Searching for Protein Biomarkers of Disease in Bronchoalveolar Lavage Fluid

Elizabeth Vi Nguyen

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2012

David Goodlett, Chair

Carlos Catalano

Sina Gharib

Lynn Schnapp

Program Authorized to Offer Degree:

Medicinal Chemistry

University of Washington

Abstract

Proteomic Profiling of Bronchoalveolar Lavage Fluid in Respiratory Disease

Elizabeth Vi Nguyen

Chair of the Supervisory Committee:

Professor David R. Goodlett

Medicinal Chemistry

Respiratory diseases minimize gas exchange at the alveoli and complicate the mechanics of breathing. Bronchoalveolar lavage fluid (BALF) is capable of sampling the components of the alveolar epithelial lining fluid (ELF) and determining the composition of the pulmonary airways. Comprehensively profiling the proteomic landscape of BALF provides a means to understanding basic pathogenic mechanisms with the potential to develop techniques for early diagnosis in diseases concerned with lung function. Combining shotgun proteomic analysis with computational methods is a powerful approach to generate large and integrated sets of data independent of investigator's biases. This integrative methodology identifies distinct protein signatures that could be utilized as diagnostic

classifiers, provide insights into the pathogenesis of lung disorder, and elucidate unsuspected mediators or pathways indicative of lung diseases.

Ventilator-associated pneumonia (VAP) is defined as pneumonia occurring after 48 hours of mechanical ventilation. It is a major cause of morbidity and mortality in critically ill patients. VAP affects 8-28% of patients receiving mechanical ventilation and has a mortality rate ranging from 24% to as high as 76% in certain settings. Clinical diagnosis of VAP is challenging since the classic symptoms of pneumonia, i.e., fever, abnormal radiographs, and elevated blood cell counts are not specific for VAP among ventilated patients. Inaccurate diagnosis of VAP leads to incorrect treatment and subsequent complications related to therapy.

Additionally, the lung is an important reservoir of human immunodeficiency virus (HIV) and site of HIV replication. HIV is uniformly detected in alveolar macrophages of pediatric patients and in the majority of adult patients. In addition to the numerous opportunistic infections that affect the lung, HIV directly causes significant pulmonary pathology, including lymphoid interstitial pneumonitis (LIP), nonspecific pneumonitis (NSIP), or lymphocytic alveolitis. LIP occurs in up to 75% of the pediatric population. Most HIV-infected adults develop an asymptomatic lymphocytic alveolitis, which can progress to NSIP or LIP. There is strong evidence that HIV itself evokes a local immune response that facilitates lymphocytic migration and infiltration in the lung. The pathogenesis of non-infectious pulmonary complications, following HIV infection is not fully understood.

Protein profiles of BALF were determined in populations of normal, HIV and ALI

($VAP^{+/-}$) volunteers. Comparative analysis identified a set of protein signatures specific to disease states. A limited protein classifier was identified and validated in the VAP study which may aid in future diagnostic endeavors. In addition, the protein profiles distinctive to HIV suggest a role that HIV-induced immunosuppression plays in the development of lung complications.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1

Lung Immune Function

1.1 How immunity works

The function of an immune system is to recognize "self" from "non-self" and eliminate pathogens that may compromise the well-being of the organism. It is essential for the immune system to detect a wide variety of entities (e.g., viruses, bacteria, germs, pollutants, parasites) and rid the body of these foreign agents. The success of a pathogen is dependent on how fast it is capable of evolving and adapting in order to mount an effective anti-immune response within the infected host. There are numerous mechanism pathogens use to evade the immune system's natural defense including but not limited to: secretion of toxins, exhibition of modulators on the surface of the pathogens, become latent or inhibit phagocytosis, subvert or kill immune cells, inhibit apoptosis, and/or inhibit acquired or complement immunity [1]. The later immunity will be discussed in detail. To counteract the pathogens strategic ability to evolve, immune systems of varying organisms have developed multiple defense mechanisms to recognize and neutralize pathogens.

The first line of defense are physical barriers to protect the organism (e.g., skin in *Homo sapiens*). Coughing, sneezing, tearing, urinating, and mucus secretions are mechanical barriers that trap and/or expel foreign agents from orifices to which pathogens and irritants may enter the body [2]. Antimicrobial peptides, enzymes, and pH fluctuation are chemical defenses that provide additional barriers for protection. For

example the skin and respiratory tract secrete antimicrobial peptide β-defensins [3] and the stomach produces gastric acid and digestive enzymes to protect against ingested pathogens [4]. Biological barriers such as the commensal flora in gastrointestinal and urinary tracts are other examples of protective barriers [4].

If these protective barriers are infiltrated an immune response is subsequently initiated. Undifferentiated stems cells from the bone marrow mature into granulocyte/macrophage or lymphoid progenitor cells. These progenitor cells can differentiate into macrophages or lymphocytes both referred to as leukocytes (i.e., white blood cells) [5]. These leukocytes are the immune response whose function is to distinguish and rid of "non-self" molecules that are a threat to the organism. The immune response breaks down into two complementary systems: innate (i.e., cellular) and adaptive (i.e., humoral) [6]. Innate immunity is the first to be activated and is usually sufficient to clear pathogens; however, the response is non-specific. When it is overwhelmed, the adaptive immune system is triggered. The adaptive immune system produces a specific immune response where antibodies (i.e., immunoglobulins) are produced that can "remember" the pathogen in case of future exposure. The adaptive immune system will be discussed later.

The innate immune system is not as well studied as the humoral immune system. The main components are T cells (i.e., T lymphocytes), named because their later stages of development occur in the thymus. T cells can mature into cytotoxic T cells (i.e., killer T cells) and helper T cells. These cells respond to receptors on host cells exhibiting that the cell has been compromised. Receptors known to trigger responses are the well-

established major histocompatibility complex (MHC I and II) [5] and pattern-recognition receptors (PRRs) [7].

MHC proteins bind peptide fragments of proteins digested in the cell and present them on the outside surface of the cell. Cellular proteins are normally digested inside a cell and the peptides are presented by MHC receptors; however, if a pathogen invades a cell then the peptide fragments will be presented by MHC receptors signaling that the cell is "non-self," also referred to as an "antigen-presenting" cell. There are two classes of MHC proteins which differ in structure, distribution in cell types, and the type of digested proteins whose peptides they display. MHC type I proteins are found on the surface of nucleated cells in all vertebrates and present endogenous antigens and activate CD8+ cytotoxic T-cells (killer T cells) which kill the cells. MHC type II occur on surfaces of a few types of specialized cells such as dendritic cells and B-cells. They present exogenous antigens and activate CD4+ helper T cells releasing cytokines that will either 1) help the cell mature, 2) spur the growth of more T-cells, and/or 3) attract macrophages, neutrophils, or lymphocytes [8]. MHC I are associated with the innate immune system while MHC II are involved in the beginning steps of the adaptive immune systems.

PRRs recognize conserved microbial signatures termed pathogen-associated molecular patterns (PAMPs) [9] as well as molecules released during cell damage designated as damage-associated molecular patterns (DAMPs) [10, 11]. Pathogens evolve rapidly and can evade the innate immune system by altering PAMPs; hence, it is essential for the immune system to recognize PAMPs that are vital to pathogens that are highly conserved or less prone to modifications [12]. PRR consist of Toll-like receptors (TLRs) which recognize extracellular microbes [13], NOD-like receptors (NLRs) senses

intracellular microbes [14], RIG-like helicases (RLH) senses viral RNA in the cytoplasm [15], and C-type lectin receptors (CLR) which are found on most classes of human pathogen [16].

The main component of the adaptive immune system are B-cells (i.e., B lymphocytes) that produce antibodies. Once a B cell encounters its antigen, it will receive signals from CD4+ T helper cells produced by MHC type II protein. There are two types of CD4+ T helper cells, Th1 and Th2, designed to eliminate different types of pathogens and trigger different adaptive responses. Th1 release interferon-gamma and induces B-cells to make opsonizing antibodies that induces innate immunity [17]. Th2 release interleukin 4 and signals for B cells to further differentiate into an effector cell referred to as a plasma cells. [18]. These cells secrete antibodies specific to the antigen triggering phagocytosis and the complement cascade. 10% of these effector cells become long-lived, “remembering” the antigen in case of future exposure.

Inflammation is the initial response of the innate immune system. It functions to move plasma and leukocytes from the blood into injured tissues and serves to remove injurious stimuli and initiate the healing process [2]. When PRRs recognize PAMPs, cytokines (i.e., IFN-ɣ, TNF-α, and IL-1) and inflammatory mediators are released . Vasodilation increases blood flow to site of infection leading to exudation of plasma proteins and fluid into the compromised tissue. The complement system and coagulation/fibrinolysis systems (to heal wounds) are also activated in parallel to initiate and propagate the inflammatory response [19].

The complement system contains over 30 different proteins and is a critical link between the innate and humoral response system [20]. The functions of the complement system is 1) opsonisation of antigens, 2) chemotaxis and 3) cell lysis. It is broken down into three pathways: 1) Classical is activated by antigen-antibody binding (humoral system), 2) Alternative is activated by serine protease and many of the surfaces of pathogens, and 3) Lectin is activated by microbial carbohydrates and ficolins. All these pathways converge at the cleavage of complement 3 protein (C3) which in turn produces the membrane attack complex (MAC) referred to as C5b-9 [21]. The MAC induces cell death by forming a transmembrane channel causing lysis of the cells by osmosis.

1.2 Lung Anatomy

The lung is the essential respiration organ whose principal function is to transport oxygen from the atmosphere and release carbon dioxide from the bloodstream at the alveoli. Air enters through the nasal or oral cavity traveling to the larynx, trachea, branching into two bronchi, which further branch into bronchioles like an upside down tree, into alveolar sacs composed of a cluster of alveoli [4]. Each individual alveolus is wrapped with blood capillaries. The membranes of the alveolar wall and capillaries compose the alveolar-capillary (also referred to as blood-air) barrier [22]. It is here where oxygen diffuses from the alveolar space into blood and is exchanged for carbon dioxide in hemoglobins.

Inspiration occurs when involuntary muscular contraction of the diaphragm pulls the diaphragm downward and increases the volume space in the lung. The pressure is decreased in the lung and causes air to flow in. Expiration occurs when the diaphragm

relaxes and pushes the air out [4]. Breathing is an essential function of humans and is constantly exposing the body to external factors (i.e., inhaled particles, airborne pollutants, pathogens, and infectious agents). The lung is the organ that protects organisms from foreign agents during respiration. It is the body's first line of immune defense (physical barrier) to foreign antigens that may be introduced from the external environment. Coughing and sneezing are mechanical barriers used to eject pathogens and other irritants from the respiratory tract [23].

The bronchi is also composed of a specialized epithelium lined with cilia. On top of the epithelium is a thin fluid film of mucus with an additional viscous film where microorganisms are trapped. The cilia propels the mucus layer in the direction of the pharynx where it is either swallowed or expelled via coughing, this process is referred to as the mucociliary transport system [24]. The environment of the lung is very moist and hospitable for pathogens and these are processes aimed to rid of microorganism on an everyday basis. Respiratory diseases are a common and important cause of illnesses and death. Bacterial or viral infections of the lungs results in inflammation that can result in narrowing of lung anatomy (e.g., bronchioles), edema, disruption of air-gas exchange, and complicate the mechanics of breathing.

1.3 Lung Inflammatory Response

The immune system of the respiratory tract is in constant demand to protect from the external environment. Pathogens can also spread to the lung through blood, in which case they first contact the alveolar-capillary membrane first. The lung has adopted a specified immune response to defend against foreign antigens in a manner that does not

interfere with its primary biological functions. The immune system must keep the inflammatory response in check and yet be prepared to respond quickly to microorganisms. If the inflammatory response is not carefully regulated or inadequate to clear the organism of invading pathogens, then chronic inflammation occurs which can lead to permanent damage to lung tissues and function. Here, the initial mediators of inflammation become detrimental. Studies have demonstrated that the respiratory tract maintains an environment that tends to decrease inflammation [25]. Inflammation can quickly compromise the lung environment; hence, mechanisms are in place for protection and control.

If the mechanical barriers of sneezing, coughing, and the mucociliary transport system are penetrated, the innate immune response is activated. This initial response of inflammation is known as acute. Neutrophils (derived from granulocyte/macrophage progenitor cells) migrate from blood vessels, interstitial tissues, to the site of infection through a process called chemotaxis [26]. The neutrophils bind to up-regulated adhesions molecules P- and E-selectins, intracellular adhesion molecules-1 (ICAM-1), and vascular cell adhesion molecule-1 (VCAM-1) on the surfaces of the vascular endothelial [27]. It is only the activated endothelium that participates in the inflammatory response. Neutrophils will not adhere to the unperturbed endothelium, only to areas expressing these relevant adhesion molecules. Adhered neutrophils then increase their level of receptors that bind to antibodies and complement-coated infected cells or pathogens and induce phagocytosis, these receptors are named Fc receptors [28] . At the same time, macrophages residing in the lungs upregulate adhesions molecules needed for chemotaxis further promoting vascular neutrophil migration [29].

Lung macrophages, epithelial cells, and fibroblasts produce cytokines important in inflammation after 30-90 minutes of exposure to an invading pathogen or pollutant [30]. The primary cytokines responsible for acute inflammation include tumor necrosis factor α (TNFα), interleukin 1 (IL-1), and interleukin 6 (IL-6) [31]. TNFα and IL-1 induces dendritic cells (DC) and antigen presenting cells such as macrophages to migrate to infected area, initiating an adaptive immune response [32, 33]. The release of these cytokines activates an inflammatory cascade of mediators including cytokines, lipid mediators, cell adhesion molecules, and acute phase proteins. TNFα and IL-1 results in the production of IL-6 from a broad range of lung cells [30]. IL-6 decreases the release of IL-1 and TNFα and the influx of inflammatory cells [34, 35]. High levels of TNFα and IL-1 can cause severe damage to the lung tissues and systemic disease compromising the organism [36]. Much of the death resulting from bacterial infection is attributed to chronic inflammation and septic shock has been linked to increase TNFα production [30, 37]. Another important function of IL-6 is inducing the adaptive immune response by differentiating B-cells to plasma cells that secrete antigens.

Interleukin 10 (IL-10) and transforming growth factor β (TGF-β) are cytokines known to reduce inflammation and modulate the immune response in the lung [38, 39]. A wide range of inflammatory cytokines from pulmonary tissues have been shown to be inhibited by IL-10 [40]. Deficiencies in IL-10 in the respiratory tract are associated with increased tissue injury and inflammation [25].

Chemokines are known to be involved in the lung inflammatory response but are still under investigation. Chemokines are small polypeptides that control adhesion, chemotaxis, and activation of leukocytes. More than 50 chemokines and 15 receptors

have been identified in mice and humans [41]. The most well studied chemokine is CXCL8 (i.e., IL-8) which has the ability to activate and recruit leukocytes although these mechanism are not well understood [42, 43].

Lung macrophages and type II cells of alveoli secrete surfactants reducing surface tension and the energy required to inflate the lungs. Surfactant A (SP-A) and surfactant D (SP-D) have numerous roles in the lung first-line defense system. Both have a carbohydrate recognition domain that allows them to bind to a variety of pathogens as bacteria, viruses, fungi, yeasts, allergens, and lipopolysaccharides (LPS) for pathogen neutralization and clearance by leukocytes [44]. These lung surfactants have been proposed to inhibit the proinflammatory cytokine production having an anti-inflammatory role in excessive immune response [45]. Studies done on transgenic mice knockout show that SP-D is more susceptible to viral infections while SP-D is susceptible to *Streptococcus* [46, 47]. Surfactant degradation or inactivation may contribute to enhanced susceptibility to lung inflammation and infection.

Distinct collections of lymphoid tissue along the respiratory tract further play a role in adaptive immune response. These tissues include the nasal-associated lymphoid tissue (NALT), bronchus-associated lymphoid tissue (BALT), and lymph nodes that receive drainage from the nose or lung [48-50]. Unlike classical lymph nodes, NALT and BALT is not encapsulated but are integrated directly in the lung tissue [51]. These lymphoid structure sample antigen directly from the lung lumen and secrete antibodies inducing an adaptive immune response [52]. The roles played by NALT, BALT, and the draining lymph nodes in the induction and expression of lung immunity is not entirely clear and needs further investigation.

The “respiratory burst” is another mechanism of the lung to defend against invading pathogens. Here, the cells of the lungs release reactive oxygen species as they come into contact with different microorganism [53-55]. NADPH oxidase on the surface of lung cells produce superoxide which can spontaneously recombine with other molecules to produce free radicals (i.e., hydroxyl, hydrogen peroxide, and hypochlorite) which in turn aids in combating foreign antigens [56]. There is evidence that these oxidants such as hypochlorite can also directly activate matrix metalloproteinases [57]. However because these enzymes are located on the surface of cells, it has been suggested that the activation of these enzymes leads to tissue matrix damaging effects [58, 59]. Another free radical released in the respiratory burst is nitric oxide from the enzyme nitric oxide synthase (NOS) [60-63]. The role of nitric oxide is an enigma and there are contradicting results on its role in the lung inflammatory response. Nitric oxide has been suggested to reduce adhesion of leukocytes to the endothelial cells giving it an anti-inflammatory role [64]. In the presence of superoxide, nitric oxide reacts to form peroxynitrite anion, which reduces nitric oxide concentration and diminishing its anti-adhesive effects making it inflammatory [65].

Investigating the mechanisms of the lung immune system and the associated inflammatory response is important to understanding the diseases associated with lung function. It is important to define lung homeostasis and understand the activation/de-activation of pathways leading to disordered inflammatory and immune responses.

1.4 Notes to Chapter 1

1. Finlay, B.B. and G. McFadden, *Anti-immunology: evasion of the host immune system by bacterial and viral pathogens.* Cell, 2006. **124**(4): p. 767-82.
2. Alberts, B.A.J., Julian Lewis, Martin Raff, Keither Roberts, Peter Walters, *Molecular Biology of the Cell*, ed. Garland. 2002, New York, NY: Garland Science. 1463.
3. Agerberth, B. and G.H. Gudmundsson, *Host antimicrobial defence peptides in human disease.* Curr Top Microbiol Immunol, 2006. **306**: p. 67-90.
4. Marieb EN, H.K., *Human anatomy and physiology*, 2010, Benjamin Cummings: San Francisco.
5. Nelson DL, C.M., *Lehninger Principles of Biochemistry*. Third ed. 2000, New York, NY: Worth Publishers. 1152.
6. Beck, G. and G.S. Habicht, *Immunity and the invertebrates.* Sci Am, 1996. **275**(5): p. 60-3, 66.
7. Medzhitov, R. and C.A. Janeway, Jr., *Decoding the patterns of self and nonself by the innate immune system.* Science, 2002. **296**(5566): p. 298-300.
8. Janeway, C.A., Jr. and R. Medzhitov, *Innate immune recognition.* Annu Rev Immunol, 2002. **20**: p. 197-216.
9. Song, W.Y., et al., *A receptor kinase-like protein encoded by the rice disease resistance gene, Xa21.* Science, 1995. **270**(5243): p. 1804-6.
10. Rubartelli, A. and M.T. Lotze, *Inside, outside, upside down: damage-associated molecular-pattern molecules (DAMPs) and redox.* Trends Immunol, 2007. **28**(10): p. 429-36.

11. Farkas, A.M., T.M. Kilgore, and M.T. Lotze, *Detecting DNA: getting and begetting cancer.* Curr Opin Investig Drugs, 2007. **8**(12): p. 981-6.

12. Janeway, C.A., Jr., *The immune system evolved to discriminate infectious nonself from noninfectious self.* Immunol Today, 1992. **13**(1): p. 11-6.

13. Takeda, K. and S. Akira, *TLR signaling pathways.* Semin Immunol, 2004. **16**(1): p. 3-9.

14. Martinon, F., A. Mayor, and J. Tschopp, *The inflammasomes: guardians of the body.* Annu Rev Immunol, 2009. **27**: p. 229-65.

15. Sumpter, R., Jr., et al., *Regulating intracellular antiviral defense and permissiveness to hepatitis C virus RNA replication through a cellular RNA helicase, RIG-I.* J Virol, 2005. **79**(5): p. 2689-99.

16. Geijtenbeek, T.B. and S.I. Gringhuis, *Signalling through C-type lectin receptors: shaping immune responses.* Nat Rev Immunol, 2009. **9**(7): p. 465-79.

17. Schoenborn, J.R. and C.B. Wilson, *Regulation of interferon-gamma during innate and adaptive immune responses.* Adv Immunol, 2007. **96**: p. 41-101.

18. Sokol, C.L., et al., *A mechanism for the initiation of allergen-induced T helper type 2 responses.* Nat Immunol, 2008. **9**(3): p. 310-8.

19. Cotran, K., *Robbins Patholgic Basis of Disease*. 1998, Philadelphi: W.B Saundes Company.

20. Rus, H., C. Cudrici, and F. Niculescu, *The role of the complement system in innate immunity.* Immunol Res, 2005. **33**(2): p. 103-12.

21. Liszewski, M.K., et al., *Control of the complement system.* Adv Immunol, 1996. **61**: p. 201-83.

22. Kindlen, S., *Physiology for Health Care and Nursing*. 2003, UK: Elsevier Health Sciences.

23. Boyton, R.J. and P.J. Openshaw, *Pulmonary defences to acute respiratory infection.* Br Med Bull, 2002. **61**: p. 1-12.

24. Sleigh, M.A., J.R. Blake, and N. Liron, *The propulsion of mucus by cilia.* Am Rev Respir Dis, 1988. **137**(3): p. 726-41.

25. Raychaudhuri, B., et al., *Interleukin 10 (IL-10)-mediated inhibition of inflammatory cytokine production by human alveolar macrophages.* Cytokine, 2000. **12**(9): p. 1348-55.

26. Downey, G.P., et al., *Neutrophil sequestration and migration in localized pulmonary inflammation. Capillary localization and migration across the interalveolar septum.* Am Rev Respir Dis, 1993. **147**(1): p. 168-76.

27. Imhof, B.A. and D. Dunon, *Leukocyte migration and adhesion.* Adv Immunol, 1995. **58**: p. 345-416.

28. Simms, H.H. and R. D'Amico, *Regulation of intracellular polymorphonuclear leukocyte Fc receptors by lipopolysaccharide.* Cell Immunol, 1994. **157**(2): p. 525-41.

29. Guo, R.F. and P.A. Ward, *Mediators and regulation of neutrophil accumulation in inflammatory responses in lung: insights from the IgG immune complex model.* Free Radic Biol Med, 2002. **33**(3): p. 303-10.

30. Cohen, J., *The immunopathogenesis of sepsis.* Nature, 2002. **420**(6917): p. 885-91.

31. Lazarus, S.C., *Role of inflammation and inflammatory mediators in airways disease.* Am J Med, 1986. **81**(5A): p. 2-7.

32. Zou, G.M. and Y.K. Tam, *Cytokines in the generation and maturation of dendritic cells: recent advances.* Eur Cytokine Netw, 2002. **13**(2): p. 186-99.

33. Stylianou, E. and J. Saklatvala, *Interleukin-1.* Int J Biochem Cell Biol, 1998. **30**(10): p. 1075-9.

34. Aderka, D., J.M. Le, and J. Vilcek, *IL-6 inhibits lipopolysaccharide-induced tumor necrosis factor production in cultured human monocytes, U937 cells, and in mice.* J Immunol, 1989. **143**(11): p. 3517-23.

35. Schindler, R., et al., *Correlations and interactions in the production of interleukin-6 (IL-6), IL-1, and tumor necrosis factor (TNF) in human blood mononuclear cells: IL-6 suppresses IL-1 and TNF.* Blood, 1990. **75**(1): p. 40-7.

36. Parbhakar, O.P., et al., *Depletion of pulmonary intravascular macrophages partially inhibits lipopolysaccharide-induced lung inflammation in horses.* Vet Res, 2005. **36**(4): p. 557-69.

37. Tracey, K.J., et al., *Shock and tissue injury induced by recombinant human cachectin.* Science, 1986. **234**(4775): p. 470-4.

38. Fiorentino, D.F., et al., *IL-10 acts on the antigen-presenting cell to inhibit cytokine production by Th1 cells.* J Immunol, 1991. **146**(10): p. 3444-51.

39. Shull, M.M., et al., *Targeted disruption of the mouse transforming growth factor-beta 1 gene results in multifocal inflammatory disease.* Nature, 1992. **359**(6397): p. 693-9.

40. Lo, C.J., M. Fu, and H.G. Cryer, *Interleukin 10 inhibits alveolar macrophage production of inflammatory mediators involved in adult respiratory distress syndrome.* J Surg Res, 1998. **79**(2): p. 179-84.

41. Nelson, P.J. and A.M. Krensky, *Chemokines, lymphocytes and viruses: what goes around, comes around.* Curr Opin Immunol, 1998. **10**(3): p. 265-70.

42. Morsey, M.A., et al., *Cytokine profiles following interaction between bovine alveolar macrophages and Pasteurella haemolytica.* Microb Pathog, 1999. **26**(6): p. 325-31.

43. Elder, A., et al., *Lung inflammation induced by endotoxin is enhanced in rats depleted of alveolar macrophages with aerosolized clodronate.* Exp Lung Res, 2005. **31**(6): p. 527-46.

44. Vaandrager, A.B. and L.M. van Golde, *Lung surfactant proteins A and D in innate immune defense.* Biol Neonate, 2000. **77 Suppl 1**: p. 9-13.

45. Lawson, P.R. and K.B. Reid, *The roles of surfactant proteins A and D in innate immunity.* Immunol Rev, 2000. **173**: p. 66-78.

46. LeVine, A.M., et al., *Distinct effects of surfactant protein A or D deficiency during bacterial infection on the lung.* J Immunol, 2000. **165**(7): p. 3934-40.

47. LeVine, A.M., et al., *Surfactant protein D enhances clearance of influenza A virus from the lung in vivo.* J Immunol, 2001. **167**(10): p. 5868-73.

48. Pabst, R. and T. Tschernig, *Lymphocytes in the lung: an often neglected cell. Numbers, characterization and compartmentalization.* Anat Embryol (Berl), 1995. **192**(4): p. 293-9.

49. Tschernig, T., W.J. Kleemann, and R. Pabst, *Bronchus-associated lymphoid tissue (BALT) in the lungs of children who had died from sudden infant death syndrome and other causes.* Thorax, 1995. **50**(6): p. 658-60.

50. Yokomise, H., et al., *Importance of intrapulmonary lymph nodes in the differential diagnosis of small pulmonary nodular shadows.* Chest, 1998. **113**(3): p. 703-6.

51. Wu, H.Y. and M.W. Russell, *Nasal lymphoid tissue, intranasal immunization, and compartmentalization of the common mucosal immune system.* Immunol Res, 1997. **16**(2): p. 187-201.

52. Moyron-Quiroz, J.E., et al., *Role of inducible bronchus associated lymphoid tissue (iBALT) in respiratory immunity.* Nat Med, 2004. **10**(9): p. 927-34.

53. Jorens, P.G., et al., *L-arginine-dependent production of nitrogen oxides by rat pulmonary macrophages.* Eur J Pharmacol, 1991. **200**(2-3): p. 205-9.

54. Rimele, T.J., et al., *Interaction of neutrophils with vascular smooth muscle: identification of a neutrophil-derived relaxing factor.* J Pharmacol Exp Ther, 1988. **245**(1): p. 102-11.

55. Bissonnette, E.Y., et al., *Potentiation of tumor necrosis factor-alpha-mediated cytotoxicity of mast cells by their production of nitric oxide.* J Immunol, 1991. **147**(9): p. 3060-5.

56. Oury, T.D., B.J. Day, and J.D. Crapo, *Extracellular superoxide dismutase in vessels and airways of humans and baboons.* Free Radic Biol Med, 1996. **20**(7): p. 957-65.

