109713.s001

Table S2 Differentially expressed proteins between early phase survivors and non survivors. doi:10.1371/journal.pone.0109713.s002

Table S3 Differentially expressed proteins between early and late phase survivors. doi:10.1371/journal.pone.0109713.s003

The remaining Supplemental Tables can be obtained by contacting Dr. David Ingbar or me.

Chapter 4

Table S4 Table S4_ARDS 1-6 Protein List

Table S5 Table S5_PAT_Master_Aligned_High_Control

Table S6 Table S6_FDR<5%

Table S7 Table S7_IPA_Biological processes

Table S8 Table S8_GO David Analysis

Chapter 5

Table S9 Table S9_Labeling_strategy_BMT

Table S10 Table S10_Protein LIst 1% FDR alias

Table S11 Table S11_Mikes_list_ Common protiens in all ITRAQs alias

Table S12 Table S12_5L_5G_Low_Control PAT Master_Aligned_High_Control

Table S13 Table S13_bmt-res1 copy alias

Table S14 Table S14_GO clusters with protein-fold change copy alias