57. Barnes, P.J., *Reactive oxygen species and airway inflammation.* Free Radic Biol Med, 1990. **9**(3): p. 235-43.

58. Mulligan, M.S., et al., *Tissue injury caused by deposition of immune complexes is L-arginine dependent.* Proc Natl Acad Sci U S A, 1991. **88**(14): p. 6338-42.

59. Ward, P.A. and M.S. Mulligan, *New insights into mechanisms of oxyradical and neutrophil mediated lung injury.* Klin Wochenschr, 1991. **69**(21-23): p. 1009-11.

60. Oury, T.D., B.J. Day, and J.D. Crapo, *Extracellular superoxide dismutase: a regulator of nitric oxide bioavailability.* Lab Invest, 1996. **75**(5): p. 617-36.

61. Goretski, J. and T.C. Hollocher, *Trapping of nitric oxide produced during denitrification by extracellular hemoglobin.* J Biol Chem, 1988. **263**(5): p. 2316-23.

62. Nathan, C., *Nitric oxide as a secretory product of mammalian cells.* FASEB J, 1992. **6**(12): p. 3051-64.

63. Xie, Q.W., et al., *Cloning and characterization of inducible nitric oxide synthase from mouse macrophages.* Science, 1992. **256**(5054): p. 225-8.

64. Ashutosh, K., *Nitric oxide and asthma: a review.* Curr Opin Pulm Med, 2000. **6**(1): p. 21-5.

65. Gaston, B., et al., *The biology of nitrogen oxides in the airways.* Am J Respir Crit Care Med, 1994. **149**(2 Pt 1): p. 538-51.

Chapter 2

Proteomic Analysis of Bronchoalveolar Lavage Fluid in Normal Subjects

This body of work was done in collaboration with Dr. Sina A. Gharib and Dr. Lynn M. Schnapp, Division of Pulmonary and Critical Care Medicine, Harborview Medical Center, University of Washington, Seattle, WA, US

2.1 Introduction

The lung is the body's first line of immune defense to inhaled foreign antigens. The lung proteome is comprised of a multifaceted group of proteins in constant flux as a result of the dynamic contact with environmental compounds [1]. BAL proteins are derived from a variety of sources including resident cells or diffusion (passive and active) through the alveolar-capillary barrier [2]. A better understanding of the normal constituents of the lung proteome is vital in assessing alterations that occur when respiratory diseases disrupt normal lung physiology.

The airways and alveoli are covered with a thin layer of epithelial lining fluid (ELF) composed of soluble components of the lung that are responsible for structural integrity of airspaces, maintaining gas-exchange, and providing immune protection. The protein composition of ELF is affected by external factors and/or diseases affecting the lung; this proteome has been demonstrated to be important in early diagnosis and has provided insight into pathogenesis for a range of lung diseases including idiopathic pulmonary fibrosis, cystic fibrosis, chronic obstructive pulmonary disease and acute respiratory distress syndrome [3]. Various external

factors (i.e., inhaled particles, airborne pollutants, pathogens, and infectious agents) and lung diseases induce detectable biochemical modifications in ELF [4]. The techniques that have been used to sample ELF include bronchoalveolar lavage (BAL) [4], nasal lavage [5], breath condensates collection [6], and the induction of sputum with hypertonic saline [7]. While the last three techniques are non-invasive, the fluids acquired represent a limited subset of the lung proteome likely to contain proteins from sources other than the airspaces. Currently, the most comprehensive sampling of lung proteins can be gained by examination of BAL fluid (BALF) which is acquired by fiberoptic bronchoscopy and represents ELF from the airspaces and small airways.

Proteomic analysis of BALF therefore represents a practical approach to understanding the expressed lung proteome [8-11]. Proteomic analysis of BALF has been studied with the intent to identify biomarkers of disease and discern biochemical mechanisms responsible for a range of lung diseases, including idiopathic pulmonary fibrosis [12], cystic fibrosis [13], sarcoidosis [3], chronic obstructive pulmonary disease (COPD) [14, 15], and acute respiratory distress syndrome (ARDS)[16-18]. In this paper, we review the existing literature describing the proteomic methodologies and their reported results in defining the normal lung BALF proteome.

2.2 Bronchoalveolar Lavage Fluid

Collection

The European Respiratory Society has drafted guidelines on the methods for performing BAL in an effort to standardize the procedure and produce results that are reproducible for comparative analysis [10]. In order to obtain BALF, selected lobes of the lung are washed five

times with 20 ml of sterile saline (0.9% w/v) during fiberoptic bronchoscopy [19]. The fluid is then centrifuged to remove cellular debris resulting in a cell-free supernatant that is referred to as BALF.

Composition

BALF is a complex mixture of soluble components such as phospholipids, neutral lipids, nucleic acids, peptides, and proteins derived from resident cells or diffusion (passive and active) through the alveolar-capillary barrier [2]. Soluble proteins account for 20-30% of surfactant weight and are a component of ELF that may aid in early diagnosis or provide insight into pathogenesis of respiratory diseases [20]. The most abundant proteins detected in BALF are plasma proteins (e.g., 50% albumin, 5.6% transferrin, 3.5% alpha 1-antitrypsin, 30% immunoglobulin A and G) that may originate from diffusion (passive and active) across the blood-air barrier, active enrichment [21] . Most interstitial lung diseases are characterized by enrichment of plasma proteins in BALF as a consequence of increased barrier permeability due to chronic injury and inflammation [22].

The alveolar-capillary barrier is comprised of pneumocytes (type I and type II), epithelial cells, in the alveolar wall and endothelial cells of the capillaries[23]. Type II pneumocytes Lung epithelial cells produce surfactant-associated proteins (SP-A, SP-B, SP-C, SP-D), Clara cell protein (CC-16), and mucin-associated antigens [24]. Detection of these proteins in BALF or serum can implicate changes in epithelial cell function and be used to assess the integrity of the alveolar-capillary barrier.

2.3 Mass spectrometry-based proteomics

The dynamic range of protein abundance in BALF has been estimated to be on the order of 10^{10} [1], yet the resolving power of MS-based proteomic platforms is limited to 10^{2}-10^{4} [25]. As a result of this 6-order of magnitude discrepancy, it has been very difficult to profile all proteins present in BALF. Currently, two main techniques are used to define proteomes using mass spectrometry (Figure 2.1). The first is a 2-step process involving two-dimensional gel electrophoresis (2-DE) to separate proteins followed by tandem mass spectrometry (MS) to identify peptides produced by proteolysis of proteins. The 2-DE approach first separates proteins in a polyacrylamide gel by isoelectric focusing in a pH gradient [26] where proteins migrate in the polyacrylamide gel to their isoelectric point (pI). The second dimension of separation in a polyacrylamide gel is based on molecular weight. The combined separations result in an ability to separate approximately 1000 protein forms. The separated proteins are visualized by various staining methods, such as Coomassie Brilliant Blue or silver stain. Proteins visualized as stained "spots" are identified by excision of the gel piece containing the protein, which is followed by enzymatic digestion and mass spectrometric analysis (MS). Often this MS step is a combination of high performance liquid chromatography (HLPC) and tandem mass spectrometry (MS/MS). The second, and recently more popular, approach involves the enzymatic digestion of unfractionated BALF followed by separation and analysis using the same HPLC-MS/MS method that was also coupled to the 2-DE method to identify proteins in individual "spots". This second method, referred to as shotgun proteomics, bypasses the 2DE protein separation approach in favor of analysis of a single sample--making it more rapid than analyzing hundreds of individual isolated proteins.

Although 2-DE gel is capable of separating sequence isoforms and post-translational modifications of proteins, it has many drawbacks. These include its labor intensive nature, difficulty in reproducing the separations, the need for a label such as fluorescent tags followed by densitometry to quantify protein levels, and poor recovery of proteins from the gel for MS identification [27]. As mentioned, the shotgun proteomic approach has several advantages over the 2-DE-MS coupled method including speed since the sample's proteome is interrogated in a single run instead of hundreds of individual runs. Additional advantages include ease of automation and the ability to quantify proteins in a label free manner [28].

Regardless of whether one uses the 2-DE-MS or shotgun proteomic approach, each approach typically utilizes HPLC-MS/MS where peptides are transferred to the mass spectrometer using electrospray ionization and automated computer driven routines for control of MS operation that allows unattended selection of peptides for collision induced dissociation (CID). This automated selection of peptides for fragmentation is often referred to as a data-dependent acquisition (DDA) because ion selection for fragmentation is based on precursor ion intensity in the MS1 scan. After fragmentation, the resulting tandem mass spectra are searched against a theoretical sequence database. The resulting matches to peptide sequences in the database scores are assigned to each match using various methods, such as [29], Phenyx [30], and SEQUEST [31]. The end result of this process is a list of the most likely protein candidates present in the sample. For the purpose of further discussion we will only consider details of the SEQUEST scoring system because the majority of normal BALF proteins have been identified by SEQUEST.

Briefly, SEQUEST uses three scores to rank the identified peptides and proteins. The cross-correlation value (Xcorr) is a value based on how similar the acquired tandem mass

spectrum is to a theoretical tandem mass spectrum produced from a database of sequence, a Delta-correlation (ΔCn) that represents how far the top Xcorr is from the next highest Xcorr score matched to other theoretical tandem mass spectra, and an Sp score that is the preliminary score based on the number of experimental ions that matched to values of the theoretical peptides for a protein. It is important to note that each tandem mass spectrum has a best fit sequence match out of many other possibly matches. In an attempt to bring clarity to the process of identifying proteins from peptide tandem mass spectra, various statistical tools have been developed to aid interpretation. In the case of SEQUEST two software programs, PeptideProphet [32] and ProteinProphet [33] were developed at the Institute for Systems Biology to assign objectively a score to the best matching peptide sequence. Together they offer a mechanism to determine the most statistically probable peptide sequence match and a statistical model to estimate the likelihood that a peptide is derived from a given parent protein in the event of detection of multiple peptides from a single protein (some of which may also be included in multiple protein sequences). These details may seem esoteric, but in fact play an important role in the resulting list of proteins that are claimed to be "identified" in any given experiment.

2.4 Normal BALF proteome

2DE-MS proteomic methods

Prior to the advent of mass spectrometry for protein identification and coining of the term proteomics [34-38], 2DE methods were used to characterize proteomes. In fact a manuscript published in 1979 described the analysis of BALF samples from normal volunteers and patients with pulmonary alveolar proteinosis [39]. Serum samples were also collected and analyzed in

parallel. BALF proteins that co-migrated (i.e. had the same pI and molecular weight) on the 2DE gel with serum proteins on a separate gel were identified by inference to the serum proteins that had been identified using monospecific antibodies. This comparison of unknown BALF proteins to previously characterized serum proteins on a 2D-PAGE gel indicated that the majority of proteins in unfractionated BALF were in fact serum derived. While 37 of the proteins visualized as dye-stained protein spots were excised for protein analysis, only 19 proteins were confidently identified by inference to migration of standards on identical gels.

The implementation of peptide mass fingerprinting (PMF) [34-38] in the early 1990's as a means to identify proteins directly from polyacrylamide gels yielded a "high-throughput" strategy that resulted in an increase in the number of identified BALF proteins [40]. In the PMF approach each stained protein spot from the 2DE gel is isolated, digested with a protease and analyzed using single stage mass spectrometry that collectively is known as 2-DE-MS. Proteins are identified by comparison of the peptide measured masses to a database of protein sequences theoretically digested with the same protease. Each set of detectable peptide masses represents a unique barcode for a given parent protein with confidence increasing as the number of peptide masses measured increases. The most recent reference gel representing BALF proteomics using this method comprised >900 stained protein spots. Of those 900, 78 of the spots could be identified as specific proteins using the 2-DE PMF technique [41]. Over 50% of the proteins identified in that study were serum-derived while the other 50% of constituents were proteins involved in anti-oxidation, tissue repair, immunological response, inflammation, lipid metabolism, and cytoskeletal functions. Thus, these preliminary studies demonstrate that the BALF proteome contains numerous different protein classes reflecting the diversity of their cellular origins and functions. The fact that 50% of the identified proteins were classified as

serum in origin represents a common conundrum in BALF proteomics. These serum-classified proteins may be present in BALF due to either active or passive transport across the alveoli or they could potentially be produced in the lungs, but having first been classified as serum in origin are thus considered serum proteins. Currently, there is no *a priori* means to determine their origin.

Shotgun proteomic methods

An increasingly popular method for proteome characterization is an approach known as shotgun proteomics. Shotgun proteomic analysis may be coupled to the 2-DE separation of proteins (Figure 2.1), but typically this separation is bypassed. Instead complex protein mixtures are first enzymatically digested into peptides, separated by high performance liquid chromatography, and ionized directly into a tandem mass spectrometer to identify proteins. The method has become popular because of the ease of use and a typical 2-fold increase in proteins identified [42] compared to 2DE methods. Another reason for its popularity over the 2-DE MS approach is that the shotgun method has a much wider detectable dynamic range than gel-based method. This is in part true because it is difficult to detect proteins on a gel with molecular weights outside the 10-150kDa range, insoluble hydrophobic proteins, and/or proteins with high pI values [43]. The detection of low abundant proteins may also be hampered by close proximity on the gel to a higher abundant protein with similar size and charge [44].

We review four manuscripts that used different shotgun proteomics approaches to define the normal BALF proteome. The work of Gharib et al. in 2009 used multidimensional protein identification technology (MudPIT) to separate peptides instead of the 2DE approach to protein

separation. The MudPIT method uses 2D-HPLC coupled directly to tandem mass spectrometry to identify proteins. As with the basic shotgun proteomic method described above, peptides from proteolytic digestion of complex protein mixtures, are first separated on a strong cation-exchange (SCX) support after which multiple fractions are analyzed using a reversed-phase C18 column [44] from which peptides elute and are ionized into a tandem mass spectrometer for fragmentation by CID. In this study BALF from four normal subjects were analyzed in parallel to BALF from eight cystic fibrosis (CF) patients utilizing the MudPIT method on a Finnigan LCQ Deca ProteomeX ion trap mass spectrometer [45]. Peptide tandem mass spectra were searched against the Human IPI database version 3.01. The criterion for protein identification was that peptide tandem mass spectra should have a peptide probability ≥ 0.9 and a protein probability ≥ 0.96, with at least three unique peptides in one sample. These probabilities are built in to the popular tools contained in the trans proteomic pipeline (TPP) available from the Institute for Systems Biology [46, 47]. From this analysis 304 proteins were identified in the normal BALF proteome with the top three most abundant proteins according to spectral counting, which is a means for determining relative quantities directly from peptide tandem mass spectra, being albumin, complement 3, and transferrin.

The next three reports of proteins detected in normal BALF were conducted using an abbreviated form of shotgun proteomics that avoided sample fractionation by SCX, using only C18-based reverse phase chromatography coupled to MS/MS analysis.

In the first of these reports by Chen et al. in 2008, BALF from six normal subjects was analyzed in quadruplicate on a linear ion trap (LTQ) hybrid Fourier transform-ion cylclotron resonance (ICR) mass spectrometer (*i.e.* LTQ-FT produced by ThermoFinnigan, CA) [42]. Tandem mass spectra were searched against the Human IPI database version 2.31 with the

criteria for matching a tandem mass spectrum to a peptide sequence being: Xcorr ≥ 1.9 for charge state 1+, Xcorr ≥ 2.2 for charge state 2+, and Xcorr ≥ 3.75 for charge state 3+ all having a ΔCn ≥ 0.1. Tandem mass spectra passing these criteria were considered matched to the top ranked peptide sequence from the database search using SEQUEST. However, further criteria were implemented including a ProteinProphet probability score ≥ 0.8 and requiring more than one unique tryptic peptide to identify a given protein. Using these criteria over 100 proteins were identified in each individual, 167 unique proteins altogether, and 42 common proteins among all four participants.

The work of Chen et al. also examined three different semi-quantitative methods to assess relative intra- and inter-subject proteome variability. These included: 1) simple protein-sequence coverage, 2) peptide spectral count which is based on statistical ranking of the number of times a given peptide is selected for CID and produces a spectrum of high enough quality for a database match, and 3) an area under the curve (AUC) method that used extracted peptide ion current traces referred to as peptide ion current areas (PICA). All these methods were used to evaluate both intra- and inter-subject variability in the rank order of proteins detected. For the 42 common proteins they found that inter-sample variability ranged between 1- to 6-fold demonstrating a large degree of heterogeneity among different normal human subjects. This shotgun proteomic report identified over a two-fold increase in protein identifications compared to the previously discussed 2-DE MS [48, 49] approach and provided quantitative assessments of intra-and inter-subject variability.

The relatively unbiased and comprehensive nature of shotgun proteomics data sets allow for higher-level functional analysis of identified proteins. Thus, rather than solely identifying single proteins, it is common to provide information on the biochemical functions of identified

proteins found to be statistically relevant by comparison to the entire human proteome in order to represent pathways populating the BALF proteome. One common tool for enrichment analysis known as Database for Annotation, Visualization and Integrated Discovery (DAVID) [50] was used by Chen et al. to functionally annotate the identified proteins based on the Gene Ontology (GO) database (http://www.geneontology.org) [51]. While there was variation in specific proteins identified in each subject with only 42 proteins being common to all, there was significant consistency in the highly enriched functional processes in all six subjects. These functional classifications confirmed known lung functions including immune response, regulation of liquid surface tension, inflammatory response, and response to pathogen, but also identified carbohydrate binding/metabolism as possibly being associated with normal lung function.

Another publication from the same group extended the dynamic range of detection using gas phase fractionation (GPF), and was applied to pooled BALF from four normal human subjects and six mice on the same LTQ-FT mass spectrometer used in the prior report (*i.e.* LTQ-FT ThermoFinnigan, CA) [52]. By fractionating samples in the MS rather than in solution [53], GPF takes advantage of the fact that the mass spectrometer is a very sensitive device with a large dynamic range of 5×10^4 on a standard peptide mixture [54]. As with all shotgun proteomic methods discussed so far, the mass spectrometer is operated in a data-dependent mode where the most intense precursor ions are sequentially isolated and subjected to CID in a series of narrow *m/z* ranges during HPLC-MS/MS analysis. By limiting the *m/z* range for DDA selection of ions to a narrow subset of all available ions, precursor ions of lower intensity, which are masked in a wide *m/z* range experiment by higher intensity precursor ions, are readily selected thus expanding detectable dynamic range and the number of proteins identified.

To achieve full coverage of the available m/z range of the LTQ-FT instrument (350-2,000) the samples were iteratively injected using the following *m/z* ranges for GPF: 350-550, 500-700, 650-850, 800-1,000, 950-1,500 and 1,450-2000. The criteria for a valid protein identification and functional annotation were the same as reported in the prior report of Gharib et al. In this GPF-based study, 91 unique proteins were identified in the pooled human BALF and 117 unique proteins were identified in the mouse BALF. Notable from this study was that functional analysis between human and murine BALF proteomes revealed enrichment in oxidative stress proteins in the mouse compared to the human samples. The authors concluded that using murine models to study human lung function should be undertaken with an understanding of potentially significant differences between their respective BALF proteomes.

The final of the four reported data sets, from Nguyen et al., examined BALF proteomes from five normal subjects each analyzed in triplicate on a hybrid LTQ Velos mass spectrometer (Thermo Fisher, San Jose, CA) with GPF *m/z* ranges based on the theoretical human tryptic peptide proteome. [55]. Unlike the prior version of the LTQ, the LTQ Velos linear ion trap mass spectrometer used in this study incorporated an S-lens ion optic guide and a dual-pressure linear ion trap [56] that together increased the amount of data acquired in the same amount of time over the LTQ-FT. In fact, the rate of CID acquisition in the LTQ Velos is roughly twice that of the prior LTQ ion trap because of the new ion optics. Additionally, the use of a dual ion trap provides some improvement in data quality. Specifically, the first cell is held at a higher pressure (~5 x 10^{-3} Torr) which improves ion trapping, isolation, and fragmentation efficiencies while the second cell is held at lower pressure (4 x 10^{-4} Torr), allowing attainment of higher mass resolution for a given scan rate. The practical result for proteomics is that approximately twice as many scans may be acquired per second of chromatographic time than in a standard ion trap. For

example, a whole cell digest of *Saccharomyces cerevisiae* acquired in triplicate on a LTQ XL linear ion trap and a LTQ Velos dual-pressure linear ion trap mass spectrometer resulted in 2-4 times more medium and abundant unique peptides and proteins [56] identified by the LTQ Velos. This study incorporated a genome-based approach to design of the m/z ranges used in the GPF experiment. This genome-based GPF approach produced *m/z* ranges with approximately equal ion densities based on theoretical distributions of tryptic peptides for the human genome [57]. In theory this approach should result in more efficient data acquisition for each *m/z* range.

A combination of the LTQ Velos (Thermo Fisher, San Jose, CA) and genome-based GPF was used by Nguyen et al. to examine BALF from five normal human subjects in triplicate [[42, 55] using the following three *m/z* ranges: 400-559, 559-846, 846-2000. The resulting tandem mass spectra were searched against Human IPI database version 3.53 with criteria for a valid protein identification and functional annotation being the same as above (section 3.3a) with a further criterion requiring a protein to be detected in 50% of the patient population prior to functional classification. This led to 251 unique proteins identified in the normal BALF proteome, the highest number identified to date. Functional classification of these 251 proteins confirmed the known lung function matching previous results [58] that included inflammatory response, response to wounding, immune response, and carbohydrate binding/metabolism with the most abundant proteins again being albumin, complement 3, and mucin.

2.5 Re-analysis of prior data

While all the studies described above used different technologies and informatics methods to identify proteins spanning nearly ten years, we wondered if this made any difference

in the types of proteins identified or functional categories of proteins identified. Three of the four data sets were publically available and could be re-examined. Unfortunately, the MUDPIT data was not one of these and was heretofore omitted. To remove differences due to varied database search strategies, we re-analyzed data from the following reports: 1) straight DDA without GPF, 2) GPF of pooled samples, and 3) genome-based GPF of individuals. Tandem mass spectra were searched against the Human IPI database version 3.53. In contrast to the original reports, all search criteria for matching a tandem mass spectrum to a peptide sequence were the same, specifically Xcorr $\geq$ 1.9 with charge state 1+, Xcorr $\geq$ 2.2 with charge state 2+, or Xcorr $\geq$ 3.75 with charge state 3+, as well as ΔCn $\geq$ 0.1. Tandem mass spectra passing these criteria were passed to ProteinProphet where a protein was considered to be identified only if a ProteinProphet probability $\geq$ 0.8 was produced and only if more than one unique tryptic peptide was found for each protein. The results generated are summarized in Table 2.1 (See Table 2.2 for protein identifications of re-analysis of 3 provided methods). Functional annotation of the normal BALF proteome was conducted using the DAVID software [51]. Functional enrichment was statistically based on Gene Ontology (GO) annotation relative to the entire human proteome using a variant of the one-tailed Fisher exact probability [59]. Functional categories were deemed significantly "enriched" with a false discovery rate (FDR) cutoff of 0.01. The top 10 biological processes of the normal BALF proteome generated for all four reports are summarized in Figure 2.2 (see Table 2.3 for additional information).

2.6 Discussion

Considerable progress has been made in proteomic analysis of the normal BALF proteome over the last 33 years since Bell and Hook's 2-DE study in 1979. Since then improvements have occurred in protein separation that increased resolving power, data

acquisition strategies that increased detectable dynamic range, and instrumentation that provided more tandem mass spectra. In addition, standardization of BALF collection has aided our current knowledge of the proteomic profile of BALF and inferences of the surrounding airspace. Still there are defined limits to our understanding that are framed by the technology used to describe the BALF proteome of normal individuals. For example, protein identifications obtained by mass spectrometry are dependent on the methods of separation, the data acquisition process, the version (i.e. content and annotation) of database used, statistical data analysis tools used to assign confident identifications, and of course the mass spectrometry technology employed. Hence, it is crucial to understand both the biological and the methodological contexts around which data are produced in order to carry out valid comparisons between datasets. Currently, efforts are underway to provide advice on contextualizing 'metadata' (i.e. data about the data) known as the "minimum information about a proteomics experiment" (MIAPE) guidelines [60, 61], making explicit both where samples came from and how analyses were performed [41]. These efforts are still ongoing and to date there is no defined "minimal description of the experiment" or "validation criteria." Here we attempted to standardize the "validation criteria" for protein identifications among three shotgun MS data from the BALF proteome of normal volunteers. The results drastically differed from the original published results when the validation criteria was adjusted, reaffirming the necessity of a MIAPE guideline. All these efforts will increase confidence in MS-based comparative proteomic investigation, which in turn will provide important insights to mediators and pathways indicative of lung disease.

Presently, all shotgun MS data on normal BALF was generated using what is termed a "bottom-up" proteomic approach. In this method, proteins are digested into peptides that are subsequently fragmented and from which protein identifications are inferred using statistical

analysis of the tandem mass spectra of the fragmented peptides. Although "bottom-up" proteomic approaches are capable of providing a great amount of information on the proteins present in a sample and their quantities, limitations remain. For example, the coverage of the proteome is biased towards high molecular weight proteins that produce a greater number of proteolytic peptides that are in turn more readily detected over smaller proteins. This is why one of the promising next phases of proteomics involves whole protein analysis in what is referred to as "top-down" proteomics [62, 63] where the enzymatic digestion step is excluded and whole intact proteins are introduced into the mass spectrometer for fragmentation.

Herein we have reviewed progress in characterizing the normal human BALF proteome. To date this work included various methods spanning > 30 years including: 2-DE with antibodies, 2-DE-MS with PMF and shotgun proteomic analysis. The first proteomic study, identifying proteins by inference on gels using antibodies, revealed that many of the proteins identified in BALF are plasma proteins [39]. These plasma proteins most likely originated from passive diffusion through the air-blood barrier. More recent proteomic studies, utilizing 2-DE-MS with PMF, further implicated proteins involved in tissue repair and proliferation, cytoskeleton, antioxidant, immunological and inflammation responses, and lipid metabolism [41].

Four different shotgun proteomic approaches each with slight variations on peptide separation technologies, data acquisition, MS instrumentation, and search/database criteria were compared here. In the original studies sample separation with multidimensional liquid chromatography using the MUDPIT method reported the most comprehensive normal BALF proteome while the study of pooled BALF samples using GPF produced the least number of proteins identified. Fractionation, whether performed in liquid phase by MUDPIT or in the gas

phase by GPF, increased protein identifications. However, when the three non MUDPIT data sets were re-analyzed using identical search parameters for matching tandem mass spectra to peptide sequence and criteria for considering protein identification valid, then the results were drastically different (Table 2.1). This re-analysis with identical search conditions for all datasets found that the genome-based GPF method detected twice as many proteins as the non-reanalyzed MUDPIT results. This result was in large part due to the removal of the requirement that a protein be detected in 50% of the patient population in the original report using genome-based GPF that was not applied to the other shotgun proteomic approaches. Furthermore the pooled samples using a standard GPF method reported more than twice the number of proteins after re-analysis in comparison to the original identified proteins. While the data from the straight DDA shotgun proteomic method (no GPF) after re-analysis reported fewer proteins, this is due to the removal of redundant proteins when running SEQUEST against a newer database. These outcomes emphasize the necessity for MIAPE guidelines when reporting the number of proteins since varying MS constraints will alter the perception, if not the reality, of proteins identified.

On the other hand, functional annotation of the top 10 biological processes of all the shotgun proteomic data sets demonstrated that the most enriched processes were relatively the same (Figure 2.2) regardless of the technology used to generate the data. Many of these processes are consistent with the lung's known multi-faceted functions such as response to wounding, inflammatory and immune responses. While advancements in techniques and instrumentation spanning the past five years increased the number of proteins identified, similar functional groups of proteins were identified in all of the shotgun proteomic data. Despite inter-individual heterogeneity and the low identification of common homologous proteins between all shotgun proteomic methods, key processes involved in immunity and inflammatory processes in

the normal BALF proteome were tightly preserved across patients, time, and technology. This indicates that the normal BALF proteome, defined by standard shotgun proteomic approaches, has provided what may be considered now to be a well-established data set of BALF proteins in the normal lung. This collective data set, while of course biased toward high abundance proteins, then should serve as a reliable reference when performing future comparative analyses in various lung disease states.

Finally, we note that the BALF proteome is a very complex mixture and high molecular weight plasma proteins were over-represented in all four data sets examined with albumin and immunoglobin proteins being the most abundant. While immunodepletion is one accepted solution to improving proteomic coverage, the methodology was not used in any of the examined studies because each sought to characterize the native normal BALF proteome. As such the data reviewed here represent an array of proteins, biased toward high abundance ones like albumin, that are found in human BALF. Notably, the focus in these studies to characterize the normal BALF proteome by circumventing depletion of high abundance proteins has produced a normal BALF proteome that lacks proteins such as soluble triggering receptor expressed on myeloid cells 1 (sTREM-1) that are known to be present. Addition of these lower abundance and lower molecular weight proteins will only be possible with enrichment strategies or with novel proteomic technologies.

Critical advances in MS technology (i.e., peptide mass fingerprinting, web-based platforms, GPF, and development of more selective mass spectrometers) have accelerated the field of proteomics. At this state, referencing the normal BALF proteome may provide insight into differential functional categories but improvements in sample preparation, data acquisition, robust statistical methods, and standardizing the criteria for protein identifications will have a

significant impact on biomarker discovery. Despite these challenges, BALF proteomics has already proved to be a promising approach providing insights for various lung diseases.

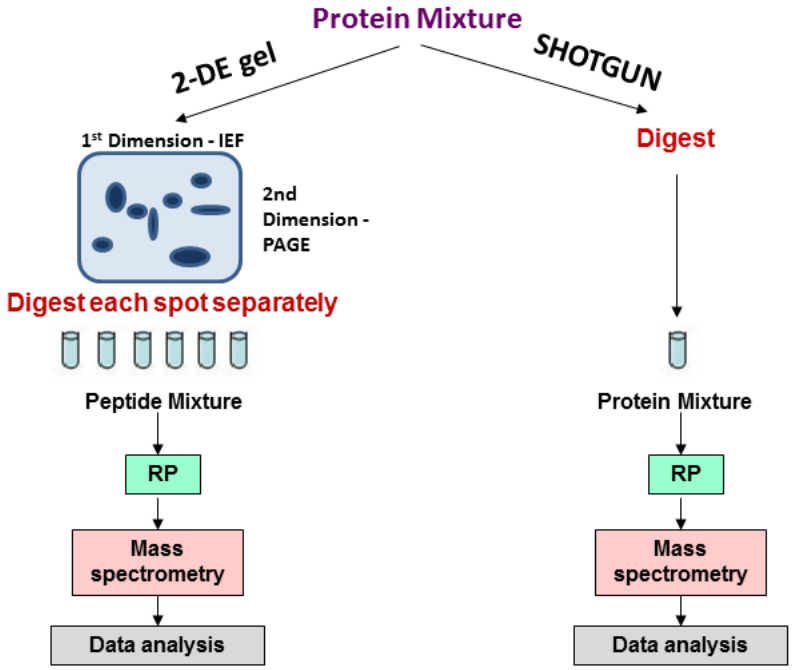

Figure 2.1 General workflow of MS-based proteomic approaches. Sample preparation and data analysis of 2-DE gel (left) and shotgun (right) proteomic MS approaches

Table 2.1 Conditions and protein identifications for four "shotgun" proteomic approaches

	3.3a MUDPIT Gharib et al. in 2009	**3.3b No GPF** Chen et al. in 2008	**3.3c GPF** Gharib et al. in 2010	**3.3d Genome-based GPF** Nguyen et al. in 2012
# of subjects	4	6	4 pooled	5
Instrument	Finnigan LCQ Deca ProteomeX IT	LTQ-FT ThermoFinnigan	LTQ-FT ThermoFinnigan	Hybrid LTQ-Velos
GPF range	None	None	350-550, 500-700, 650-850, 800-1000, 950-1500, 1450-2000	400-559, 559-846, 846-2000
Proteins Identified	**304**	**167**	**91**	**251**
Replicates	2	4	0	3
DB	IPI 3.01	IPI 2.31	IPI 2.31	IPI 3.53
minimum peptides	3	2	2	2
Proteins Identified (Reanalysis Criteria)	N/A	**123**	**205**	**598**

Table 2.2

List of proteins identified in re-analysis of normal BALF proteome in 3 shotgun proteomic approaches

No GPF (123 proteins identified) Chen et al. in 2008		
IPI	**Protein Name**	**Entrez Gene ID**
IPI00004573	polymeric immunoglobulin receptor	5284
IPI00004656	beta-2-microglobulin	567
IPI00006705	secretoglobin, family 1A, member 1 (uteroglobin)	7356
IPI00007047	S100 calcium binding protein A8	6279
IPI00010303	serpin peptidase inhibitor, clade B (ovalbumin), member 4	6318
IPI00019038	lysozyme (renal amyloidosis)	4069
IPI00019359	keratin 9 (epidermolytic palmoplantar keratoderma)	3857
IPI00020091	orosomucoid 2	5005
IPI00021854	apolipoprotein A-II	336
IPI00022394	complement component 1, q subcomponent, C chain	714
IPI00022417	leucine-rich alpha-2-glycoprotein 1	116844
IPI00022429	orosomucoid 1	5004
IPI00022488	hemopexin	3263
IPI00022974	prolactin-induced protein	5304
IPI00027462	S100 calcium binding protein A9	6280
IPI00032294	cystatin S	1472
IPI00066193	secretoglobin, family 3A, member 1	92304
IPI00163563	phosphatidylethanolamine-binding protein 4	157310
IPI00178926	immunoglobulin J polypeptide, linker protein for immunoglobulin alpha and mu polypeptides	3512
IPI00216691	profilin 1	5216
IPI00291005	malate dehydrogenase 1, NAD (soluble)	4190
IPI00291878	surfactant, pulmonary-associated protein D	6441
IPI00296654	bactericidal/permeability-increasing protein-like 1	80341
IPI00297160	CD44 molecule (Indian blood group)	960
IPI00298828	apolipoprotein H (beta-2-glycoprotein I)	350
IPI00299547	lipocalin 2	3934
IPI00301579	Niemann-Pick disease, type C2	10577
IPI00305477	cystatin SN	1469
IPI00306413	tubulin polymerization-promoting protein family member 3	51673
IPI00329801	annexin A5	308
IPI00451624	cartilage acidic protein 1	55118
IPI00555812	group-specific component (vitamin D binding protein)	2638
IPI00855918	mucin 5B, oligomeric mucus/gel-forming	727897

IPI00000874	peroxiredoxin 1	5052
IPI00003269	actin, beta-like 2	345651
IPI00007752	tubulin, beta 2C	10383
IPI00009650	lipocalin 1 (tear prealbumin)	3933
IPI00009865	keratin 10 (epidermolytic hyperkeratosis	3858
IPI00010471	lymphocyte cytosolic protein 1 (L-plastin)	3936
IPI00012011	cofilin 1 (non-muscle)	1072
IPI00012889	surfactant, pulmonary-associated protein A1B	6435
IPI00646877	surfactant, pulmonary-associated protein A1B	6435
IPI00013895	S100 calcium binding protein A11	6282
IPI00014055	napsin A aspartic peptidase	9476
IPI00015614	protease, serine, 3	5646
IPI00017601	ceruloplasmin (ferroxidase)	1356
IPI00019591	complement factor B	629
IPI00021841	apolipoprotein A-I	335
IPI00021885	fibrinogen alpha chain	2243
IPI00021891	fibrinogen gamma chain	2266
IPI00022229	apolipoprotein B (including Ag(x) antigen)	338
IPI00022418	fibronectin 1	2335
IPI00022431	alpha-2-HS-glycoprotein	197
IPI00022432	transthyretin (prealbumin, amyloidosis type I)	7276
IPI00022463	transferrin	7018
IPI00022895	alpha-1-B glycoprotein	1
IPI00023673	lectin, galactoside-binding, soluble, 3 binding protein	3959
IPI00024915	peroxiredoxin 5	25824
IPI00026197	immunoglobulin kappa variable 4-1	28908
IPI00026314	gelsolin (amyloidosis, Finnish type)	2934
IPI00029739	complement factor H	3075
IPI00032258	complement component 4A (Rodgers blood group)	720
IPI00073772	fructose-1,6-bisphosphatase 1	2203
IPI00103397	mucin 5AC, oligomeric mucus/gel-forming	4586
IPI00166729	alpha-2-glycoprotein 1, zinc-binding	563
IPI00169383	phosphoglycerate kinase 1	5230
IPI00180240	thymosin-like 3	7117
IPI00215983	carbonic anhydrase I	759
IPI00216298	thioredoxin	7295
IPI00217963	keratin 16 (focal non-epidermolytic palmoplantar keratoderma)	3868
IPI00218733	superoxide dismutase 1, soluble (amyotrophic lateral sclerosis 1 (adult))	6647
IPI00218918	annexin A1	301
IPI00219018	glyceraldehyde-3-phosphate dehydrogenase	2597

IPI00219446	phosphatidylethanolamine binding protein 1	5037
IPI00219757	glutathione S-transferase pi	2950
IPI00220327	keratin 1 (epidermolytic hyperkeratosis)	3848
IPI00021304	keratin 2 (epidermal ichthyosis bullosa of Siemens)	3849
IPI00220644	pyruvate kinase, muscle	5315
IPI00242956	Fc fragment of IgG binding protein	8857
IPI00291262	clusterin	1191
IPI00291410	chromosome 20 open reading frame 114	92747
IPI00291866	serpin peptidase inhibitor, clade G (C1 inhibitor), member 1, (angioedema, hereditary)	710
IPI00296083	surfactant, pulmonary-associated protein B	6439
IPI00296183	aldehyde dehydrogenase 3 family, memberA1	218
IPI00298497	fibrinogen beta chain	2244
IPI00300786	amylase, alpha 1A (salivary)	276
IPI00328493	similar to Ig heavy chain V-II region ARH-77 precursor	652128
IPI00382470	heat shock protein 90kDa alpha (cytosolic), class A member 1	3320
IPI00410714	hemoglobin, alpha 1	3039
IPI00418169	annexin A2	302
IPI00419585	peptidylprolyl isomerase A (cyclophilin A)	5478
IPI00465248	enolase 1, (alpha)	2023
IPI00465352	calcyphosine	828
IPI00478003	alpha-2-macroglobulin	2
IPI00550640	immunoglobulin heavy constant gamma 4 (G4m marker)	3503
IPI00785067	immunoglobulin heavy locus	3492
IPI00549330	immunoglobulin kappa variable 3D-15	28875
IPI00168728	immunoglobulin heavy constant mu	3507
IPI00642017	immunoglobulin heavy constant alpha 2 (A2m marker)	3494
IPI00784810	immunoglobulin heavy variable 4-31	28396
IPI00645363	immunoglobulin heavy constant gamma 1 (G1m marker)	3500
IPI00827488	immunoglobulin kappa constant	3514
IPI00830132	immunoglobulin heavy constant gamma 4 (G4m marker)	3503
IPI00550991	serpin peptidase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 3	12
IPI00553177	serpin peptidase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 1	5265
IPI00641737	haptoglobin	3240
IPI00154742	immunoglobulin lambda locus	3535
IPI00654755	hemoglobin, beta	3043
IPI00473011	hemoglobin, beta	3043
IPI00745872	albumin	213
IPI00783987	complement component 3	718

IPI00060800	similar to common salivary protein 1	124220
IPI00215611	cysteine-rich protein 1 (intestinal)	1396
IPI00291867	complement factor I	3426
IPI00465436	catalase	847
IPI00025447	eukaryotic translation elongation factor 1 alpha 1	1915
IPI00295777	glycerol-3-phosphate dehydrogenase 1 (soluble)	2819
IPI00019580	plasminogen	5340
IPI00022371	histidine-rich glycoprotein	3273
IPI00032179	serpin peptidase inhibitor, clade C (antithrombin), member 1	462
IPI00377025	proline-rich protein BstNI subfamily 4	5545
IPI00643920	transketolase (Wernicke-Korsakoff syndrome)	7086
IPI00328722	solute carrier family 9, member 11	284525

GPF (205 proteins identified) Gharib et al. in 2010

IPI	**Protein Name**	**Entrez Gene ID**
IPI00001754	F11 receptor	50848
IPI00002147	chitinase 3-like 1 (cartilage glycoprotein-39)	1116
IPI00003362	heat shock 70kDa protein 5 (glucose-regulated protein, 78k	3309
IPI00004656	beta-2-microglobulin	567
IPI00005721	defensin, alpha 1	1667
IPI00006705	secretoglobin, family 1A, member 1 (uteroglobin)	7356
IPI00006988	resistin	56729
IPI00007910	solute carrier family 34 (sodium phosphate), member 2	10568
IPI00008580	secretory leukocyte peptidase inhibitor	6590
IPI00009521	macrophage receptor with collagenous structure	8685
IPI00010182	diazepam binding inhibitor (GABA receptor modulator, acyl-	1622
IPI00011229	cathepsin D	1509
IPI00011252	complement component 8, alpha polypeptide	731
IPI00011302	CD59 molecule, complement regulatory protein	966
IPI00013122	cell division cycle 37 homolog (S. cerevisiae)	11140
IPI00014048	ribonuclease, RNase A family, 1 (pancreatic)	6035
IPI00018236	GM2 ganglioside activator	2760
IPI00019038	lysozyme (renal amyloidosis)	4069
IPI00019359	keratin 9 (epidermolytic palmoplantar keratoderma)	3857
IPI00019581	coagulation factor XII (Hageman factor)	2161
IPI00019943	afamin	173
IPI00020091	orosomucoid 2	5005
IPI00021828	cystatin B (stefin B)	1476
IPI00021842	apolipoprotein E	348
IPI00021854	apolipoprotein A-II	336

IPI00022204	serpin peptidase inhibitor, clade B (ovalbumin), member 3	6317
IPI00022371	histidine-rich glycoprotein	3273
IPI00022394	complement component 1, q subcomponent, C chain	714
IPI00022417	leucine-rich alpha-2-glycoprotein 1	116844
IPI00022420	retinol binding protein 4, plasma	5950
IPI00022426	alpha-1-microglobulin/bikunin precursor	259
IPI00022429	orosomucoid 1	5004
IPI00022488	hemopexin	3263
IPI00022774	valosin-containing protein	7415
IPI00022974	prolactin-induced protein	5304
IPI00022990	statherin	6779
IPI00024095	annexin A3	306
IPI00027444	serpin peptidase inhibitor, clade B (ovalbumin), member 1	1992
IPI00027462	S100 calcium binding protein A9	6280
IPI00027547	dermcidin	117159
IPI00027848	mannose receptor, C type 1	4360
IPI00029260	CD14 molecule	929
IPI00032179	serpin peptidase inhibitor, clade C (antithrombin), member	462
IPI00032291	complement component 5	727
IPI00066193	secretoglobin, family 3A, member 1	92304
IPI00104074	CD163 molecule	9332
IPI00163207	peptidoglycan recognition protein 2	114770
IPI00163563	phosphatidylethanolamine-binding protein 4	157310
IPI00165972	complement factor D (adipsin)	1675
IPI00178926	immunoglobulin J polypeptide, linker protein for immunoglo	3512
IPI00215611	cysteine-rich protein 1 (intestinal)	1396
IPI00216691	profilin 1	5216
IPI00218733	superoxide dismutase 1, soluble (amyotrophic lateral scler	6647
IPI00219219	lectin, galactoside-binding, soluble, 1 (galectin 1)	3956
IPI00246058	programmed cell death 6 interacting protein	10015
IPI00257508	dihydropyrimidinase-like 2	1808
IPI00291867	complement factor I	3426
IPI00291878	surfactant, pulmonary-associated protein D	6441
IPI00295741	cathepsin B	1508
IPI00297160	CD44 molecule (Indian blood group)	960
IPI00298971	vitronectin	7448
IPI00299547	lipocalin 2	3934
IPI00299729	transcobalamin I (vitamin B12 binding protein, R binder fa	6947
IPI00301579	Niemann-Pick disease, type C2	10577
IPI00304273	apolipoprotein A-IV	337
IPI00306413	tubulin polymerization-promoting protein family member 3	51673

IPI00329801	annexin A5	308
IPI00375676	ferritin, light polypeptide	2512
IPI00399260	proline-rich protein BstNI subfamily 1	5542
IPI00410714	hemoglobin, alpha 1	3039
IPI00456158	proline-rich protein BstNI subfamily 1	5542
IPI00465436	catalase	847
IPI00477992	complement component 1, q subcomponent, B chain	713
IPI00550069	ribonuclease/angiogenin inhibitor 1	6050
IPI00555812	group-specific component (vitamin D binding protein)	2638
IPI00844156	serpin peptidase inhibitor, clade C (antithrombin), member	462
IPI00855918	mucin 5B, oligomeric mucus/gel-forming	727897
IPI00000861	LIM and SH3 protein 1	3927
IPI00000874	peroxiredoxin 1	5052
IPI00003590	quiescin Q6 sulfhydryl oxidase 1	5768
IPI00003817	Rho GDP dissociation inhibitor (GDI) beta	397
IPI00003865	heat shock 70kDa protein 8	3312
IPI00304925	heat shock 70kDa protein 1A	3303
IPI00004573	polymeric immunoglobulin receptor	5284
IPI00006114	serpin peptidase inhibitor, clade F (alpha-2 antiplasmin,	5176
IPI00007797	fatty acid binding protein 5 (psoriasis-associated)	2171
IPI00008494	intercellular adhesion molecule 1 (CD54), human rhinovirus	3383
IPI00009650	lipocalin 1 (tear prealbumin)	3933
IPI00009865	keratin 10 (epidermolytic hyperkeratosis	3858
IPI00217963	keratin 16 (focal non-epidermolytic palmoplantar keratoder	3868
IPI00384444	keratin 14 (epidermolysis bullosa simplex, Dowling-Meara,	3861
IPI00010397	major histocompatibility complex, class II, DQ beta 1	3119
IPI00010471	lymphocyte cytosolic protein 1 (L-plastin)	3936
IPI00010896	chloride intracellular channel 1	1192
IPI00012011	cofilin 1 (non-muscle)	1072
IPI00012303	selenium binding protein 1	8991
IPI00012889	surfactant, pulmonary-associated protein A1B	6435
IPI00646877	surfactant, pulmonary-associated protein A1B	6435
IPI00013895	S100 calcium binding protein A11	6282
IPI00014055	napsin A aspartic peptidase	9476
IPI00017601	ceruloplasmin (ferroxidase)	1356
IPI00018219	transforming growth factor, beta-induced, 68kDa	7045
IPI00018909	trefoil factor 3 (intestinal)	7033
IPI00019502	myosin, heavy chain 9, non-muscle	4627
IPI00019568	coagulation factor II (thrombin)	2147
IPI00019580	plasminogen	5340
IPI00019591	complement factor B	629

IPI00303963	complement component 2	717
IPI00020986	lumican	4060
IPI00021263	tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activa	7534
IPI00021302	sushi domain containing 2	56241
IPI00021841	apolipoprotein A-I	335
IPI00021885	fibrinogen alpha chain	2243
IPI00021891	fibrinogen gamma chain	2266
IPI00022431	alpha-2-HS-glycoprotein	197
IPI00022432	transthyretin (prealbumin, amyloidosis type I)	7276
IPI00022463	transferrin	7018
IPI00789477	lactotransferrin	4057
IPI00022895	alpha-1-B glycoprotein	1
IPI00023011	submaxillary gland androgen regulated protein 3B	10879
IPI00023673	lectin, galactoside-binding, soluble, 3 binding protein	3959
IPI00024915	peroxiredoxin 5	25824
IPI00025252	protein disulfide isomerase family A, member 3	2923
IPI00026314	gelsolin (amyloidosis, Finnish type)	2934
IPI00027019	proline-rich protein BstNI subfamily 4	5545
IPI00027497	glucose phosphate isomerase	2821
IPI00029658	EGF-containing fibulin-like extracellular matrix protein 1	2202
IPI00029739	complement factor H	3075
IPI00032220	angiotensinogen (serpin peptidase inhibitor, clade A, memb	183
IPI00032294	cystatin S	1472
IPI00305477	cystatin SN	1469
IPI00073772	fructose-1,6-bisphosphatase 1	2203
IPI00099110	deleted in malignant brain tumors 1	1755
IPI00103397	mucin 5AC, oligomeric mucus/gel-forming	4586
IPI00152418	CD55 molecule, decay accelerating factor for complement (C	1604
IPI00154742	immunoglobulin lambda locus	3535
IPI00166729	alpha-2-glycoprotein 1, zinc-binding	563
IPI00169383	phosphoglycerate kinase 1	5230
IPI00180240	thymosin-like 3	7117
IPI00186290	eukaryotic translation elongation factor 2	1938
IPI00215894	kininogen 1	3827
IPI00218319	tropomyosin 3	7170
IPI00218914	aldehyde dehydrogenase 1 family, member A1	216
IPI00218918	annexin A1	301
IPI00219018	glyceraldehyde-3-phosphate dehydrogenase	2597
IPI00219077	leukotriene A4 hydrolase	4048
IPI00219217	lactate dehydrogenase B	3945
IPI00219365	moesin	4478

IPI00746388	ezrin	7430
IPI00219446	phosphatidylethanolamine binding protein 1	5037
IPI00219757	glutathione S-transferase pi	2950
IPI00220271	aldo-keto reductase family 1, member A1 (aldehyde reductas	10327
IPI00220327	keratin 1 (epidermolytic hyperkeratosis)	3848
IPI00021304	keratin 2 (epidermal ichthyosis bullosa of Siemens)	3849
IPI00300725	keratin 6A	3853
IPI00009867	keratin 5 (epidermolysis bullosa simplex, Dowling-Meara/Ko	3852
IPI00242956	Fc fragment of IgG binding protein	8857
IPI00291262	clusterin	1191
IPI00291410	chromosome 20 open reading frame 114	92747
IPI00291488	WAP four-disulfide core domain 2	10406
IPI00291866	serpin peptidase inhibitor, clade G (C1 inhibitor), member	710
IPI00292530	inter-alpha (globulin) inhibitor H1	3697
IPI00294004	protein S (alpha)	5627
IPI00296083	surfactant, pulmonary-associated protein B	6439
IPI00297487	cathepsin H	1512
IPI00298497	fibrinogen beta chain	2244
IPI00298828	apolipoprotein H (beta-2-glycoprotein I)	350
IPI00299078	proline-rich protein BstNI subfamily 4	5545
IPI00300786	amylase, alpha 1A (salivary)	276
IPI00302592	filamin A, alpha (actin binding protein 280)	2316
IPI00305461	inter-alpha (globulin) inhibitor H2	3698
IPI00382470	heat shock protein 90kDa alpha (cytosolic), class A member	3320
IPI00026197	immunoglobulin kappa variable 4-1	28908
IPI00418471	vimentin	7431
IPI00419585	peptidylprolyl isomerase A (cyclophilin A)	5478
IPI00451624	cartilage acidic protein 1	55118
IPI00465138	cytochrome P450, family 3, subfamily A, polypeptide 4	1576
IPI00465248	enolase 1, (alpha)	2023
IPI00465352	calcyphosine	828
IPI00465439	aldolase A, fructose-bisphosphate	226
IPI00478003	alpha-2-macroglobulin	2
IPI00549330	immunoglobulin kappa variable 3D-15	28875
IPI00550363	transgelin 2	8407
IPI00550640	immunoglobulin heavy constant gamma 4 (G4m marker)	3503
IPI00477090	immunoglobulin heavy constant mu	3507
IPI00647704	immunoglobulin heavy constant alpha 1	3493
IPI00426051	similar to hCG2038920	1E+08
IPI00816314	immunoglobulin heavy constant mu	3507
IPI00845354	immunoglobulin kappa constant	3514

IPI00784430	immunoglobulin kappa variable 3D-11	28876
IPI00827488	immunoglobulin kappa constant	3514
IPI00061977	immunoglobulin heavy constant alpha 1	3493
IPI00645363	immunoglobulin heavy constant gamma 1 (G1m marker)	3500
IPI00550991	serpin peptidase inhibitor, clade A (alpha-1 antiproteinas	12
IPI00553177	serpin peptidase inhibitor, clade A (alpha-1 antiproteinas	5265
IPI00641737	haptoglobin	3240
IPI00643567	CAP, adenylate cyclase-associated protein 1 (yeast)	10487
IPI00643920	transketolase (Wernicke-Korsakoff syndrome)	7086
IPI00654755	hemoglobin, beta	3043
IPI00745872	albumin	213
IPI00783987	complement component 3	718
IPI00791350	C-type lectin domain family 3, member B	7123
IPI00027341	capping protein (actin filament), gelsolin-like	822
IPI00442909	immunoglobulin heavy variable 4-31	28396
IPI00296654	bactericidal/permeability-increasing protein-like 1	80341

Rational GPF (598 proteins identified) Nguyen et al. in 2012

IPI	**Protein Name**	**Entrez Gene ID**
IPI00000190	CD81 molecule	975
IPI00000581	OTU domain, ubiquitin aldehyde binding 1	55611
IPI00000949	crystallin, mu	1428
IPI00001639	karyopherin (importin) beta 1	3837
IPI00003362	heat shock 70kDa protein 5 (glucose-regulated protein, 78kDa)	3309
IPI00004433	contactin 6	27255
IPI00004573	polymeric immunoglobulin receptor	5284
IPI00005160	actin related protein 2/3 complex, subunit 1B, 41kDa	10095
IPI00005721	defensin, alpha 1	1667
IPI00005969	capping protein (actin filament) muscle Z-line, alpha 1	829
IPI00006705	secretoglobin, family 1A, member 1 (uteroglobin)	7356
IPI00007047	S100 calcium binding protein A8	6279
IPI00007682	ATPase, H+ transporting, lysosomal 70kDa, V1 subunit A	523
IPI00007910	solute carrier family 34 (sodium phosphate), member 2	10568
IPI00008164	prolyl endopeptidase	5550
IPI00008455	myosin VI	4646
IPI00009342	IQ motif containing GTPase activating protein 1	8826
IPI00009856	palate, lung and nasal epithelium associated	51297
IPI00009904	protein disulfide isomerase family A, member 4	9601
IPI00010720	chaperonin containing TCP1, subunit 5 (epsilon)	22948
IPI00011302	CD59 molecule, complement regulatory protein	966

IPI00012007	S-adenosylhomocysteine hydrolase	191
IPI00012503	prosaposin (variant Gaucher disease and variant metachromatic leukodystrophy)	5660
IPI00012540	prominin 1	8842
IPI00013382	cystatin SA	1470
IPI00013895	S100 calcium binding protein A11	6282
IPI00013933	desmoplakin	1832
IPI00016342	RAB7A, member RAS oncogene family	7879
IPI00017704	coactosin-like 1 (Dictyostelium)	23406
IPI00018451	calcium and integrin binding 1 (calmyrin)	10519
IPI00018931	vacuolar protein sorting 35 homolog (S. cerevisiae)	55737
IPI00018953	dipeptidyl-peptidase 4 (CD26, adenosine deaminase complexing protein 2)	1803
IPI00019359	keratin 9 (epidermolytic palmoplantar keratoderma)	3857
IPI00019580	plasminogen	5340
IPI00021070	neutrophil cytosolic factor 2 (65kDa, chronic granulomatous disease, autosomal 2)	4688
IPI00022371	histidine-rich glycoprotein	3273
IPI00022394	complement component 1, q subcomponent, C chain	714
IPI00022395	complement component 9	735
IPI00022417	leucine-rich alpha-2-glycoprotein 1	116844
IPI00022420	retinol binding protein 4, plasma	5950
IPI00022426	alpha-1-microglobulin/bikunin precursor	259
IPI00022462	transferrin receptor (p90, CD71)	7037
IPI00022488	hemopexin	3263
IPI00022624	G protein-coupled receptor, family C, group 5, member A	9052
IPI00022774	valosin-containing protein	7415
IPI00022810	cathepsin C	1075
IPI00022974	prolactin-induced protein	5304
IPI00024095	annexin A3	306
IPI00024689	aquaporin 1 (Colton blood group)	358
IPI00025084	calpain, small subunit 1	826
IPI00027019	proline-rich protein BstNI subfamily 4	5545
IPI00027175	sorcin	6717
IPI00027223	isocitrate dehydrogenase 1 (NADP+), soluble	3417
IPI00027442	alanyl-tRNA synthetase	16
IPI00027444	serpin peptidase inhibitor, clade B (ovalbumin), member 1	1992
IPI00027462	S100 calcium binding protein A9	6280
IPI00027482	serpin peptidase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 6	866
IPI00027626	chaperonin containing TCP1, subunit 6A (zeta 1)	908
IPI00028004	proteasome (prosome, macropain) subunit, beta type, 3	5691

IPI00029260	CD14 molecule	929
IPI00030936	tetraspanin 1	10103
IPI00032179	serpin peptidase inhibitor, clade C (antithrombin), member 1	462
IPI00032294	cystatin S	1472
IPI00037448	glyoxylate reductase/hydroxypyruvate reductase	9380
IPI00060800	similar to common salivary protein 1	124220
IPI00066193	secretoglobin, family 3A, member 1	92304
IPI00147874	N-acetylneuraminic acid synthase (sialic acid synthase)	54187
IPI00163207	peptidoglycan recognition protein 2	114770
IPI00163563	phosphatidylethanolamine-binding protein 4	157310
IPI00166766	hypothetical protein MGC45438	146556
IPI00170979	Enah/Vasp-like	51466
IPI00178926	immunoglobulin J polypeptide, linker protein for immunoglobulin alpha and mu polypeptides	3512
IPI00180707	FRAS1 related extracellular matrix protein 2	341640
IPI00215767	UDP-Gal:betaGlcNAc beta 1,4- galactosyltransferase, polypeptide 1	2683
IPI00215997	CD9 molecule	928
IPI00216048	phosphatidylinositol transfer protein, alpha	5306
IPI00216057	sorbitol dehydrogenase	6652
IPI00216106	Obg-like ATPase 1	29789
IPI00216691	profilin 1	5216
IPI00216699	fermitin family homolog 3 (Drosophila)	83706
IPI00219219	lectin, galactoside-binding, soluble, 1 (galectin 1)	3956
IPI00219622	proteasome (prosome, macropain) subunit, alpha type, 2	5683
IPI00219678	eukaryotic translation initiation factor 2, subunit 1 alpha, 35kDa	1965
IPI00220906	acyl-CoA thioesterase 2	10965
IPI00246058	programmed cell death 6 interacting protein	10015
IPI00289499	5-aminoimidazole-4-carboxamide ribonucleotide formyltransferase/IMP cyclohydrolase	471
IPI00291005	malate dehydrogenase 1, NAD (soluble)	4190
IPI00291878	surfactant, pulmonary-associated protein D	6441
IPI00291922	proteasome (prosome, macropain) subunit, alpha type, 5	5686
IPI00292657	prostaglandin reductase 1	22949
IPI00292858	thymidine phosphorylase	1890
IPI00292946	serpin peptidase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 7	6906
IPI00292950	serpin peptidase inhibitor, clade D (heparin cofactor), member 1	3053
IPI00294158	biliverdin reductase A	644
IPI00294495	ubiquitin-fold modifier conjugating enzyme 1	51506
IPI00294739	SAM domain and HD domain 1	25939
IPI00296083	surfactant, pulmonary-associated protein B	6439

IPI00296353	Rho GTPase activating protein 18	93663
IPI00296526	N-acetylglucosamine kinase	55577
IPI00297487	cathepsin H	1512
IPI00297779	chaperonin containing TCP1, subunit 2 (beta)	10576
IPI00298497	fibrinogen beta chain	2244
IPI00298971	vitronectin	7448
IPI00299547	lipocalin 2	3934
IPI00299571	protein disulfide isomerase family A, member 6	10130
IPI00302927	chaperonin containing TCP1, subunit 4 (delta)	10575
IPI00304273	apolipoprotein A-IV	337
IPI00304557	chromosome 20 open reading frame 70	140683
IPI00305477	cystatin SN	1469
IPI00306413	tubulin polymerization-promoting protein family member 3	51673
IPI00328350	family with sequence similarity 129, member A	116496
IPI00329331	UDP-glucose pyrophosphorylase 2	7360
IPI00329801	annexin A5	308
IPI00374315	chromosome 6 open reading frame 58	352999
IPI00384051	proteasome (prosome, macropain) activator subunit 2 (PA28 beta)	5721
IPI00410600	calcium channel, voltage-dependent, alpha 2/delta subunit 2	9254
IPI00410714	hemoglobin, alpha 1	3039
IPI00411706	esterase D/formylglutathione hydrolase	2098
IPI00418495	CD36 molecule (thrombospondin receptor)	948
IPI00441498	folate receptor 1 (adult)	2348
IPI00451624	cartilage acidic protein 1	55118
IPI00456750	family with sequence similarity 129, member B	64855
IPI00456969	dynein, cytoplasmic 1, heavy chain 1	1778
IPI00477992	complement component 1, q subcomponent, B chain	713
IPI00550991	serpin peptidase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 3	12
IPI00555812	group-specific component (vitamin D binding protein)	2638
IPI00644472	haloacid dehalogenase-like hydrolase domain containing 2	84064
IPI00783313	phosphorylase, glycogen	5836
IPI00798401	coronin, actin binding protein, 1C	23603
IPI00021263	tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, zeta polypeptide	7534
IPI00000874	peroxiredoxin 1	5052
IPI00002966	heat shock 70kDa protein 4	3308
IPI00003590	quiescin Q6 sulfhydryl oxidase 1	5768
IPI00005118	hexokinase 3 (white cell)	3101
IPI00006114	serpin peptidase inhibitor, clade F (alpha-2 antiplasmin, pigment epithelium derived factor), member 1	5176

IPI00007058	coronin, actin binding protein, 1B	57175
IPI00007427	anterior gradient homolog 2 (Xenopus laevis)	10551
IPI00007797	fatty acid binding protein 5 (psoriasis-associated)	2171
IPI00008274	CAP, adenylate cyclase-associated protein 1 (yeast)	10487
IPI00008485	aconitase 1, soluble	48
IPI00008494	intercellular adhesion molecule 1 (CD54), human rhinovirus receptor	3383
IPI00009104	RuvB-like 2 (E. coli)	10856
IPI00009865	keratin 10 (epidermolytic hyperkeratosis	3858
IPI00010133	coronin, actin binding protein, 1A	11151
IPI00010180	carboxylesterase 1 (monocyte/macrophage serine esterase 1)	1066
IPI00010471	lymphocyte cytosolic protein 1 (L-plastin)	3936
IPI00010896	chloride intracellular channel 1	1192
IPI00011229	cathepsin D	1509
IPI00011285	calpain 1, (mu/I) large subunit	823
IPI00012011	cofilin 1 (non-muscle)	1072
IPI00012303	selenium binding protein 1	8991
IPI00012837	kinesin family member 5B	3799
IPI00013698	N-acylsphingosine amidohydrolase (acid ceramidase) 1	427
IPI00013808	actinin, alpha 4	81
IPI00014055	napsin A aspartic peptidase	9476
IPI00015475	solute carrier family 1 (neuronal/epithelial high affinity glutamate transporter, system Xag), member 1	6505
IPI00016610	poly(rC) binding protein 1	5093
IPI00016786	cell division cycle 42 (GTP binding protein, 25kDa)	998
IPI00017184	EH-domain containing 1	10938
IPI00017469	sepiapterin reductase (7,8-dihydrobiopterin:NADP+ oxidoreductase)	6697
IPI00017601	ceruloplasmin (ferroxidase)	1356
IPI00018219	transforming growth factor, beta-induced, 68kDa	7045
IPI00018246	hexokinase 1	3098
IPI00018342	adenylate kinase 1	203
IPI00019502	myosin, heavy chain 9, non-muscle	4627
IPI00019568	coagulation factor II (thrombin)	2147
IPI00019591	complement factor B	629
IPI00020557	low density lipoprotein-related protein 1 (alpha-2-macroglobulin receptor)	4035
IPI00021187	RuvB-like 1 (E. coli)	8607
IPI00021302	sushi domain containing 2	56241
IPI00021727	complement component 4 binding protein, alpha	722
IPI00021841	apolipoprotein A-I	335
IPI00021885	fibrinogen alpha chain	2243
IPI00021891	fibrinogen gamma chain	2266
IPI00022204	serpin peptidase inhibitor, clade B (ovalbumin), member 3	6317

IPI00022213	progastricsin (pepsinogen C)	5225
IPI00022392	complement component 1, q subcomponent, A chain	712
IPI00022431	alpha-2-HS-glycoprotein	197
IPI00022432	transthyretin (prealbumin, amyloidosis type I)	7276
IPI00022463	transferrin	7018
IPI00022895	alpha-1-B glycoprotein	1
IPI00022977	creatine kinase, brain	1152
IPI00023673	lectin, galactoside-binding, soluble, 3 binding protein	3959
IPI00024067	clathrin, heavy chain (Hc)	1213
IPI00024915	peroxiredoxin 5	25824
IPI00025019	proteasome (prosome, macropain) subunit, beta type, 1	5689
IPI00025252	protein disulfide isomerase family A, member 3	2923
IPI00025512	heat shock 27kDa protein 1	3315
IPI00026185	capping protein (actin filament) muscle Z-line, beta	832
IPI00026216	aminopeptidase puromycin sensitive	9520
IPI00026268	guanine nucleotide binding protein (G protein), beta polypeptide 1	2782
IPI00026272	histone cluster 1, H2ae	3012
IPI00026314	gelsolin (amyloidosis, Finnish type)	2934
IPI00027230	heat shock protein 90kDa beta (Grp94), member 1	7184
IPI00027848	mannose receptor, C type 1	4360
IPI00028091	ARP3 actin-related protein 3 homolog (yeast)	10096
IPI00029658	EGF-containing fibulin-like extracellular matrix protein 1	2202
IPI00029739	complement factor H	3075
IPI00029997	6-phosphogluconolactonase	25796
IPI00030781	signal transducer and activator of transcription 1, 91kDa	6772
IPI00031461	GDP dissociation inhibitor 2	2665
IPI00032220	angiotensinogen (serpin peptidase inhibitor, clade A, member 8)	183
IPI00032291	complement component 5	727
IPI00033022	dynamin 2	1785
IPI00033494	myosin regulatory light chain MRLC2	103910
IPI00060715	potassium channel tetramerisation domain containing 12	115207
IPI00073772	fructose-1,6-bisphosphatase 1	2203
IPI00100160	cullin-associated and neddylation-dissociated 1	55832
IPI00152418	CD55 molecule, decay accelerating factor for complement (Cromer blood group)	1604
IPI00156689	vesicle amine transport protein 1 homolog (T. californica)	10493
IPI00166729	alpha-2-glycoprotein 1, zinc-binding	563
IPI00169383	phosphoglycerate kinase 1	5230
IPI00177728	CNDP dipeptidase 2 (metallopeptidase M20 family)	55748
IPI00183046	protein tyrosine phosphatase, non-receptor type 6	5777
IPI00186290	eukaryotic translation elongation factor 2	1938

IPI00215746	fatty acid binding protein 4, adipocyte	2167
IPI00216008	glucose-6-phosphate dehydrogenase	2539
IPI00216049	heterogeneous nuclear ribonucleoprotein K	3190
IPI00217906	guanine nucleotide binding protein (G protein), alpha inhibiting activity polypeptide 2	2771
IPI00217920	aldehyde dehydrogenase 16 family, member A1	126133
IPI00218192	inter-alpha (globulin) inhibitor H4 (plasma Kallikrein-sensitive glycoprotein)	3700
IPI00218413	biotinidase	686
IPI00218693	adenine phosphoribosyltransferase	353
IPI00218914	aldehyde dehydrogenase 1 family, member A1	216
IPI00218918	annexin A1	301
IPI00219077	leukotriene A4 hydrolase	4048
IPI00219217	lactate dehydrogenase B	3945
IPI00219365	moesin	4478
IPI00219446	phosphatidylethanolamine binding protein 1	5037
IPI00219525	phosphogluconate dehydrogenase	5226
IPI00219526	phosphoglucomutase 1	5236
IPI00219682	stomatin	2040
IPI00219757	glutathione S-transferase pi	2950
IPI00220271	aldo-keto reductase family 1, member A1 (aldehyde reductase)	10327
IPI00220301	peroxiredoxin 6	9588
IPI00220327	keratin 1 (epidermolytic hyperkeratosis)	3848
IPI00221221	arachidonate 15-lipoxygenase	246
IPI00242956	Fc fragment of IgG binding protein	8857
IPI00257508	dihydropyrimidinase-like 2	1808
IPI00288947	guanine nucleotide binding protein (G protein), q polypeptide	2776
IPI00289758	calpain 2, (m/II) large subunit	824
IPI00291175	vinculin	7414
IPI00291262	clusterin	1191
IPI00291410	chromosome 20 open reading frame 114	92747
IPI00291866	serpin peptidase inhibitor, clade G (C1 inhibitor), member 1, (angioedema, hereditary)	710
IPI00291867	complement factor I	3426
IPI00292530	inter-alpha (globulin) inhibitor H1	3697
IPI00294578	transglutaminase 2 (C polypeptide, protein-glutamine-gamma-glutamyltransferase)	7052
IPI00295386	carbonyl reductase 1	873
IPI00295400	tryptophanyl-tRNA synthetase	7453
IPI00295777	glycerol-3-phosphate dehydrogenase 1 (soluble)	2819
IPI00296608	complement component 7	730

IPI00296654	bactericidal/permeability-increasing protein-like 1	80341
IPI00297160	CD44 molecule (Indian blood group)	960
IPI00298961	exportin 1 (CRM1 homolog, yeast)	7514
IPI00298994	talin 1	7094
IPI00302925	chaperonin containing TCP1, subunit 8 (theta)	10694
IPI00304925	heat shock 70kDa protein 1A	3303
IPI00305461	inter-alpha (globulin) inhibitor H2	3698
IPI00382470	heat shock protein 90kDa alpha (cytosolic), class A member 1	3320
IPI00413451	serpin peptidase inhibitor, clade B (ovalbumin), member 6	5269
IPI00418163	complement component 4B (Childo blood group)	721
IPI00418471	vimentin	7431
IPI00419237	leucine aminopeptidase 3	51056
IPI00419585	peptidylprolyl isomerase A (cyclophilin A)	5478
IPI00437751	angiotensin I converting enzyme (peptidyl-dipeptidase A) 1	1636
IPI00465248	enolase 1, (alpha)	2023
IPI00465279	major histocompatibility complex, class II, DQ beta 1	3119
IPI00465352	calcyphosine	828
IPI00465439	aldolase A, fructose-bisphosphate	226
IPI00478003	alpha-2-macroglobulin	2
IPI00479722	proteasome (prosome, macropain) activator subunit 1 (PA28 alpha)	5720
IPI00550069	ribonuclease/angiogenin inhibitor 1	6050
IPI00550363	transgelin 2	8407
IPI00553177	serpin peptidase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 1	5265
IPI00554811	actin related protein 2/3 complex, subunit 4, 20kDa	10093
IPI00642211	arginyl aminopeptidase (aminopeptidase B)	6051
IPI00643920	transketolase (Wernicke-Korsakoff syndrome)	7086
IPI00645078	ubiquitin-like modifier activating enzyme 1	7317
IPI00646877	surfactant, pulmonary-associated protein A1B	6435
IPI00654755	hemoglobin, beta	3043
IPI00657682	glutathione S-transferase A1	2938
IPI00745872	albumin	213
IPI00783987	complement component 3	718
IPI00793199	annexin A4	307
IPI00847989	pyruvate kinase, muscle	5315
IPI00848090	capping protein (actin filament), gelsolin-like	822
IPI00855918	mucin 5B, oligomeric mucus/gel-forming	727897
IPI00002147	chitinase 3-like 1 (cartilage glycoprotein-39)	1116
IPI00002851	cystatin D	1473
IPI00005161	actin related protein 2/3 complex, subunit 2, 34kDa	10109
IPI00006907	chromosome 12 open reading frame 5	57103

IPI00007244	myeloperoxidase	4353
IPI00018909	trefoil factor 3 (intestinal)	7033
IPI00020996	insulin-like growth factor binding protein, acid labile subunit	3483
IPI00032313	S100 calcium binding protein A4	6275
IPI00165972	complement factor D (adipsin)	1675
IPI00220637	seryl-tRNA synthetase	6301
IPI00293975	glutathione peroxidase 1	2876
IPI00294004	protein S (alpha)	5627
IPI00453473	histone cluster 1, H4i	8294
IPI00550364	phosphoglucomutase 2	55276
IPI00551024	dihydroxyacetone kinase 2 homolog (S. cerevisiae)	26007
IPI00844156	serpin peptidase inhibitor, clade C (antithrombin), member 1	462
IPI00879709	complement component 6	729
IPI00005162	actin related protein 2/3 complex, subunit 3, 21kDa	10094
IPI00023860	nucleosome assembly protein 1-like 1	4673
IPI00305010	hypothetical protein FLJ11151	55313
IPI00337741	N-acylaminoacyl-peptide hydrolase	327
IPI00465085	nicotinate phosphoribosyltransferase domain containing 1	93100
IPI00002856	STEAP family member 4	79689
IPI00003949	ubiquitin-conjugating enzyme E2N (UBC13 homolog, yeast)	7334
IPI00018398	proteasome (prosome, macropain) 26S subunit, ATPase, 3	5702
IPI00021828	cystatin B (stefin B)	1476
IPI00217493	myoglobin	4151
IPI00020436	RAB11B, member RAS oncogene family	9230
IPI00220580	sperm associated antigen 6	9576
IPI00009032	Sjogren syndrome antigen B (autoantigen La)	6741
IPI00024664	ubiquitin specific peptidase 5 (isopeptidase T)	8078
IPI00303476	ATP synthase, H+ transporting, mitochondrial F1 complex, beta polypeptide	506
IPI00306960	asparaginyl-tRNA synthetase	4677
IPI00023019	sex hormone-binding globulin	6462
IPI00414315	EPS8-like 2	64787
IPI00643041	RAN, member RAS oncogene family	5901
IPI00003935	histone cluster 2, H2be	8349
IPI00295209	sorting nexin 5	27131
IPI00783097	glycyl-tRNA synthetase	2617
IPI00031023	flightless I homolog (Drosophila)	2314
IPI00328609	serpin peptidase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 4	5267
IPI00019755	glutathione S-transferase omega 1	9446
IPI00220219	coatomer protein complex, subunit beta 2 (beta prime)	9276
IPI00479877	aldehyde dehydrogenase 9 family, member A1	223

IPI00328257	adaptor-related protein complex 1, beta 1 subunit	162
IPI00465233	eukaryotic translation initiation factor 3, subunit E interacting protein	51386
IPI00019912	hydroxysteroid (17-beta) dehydrogenase 4	3295
IPI00217766	scavenger receptor class B, member 2	950
IPI00008234	cytochrome b5 reductase 2	51700
IPI00029012	eukaryotic translation initiation factor 3, subunit A	8661
IPI00328243	phospholipase D family, member 3	23646
IPI00022975	arachidonate 5-lipoxygenase-activating protein	241
IPI00027996	mitochondria-associated protein involved in granulocyte-macrophage colony-stimulating factor signal transduction	51025
IPI00376344	myosin IB	4430
IPI00023283	titin	7273
IPI00783982	coatomer protein complex, subunit gamma	22820
IPI00290566	t-complex 1	6950
IPI00442909	immunoglobulin heavy variable 4-31	28396
IPI00017763	nucleosome assembly protein 1-like 4	4676
IPI00018465	chaperonin containing TCP1, subunit 7 (eta)	10574
IPI00021854	apolipoprotein A-II	336
IPI00044550	secretoglobin, family 3A, member 2	117156
IPI00157757	microtubule associated monoxygenase, calponin and LIM domain containing 1	64780
IPI00218414	carbonic anhydrase II	760
IPI00419916	alkaline phosphatase, liver/bone/kidney	249
IPI00744692	transaldolase 1	6888
IPI00010270	ras-related C3 botulinum toxin substrate 2 (rho family, small GTP binding protein Rac2)	5880
IPI00019038	lysozyme (renal amyloidosis)	4069
IPI00031564	chromosome 7 open reading frame 24	79017
IPI00215998	CD63 molecule	967
IPI00293590	monoglyceride lipase	11343
IPI00295851	coatomer protein complex, subunit beta 1	1315
IPI00220267	argininosuccinate lyase	435
IPI00165949	endoplasmic reticulum aminopeptidase 1	51752
IPI00023048	eukaryotic translation elongation factor 1 delta (guanine nucleotide exchange protein)	1936
IPI00293074	solute carrier family 44, member 2	57153
IPI00003817	Rho GDP dissociation inhibitor (GDI) beta	397
IPI00005171	major histocompatibility complex, class II, DR alpha	3122
IPI00020986	lumican	4060
IPI00030023	histamine N-methyltransferase	3176
IPI00103067	calcyphosine	828

IPI00218646	cytochrome b-245, beta polypeptide (chronic granulomatous disease)	1536
IPI00258833	sorting nexin 6	58533
IPI00294395	complement component 8, beta polypeptide	732
IPI00003865	heat shock 70kDa protein 8	3312
IPI00005578	EH-domain containing 4	30844
IPI00217963	keratin 16 (focal non-epidermolytic palmoplantar keratoderma)	3868
IPI00032304	plastin 1 (I isoform)	5357
IPI00013475	tubulin, beta 2A	7280
IPI00018873	nicotinamide phosphoribosyltransferase	10135
IPI00021842	apolipoprotein E	348
IPI00027341	capping protein (actin filament), gelsolin-like	822
IPI00027497	glucose phosphate isomerase	2821
IPI00099110	deleted in malignant brain tumors 1	1755
IPI00168184	protein phosphatase 2 (formerly 2A), regulatory subunit A , alpha isoform	5518
IPI00216256	WD repeat domain 1	9948
IPI00218604	protein tyrosine phosphatase, non-receptor type 6	5777
IPI00299095	sorting nexin 2	6643
IPI00303882	mannose-6-phosphate receptor binding protein 1	10226
IPI00479708	immunoglobulin heavy constant mu	3507
IPI00641737	haptoglobin	3240
IPI00013122	cell division cycle 37 homolog (S. cerevisiae)	11140
IPI00021857	apolipoprotein C-III	345
IPI00027493	solute carrier family 3 (activators of dibasic and neutral amino acid transport), member 2	6520
IPI00011261	complement component 8, gamma polypeptide	733
IPI00020091	orosomucoid 2	5005
IPI00217296	protein phosphatase 2A activator, regulatory subunit 4	5524
IPI00063827	abhydrolase domain containing 14B	84836
IPI00215637	DEAD (Asp-Glu-Ala-Asp) box polypeptide 3, X-linked	1654
IPI00414320	annexin A11	311
IPI00219953	cytidine monophosphate (UMP-CMP) kinase 1, cytosolic	51727
IPI00010779	tropomyosin 4	7171
IPI00060181	EF-hand domain family, member D2	79180
IPI00018140	synaptotagmin binding, cytoplasmic RNA interacting protein	10492
IPI00641950	guanine nucleotide binding protein (G protein), beta polypeptide 2-like 1	10399
IPI00017025	similar to peptidyl-Pro cis trans isomerase	128192
IPI00007812	ATPase, H+ transporting, lysosomal 56/58kDa, V1 subunit B2	526
IPI00793443	importin 5	3843
IPI00031420	UDP-glucose dehydrogenase	7358
IPI00015148	RAP1B, member of RAS oncogene family	5908

IPI00644231	cytoplasmic FMR1 interacting protein 1	23191
IPI00004307	lysosomal-associated membrane protein 3	27074
IPI00008380	protein phosphatase 2 (formerly 2A), catalytic subunit, alpha isoform	5515
IPI00013004	pyridoxal (pyridoxine, vitamin B6) kinase	8566
IPI00014898	plectin 1, intermediate filament binding protein 500kDa	5339
IPI00019329	dynein, light chain, LC8-type 1	8655
IPI00030876	diaphanous homolog 1 (Drosophila)	1729
IPI00103356	integrin, beta 2 (complement component 3 receptor 3 and 4 subunit)	3689
IPI00165936	chloride intracellular channel 6	54102
IPI00215768	glutamate-cysteine ligase, catalytic subunit	2729
IPI00218638	myosin IF	4542
IPI00218732	paraoxonase 1	5444
IPI00219029	glutamic-oxaloacetic transaminase 1, soluble (aspartate aminotransferase 1)	2805
IPI00000105	major vault protein	9961
IPI00000816	tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, epsilon polypeptide	7531
IPI00003269	actin, beta-like 2	345651
IPI00004358	phosphorylase, glycogen	5834
IPI00006662	apolipoprotein D	347
IPI00006663	aldehyde dehydrogenase 2 family (mitochondrial)	217
IPI00010271	ras-related C3 botulinum toxin substrate 1 (rho family, small GTP binding protein Rac1)	5879
IPI00019971	syntaxin binding protein 2	6813
IPI00032959	glycerol-3-phosphate dehydrogenase 1-like	23171
IPI00298828	apolipoprotein H (beta-2-glycoprotein I)	350
IPI00299086	syndecan binding protein (syntenin)	6386
IPI00302688	enoyl Coenzyme A hydratase domain containing 1	55862
IPI00305978	aldo-keto reductase family 7, member A2 (aflatoxin aldehyde reductase)	8574
IPI00329236	protein kinase C, delta	5580
IPI00332371	phosphofructokinase, liver	5211
IPI00410214	3'(2'), 5'-bisphosphate nucleotidase 1	10380
IPI00431645	haptoglobin	3240
IPI00553012	major histocompatibility complex, class II, DQ beta 1	3119
IPI00719373	immunoglobulin lambda locus	3535
IPI00009253	N-ethylmaleimide-sensitive factor attachment protein, alpha	8775
IPI00027193	chloride intracellular channel 5	53405
IPI00009901	nuclear transport factor 2	10204
IPI00026087	barrier to autointegration factor 1	8815
IPI00218728	platelet-activating factor acetylhydrolase, isoform Ib, alpha subunit 45kDa	5048
IPI00218916	arachidonate 5-lipoxygenase	240

IPI00419979	p21 protein (Cdc42/Rac)-activated kinase 2	5062
IPI00220421	centaurin, delta 2	116985
IPI00018871	ADP-ribosylation factor-like 8B	55207
IPI00022664	Rab geranylgeranyltransferase, alpha subunit	5875
IPI00027463	S100 calcium binding protein A6	6277
IPI00375380	proteasome (prosome, macropain) 26S subunit, non-ATPase, 13	5719
IPI00218342	methylenetetrahydrofolate dehydrogenase (NADP+ dependent) 1, methenyltetrahydrofolate cyclohydrolase, formyltetrahydrofolate synthetase	4522
IPI00029629	tripartite motif-containing 25	7706
IPI00219913	ubiquitin specific peptidase 14 (tRNA-guanine transglycosylase)	9097
IPI00473014	destrin (actin depolymerizing factor)	11034
IPI00029468	ARP1 actin-related protein 1 homolog A, centractin alpha (yeast)	10121
IPI00002459	annexin A6	309
IPI00001592	glycoprotein (transmembrane) nmb	10457
IPI00002230	arylacetamide deacetylase-like 1	57552
IPI00002405	2'-5'-oligoadenylate synthetase 3, 100kDa	4940
IPI00002460	annexin A7	310
IPI00004860	arginyl-tRNA synthetase	5917
IPI00008433	ribosomal protein S5	6193
IPI00009305	glucosamine-6-phosphate deaminase 1	10007
IPI00011603	proteasome (prosome, macropain) 26S subunit, non-ATPase, 3	5709
IPI00015944	echinoderm microtubule associated protein like 2	24139
IPI00017342	ras homolog gene family, member G (rho G)	391
IPI00017510	cytochrome c oxidase II	4513
IPI00020672	Bardet-Biedl syndrome 1	582
IPI00021290	ATP citrate lyase	47
IPI00022256	adaptor-related protein complex 2, mu 1 subunit	1173
IPI00024466	UDP-glucose ceramide glucosyltransferase-like 1	56886
IPI00027252	prohibitin 2	11331
IPI00029750	ribosomal protein S24	6229
IPI00030706	AHA1, activator of heat shock 90kDa protein ATPase homolog 1 (yeast)	10598
IPI00031583	USO1 homolog, vesicle docking protein (yeast)	8615
IPI00102864	hexokinase 2	3099
IPI00103994	leucyl-tRNA synthetase	51520
IPI00140420	staphylococcal nuclease and tudor domain containing 1	27044
IPI00215743	ribosome binding protein 1 homolog 180kDa (dog)	6238
IPI00216951	aspartyl-tRNA synthetase	1615
IPI00217872	phosphoglucomutase 1	5236
IPI00291928	RAB14, member RAS oncogene family	51552
IPI00295857	coatomer protein complex, subunit alpha	1314

IPI00302436	haloacid dehalogenase-like hydrolase domain containing 1A	8226
IPI00306436	signal transducer and activator of transcription 3 (acute-phase response factor)	6774
IPI00412216	vacuolar protein sorting 13 homolog C (S. cerevisiae)	54832
IPI00641073	solute carrier family 44, member 4	80736
IPI00884105	lysosomal-associated membrane protein 1	3916
IPI00001466	echinoderm microtubule associated protein like 4	27436
IPI00739539	ANKRD26-like family C, member 1B	728378
IPI00003348	guanine nucleotide binding protein (G protein), beta polypeptide 2	2783
IPI00007321	lysophospholipase I	10434
IPI00011454	glucosidase, alpha	23193
IPI00012889	surfactant, pulmonary-associated protein A1B	6435
IPI00013452	glutamyl-prolyl-tRNA synthetase	2058
IPI00014338	neutrophil cytosolic factor 4, 40kDa	4689
IPI00015018	pyrophosphatase (inorganic) 1	5464
IPI00016339	RAB5C, member RAS oncogene family	5878
IPI00020984	calnexin	821
IPI00022143	family with sequence similarity 62 (C2 domain containing), member A	23344
IPI00025049	mannose-6-phosphate receptor (cation dependent)	4074
IPI00025491	eukaryotic translation initiation factor 4A, isoform 1	1973
IPI00028031	acyl-Coenzyme A dehydrogenase, very long chain	37
IPI00031008	tenascin C (hexabrachion)	3371
IPI00031522	hydroxyacyl-Coenzyme A dehydrogenase/3-ketoacyl-Coenzyme A thiolase/enoyl-Coenzyme A hydratase (trifunctional protein), alpha subunit	3030
IPI00072377	SET nuclear oncogene	6418
IPI00155168	protein tyrosine phosphatase, receptor type, C	5788
IPI00216139	septin 6	23157
IPI00216184	phosphatidylinositol binding clathrin assembly protein	8301
IPI00216308	voltage-dependent anion channel 1	7416
IPI00216587	ribosomal protein S8	6202
IPI00218319	tropomyosin 3	7170
IPI00219604	mitogen-activated protein kinase kinase 1	5604
IPI00257882	peptidase D	5184
IPI00290770	chaperonin containing TCP1, subunit 3 (gamma)	7203
IPI00295105	carbonic anhydrase VI	765
IPI00296191	ATPase, H+ transporting, lysosomal 50/57kDa, V1 subunit H	51606
IPI00298289	reticulon 4	57142
IPI00300786	amylase, alpha 1A (salivary)	276
IPI00303318	family with sequence similarity 49, member B	51571
IPI00329633	threonyl-tRNA synthetase	6897

IPI00440493	ATP synthase, H+ transporting, mitochondrial F1 complex, alpha subunit 1, cardiac muscle	498
IPI00470467	P450 (cytochrome) oxidoreductase	5447
IPI00641829	HLA-B associated transcript 1	7919
IPI00644712	X-ray repair complementing defective repair in Chinese hamster cells 6 (Ku autoantigen, 70kDa)	2547
IPI00784807	immunoglobulin heavy constant gamma 2 (G2m marker)	3501
IPI00024403	copine III	8895
IPI00025307	cytochrome P450, family 27, subfamily A, polypeptide 1	1593
IPI00027107	Tu translation elongation factor, mitochondrial	7284
IPI00171199	proteasome (prosome, macropain) subunit, alpha type, 3	5684
IPI00221224	alanyl (membrane) aminopeptidase (aminopeptidase N, aminopeptidase M, microsomal aminopeptidase, CD13, p150)	290
IPI00465429	RUN and FYVE domain containing 1	80230
IPI00008240	methionyl-tRNA synthetase	4141
IPI00011253	ribosomal protein S3	6188
IPI00170972	chromosome 9 open reading frame 64	84267
IPI00217049	2'-5'-oligoadenylate synthetase 2, 69/71kDa	4939
IPI00022733	phospholipid transfer protein	5360
IPI00168262	glycosyltransferase 25 domain containing 1	79709
IPI00004312	signal transducer and activator of transcription 2, 113kDa	6773
IPI00007926	chromosome 6 open reading frame 108	10591
IPI00295618	platelet/endothelial cell adhesion molecule (CD31 antigen)	5175
IPI00031131	chromosome 20 open reading frame 3	57136
IPI00029469	ARP1 actin-related protein 1 homolog B, centractin beta (yeast)	10120
IPI00328415	cytochrome b5 reductase 3	1727
IPI00328867	v-src sarcoma (Schmidt-Ruppin A-2) viral oncogene homolog (avian)	6714
IPI00018298	interferon-induced protein with tetratricopeptide repeats 2	3433
IPI00037283	dynamin 1-like	10059
IPI00221088	ribosomal protein S9	6203
IPI00002270	chromosome 6 open reading frame 211	79624
IPI00300026	sulfotransferase family, cytosolic, 1A, phenol-preferring, member 1	6817
IPI00024660	KIAA0174	9798
IPI00027009	protein kinase C and casein kinase substrate in neurons 2	11252
IPI00023987	phosphopantothenoylcysteine synthetase	79717
IPI00030116	phosphoglucomutase 3	5238
IPI00216298	thioredoxin	7295
IPI00008787	N-acetylglucosaminidase, alpha- (Sanfilippo disease IIIB)	4669
IPI00006451	N-ethylmaleimide-sensitive factor	4905
IPI00004406	uridine phosphorylase 1	7378
IPI00017376	Sec23 homolog B (S. cerevisiae)	10483

IPI00104074	CD163 molecule	9332
IPI00221231	nudix (nucleoside diphosphate linked moiety X)-type motif 2	318
IPI00027851	hexosaminidase A (alpha polypeptide)	3073
IPI00024145	voltage-dependent anion channel 2	7417
IPI00059366	H2A histone family, member Y	9555
IPI00219663	T-cell, immune regulator 1, ATPase, H+ transporting, lysosomal V0 subunit A3	10312
IPI00375676	ferritin, light polypeptide	2512
IPI00009634	sulfide quinone reductase-like (yeast)	58472
IPI00025380	integrin, alpha L (antigen CD11A (p180), lymphocyte function-associated antigen 1	3683
IPI00429191	eukaryotic translation termination factor 1	2107
IPI00216514	CD47 molecule	961
IPI00014238	lysyl-tRNA synthetase	3735
IPI00032516	adaptor-related protein complex 1, mu 1 subunit	8907
IPI00218200	B-cell receptor-associated protein 31	10134
IPI00030702	isocitrate dehydrogenase 3 (NAD+) alpha	3419
IPI00028006	proteasome (prosome, macropain) subunit, beta type, 2	5690
IPI00456635	unc-13 homolog D (C. elegans)	201294
IPI00021954	golgi-specific brefeldin A resistance factor 1	8729
IPI00013068	eukaryotic translation initiation factor 3, subunit E	3646
IPI00183526	nucleolin	4691
IPI00025023	lactoperoxidase	4025
IPI00152653	dynein, axonemal, heavy chain 5	1767
IPI00103552	mucin 16, cell surface associated	94025
IPI00178316	mucin 4, cell surface associated	4585
IPI00302592	filamin A, alpha (actin binding protein 280)	2316
IPI00335168	myosin, light chain 6, alkali, smooth muscle and non-muscle	4637
IPI00410096	intraflagellar transport 172 homolog (Chlamydomonas)	26160
IPI00410380	PARK2 co-regulated	135138
IPI00788188	dynein, cytoplasmic 2, heavy chain 1	79659
IPI00032328	kininogen 1	3827
IPI00607861	hexose-6-phosphate dehydrogenase (glucose 1-dehydrogenase)	9563
IPI00021405	lamin A/C	4000
IPI00027434	ras homolog gene family, member C	389
IPI00396243	WD repeat domain 19	57728
IPI00004397	v-ral simian leukemia viral oncogene homolog B (ras related	5899
IPI00028064	cathepsin G	1511
IPI00009650	lipocalin 1 (tear prealbumin)	3933

Figure 2.2 Functional annotation of the top 10 biological processes of the normal BALF proteome as defined by four different "shotgun" proteomic approaches

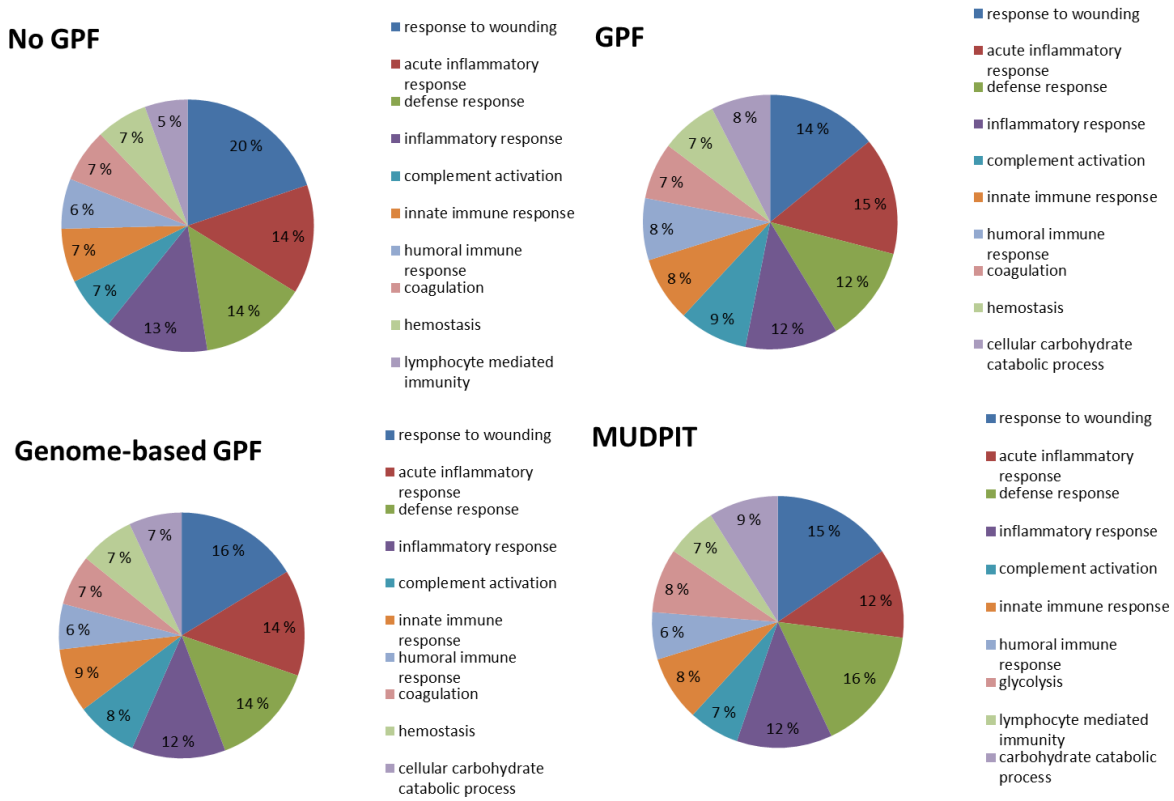

Table 2.3 Functional annotation, gene count, and p-value of the top 10-12 biological processes of the normal BALF proteome as defined by four different "shotgun" MS approaches

No GPF (Chen et al. in 2008)	Count	%	PValue
response to wounding	30	28	$2.4x10^{-19}$
acute inflammatory response	14	13	$6.9x10^{-14}$
defense response	26	24	$1.2x10^{-13}$
inflammatory response	20	19	$3.1x10^{-13}$
wound healing	12	11	$5.3x10^{-8}$
complement activation	7	7	$3.2x10^{-7}$
innate immune response	10	9	$3.2x10^{-7}$
coagulation	9	8	$3.7x10^{-7}$
hemostasis	9	8	$5.7x10^{-7}$
humoral immune response	8	7	$8.9x10^{-7}$

GPF (Gharib et al. in 2010)	Count	%	PValue
acute inflammatory response	24	10	$9.8x10^{-21}$
response to wounding	45	18	$1.2x10^{-19}$
defense response	45	18	$3.3x10^{-17}$
inflammatory response	33	13	$1.4x10^{-16}$
complement activation	13	5	$1.7x10^{-12}$
innate immune response	19	8	$8.4x10^{-12}$
complement activation, classical path	11	4	$1.5x10^{-11}$
humoral immune response	15	6	$2.8x10^{-11}$
cellular carbohydrate catabolic proce	15	6	$7.7x10^{-11}$
hemostasis	16	7	$2x10^{-10}$

Genome-based GPF (Nguyen et al. In 2012)	Count	%	PValue
response to wounding	44	24	$2x10^{-24}$
acute inflammatory response	22	12	$5.7x10^{-21}$
defense response	43	23	$5.7x10^{-21}$
inflammatory response	31	17	$1.3x10^{-18}$
humoral immune response	15	8	$4.1x10^{-13}$
innate immune response	18	10	$5.7x10^{-13}$
complement activation	12	7	$1.5x10^{-12}$
hemostasis	15	8	$3.4x10^{-11}$
coagulation	14	8	$2.2x10^{-10}$
humoral immune response m	9	5	$2.1x10^{-9}$
leukocyte mediated immunit	12	7	$5.5x10^{-9}$

MUDPIT (Gharib et al. in 2009)	Count	%	PValue
defense response	49	19	$3.9x10^{-19}$
response to wounding	45	17	$1.4x10^{-18}$
acute inflammatory response	19	7	$5.6x10^{-14}$
carbohydrate catabolic process	17	6	$5.2x10^{-11}$
innate immune response	18	7	$2.2x10^{-10}$
glycolysis	12	5	$3.2x10^{-10}$
leukocyte mediated immunity	13	5	$2.6x10^{-8}$
lymphocyte mediated immunity	12	5	$2.8x10^{-8}$
complement activation	10	4	$3.1x10^{-8}$
humoral immune response	12	5	$1x10^{-7}$
wound healing	17	6	$1.9x10^{-7}$
inflammatory response	32	12	$7.2x10^{-15}$

2.6 Notes to Chapter 2

1. Hirsch, J., et al., *Proteomics: current techniques and potential applications to lung disease.* Am J Physiol Lung Cell Mol Physiol, 2004. **287**(1): p. L1-23.

2. Murray J.F., N.J.A., *Murray and Nadel's textbook of respiratory medicine*. 4th ed. 2005, Philadelphia, Pennslyvania: Saunders.

3. Griese, M., *Pulmonary surfactant in health and human lung diseases: state of the art.* Eur Respir J, 1999. **13**(6): p. 1455-76.

4. Reynolds, H.Y., *Use of bronchoalveolar lavage in humans--past necessity and future imperative.* Lung, 2000. **178**(5): p. 271-93.

5. Lindahl, M., B. Stahlbom, and C. Tagesson, *Newly identified proteins in human nasal and bronchoalveolar lavage fluids: potential biomedical and clinical applications.* Electrophoresis, 1999. **20**(18): p. 3670-6.

6. Griese, M., J. Noss, and C. von Bredow, *Protein pattern of exhaled breath condensate and saliva.* Proteomics, 2002. **2**(6): p. 690-6.

7. Beier, J., et al., *Induced sputum methodology: Validity and reproducibility of total glutathione measurement in supernatant of healthy and asthmatic individuals.* J Lab Clin Med, 2004. **144**(1): p. 38-44.

8. Baughman, R.P. and M. Drent, *Role of bronchoalveolar lavage in interstitial lung disease.* Clin Chest Med, 2001. **22**(2): p. 331-41.

9. Rottoli, P. and E. Bargagli, *Is bronchoalveolar lavage obsolete in the diagnosis of interstitial lung disease?* Curr Opin Pulm Med, 2003. **9**(5): p. 418-25.

10. Haslam, P.L. and R.P. Baughman, *Report of ERS Task Force: guidelines for measurement of acellular components and standardization of BAL.* Eur Respir J, 1999. **14**(2): p. 245-8.

11. Meyer, K.C., *The role of bronchoalveolar lavage in interstitial lung disease.* Clin Chest Med, 2004. **25**(4): p. 637-49, v.

12. Lenz, A.G., et al., *Bronchoalveolar lavage fluid proteins in human lung disease: analysis by two-dimensional electrophoresis.* Electrophoresis, 1993. **14**(3): p. 242-4.

13. von Bredow, C., P. Birrer, and M. Griese, *Surfactant protein A and other bronchoalveolar lavage fluid proteins are altered in cystic fibrosis.* Eur Respir J, 2001. **17**(4): p. 716-22.

14. Soler, N., et al., *Bronchial microbial patterns in severe exacerbations of chronic obstructive pulmonary disease (COPD) requiring mechanical ventilation.* Am J Respir Crit Care Med, 1998. **157**(5 Pt 1): p. 1498-505.

15. Plymoth, A., et al., *Human bronchoalveolar lavage: biofluid analysis with special emphasis on sample preparation.* Proteomics, 2003. **3**(6): p. 962-72.

16. Schnapp, L.M., et al., *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**(1): p. 86-95.

17. Pittet, J.F., et al., *Biological markers of acute lung injury: prognostic and pathogenetic significance.* Am J Respir Crit Care Med, 1997. **155**(4): p. 1187-205.

18. Bowler, R.P., et al., *Proteomic analysis of pulmonary edema fluid and plasma in patients with acute lung injury.* Am J Physiol Lung Cell Mol Physiol, 2004. **286**(6): p. L1095-104.

19. Capron, F., *[Bronchoalveolar lavage].* Arch Anat Cytol Pathol, 1997. **45**(5): p. 255-60.

20. Wattiez, R., et al., *Human bronchoalveolar lavage fluid: two-dimensional gel electrophoresis, amino acid microsequencing and identification of major proteins.* Electrophoresis, 1999. **20**(7): p. 1634-45.

21. Bell, D.Y., et al., *Plasma proteins of the bronchoalveolar surface of the lungs of smokers and nonsmokers.* Am Rev Respir Dis, 1981. **124**(1): p. 72-9.

22. Noel-Georis, I., et al., *Proteomics as the tool to search for lung disease markers in bronchoalveolar lavage.* Dis Markers, 2001. **17**(4): p. 271-84.

23. Kindlen, S., *Physiology for Health Care and Nursing*, S. Kindlen, Editor 2003, Elsevier Health Sciences: UK.

24. Hermans, C. and A. Bernard, *Lung epithelium-specific proteins: characteristics and potential applications as markers.* Am J Respir Crit Care Med, 1999. **159**(2): p. 646-78.

25. Anderson, N.L. and N.G. Anderson, *The human plasma proteome: history, character, and diagnostic prospects.* Mol Cell Proteomics, 2002. **1**(11): p. 845-67.

26. Barnouin, K., *Two-dimensional gel electrophoresis for analysis of protein complexes.* Methods Mol Biol, 2004. **261**: p. 479-98.

27. Issaq, H.J., *The role of separation science in proteomics research.* Electrophoresis, 2001. **22**(17): p. 3629-38.

28. Washburn, M.P., D. Wolters, and J.R. Yates, 3rd, *Large-scale analysis of the yeast proteome by multidimensional protein identification technology.* Nat Biotechnol, 2001. **19**(3): p. 242-7.

29. Perkins, D.N., et al., *Probability-based protein identification by searching sequence databases using mass spectrometry data.* Electrophoresis, 1999. **20**(18): p. 3551-67.

30. Colinge, J., et al., *OLAV: towards high-throughput tandem mass spectrometry data identification.* Proteomics, 2003. **3**(8): p. 1454-63.

31. Yates, J.R., 3rd, et al., *Method to correlate tandem mass spectra of modified peptides to amino acid sequences in the protein database.* Anal Chem, 1995. **67**(8): p. 1426-36.

32. Keller, A., et al., *Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search.* Anal Chem, 2002. **74**(20): p. 5383-92.

33. Nesvizhskii, A.I., et al., *A statistical model for identifying proteins by tandem mass spectrometry.* Anal Chem, 2003. **75**(17): p. 4646-58.

34. Pappin, D.J., P. Hojrup, and A.J. Bleasby, *Rapid identification of proteins by peptide-mass fingerprinting.* Curr Biol, 1993. **3**(6): p. 327-32.

35. Henzel, W.J., et al., *Identifying proteins from two-dimensional gels by molecular mass searching of peptide fragments in protein sequence databases.* Proc Natl Acad Sci U S A, 1993. **90**(11): p. 5011-5.

36. Mann, M., P. Hojrup, and P. Roepstorff, *Use of mass spectrometric molecular weight information to identify proteins in sequence databases.* Biol Mass Spectrom, 1993. **22**(6): p. 338-45.

37. James, P., et al., *Protein identification by mass profile fingerprinting.* Biochem Biophys Res Commun, 1993. **195**(1): p. 58-64.

38. Yates, J.R., 3rd, et al., *Peptide mass maps: a highly informative approach to protein identification.* Anal Biochem, 1993. **214**(2): p. 397-408.

39. Bell, D.Y. and G.E. Hook, *Pulmonary alveolar proteinosis: analysis of airway and alveolar proteins.* Am Rev Respir Dis, 1979. **119**(6): p. 979-90.

40. Sabounchi-Schutt, F., et al., *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**(9): p. 1851-60.

41. Wattiez, R. and P. Falmagne, *Proteomics of bronchoalveolar lavage fluid.* J Chromatogr B Analyt Technol Biomed Life Sci, 2005. **815**(1-2): p. 169-78.

42. Chen, J., et al., *Exploration of the normal human bronchoalveolar lavage fluid proteome.* Proteomics Clin Appl, 2008. **2**(4): p. 585-95.

43. Giorgianni, F., D.M. Desiderio, and S. Beranova-Giorgianni, *Proteome analysis using isoelectric focusing in immobilized pH gradient gels followed by mass spectrometry.* Electrophoresis, 2003. **24**(1-2): p. 253-9.

44. Griffin, T.J. and R. Aebersold, *Advances in proteome analysis by mass spectrometry.* J Biol Chem, 2001. **276**(49): p. 45497-500.

45. Gharib, S.A., et al., *Mapping the lung proteome in cystic fibrosis.* J Proteome Res, 2009. **8**(6): p. 3020-8.

46. Shteynberg, D., et al., *iProphet: multi-level integrative analysis of shotgun proteomic data improves peptide and protein identification rates and error estimates.* Mol Cell Proteomics, 2011. **10**(12): p. M111.007690.

47. Keller, A. and D. Shteynberg, *Software pipeline and data analysis for MS/MS proteomics: the trans-proteomic pipeline.* Methods Mol Biol, 2011. **694**: p. 169-89.

48. Noel-Georis, I., et al., *Database of bronchoalveolar lavage fluid proteins.* J Chromatogr B Analyt Technol Biomed Life Sci, 2002. **771**(1-2): p. 221-36.

49. Sabounchi-Schutt, F., et al., *Changes in bronchoalveolar lavage fluid proteins in sarcoidosis: a proteomics approach.* Eur Respir J, 2003. **21**(3): p. 414-20.

50. Dennis, G., Jr., et al., *DAVID: Database for Annotation, Visualization, and Integrated Discovery.* Genome Biol, 2003. **4**(5): p. P3.

51. Ashburner, M., et al., *Gene ontology: tool for the unification of biology. The Gene Ontology Consortium.* Nat Genet, 2000. **25**(1): p. 25-9.

52. Gharib, S.A., et al., *Of mice and men: comparative proteomics of bronchoalveolar fluid.* Eur Respir J, 2010. **35**(6): p. 1388-95.

53. Spahr, C.S., et al., *Towards defining the urinary proteome using liquid chromatography-tandem mass spectrometry. I. Profiling an unfractionated tryptic digest.* Proteomics, 2001. **1**(1): p. 93-107.

54. Makarov, A., et al., *Dynamic range of mass accuracy in LTQ Orbitrap hybrid mass spectrometer.* J Am Soc Mass Spectrom, 2006. **17**(7): p. 977-82.

55. Nguyen, E.V., et al., *Proteomic Profiling of Bronchoalveolar Lavage Fluid in Critically Ill Patients with Ventilator-associated Pneumonia.* PLOS ONE, 2012.

56. Second, T.P., et al., *Dual-pressure linear ion trap mass spectrometer improving the analysis of complex protein mixtures.* Anal Chem, 2009. **81**(18): p. 7757-65.

57. Scherl, A., et al., *Genome-specific gas-phase fractionation strategy for improved shotgun proteomic profiling of proteotypic peptides.* Anal Chem, 2008. **80**(4): p. 1182-91.

58. Huang, d.W., B.T. Sherman, and R.A. Lempicki, *Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources.* Nat Protoc, 2009. **4**(1): p. 44-57.

59. Taylor, C.F., et al., *The minimum information about a proteomics experiment (MIAPE).* Nat Biotechnol, 2007. **25**(8): p. 887-93.

60. Orchard, S., H. Hermjakob, and R. Apweiler, *The proteomics standards initiative.* Proteomics, 2003. **3**(7): p. 1374-6.

61. Hermjakob, H., *The HUPO proteomics standards initiative--overcoming the fragmentation of proteomics data.* Proteomics, 2006. **6 Suppl 2**: p. 34-8.

62. Sze, S.K., et al., *Top-down mass spectrometry of a 29-kDa protein for characterization of any posttranslational modification to within one residue.* Proc Natl Acad Sci U S A, 2002. **99**(4): p. 1774-9.

63. Kelleher, N.L., *Top-down proteomics.* Anal Chem, 2004. **76**(11): p. 197A-203A.

Chapter 3

Proteomic Profiling of Bronchoalveolar Lavage Fluid in Critically Ill Patients with Ventilator Associated Pneumonia

This body of work was done in collaboration with Dr. Sina A. Gharib and Dr. Lynn M. Schnapp, Division of Pulmonary and Critical Care Medicine, Harborview Medical Center, University of Washington, Seattle, WA, US

3.1 Introduction

Ventilator-associated pneumonia (VAP), defined as pneumonia occurring after 48 hours of intubation and mechanical ventilation, is a major cause of increased mortality and morbidity in critically ill patients (1). Clinical diagnosis of VAP is especially challenging in the setting of acute lung injury (ALI) since many signs and symptoms of pneumonia such as fever, abnormal radiographs, and elevated white blood cell counts are common in all patients with ALI. Culturing samples from airways and airspaces using bronchoscopy with lavage or protected brushings is the key step in identifying VAP, but can also delay diagnostic and treatment plans. Therefore, identifying biomarkers for VAP in serum or bronchoalveolar lavage fluid (BALF) may provide a rapid, accurate and clinically useful test for early detection and risk stratification of critically ill patients. A number of individual proteins have been proposed as biomarkers for the presence of VAP, but single biochemical measurements are not consistent predictors of either onset or severity of VAP. Serum procalcitonin (2), serum C-reactive protein (3), and BALF soluble

triggering receptor expressed on myeloid cells-1 (sTREM-1) (4) are several proposed VAP candidates. However, predictive and quantitative threshold levels for these single biomarkers of VAP have been disappointingly nonspecific when follow-up validation studies were attempted (5, 6).

We hypothesized that comprehensively profiling the proteomic landscape of BALF in ALI patients would identify distinct protein signatures that could be utilized as diagnostic classifiers for VAP, and provide insights into the pathogenesis of this complex disorder. To this end, we integrated tandem mass spectrometry-based proteomics with statistical and computational methods to identify differentially expressed proteins between patients with and without VAP (VAP^+, VAP^-). Using classification algorithms, we narrowed our candidate list to three proteins and validated its predictive robustness in an independent cohort of critically ill patients undergoing evaluation for clinical suspicion of VAP.

3.2 Materials and Methods

BALF collection

The protocol for collecting human BALF was approved by the Institutional Review Board at the University of Washington. Patients with ALI undergoing bronchoscopy for clinically suspected VAP were enrolled. Written informed consent was obtained from patients or a legal next of kin. ALI was defined using standard criteria proposed by the American-European Consensus Conference (7). Clinical criteria for suspected VAP included: 48 hours of mechanical ventilation, new or progressive bilateral pulmonary infiltrates on radiograph, and one or more of the following: fever, leukocytosis, leukopenia, or an increase in purulent endotracheal secretions.

BALF was collected from healthy and ALI subjects as previously described (8, 9). Subject and BALF characteristics are shown in Table 3.1 and Table 3.2.

Shotgun proteomics analysis

Tryptic digests of each BALF sample were analyzed by HPLC-MS/MS using a NanoAquity HPLC system (Milford, MA, USA) via electrospray ionization on-line to a hybrid linear ion trap-orbitrap mass spectrometer (Thermo Fisher, San Jose, CA). Normal BALF samples were analyzed in an identical manner on a hybrid LTQ-Velos mass spectrometer (Thermo Fisher, San Jose, CA). The experiment was repeated in triplicate using gas phase fraction over the following three *m/z* ranges: 400-559, 559-846, 846-2000 (10). The tandem mass spectra were then matched to a protein sequence in the IPI Human 3.53 database using SEQUEST (11). For each pair-wise group comparison (i.e., normal, VAP^+, VAP^-), we restricted analysis to those proteins that were present in at least half the subjects of one group.

Correspondence and cluster analyses

Correspondence analysis was performed based on the variability in protein expression in normal and ALI populations. We performed hierarchical clustering of differentially expressed proteins between ALI subsets (VAP^+ vs. VAP^-). K-medians cluster analysis was performed on the expression of the limited proteomic classifier across all 30 ALI patients (MeV (12)).

Differential protein expression

To assess differences in relative protein abundance between subject populations, individual protein spectral counts were normalized using the spectral index (SI) metric as

previously described (13). We chose a 95% confidence threshold for significant differential expression.

Functional analysis

Functional annotation of the BALF proteomes for normal and ALI patients was conducted using database for annotation, visualization, and integrated discovery (DAVID) software (14).

Protein network analysis

We utilized experimentally verified protein interactions from several resources including Ingenuity (15) and STRING (16) to construct a relational network comprised of differentially expressed proteins in patients with VAP.

Western blot analysis

Equal volumes of BALF samples from a set of subjects not previously used in the shotgun proteomic analysis were separated by SDS-PAGE, blocked and incubated with polyclonal antibody S100A8 (Diagnostic Standards Laboratory). Relative band intensities were quantified and analyzed with ImageJ (17). Difference between groups was assessed with the Mann-Whitney test.

ELISA

Lactotransferrin levels in the BALF of a cohort of ALI patients not previously used in the shotgun proteomic analysis were measured in duplicate by ELISA (MyBioSource, San Diego,

CA) per manufacturer's instructions. Difference between groups was assessed using the Mann-Whitney test.

3.3 Results

Overview of BALF proteome in healthy subjects and patients with ALI

We identified 1288 unique proteins in the BALF of all 35 subjects (5 normal, 14 ALI: VAP^+, 16 ALI: VAP^-). Many of these proteins were not detected in most subjects, suggesting very low abundance. To improve confidence and ensure biologic relevance, we restricted our analysis to proteins that were detected in greater than 50% of subjects in at least one of the subgroups. We identified 251 unique proteins in the BALF of 5 normal volunteers and 394 unique proteins in 30 ALI subjects (i.e., VAP^+ and VAP^-). Of these, 184 proteins were common to both groups. The most abundant protein based on raw spectral counts across all BALF samples was albumin. Tables 3.3 and 3.4 contain lists of proteins found in normal and ALI subjects. Correspondence analysis of all identified BALF proteins distinctly separated normal subjects from ALI patients, implying that lung injury causes global changes in the expression levels of proteins in the airspaces (Figure 3.1).

BALF proteome in ALI

We applied a statistical test based on protein spectral counts known as the spectral index to identify 166 differentially expressed proteins between normal and ALI subjects. Of these, 47 proteins were significantly more abundant in normal BALF, whereas 119 were differentially upregulated in the BALF of ALI patients (Table 3.5). To better elucidate the pathways and mechanisms activated in ALI, we performed Gene Ontology analysis on the differentially

expressed BALF proteins. As shown in Table 3.7, the proteins upregulated during ALI mapped to a number of highly overrepresented functional categories including defense response (*P*-value: 2×10^{-15}), inflammatory response (*P*-value: 5×10^{-14}), wound healing (*P*-value: 7×10^{-13}), immunity (*P*-value: 1×10^{-9}). In contrast, downregulated proteins in ALI (and therefore more abundant in normal BALF) mapped to enriched processes involved in endopeptidase inhibitor activity (*P*-value: 3×10^{-10}) and metabolic processes (*P*-value: $\sim1\times10^{-5}$).

BALF proteome in VAP

As shown in Figure 3.1, correspondence analysis of the BALF proteome of all subjects did not segregate the subset of ALI patients with VAP (VAP^+, n = 14) from those without it (VAP^-, n = 16). This is likely due to the dominant effect of lung injury on global protein expression and variability in the airspaces, regardless of the presence or absence of pneumonia. Therefore, we applied the spectral index method to compare the BALF proteome between VAP^+ and VAP^- patients, and identified 75 differentially expressed proteins out of a total of 394 unique proteins (Table 3.6). The majority of the differentially expressed proteins were upregulated in BALF of VAP^+ subjects (n = 60 proteins), whereas 15 proteins were more abundantly expressed in VAP^- patients. We subsequently performed unsupervised hierarchical cluster analysis on the differentially expressed proteomic profiles of all 30 ALI patients and demonstrated that this signature is a reasonably powerful discriminator between VAP^+ and VAP^- subjects (Figure 3.2). Functional categorization of differentially expressed proteins revealed that mechanisms involved in defense response, immunity, response to bacterium and leukocyte migration were enriched in the VAP^+ subjects, whereas fibrinogen complex, cell surface binding, wound healing and developmental processes were overrepresented in the VAP^- patients (Table 3.8). These findings

suggest that there is persistent activation of anti-bacterial and immunologic pathways in the airspace milieu of critically ill patients with VAP, whereas reparative mechanisms have been initiated in the lungs of ALI patients without VAP.

Protein interaction network in VAP

To better elucidate functional relationships among differentially expressed proteins in VAP, we performed a network analysis using available protein-protein interaction databases (Figure 3.3). This putative network highlights complex partnerships between proteins that are functionally active in the extracellular space (e.g., MMP9, ELA2, LTF, PLUNC), plasma membrane (e.g., ITGAM, ITGB2), and the cytoplasm (e.g., S100A8, S100A9, MPO, CTSG).

Development and validation of a proteomic classifier for VAP

We employed a two-step strategy to discover a limited proteomic predictor for VAP. Initially we applied a supervised classification algorithm known as prediction analysis of microarrays (PAM) to the entire ALI BALF proteomic dataset of 394 proteins (18). PAM was used to train and cross validate a set of proteins with robust properties to discriminate VAP^+ and VAP^- patients. However, this list was comprised of 28 proteins. Since our aim was to develop a much smaller classifier, we selected the top three proteins with (i) the highest discrimination scores between the groups and (ii) spectral index values reaching 95% confidence levels. This limited proteomic predictor included S100A8, lactotransferrin (LTF), and actinin 1 (ACTN1). As shown in Figure 3.4, when the 30 ALI patients underwent unsupervised K-median clustering based on the expression profiles of these 3 proteins, 90% of the subjects were correctly classified. To validate the predictive power of the VAP proteomic signature, we biochemically

assessed the levels of S100A8 and LTF in the BALF of an independent group of ALI patients with clinical suspicion for VAP (Figure 3.5).

3.4 Discussion

Ventilator associated pneumonia remains a common complication in critically ill patients and adversely influences clinical outcomes. Early diagnosis and treatment may reduce the morbidity and mortality associated with VAP, but robust biomarkers have not been identified. Our study represents the largest investigation of the airspace proteome in patients with acute lung injury and VAP. Using a shotgun proteomics approach, we not only identified differentially expressed proteins in the BALF of ALI subjects with VAP, but also developed and validated a limited proteomic signature to discriminate critically ill patients with VAP from those without it.

We established that compared to normal individuals, the BALF proteome of ALI patients was highly enriched in proteins involved in inflammation, defense response and immunity (Table 3.7). However, global proteomic profiling of normal and ALI subjects did not segregate ALI subsets (VAP^+ and VAP^-, Figure 3.1), most likely because acute lung injury (regardless of VAP status) is the dominant effector of protein expression in BALF. We therefore limited our analysis to the 30 ALI patients (14 VAP^+ and 16 VAP^-) and identified 76 differentially expressed proteins (Figure 3.2). The majority of these proteins were more abundant in the BALF of VAP^+ patients, and mapped to distinct functional categories that included defense response and response to bacterium (Table 3.8)—highlighting ongoing activation of these pro-inflammatory processes. In contrast, the BALF of VAP^- subjects was enriched in proteins involved in wound healing and development, suggesting that in the absence of active infection, reparative programs were induced following acute lung injury.

We further explored functional relationships among differentially expressed proteins in VAP using gene product interaction network analysis (Figure 3.3). Complex biological networks possess topologic properties that are functionally informative (19). For example, highly connected nodes—known as hubs, disproportionately affect the stability of a network and may be important orchestrators of biological processes (19, 20). Figure 3 depicts several such hubs, including integrin beta 2 (ITGB2), integrin alpha M (ITGAM), and myeloperoxidase (MPO)—all of which were up-regulated in VAP^+ patients and directly interacted with each other. The β_2 integrins ITGB2 and ITGAM (also known as CD18 and CD11b) are critical mediators of neutrophil adhesion and transmigration across the endothelium and promote bacterial phagocytosis (21). Their up-regulation in VAP^+ subjects suggests persistent activation and infiltration of leukocytes into the airspaces. Furthermore, studies demonstrate that MPO, a potent phagocyte-derived peroxidase, suppresses neutrophil apoptosis through signaling via ITGB2 and ITGAM thereby prolonging inflammation and contributing to persistent lung injury (22). In contrast, another network hub, fibronectin 1 (FN1), was downregulated in VAP^+ subjects. FN1, a principal component of the extracellular matrix, is a key regulator of epithelial wound repair (23). FN1 was recently reported to be down-regulated in airway epithelial cells of asthmatics and this suppression contributed to impaired airway repair in these patients (24). These findings suggest that VAP is associated with an imbalance between overabundance of pro-inflammatory mediators and reduced expression of reparative proteins. Although our analysis was primarily descriptive, integrating shotgun proteomics with bioinformatics and network analysis provided unique insights into the pathogenesis of VAP in critically ill patients and identified putative mechanisms whereby VAP promotes further lung injury and delays healing processes.

A powerful feature of our shotgun proteomics approach was its amenability to supervised discriminatory analysis for biomarker discovery. Using a two-step approach, we identified a limited proteomic signature that accurately separated VAP^+ and VAP^- patients (Figure 3.4). We validated the robustness of the classifier by prospectively measuring two member proteins, lactotransferrin (LTF) and S100 calcium binding protein A8 (S100A8), in a group of ALI patients undergoing bronchoscopy for suspicion of VAP. We confirmed that the expression patterns of LTF and S100A8 in the validation cohort mimicked those of the original 30 ALI subjects, with both proteins being significantly more abundant in BALF of VAP^+ patients (Figure 3.5). Although the utility of such signatures is solely due to their ability to distinguish classes based on distinct expression patterns, we noted that two members of the classifier (LTF and S100A8) mapped to the VAP interactome (Figure 3.3). This observation implied that these candidates, beyond their discriminatory properties, may also have biologic relevance in the pathogenesis of VAP.

LTF is a key component of the respiratory tract antimicrobial defense system and innate immunity (25). Elevated LTF levels have been reported in the BALF of patients with cystic fibrosis (26), chronic bronchitis (27), and acute respiratory distress syndrome (28). S100A8 is an antimicrobial protein expressed by leukocytes and epithelial cells, often partnering with S100A9 to inhibit bacterial adhesion to epithelium and limit infectivity. S100A8 is a potent neutrophil chemoattractant and recruits these cells to sites of inflammation in response to lipopolysaccharide (29). Animal studies have demonstrated that S100A8 plays a critical role in promoting the transepithelial migration of leukocytes into the airspaces during bacterial pneumonia, whereas its blockade dramatically diminishes this inflammatory process (30).

Therefore, our BALF proteomic signature robustly discriminated ALI patients with VAP from those with pneumonia and identified putative effectors of disease propagation.

Our study has a number of limitations. First, our findings need to be confirmed independently in larger cohorts of critically ill patients at risk for developing VAP. Secondly, despite its comprehensive coverage, shotgun proteomics is still biased towards detecting proteins with larger molecular weights and higher abundance. Since we do not know the relative contribution of different cell types in airspaces to the BALF proteome, our network analysis and proposed mechanisms of VAP pathogenesis must be considered speculative. Follow-up corroboration of the roles played by candidate targets in VAP is necessary. Finally, the VAP proteomic classifier identified in our study is not unique, and other statistical methodologies based on the BALF proteome may yield different yet highly informative biomarker panels.

In summary, we have outlined a systematic, proteomics-based approach to segregate ALI patients with and without VAP based on differentially expressed proteins in the BALF. We identified and validated a limited protein classifier that possessed robust discriminatory properties while maintaining biological relevance. Collectively, our findings suggest that integrating computational tools with proteomics of BALF is a promising venue for classifying and prognosticating pulmonary disorders, and may provide mechanistic insights into disease pathogenesis.

Table 3.1 Characteristics of ALI patients.

Subject characteristics	**VAP^+ (n = 14)**	**VAP^- (n = 16)**
Age (mean ± SD)	53 ± 31	48 ± 28
Gender	14 male, 0 female	10 male, 6 female
Ventilator days* (mean ± SD)	10 ± 25	15 ± 18
P_aO_2/F_iO_2** (mean ± SD)	206 ± 54 mmHg	212 ± 48 mmHg

*Number of days on the ventilator prior to bronchoscopy
**Ratio of arterial oxygen partial pressure (P_aO_2) to fraction of inspired oxygen (F_iO_2) at time of bronchoscopy

Table 3.2 Characteristics of ALI subjects and their BALF samples. State of discharge, age, sex, whether patient was diagnosed with VAP, microbiology of bacterial strain detected in cell culture, number of days on ventilator, P/F ratio, and risk factor provided.

Subject	State at discharge	Age	Sex	VAP	Microbiology	Ventdays *	P/F Ŧ	ALI risk factor
1	Alive	60	M	Yes	Bifidobacterium scardovii	5/18	162	Sepsis
2	Alive	40	M	Yes	Staphylococcus aureus, Streptococci	2/7	185	Trauma
3	Alive	76	M	Yes	Escherichia coli, Enterococcus	7/11	148	Trauma
4	Alive	22	M	Yes	Alpha streptococci, MRSA	6/8	247	Sepsis
5	Alive	56	M	Yes	H.influenzae	3/13	240	Trauma
6	Alive	56	M	Yes	Streptococcus pseudopneumoniae, Staphylococcus aureus, H.influenzae	6/19	247	Burns, Sepsis
7	Alive	25	M	Yes	Alpha streptococcus, Diphtheroids	19/38	280	Trauma
8	Alive	69	M	Yes	Staphylococcus aureus	5/13	310	Sepsis
9	Alive	25	M	Yes	Staphylococcus aureus, acenitobacter	11/26	138	Trauma
10	Dead	79	M	Yes	Pseudomonas	16/35	185	Trauma
11	Alive	66	M	Yes	Staphylococcus aureus	14/15	173	Trauma
12	Dead	62	M	Yes	Staphylococcus aureus	7/9	187	Cardiac shock
13	Alive	46	M	Yes	Streptococcus milleri	4/21	141	Trauma
14	Dead	65	M	Yes	Pseudomonas	12/17	240	Sepsis
15	Alive	70	M	NO	N/A	18/33	212	Trauma
16	Alive	38	F	NO	N/A	5/13	144	Pancreatitis
17	Alive	22	M	NO	N/A	3/5	138	Trauma
18	Alive	20	M	NO	N/A	10/30	185	Trauma
19	Alive	62	F	NO	N/A	7/11	254	Sepsis
20	Alive	53	M	NO	N/A	4/10	223	Sepsis, Shock
21	Alive	61	M	NO	N/A	9/15	208	Sepsis
22	Alive	56	F	NO	N/A	2/5	250	Sepsis

23	Dead	60	M	NO	N/A	10/25	220	Cancer
24	Alive	62	F	NO	N/A	1/10	234	Sepsis
25	Alive	49	M	NO	N/A	4/11	168	Trauma
26	Alive	59	F	NO	N/A	6/11	300	Trauma
27	Alive	40	M	NO	N/A	11/25	300	Trauma
28	Alive	49	F	NO	N/A	5/17	183	Sepsis
29	Alive	42	M	NO	N/A	5/35	170	Burns
30	Alive	25	M	NO	N/A	13/38	203	Trauma

***** Days on ventilator prior to bronchoscopy/Total number of days on the ventilator
[Ŧ] P/F ratio closest to time of bronchoscopy

Table 3.3 List of common proteins identified in BALF of normal volunteers and ALI subjects ranked by spectral counts.

IPI	Protein Name	Entrez Gene ID	Spectral Counts
IPI00745872	Albumin	213	49443
IPI00012889	surfactant, pulmonary-associated protein A1B; surfactant, pulmonary-associated protein A1	6435	32627
IPI00298497	fibrinogen beta chain	2244	7107
IPI00032258	complement component 4A (Rodgers blood group)	720	4454
IPI00478003	alpha-2-macroglobulin	2	4227
IPI00291410	chromosome 20 open reading frame 114	92747	3656
IPI00553177	serpin peptidase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 1	5265	3591
IPI00021885	fibrinogen alpha chain	2243	3500
IPI00739237	similar to Complement C3 precursor	653879	3334
IPI00019502	myosin, heavy chain 9, non-muscle	4627	2733
IPI00021891	fibrinogen gamma chain	2266	2573
IPI00855918	mucin 5B, oligomeric mucus/gel-forming	727897	2035
IPI00007047	S100 calcium binding protein A8	6279	1959
IPI00027462	S100 calcium binding protein A9	6280	1947
IPI00654755	hemoglobin, beta	3043	1558
IPI00298860	Lactotransferrin	4057	1525
IPI00032291	complement component 5	727	1428
IPI00296083	surfactant, pulmonary-associated protein B	6439	1385
IPI00022229	apolipoprotein B (including Ag(x) antigen)	338	1172
IPI00017601	ceruloplasmin (ferroxidase)	1356	1075
IPI00021841	apolipoprotein A-I	335	970
IPI00219365	Moesin	4478	962
IPI00431645	Haptoglobin	3240	944
IPI00292530	inter-alpha (globulin) inhibitor H1	3697	923
IPI00007244	Myeloperoxidase	4353	885
IPI00103397	mucin 5AC, oligomeric mucus/gel-forming	4586	885
IPI00382470	heat shock protein 90kDa alpha (cytosolic),class A member 1	3320	874
IPI00298994	talin 1	7094	869
IPI00220327	keratin 1 (epidermolytic hyperkeratosis)	3848	853
IPI00302592	filamin A, alpha (actin binding protein 280)	2316	791
IPI00219018	glyceraldehyde-3-phosphate dehydrogenase	2597	776
IPI00219077	leukotriene A4 hydrolase	4048	774
IPI00410714	hemoglobin, alpha 1; hemoglobin, alpha 2	3039	766
IPI00643920	transketolase (Wernicke-Korsakoff syndrome)	7086	746
IPI00004573	polymeric immunoglobulin receptor	5284	721

IPI00218914	aldehyde dehydrogenase 1 family, member A1	216	718
IPI00305461	inter-alpha (globulin) inhibitor H2	3698	710
IPI00218918	annexin A1	301	659
IPI00022395	complement component 9	735	650
IPI00099110	deleted in malignant brain tumors 1	1755	596
IPI00219525	phosphogluconate dehydrogenase	5226	583
IPI00783313	phosphorylase, glycogen; liver	5836	583
IPI00022463	Transferrin	7018	581
IPI00009342	IQ motif containing GTPase activating protein 1	8826	579
IPI00019591	complement factor B	629	578
IPI00186290	eukaryotic translation elongation factor 2	1938	554
IPI00453473	histone cluster 1, H4i; histone cluster 1, H4a; histone cluster 1, H4d; histone cluster 1, H4f; histone cluster 1, H4k; histone cluster 1, H4j; histone cluster 1, H4c; histone cluster 1, H4h; histone cluster 1, H4b; histone cluster 1, H4e; histone cluster 1, H4l; histone cluster 2, H4a; histone cluster 4, H4; histone cluster 2, H4b	8294	549
IPI00418471	Vimentin	7431	544
IPI00025447	eukaryotic translation elongation factor 1 alpha 1	1915	489
IPI00022774	valosin-containing protein	7415	485
IPI00021727	complement component 4 binding protein, alpha	722	443
IPI00029739	complement factor H	3075	432
IPI00010133	coronin, actin binding protein, 1A	11151	427
IPI00418169	annexin A2	302	425
IPI00013808	actinin, alpha 4	81	421
IPI00217966	lactate dehydrogenase A	3939	420
IPI00746388	Ezrin	7430	414
IPI00219682	Stomatin	2040	405
IPI00019359	keratin 9 (epidermolyticpalmoplantarkeratoderma)	3857	404
IPI00024067	clathrin, heavy chain (Hc)	1213	402
IPI00298971	Vitronectin	7448	400
IPI00026314	gelsolin (amyloidosis, Finnish type)	2934	382
IPI00009865	keratin 10 (epidermolytic hyperkeratosis; keratosis palmaris et plantaris)	3858	380
IPI00169383	phosphoglycerate kinase1	5230	377
IPI00414676	heat shock protein 90kDa alpha (cytosolic),class B member 1	3326	371
IPI00028091	ARP3 actin-related protein 3 homolog (yeast)	10096	359
IPI00216008	glucose-6-phosphate dehydrogenase	2539	358
IPI00002459	annexin A6	309	342
IPI00550991	serpin peptidase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 3	12	331
IPI00291262	Clusterin	1191	329
IPI00003865	heat shock 70kDa protein 8	3312	326

IPI00021263	tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, zeta polypeptide	7534	325
IPI00298828	apolipoprotein H (beta-2-glycoprotein I)	350	323
IPI00032311	lipopolysaccharide binding protein	3929	314
IPI00555812	group-specific component (vitamin D binding protein)	2638	314
IPI00027223	isocitrate dehydrogenase 1 (NADP+), soluble	3417	304
IPI00304273	apolipoprotein A-IV	337	298
IPI00018873	nicotinamidephosphoribosyltransferase	10135	290
IPI00019580	Plasminogen	5340	289
IPI00294395	complement component 8, beta polypeptide	732	286
IPI00217987	integrin, alpha M (complement component 3 receptor 3 subunit)	3684	280
IPI00022371	histidine-rich glycoprotein	3273	277
IPI00019568	coagulation factor II (thrombin)	2147	269
IPI00329801	annexin A5	308	265
IPI00005159	ARP2 actin-related protein 2 homolog (yeast)	10097	253
IPI00021842	apolipoprotein E	348	245
IPI00453476			244
IPI00218251	transglutaminase 2 (C polypeptide, protein-glutamine-gamma-glutamyltransferase)	7052	243
IPI00304925	heat shock 70kDa protein 1A; heat shock 70kDa protein 1B	3303	240
IPI00021854	apolipoprotein A-II	336	240
IPI00014055	napsin A aspartic peptidase	9476	239
IPI00216699	fermitin family homolog 3 (Drosophila)	83706	239
IPI00005721	defensin, alpha 1; defensin, alpha 1	1667	239
IPI00171611	histone cluster 2, H3c; histone cluster 2, H3a; histone cluster 2, H3d	126961	233
IPI00003935	histone cluster 2, H2be	8349	230
IPI00026781	fatty acid synthase	2194	228
IPI00013895	S100 calcium binding protein A11	6282	228
IPI00291878	surfactant, pulmonary-associated protein D	6441	225
IPI00005118	hexokinase 3 (white cell)	3101	222
IPI00289758	calpain 2, (m/II) large subunit	824	221
IPI00291866	serpin peptidase inhibitor, clade G (C1 inhibitor), member 1, (angioedema, hereditary)	710	219
IPI00027769	elastase 2, neutrophil	1991	217
IPI00032328	kininogen 1	3827	215
IPI00031461	GDP dissociation inhibitor 2	2665	212
IPI00218192	inter-alpha (globulin) inhibitor H4 (plasma Kallikrein-sensitive glycoprotein)	3700	211
IPI00552578	serum amyloid A1; serum amyloid A2	6288	210
IPI00295400	tryptophanyl-tRNAsynthetase	7453	210
IPI00219757	glutathione S-transferase pi	2950	203

IPI00007910	solute carrier family 34 (sodium phosphate), member 2	10568	201
IPI00000816	tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, epsilon polypeptide	7531	200
IPI00026119	ubiquitin-like modifier activating enzyme 1	7317	198
IPI00216691	profilin 1	5216	198
IPI00011285	calpain 1, (mu/I) large subunit	823	195
IPI00291175	Vinculin	7414	195
IPI00025084	calpain, small subunit 1	826	190
IPI00000105	major vault protein	9961	189
IPI00554811	actin related protein 2/3 complex, subunit 4, 20kDa; tubulin tyrosine ligase-like family, member 3	10093	189
IPI00218646	cytochrome b-245, beta polypeptide (chronic granulomatous disease)	1536	188
IPI00103356	integrin, beta 2 (complement component 3 receptor 3 and 4 subunit)	3689	187
IPI00022255	olfactomedin 4	10562	186
IPI00009856	palate, lung and nasal epithelium associated	51297	183
IPI00216318	tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, beta polypeptide	7529	182
IPI00552280	bactericidal/permeability-increasing protein	671	181
IPI00550069	ribonuclease/angiogenin inhibitor 1	6050	178
IPI00793199	annexin A4	307	170
IPI00477992	complement component 1, q subcomponent, B chain	713	170
IPI00022426	alpha-1-microglobulin/bikunin precursor	259	170
IPI00219217	lactate dehydrogenase B	3945	162
IPI00023673	lectin, galactoside-binding, soluble, 3 binding protein	3959	162
IPI00008274	CAP, adenylatecyclase-associated protein 1 (yeast)	10487	159
IPI00005161	actin related protein 2/3 complex, subunit 2, 34kDa	10109	159
IPI00456969	dynein, cytoplasmic 1, heavy chain 1	1778	154
IPI00789605	myosin, light chain 6, alkali, smooth muscle and non-muscle; myosin, light chain 6B, alkali, smooth muscle and non-muscle	4637	154
IPI00011654	tubulin, beta	203068	153
IPI00018931	vacuolar protein sorting 35 homolog (S. cerevisiae)	55737	153
IPI00216256	WD repeat domain 1	9948	149
IPI00299547	lipocalin 2	3934	147
IPI00022431	alpha-2-HS-glycoprotein	197	146
IPI00384051	proteasome (prosome, macropain) activator subunit 2 (PA28 beta)	5721	143
IPI00027230	heat shock protein 90kDa beta (Grp94), member 1	7184	138
IPI00657682	glutathione S-transferase A1	2938	136
IPI00847989	pyruvate kinase, muscle	5315	136
IPI00001639	karyopherin (importin) beta 1	3837	136
IPI00010896	chloride intracellular channel 1	1192	127

IPI00183046	protein tyrosine phosphatase, non-receptor type 6	5777	124
IPI00168184	protein phosphatase 2 (formerly 2A), regulatory subunit A , alpha isoform	5518	116
IPI00021304	keratin 2 (epidermal ichthyosisbullosa of Siemens)	3849	116
IPI00012011	cofilin 1 (non-muscle)	1072	115
IPI00025491	eukaryotic translation initiation factor 4A, isoform 1	1973	114
IPI00246058	programmed cell death 6 interacting protein	10015	114
IPI00022389	C-reactive protein, pentraxin-related	1401	113
IPI00294739	SAM domain and HD domain 1	25939	111
IPI00030936	tetraspanin 1	10103	106
IPI00419585	peptidylprolylisomerase A (cyclophilin A)	5478	106
IPI00478231	ras homolog gene family, member A	387	105
IPI00220642	tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, gamma polypeptide	7532	103
IPI00292858	thymidine phosphorylase	1890	100
IPI00215914	ADP-ribosylation factor 1	375	99
IPI00220644	pyruvate kinase, muscle	5315	98
IPI00022429	orosomucoid 1	5004	96
IPI00219038	H3 histone, family 3A; H3 histone, family 3B (H3.3B)	3020	92
IPI00026185	capping protein (actin filament) muscle Z-line, beta	832	91
IPI00007058	coronin, actin binding protein, 1B	57175	87
IPI00013955	mucin 1, cell surface associated	4582	84
IPI00329331	UDP-glucose pyrophosphorylase 2	7360	84
IPI00022977	creatine kinase, brain	1152	81
IPI00004358	phosphorylase, glycogen; brain	5834	78
IPI00218638	myosin IF	4542	78
IPI00003362	heat shock 70kDa protein 5 (glucose-regulated protein, 78kDa)	3309	72
IPI00010779	tropomyosin 4	7171	71
IPI00178926	immunoglobulin J polypeptide, linker protein for immunoglobulin alpha and mu polypeptides	3512	66
IPI00026216	aminopeptidasepuromycin sensitive	9520	66
IPI00103552	mucin 16, cell surface associated	94025	63
IPI00414320	annexin A11	311	62
IPI00021428	actin, alpha 1, skeletal muscle	58	61
IPI00332371	phosphofructokinase, liver	5211	56
IPI00298961	exportin 1 (CRM1 homolog, yeast)	7514	52
IPI00643567	CAP, adenylatecyclase-associated protein 1 (yeast)	10487	51
IPI00000874	peroxiredoxin 1	5052	41
IPI00003348	guanine nucleotide binding protein (G protein), beta polypeptide 2	2783	41
IPI00027497	glucose phosphate isomerase	2821	38

Table 3.4 List of 166 differential proteins found in BALF samples of 5 normal volunteers to 30 ALI patients. Orange highlighted IPI identifiers represent proteins enriched in BALF of ALI subjects.

IPI	Protein Name	Entrez Gene ID	Spectral Index
IPI00783987	complement component 3	718	-1.00
IPI00418163	complement component 4B	721	-1.00
IPI00295777	glycerol-3-phosphate dehydrogenase 1	2819	-1.00
IPI00451624	cartilage acidic protein 1	55118	-1.00
IPI00027848	mannose receptor, C type 1	4360	-0.96
IPI00413451	serpin peptidase inhibitor, clade B, member 6	5269	-0.95
IPI00011302	CD59 molecule, complement regulatory protein	966	-0.84
IPI00022462	transferrin receptor (p90, CD71)	7037	-0.80
IPI00027482	serpin peptidase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 6	866	-0.80
IPI00297487	cathepsin H	1512	-0.80
IPI00215767	UDP-Gal:betaGlcNAc beta 1,4- galactosyltransferase, polypeptide 1	2683	-0.80
IPI00012303	selenium binding protein 1	8991	-0.80
IPI00302925	4 chaperonin containing TCP1, subunit 8 (theta)	10694	-0.80
IPI00009104	RuvB-like 2 (E. coli)	10856	-0.78
IPI00018246	hexokinase 1	3098	-0.76
IPI00021302	sushi domain containing 2	56241	-0.74
IPI00017704	coactosin-like 1 (Dictyostelium)	23406	-0.73
IPI00008485	aconitase 1, soluble	48	-0.73
IPI00032179	serpin peptidase inhibitor, clade C (antithrombin), member 1	462	-0.71
IPI00291005	malate dehydrogenase 1, NAD (soluble)	4190	-0.71
IPI00374315	chromosome 6 open reading frame 58	352999	-0.68
IPI00022488	hemopexin	3263	-0.64
IPI00295386	carbonyl reductase 1	873	-0.63
IPI00220301	peroxiredoxin 6	9588	-0.61
IPI00022810	cathepsin C	1075	-0.60
IPI00305477	cystatin SN	1469	-0.60
IPI00013382	cystatin SA	1470	-0.60
IPI00032294	cystatin S	1472	-0.60
IPI00257508	dihydropyrimidinase-like 2	1808	-0.60
IPI00215746	fatty acid binding protein 4, adipocyte	2167	-0.60
IPI00292946	serpin peptidase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 6	6906	-0.60
IPI00020436	RAB11B, member RAS oncogene family	9230	-0.60

IPI00304557	83 chromosome 20 open reading frame 70	140683	-0.60
IPI00419237	leucine aminopeptidase 3	51056	-0.60
IPI00031420	UDP-glucose dehydrogenase	7358	-0.58
IPI00744692	transaldolase 1	6888	-0.57
IPI00008164	prolyl endopeptidase	5550	-0.57
IPI00163563	phosphatidylethanolamine-binding protein 4	157310	-0.56
IPI00479877	aldehyde dehydrogenase 9 family, member A1	223	-0.55
IPI00010180	carboxylesterase 1 (monocyte/macrophage serine esterase 1)	1066	-0.55
IPI00006114	serpin peptidase inhibitor, clade F (alpha-2 antiplasmin, pigment epithelium derived factor), member 1	5176	-0.54
IPI00000874	peroxiredoxin 1	5052	-0.54
IPI00022432	transthyretin (prealbumin, amyloidosis type I)	7276	-0.54
IPI00060715	potassium channel tetramerisation domain containing 12	115207	-0.53
IPI00022974	prolactin-induced protein	5304	-0.51
IPI00294158	biliverdin reductase A	644	-0.50
IPI00296353	Rho GTPase activating protein 18	93663	-0.50
IPI00847989	pyruvate kinase, muscle	5315	0.55
IPI00216256	WD repeat domain 1	9948	0.56
IPI00014338	neutrophil cytosolic factor 4, 40kDa	4689	0.56
IPI00298994	talin 1	7094	0.56
IPI00879709	complement component 6	729	0.56
IPI00552280	bactericidal/permeability-increasing protein	671	0.57
IPI00000875	eukaryotic translation elongation factor 1 gamma	1937	0.57
IPI00892671	immunoglobulin heavy constant gamma 1 (G1m marker)	3500	0.57
IPI00029863	serpin peptidase inhibitor, clade F (alpha-2 antiplasmin, pigment epithelium derived factor), member 2	5345	0.57
IPI00293925	ficolin (collagen/fibrinogen domain containing) 3 (Hakata antigen)	8547	0.57
IPI00012500	BRCA1 interacting protein C-terminal helicase 1	83990	0.57
IPI00796878	ATP-binding cassette, sub-family C (CFTR/MRP), member 12	94160	0.57
IPI00010779	tropomyosin 4	7171	0.57
IPI00384051	proteasome (prosome, macropain) activator subunit 2 (PA28 beta)	5721	0.57
IPI00022395	complement component 9	735	0.57
IPI00219682	stomatin	2040	0.57

IPI00218319	tropomyosin 3	7170	0.58
IPI00155168	protein tyrosine phosphatase, receptor type, C	5788	0.58
IPI00029739	complement factor H	3075	0.59
IPI00025084	calpain, small subunit 1	826	0.59
IPI00021854	apolipoprotein A-II	336	0.59
IPI00641229	immunoglobulin heavy constant alpha 2 (A2m marker)	3494	0.60
IPI00255052	NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 9, 22kDa	4715	0.60
IPI00003909	solute carrier family 2 (facilitated glucose transporter), member 3	6515	0.60
IPI00016513	RAB10, member RAS oncogene family	10890	0.60
IPI00165579	CNDP dipeptidase 2 (metallopeptidase M20 family)	55748	0.60
IPI00183046	protein tyrosine phosphatase, non-receptor type 6	5777	0.61
IPI00003817	Rho GDP dissociation inhibitor (GDI) beta	397	0.61
IPI00026314	gelsolin (amyloidosis, Finnish type)	2934	0.61
IPI00305461	inter-alpha (globulin) inhibitor H2	3698	0.62
IPI00299547	lipocalin 2	3934	0.62
IPI00168184	protein phosphatase 2 (formerly 2A), regulatory subunit A , alpha isoform	5518	0.63
IPI00019580	plasminogen	5340	0.63
IPI00022389	C-reactive protein, pentraxin-related	1401	0.63
IPI00027769	elastase 2, neutrophil	1991	0.63
IPI00027509	matrix metallopeptidase 9 (gelatinase B, 92kDa gelatinase, 92kDa type IV collagenase)	4318	0.63
IPI00010796	procollagen-proline, 2-oxoglutarate 4-dioxygenase (proline 4-hydroxylase), beta polypeptide	5034	0.63
IPI00008274	CAP, adenylate cyclase-associated protein 1 (yeast)	10487	0.65
IPI00304925	heat shock 70kDa protein 1A	3303	0.65
IPI00298971	vitronectin	7448	0.66
IPI00219038	H3 histone, family 3A	3020	0.67
IPI00216319	tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, eta polypeptide	7533	0.67
IPI00004524	grancalcin, EF-hand calcium binding protein	25801	0.67
IPI00218918	annexin A1	301	0.67
IPI00018873	nicotinamide phosphoribosyltransferase	10135	0.69
IPI00291410	chromosome 20 open reading frame 114	92747	0.70
IPI00291175	vinculin	7414	0.71
IPI00478003	alpha-2-macroglobulin	2	0.71
IPI00419585	peptidylprolyl isomerase A (cyclophilin A)	5478	0.71

IPI00025491	eukaryotic translation initiation factor 4A, isoform 1	1973	0.72
IPI00007244	myeloperoxidase	4353	0.72
IPI00004656	beta-2-microglobulin	567	0.73
IPI00026781	fatty acid synthase	2194	0.73
IPI00032311	lipopolysaccharide binding protein	3929	0.73
IPI00299024	brain abundant, membrane attached signal protein 1	10409	0.73
IPI00019502	myosin, heavy chain 9, non-muscle	4627	0.73
IPI00410714	hemoglobin, alpha 1	3039	0.76
IPI00168728	immunoglobulin heavy constant mu	3507	0.77
IPI00217987	integrin, alpha M (complement component 3 receptor 3 subunit)	3684	0.77
IPI00027423	protein phosphatase 1, catalytic subunit, alpha isoform	5499	0.77
IPI00171611	histone cluster 2, H3c	126961	0.77
IPI00013808	actinin, alpha 4	81	0.77
IPI00000105	major vault protein	9961	0.77
IPI00003935	histone cluster 2, H2be	8349	0.78
IPI00553177	serpin peptidase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 1	5265	0.78
IPI00002459	annexin A6	309	0.79
IPI00453473	histone cluster 1, H4i	8294	0.80
IPI00216008	glucose-6-phosphate dehydrogenase	2539	0.80
IPI00478231	ras homolog gene family, member A	387	0.80
IPI00021304	keratin 2 (epidermal ichthyosis bullosa of Siemens)	3849	0.80
IPI00022429	orosomucoid 1	5004	0.80
IPI00027230	heat shock protein 90kDa beta (Grp94), member 1	7184	0.80
IPI00220642	tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, gamma polypeptide	7532	0.80
IPI00797709	coronin, actin binding protein, 1C	23603	0.80
IPI00028064	cathepsin G	1511	0.80
IPI00218646	cytochrome b-245, beta polypeptide (chronic granulomatous disease)	1536	0.82
IPI00555812	group-specific component (vitamin D binding protein)	2638	0.82
IPI00554811	actin related protein 2/3 complex, subunit 4, 20kDa	10093	0.82
IPI00643920	transketolase (Wernicke-Korsakoff syndrome)	7086	0.83
IPI00654755	hemoglobin, beta	3043	0.83
IPI00179330	ribosomal protein S27a	6233	0.83
IPI00783313	phosphorylase, glycogen; liver (Hers disease, glycogen storage disease type VI)	5836	0.83
IPI00431645	haptoglobin	3240	0.84

IPI00298860	lactotransferrin	4057	0.85
IPI00021891	fibrinogen gamma chain	2266	0.85
IPI00010133	coronin, actin binding protein, 1A	11151	0.86
IPI00298497	fibrinogen beta chain	2244	0.86
IPI00022229	apolipoprotein B (including Ag(x) antigen)	338	0.87
IPI00025447	eukaryotic translation elongation factor 1 alpha 1	1915	0.87
IPI00216318	tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, beta polypeptide	7529	0.87
IPI00021263	tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, zeta polypeptide	7534	0.87
IPI00007047	S100 calcium binding protein A8	6279	0.88
IPI00219077	leukotriene A4 hydrolase	4048	0.89
IPI00000816	tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, epsilon polypeptide	7531	0.90
IPI00005159	ARP2 actin-related protein 2 homolog (yeast)	10097	0.90
IPI00455174	chromosome 20 open reading frame 74	57186	0.90
IPI00166768	tubulin, alpha 1c	84790	0.90
IPI00011654	tubulin, beta	203068	0.90
IPI00302592	filamin A, alpha (actin binding protein 280)	2316	0.92
IPI00003865	heat shock 70kDa protein 8	3312	0.93
IPI00027462	S100 calcium binding protein A9	6280	0.93
IPI00745872	albumin	213	0.93
IPI00418169	annexin A2	302	0.93
IPI00013955	mucin 1, cell surface associated	4582	0.93
IPI00099110	deleted in malignant brain tumors 1	1755	0.93
IPI00013508	actinin, alpha 1	87	0.97
IPI00215914	ADP-ribosylation factor 1	375	0.97
IPI00032258	complement component 4A (Rodgers blood group)	720	0.97
IPI00219018	glyceraldehyde-3-phosphate dehydrogenase	2597	0.97
IPI00217966	lactate dehydrogenase A	3939	0.97
IPI00103397	mucin 5AC, oligomeric mucus/gel-forming	4586	0.97
IPI00789605	myosin, light chain 6, alkali, smooth muscle and non-muscle	4637	0.97
IPI00552578	serum amyloid A1	6288	0.97
IPI00012889	surfactant, pulmonary-associated protein A1	653509	0.99
IPI00021885	fibrinogen alpha chain	2243	0.99
IPI00021439	actin, beta	60	1.00
IPI00414676	heat shock protein 90kDa alpha (cytosolic), class B member 1	3326	1.00

IPI00746388	ezrin	7430	1.00
IPI00739237	similar to Complement C3 precursor	653879	1.00

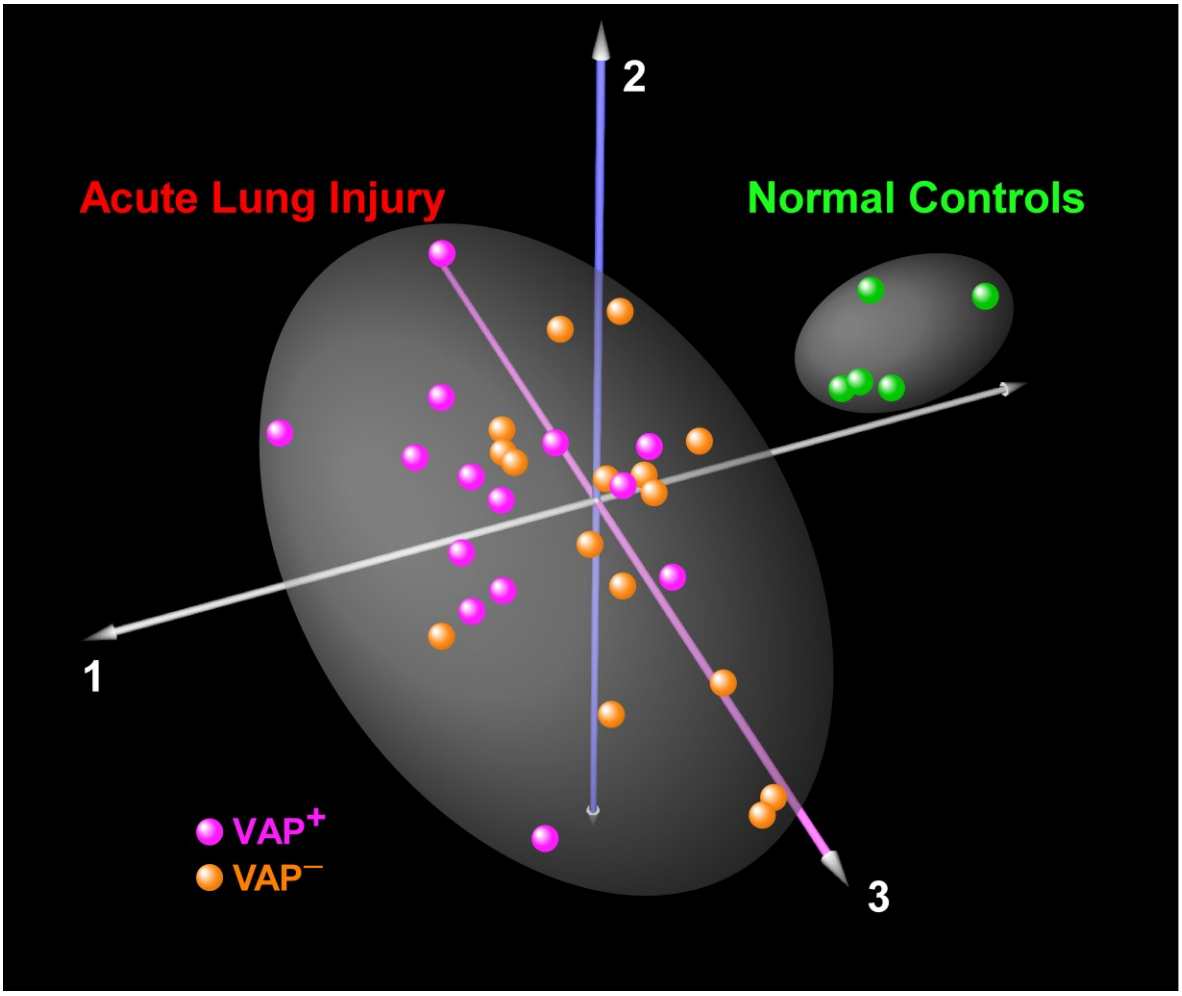

Figure 3.1: Principle component analysis on proteins identified by shotgun proteomics in the BALF of 5 normal volunteers and 30 ALI. Pink spheres, normal volunteers; white spheres, ALI patients.

Table 3.5 List of overrepresented functional categories found in BALF samples of normal and ALI subjects

NORMAL			**ALI**		
Functional Category	**P-value**	**# of proteins mapped**	**Functional Category**	**P-value**	**# of proteins mapped**
Endopeptidase inhibitor activity	8.60E-10	10	Defense response	2.12E-14	30
Secreted	1.50E-07	18	Acute inflammatory response	8.22E-14	15
Cofactor metabolic process	7.98E-06	7	Response to wounding	1.81E-12	26
Hexose metabolic process	1.09E-04	6	Actin binding cytoskeletal protein	5.89E-10	19
protease inhibitor	1.23E-04	5	Platelet alpha granule lumen	1.89E-09	9
Monosaccharide metabolic process	2.15E-04	6	Melanosome	4.64E-09	11
Serine-type endopeptidase inhibitor activity	2.36E-04	5	Immune effector process	2.00E-07	11
oxidoreductase	3.49E-04	8	Innate immune response	2.64E-07	11
Proteinase inhibitor I25, cystatin	4.84E-04	3	Humoral immune response	3.37E-07	9
Glucose metabolic process	5.60E-04	5	Acute-phase response	1.00E-06	7
Cystatin	6.33E-04	3	Complement activation	1.35E-06	7
NAD(P)-binding domain	7.68E-04	5	Protein polymerization	3.44E-06	7
Oxidation reduction	8.94E-04	8	Negative regulation of protein metabolic process	3.03E-05	10
Regulation of inflammatory response	9.00E-04	4	Regulation of apoptosis	4.45E-05	20

Table 3.6 List of 75 differential proteins found in BALF samples of ALI patients. Orange highlighted IPI identifiers represent proteins enriched in BALF of VAP^-, while pink highlighted IPI identifiers represents proteins enriched in BALF of VAP^+.

IPI	Protein Name	Entrez Gene ID	Spectral Index
IPI00021885	fibrinogen alpha chain	2243	-0.59
IPI00298237	tripeptidyl peptidase I	1200	-0.58
IPI00028306	DIX domain containing 1	85458	-0.56
IPI00009865	keratin 10	3858	-0.52
IPI00026320	ubiquitin protein ligase E3 component n-recognin 5	51366	-0.52
IPI00006705	secretoglobin, family 1A, member 1 (uteroglobin)	7356	-0.49
IPI00022895	alpha-1-B glycoprotein	1	-0.48
IPI00796878	ATP-binding cassette, sub-family C (CFTR/MRP), member 12	94160	-0.47
IPI00027019	proline-rich protein BstNI subfamily 4	5545	-0.45
IPI00298828	apolipoprotein H (beta-2-glycoprotein I)	350	-0.44
IPI00414283	fibronectin 1	2335	-0.44
IPI00006114	serpin peptidase inhibitor, clade F, member 1	5176	-0.43
IPI00298497	fibrinogen beta chain	2244	-0.43
IPI00021304	keratin 2 (epidermal ichthyosis bullosa of Siemens)	3849	-0.42
IPI00654755	hemoglobin, beta	3043	-0.42
IPI00303318	family with sequence similarity 49, member B	51571	0.42
IPI00218638	myosin IF	4542	0.42
IPI00022391	amyloid P component, serum	325	0.42
IPI00017184	EH-domain containing 1	10938	0.43
IPI00328257	adaptor-related protein complex 1, beta 1 subunit	162	0.43
IPI00155168	protein tyrosine phosphatase, receptor type, C	5788	0.43
IPI00295857	coatomer protein complex, subunit alpha	1314	0.44
IPI00242956	Fc fragment of IgG binding protein	8857	0.44
IPI00297444	CD177 molecule	57126	0.45
IPI00021428	actin, alpha 1, skeletal muscle	58	0.45
IPI00010796	procollagen-proline, 2-oxoglutarate 4-dioxygenase, beta polypeptide	5034	0.45
IPI00413587	BH3 interacting domain death agonist	637	0.45
IPI00021290	ATP citrate lyase	47	0.45
IPI00744706	spectrin, alpha, non-erythrocytic 1 (alpha-fodrin)	6709	0.45
IPI00008380	protein phosphatase 2 (formerly 2A), catalytic subunit, alpha isoform	5515	0.45
IPI00027497	glucose phosphate isomerase	2821	0.46
IPI00028064	cathepsin G	1511	0.46

IPI00013508	actinin, alpha 1	87	0.46
IPI00024067	clathrin, heavy chain (Hc)	1213	0.47
IPI00012728	acyl-CoA synthetase long-chain family member 1	2180	0.48
IPI00016610	poly(rC) binding protein 1	5093	0.48
IPI00007750	tubulin, alpha 4a	7277	0.49
IPI00027409	proteinase 3	5657	0.50
IPI00022255	olfactomedin 4	10562	0.50
IPI00020618	histone cluster 1, H4g	8369	0.50
IPI00220219	coatomer protein complex, subunit beta 2 (beta prime)	9276	0.50
IPI00296635	glucan (1,4-alpha-), branching enzyme 1	2632	0.51
IPI00295851	coatomer protein complex, subunit beta 1	1315	0.52
IPI00011454	glucosidase, alpha; neutral AB	23193	0.52
IPI00027423	protein phosphatase 1, catalytic subunit, alpha isoform	5499	0.53
IPI00004524	grancalcin, EF-hand calcium binding protein	25801	0.53
IPI00003909	solute carrier family 2 (facilitated glucose transporter), member 3	6515	0.53
IPI00783313	phosphorylase, glycogen; liver	5836	0.54
IPI00027462	S100 calcium binding protein A9	6280	0.54
IPI00003269	actin, beta-like 2	345651	0.54
IPI00030872	vanin 2	8875	0.54
IPI00216691	profilin 1	5216	0.55
IPI00221091	ribosomal protein S15a	6210	0.55
IPI00000875	eukaryotic translation elongation factor 1 gamma	1937	0.55
IPI00292532	cathelicidin antimicrobial peptide	820	0.57
IPI00298860	lactotransferrin	4057	0.57
IPI00020210	dysferlin, limb girdle muscular dystrophy 2B (autosomal recessive)	8291	0.57
IPI00021926	proteasome (prosome, macropain) 26S subunit, ATPase, 6	5706	0.57
IPI00102685	myeloid-associated differentiation marker	91663	0.57
IPI00007047	S100 calcium binding protein A8	6279	0.58
IPI00654777	eukaryotic translation initiation factor 3, subunit F	8665	0.58
IPI00025252	protein disulfide isomerase family A, member 3	2923	0.58
IPI00103356	integrin, beta 2 (complement component 3 receptor 3 and 4 subunit)	3689	0.58
IPI00218916	arachidonate 5-lipoxygenase	240	0.59
IPI00027509	matrix metallopeptidase 9	4318	0.60
IPI00009856	palate, lung and nasal epithelium associated	51297	0.60
IPI00217987	integrin, alpha M (complement component 3 receptor 3 subunit)	3684	0.60

IPI00218433	cytochrome b-245, alpha polypeptide	1535	0.61
IPI00552280	bactericidal/permeability-increasing protein	671	0.62
IPI00007244	myeloperoxidase	4353	0.63
IPI00003817	Rho GDP dissociation inhibitor (GDI) beta	397	0.65
IPI00020984	calnexin	821	0.65
IPI00022975	arachidonate 5-lipoxygenase-activating protein	241	0.66
IPI00332371	phosphofructokinase, liver	5211	0.69
IPI00027769	elastase 2, neutrophil	1991	0.70

Table 3.7 Functional analysis of differentially expressed proteins in the BALF of patients with ALI and Controls.

Enriched in BALF of ALI patients

Gene Ontology Category	Number of Proteins	Enrichment *P*-value
Protein binding	101	2.1×10^{-15}
Defense response	30	7.2×10^{-15}
Acute inflammatory response	15	4.7×10^{-14}
Response to wounding	26	7.2×10^{-13}
Response to stress	44	1.6×10^{-12}
Immune system process	30	1.1×10^{-9}
Response to stimulus	59	2.6×10^{-9}
Cytoskeletal protein binding	20	5.5×10^{-9}
Innate immune response	11	1.8×10^{-7}
Humoral immune response	9	2.5×10^{-7}

Enriched in BALF of Control subjects

Gene Ontology Category	Number of Proteins	Enrichment *P*-value
Endopeptidase inhibitor activity	10	2.8×10^{-10}
Cofactor metabolic process	7	1.1×10^{-5}
Enzyme regulator activity	11	1.2×10^{-4}
Hexose metabolic process	6	1.4×10^{-4}

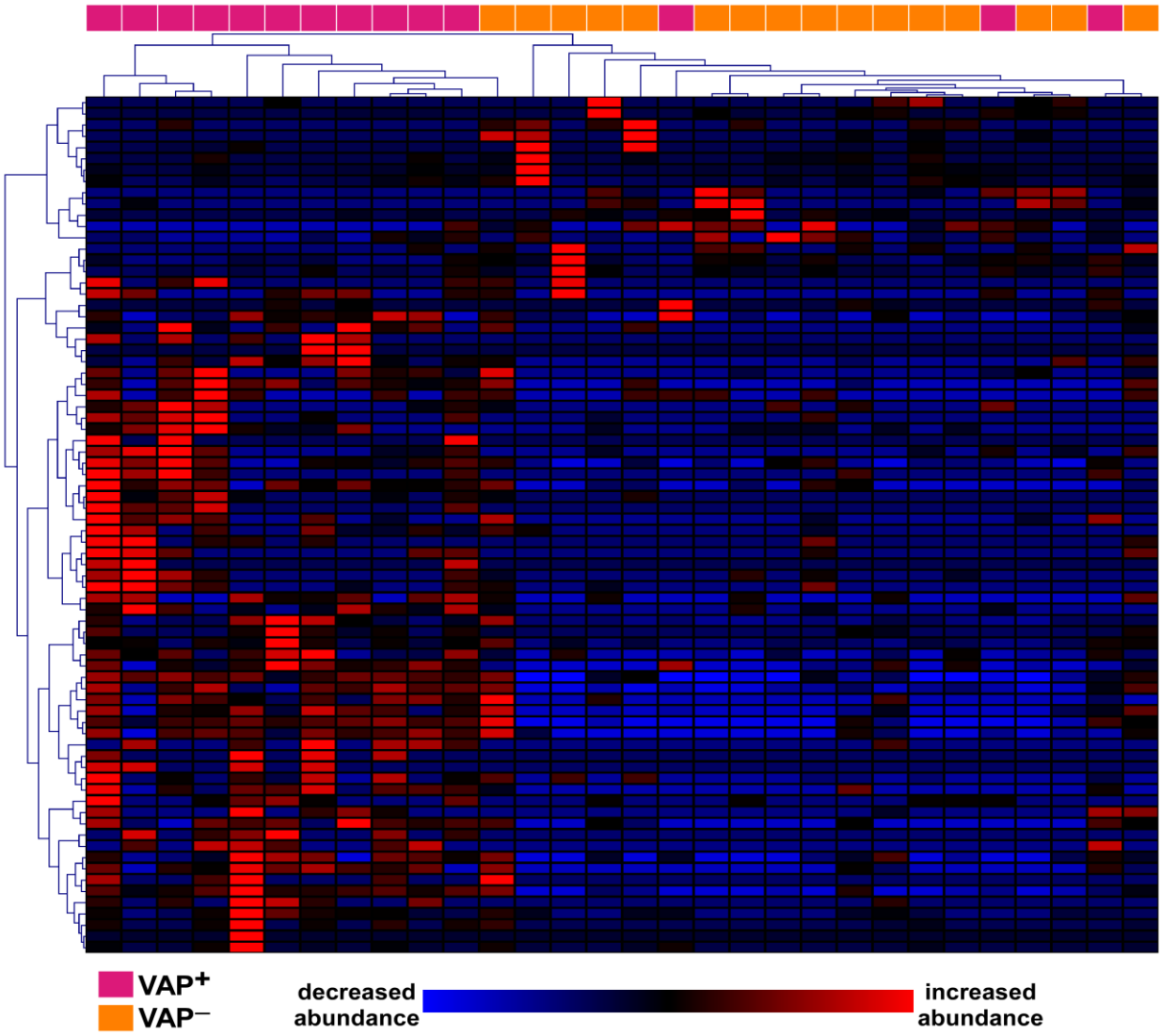

Figure 3.2: Hierarchical clustering of differentially expressed proteins in BALF of ALI subjects. Protein abundance is represented on a gradient from blue to red, with proteins showing high spectral counts represented by red and proteins showing low spectral counts represented by blue. Each column represents a patient sample, pink boxes represents patients diagnosed with VAP and orange boxes represents patients not diagnosed with VAP. Individual proteins are displayed in each row.

Table 3.8 Functional analysis of differentially expressed proteins in the BALF of VAP^+ and VAP^- patients.

Enriched in BALF of VAP^+ patients

Gene Ontology Category	Number of Proteins	Enrichment *P*-value
Defense response	14	4.8×10^{-7}
Protein binding	48	6.9×10^{-7}
Calcium ion binding	13	9.0×10^{-5}
Immune system process	14	9.1×10^{-5}
Response to other organism	7	8.9×10^{-4}
Defense response to bacterium	5	9.4×10^{-4}
Leukocyte migration	4	1.4×10^{-3}

Enriched in BALF of VAP^- patients

Gene Ontology Category	Number of Proteins	Enrichment *P*-value
Fibrinogen complex	3	1.2×10^{-3}
Developmental process	9	2.4×10^{-3}
Cell surface binding	13	3.2×10^{-3}
Wound healing	4	7.5×10^{-3}

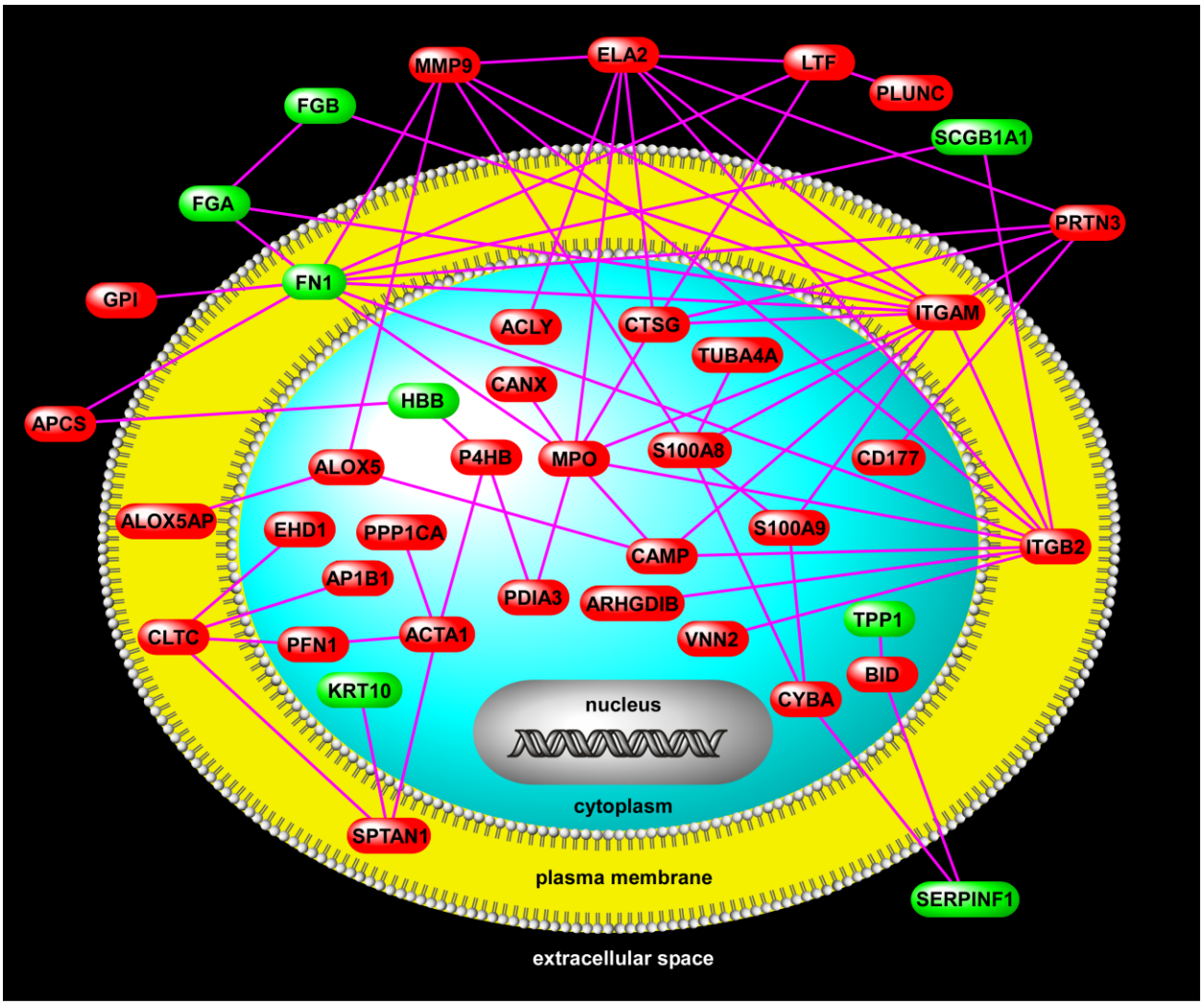

Figure 3.3: Protein network analysis of differentially expressed proteins in BALF of ALI subjects. Red are hubs enriched in VAP^{+} and green hubs are enriched in VAP^{-}.

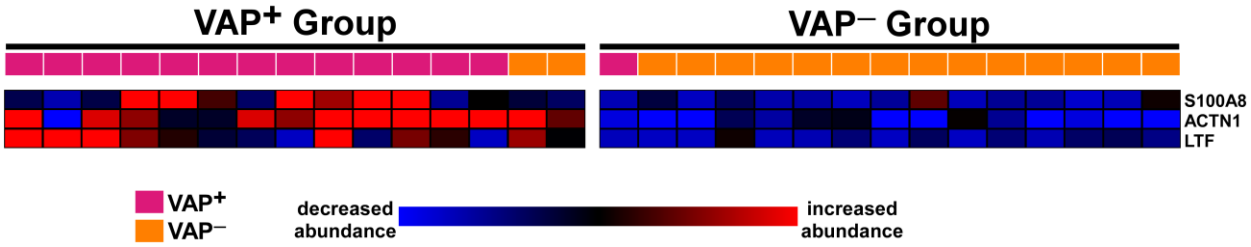

Figure 3.4 K-median clustering based on Euclidean distance was performed to discern protein profiles that may be specific to VAP indicating S100A8, Actin, and lactoferrin as candidate proteins.

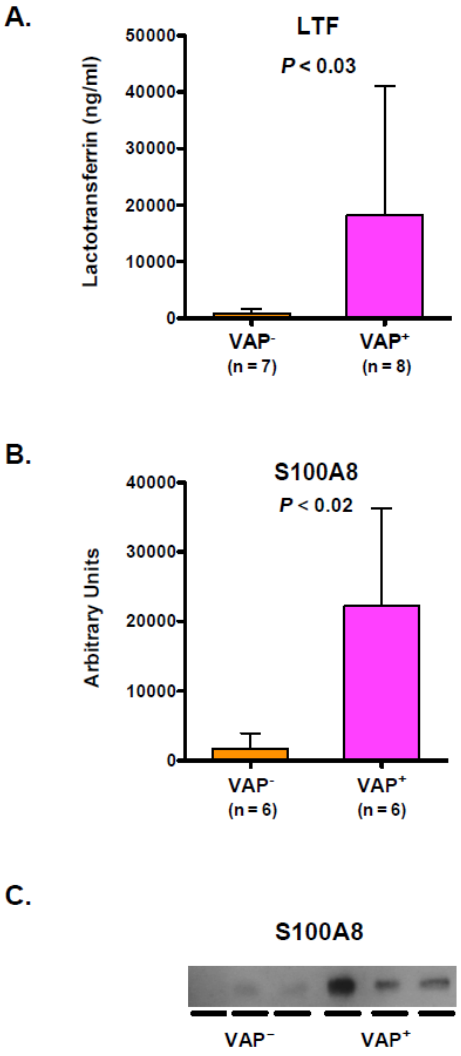

Figure 3.5: Western blot and densitrometry analysis of BALF from ALI patients A) Lactoferrin B) S100A8 C) S100A8 Western blot

3.5 Notes to Chapter 3

1. Chastre J, Fagon JY. Ventilator-associated pneumonia. *Am J Respir Crit Care Med* 2002;165:867-903.
2. Oppert M, Reinicke A, Muller C, Barckow D, Frei U, Eckardt KU. Elevations in procalcitonin but not c-reactive protein are associated with pneumonia after cardiopulmonary resuscitation. *Resuscitation* 2002;53:167-170.
3. Povoa P, Coelho L, Almeida E, Fernandes A, Mealha R, Moreira P, Sabino H. C-reactive protein as a marker of ventilator-associated pneumonia resolution: A pilot study. *Eur Respir J* 2005;25:804-812.
4. Determann RM, Millo JL, Gibot S, Korevaar JC, Vroom MB, van der Poll T, Garrard CS, Schultz MJ. Serial changes in soluble triggering receptor expressed on myeloid cells in the lung during development of ventilator-associated pneumonia. *Intensive Care Med* 2005;31:1495-1500.
5. Rea-Neto A, Youssef NC, Tuche F, Brunkhorst F, Ranieri VM, Reinhart K, Sakr Y. Diagnosis of ventilator-associated pneumonia: A systematic review of the literature. *Crit Care* 2008;12:R56.
6. Oudhuis GJ, Beuving J, Bergmans D, Stobberingh EE, ten Velde G, Linssen CF, Verbon A. Soluble triggering receptor expressed on myeloid cells-1 in bronchoalveolar lavage fluid is not predictive for ventilator-associated pneumonia. *Intensive Care Med* 2009;35:1265-1270.
7. 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.

8. Steinberg KP, Milberg JA, Martin TR, Maunder RJ, Cockrill BA, Hudson LD. Evolution of bronchoalveolar cell populations in the adult respiratory distress syndrome. *Am J Respir Crit Care Med* 1994;150:113-122.
9. Chen J, Ryu S, Gharib SA, Goodlett DR, Schnapp LM. Exploration of the normal human bronchoalveolar lavage fluid proteome. *Proteomics Clin Appl* 2008;2:585-595.
10. Scherl A, Shaffer SA, Taylor GK, Kulasekara HD, Miller SI, Goodlett DR. Genome-specific gas-phase fractionation strategy for improved shotgun proteomic profiling of proteotypic peptides. *Anal Chem* 2008;80:1182-1191.
11. Link AJ, Eng J, Schieltz DM, Carmack E, Mize GJ, Morris DR, Garvik BM, Yates JR, 3rd. Direct analysis of protein complexes using mass spectrometry. *Nat Biotechnol* 1999;17:676-682. java/Propub/biotech/nbt0799_0676.fulltext java/Propub/biotech/nbt0799_0676.abstract.
12. Saeed AI, Sharov V, White J, Li J, Liang W, Bhagabati N, Braisted J, Klapa M, Currier T, Thiagarajan M, et al. Tm4: A free, open-source system for microarray data management and analysis. *Biotechniques* 2003;34:374-378.
13. Fu X, Gharib SA, Green PS, Aitken ML, Frazer DA, Park DR, Vaisar T, Heinecke JW. Spectral index for assessment of differential protein expression in shotgun proteomics. *J Proteome Res* 2008;7:845-854.
14. 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.
15. Calvano SE, Xiao W, Richards DR, Felciano RM, Baker HV, Cho RJ, Chen RO, Brownstein BH, Cobb JP, Tschoeke SK, et al. A network-based analysis of systemic inflammation in humans. *Nature* 2005;437:1032-1037.

16. Jensen LJ, Kuhn M, Stark M, Chaffron S, Creevey C, Muller J, Doerks T, Julien P, Roth A, Simonovic M, et al. String 8--a global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Res* 2009;37:D412-416.
17. Rasband WS. Image j. US National Institutes of Health, Bethesda, Maryland USA, http://imagej.nih.gov/ij/; 1997-2011.
18. Tibshirani R, Hastie T, Narasimhan B, Chu G. Diagnosis of multiple cancer types by shrunken centroids of gene expression. *Proc Natl Acad Sci U S A* 2002;99:6567-6572.
19. Luscombe NM, Babu MM, Yu H, Snyder M, Teichmann SA, Gerstein M. Genomic analysis of regulatory network dynamics reveals large topological changes. *Nature* 2004;431:308-312.
20. Vidal M, Cusick ME, Barabasi AL. Interactome networks and human disease. *Cell* 2011;144:986-998.
21. Grommes J, Soehnlein O. Contribution of neutrophils to acute lung injury. *Mol Med* 2011;17:293-307.
22. El Kebir D, Jozsef L, Pan W, Filep JG. Myeloperoxidase delays neutrophil apoptosis through cd11b/cd18 integrins and prolongs inflammation. *Circ Res* 2008;103:352-359.
23. Herard AL, Pierrot D, Hinnrasky J, Kaplan H, Sheppard D, Puchelle E, Zahm JM. Fibronectin and its alpha 5 beta 1-integrin receptor are involved in the wound-repair process of airway epithelium. *Am J Physiol* 1996;271:L726-733.
24. Kicic A, Hallstrand TS, Sutanto EN, Stevens PT, Kobor MS, Taplin C, Pare PD, Beyer RP, Stick SM, Knight DA. Decreased fibronectin production significantly contributes to dysregulated repair of asthmatic epithelium. *Am J Respir Crit Care Med* 2010;181:889-898.

25. Parker D, Prince A. Innate immunity in the respiratory epithelium. *Am J Respir Cell Mol Biol* 2011;45:189-201.

26. Sagel SD, Sontag MK, Accurso FJ. Relationship between antimicrobial proteins and airway inflammation and infection in cystic fibrosis. *Pediatr Pulmonol* 2009;44:402-409.

27. Thompson AB, Bohling T, Payvandi F, Rennard SI. Lower respiratory tract lactoferrin and lysozyme arise primarily in the airways and are elevated in association with chronic bronchitis. *J Lab Clin Med* 1990;115:148-158.

28. 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. *Crit Care Med* 2003;31:395-400.

29. Vandal K, Rouleau P, Boivin A, Ryckman C, Talbot M, Tessier PA. Blockade of s100a8 and s100a9 suppresses neutrophil migration in response to lipopolysaccharide. *J Immunol* 2003;171:2602-2609.

30. Raquil MA, Anceriz N, Rouleau P, Tessier PA. Blockade of antimicrobial proteins s100a8 and s100a9 inhibits phagocyte migration to the alveoli in streptococcal pneumonia. *J Immunol* 2008;180:3366-3374.

Chapter 4

Proteomic Profiling of Bronchoalveolar Lavage Fluid during Human Immunodeficiency Virus Infection

This body of work was done in collaboration with Dr. Sina A. Gharib and Dr. Lynn M. Schnapp, Division of Pulmonary and Critical Care Medicine, Harborview Medical Center, University of Washington, Seattle, WA, US

4.1 Introduction

The lung is an important reservoir of HIV and site of HIV replication [1]. HIV is uniformly detected in alveolar macrophages of pediatric patients and in the majority of adult patients [2]. In addition to the numerous opportunistic infections that affect the lung, HIV directly causes significant pulmonary pathology, including lymphoid interstitial pneumonitis (LIP), nonspecific pneumonitis (NSIP), or lymphocytic alveolitis [3]. LIP occurs in up to 75% of the pediatric population. HIV-infected adults can develop an asymptomatic lymphocytic alveolitis, which can progress to NSIP or LIP [3]. There is strong evidence that HIV itself evokes a local immune response that facilitates lymphocytic migration and infiltration in the lung. In a transgenic mouse model of HIV, mice developed lymphocytic infiltration in the alveolar walls in the absence of any opportunistic infection [4]. In a mouse model of HIV-induced interstitial pneumonitis, increased viral replication correlated with the development of interstitial pneumonitis [5]. Abnormal elaboration of cytokines in the lung may be involved in the development of lymphocytic alveolitis [6, 7].

We hypothesized that comprehensively profiling the proteomic landscape of BALF in asymptomatic HIV patients would identify distinct protein signatures during HIV infection. To this end, we integrated tandem mass spectrometry-based proteomics with statistical and computational methods to identify differentially expressed proteins between healthy volunteers and HIV patients. We focused on these differentially expressed proteins to give insight into how HIV alters the airspace milieu and leads to lung pathogenesis.

4.2 Materials and Methods

BALF Collection

The protocol for collecting human BALF was approved by the Institutional Review board at the University of Washington. Written informed consent was obtained from all subjects[8]. BALF was obtained from 5 healthy volunteers and 6 patients diagnosed with HIV not on highly active antiretroviral therapy (HAART). BALF was performed as previously described [9, 10]. All subjects were non-smokers. Subject and BALF characteristics are shown in Table 4.1. More detailed patient characteristics provided in Table 4.2

Shotgun proteomic analysis

Tryptic digests of each BALF samples were analyzed by HPLC-MS using a NanoAquity HPLC system (Milford, MA, USA) via electrospray ionization on-line to a hybrid LTQ-Velos mass spectrometer (Thermo Fisher, San Jose, CA). The experiment was repeated in triplicate using gas phase fraction over the following three *m/z* ranges: 400-559, 559-846, 846-2000. The tandem mass spectra were then matched to a protein sequence in the IPI Human 3.53 database (http://www.ebi.ac.uk/IPI) using SEQUEST [11]. Criteria for matching a peptide tandem mass

spectrum to a peptide sequence were: Xcorr > 1.9 with charge state 1+, Xcorr>2.22 with charge state 2+, or Xcorr>3.75 with charge state 3+, as well as ΔCn >0.1. Peptide tandem mass spectra passing these criteria were utilized for protein identifications. A protein was considered to be identified only if ProteinProphet [12] probability >0.8 and if more than one unique peptide was found for each protein. Protein identifications passing criteria were then ran through BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi). Additional proteins that had 99% FASTA sequence similarity were removed to rid of redundant proteins. Relative protein levels were determined using spectral counting [13, 14]. Peptide spectral counts were determined by the number of times a peptide that matched a given protein was selected for CID, including all repeated selections of the sample peptide. A protein's spectral count value was calculated by summing the average spectral counts for each GPF. Homologous peptides that mapped to multiple proteins were discarded.

Correspondence and cluster analyses

Correspondence analysis was performed based on the variability in protein expression in normal and HIV populations. We performed hierarchical clustering of differentially expressed proteins between normal and HIV subjects.

Differential Protein Expression

To assess differences in relative protein abundance between subject populations, individual protein spectral counts were normalized using the spectral index (SI) metric as previously described. We chose a 95% confidence threshold for significant differential expression.

Functional Analysis

Functional annotation of the BALF proteomes for normal and HIV patients was conducted using database for annotation, visualization, and integrated discovery (DAVID [15]) software.

Protein network analysis

We utilized experimentally verified protein interactions from several resources including Ingenuity [16] and STRING [17] to construct a relational network comprised of differentially expressed proteins in patients with HIV.

4.3 Results

Overview of BALF proteome in healthy subjects and patients with HIV

We identified 251 unique proteins in the BALF of 5 normal volunteers and 223 unique proteins in 6 HIV subjects not on HAART. Of these, 154 proteins were common to both groups. The most abundant proteins based on raw spectral counts across all BALF samples were proteins known to be enriched in the normal lung proteome such as albumin, complement C3, and serotransferrin. Table 4.3 contains a list of proteins found in both normal and HIV subjects. Correspondence analysis of all identified BALF proteins separated normal subjects from HIV patients, demonstrating that HIV alters the protein profiles in the airspaces of asymptomatic patients (Figure 4.1).

BALF proteome in HIV

We applied a statistical test based on protein spectral counts known as the spectral index to identify 88 differentially expressed proteins between normal and HIV subjects. Of these, 62 proteins were significantly enriched in normal BALF, whereas 26 were significantly enriched in

the BALF of HIV patients (Table 4.4). To elucidate the pathways and mechanisms activated and/or deactivated in HIV, we performed Gene Ontology analysis on the differentially expressed BALF proteins. As shown in Figure 4.2, the proteins more abundant in normal BALF (and therefore downregulated during HIV) mapped to functional categories including humoral immune response (P-value: $2x10^{-5}$), complement activation (P-value: $2x10^{-4}$), innate immune response (P-value: $3x10^{-4}$), and defense response (P-value: $9x10^{-4}$). In contrast, upregulated proteins in HIV mapped to enriched processes involved in the extracellular region (P-value: $2x10^{-10}$) and regulation of fibrinolysis (P-value: $8x10^{-5}$). Response to wounding and acute inflammatory response were found to be enriched in both populations. More details of gene ontology analysis provided in Table 4.5.

Subsequently unsupervised hierarchical cluster analysis on the differentially expressed proteomic profiles of patients diagnosed with HIV and healthy volunteers demonstrated that this signature is a reasonably powerful discriminator between the two populations (Figure 4.3). An NIH database was further applied to systematically search for experimentally validated interactions between these BALF proteins and HIV-1 proteome. These relationships are summarized via the green-colored heatmap with confirmed relationships represented with green boxes.

Protein interaction network in HIV

To better elucidate functional relationships among differentially expressed proteins in VAP, we performed a network analysis using available protein-protein interaction databases (Figure 4.4). This putative network highlights complex partnerships between proteins that are

functionally active in the extracellular space (e.g., LTF, C4B, PLG, FGB), plasma membrane (e.g., CD36, FLNA) and the cytoplasm (e.g., LDL, HDL, CDC42).

DISCUSSION

The lung is particularly susceptible to opportunistic infections among HIV infected individuals. Our study is the first to investigate the airspace proteome in asymptomatic patients infected with HIV. By identifying protein profiles specific to HIV in BALF, we can elucidate the role of HIV in lung pathogenesis during the asymptomatic stage.

Studies suggest that pulmonary infections play a key role in the virulence and enhancement of HIV in the lungs. Lung complications in HIV patients are exhibited during the asymptomatic stage of infection. The asymptomatic stage has been associated with high levels of viral replication [18, 19]. HIV RNA levels in BALF are increased during opportunisitic infections of *Mycobacterium tuberculosis* and *Pneumocystis carinii* pneumonia [20, 21]. In addition HIV RNA levels were significantly increased in BALF of HIV infected patients with active pulmonary infections in comparison to asymptomatic individuals [22]. HIV RNA levels in serum were indistinguishable between both study groups providing evidence for active HIV replication in the lungs. The source of HIV is unknown although macrophages has been indicated as they are among the first cells infected by HIV forming reservoirs that spread disease to other tissues [16,21,30,31]. Innate immune response CD4 T-cells and frequencies of cells expressing interferon (IFN)-ɣ, tumor necrosis factor (TNF)-α, and interleukin (IL)-2 are also lower in BALF of HIV-1-infected subjects in comparison to controls [23]. These results

demonstrate that the HIV virion in BALF alters immune response, is influenced by lung-specific factors such as infection, and is compartmentalized in the lung.

We established that the BALF proteome between normal volunteers and patients infected with HIV not on HAART therapy is distinguishable (Figure 4.1). To better characterize the roles of the protein clusters between the populations, we identified 88 differentially expressed proteins. The majority of these proteins was more abundant in the BALF of normal volunteers, and mapped to distinct functional categories that included humoral, innate, and complement immune response. In contrast, the BALF of HIV infected subjects was enriched in proteins involved in the extracellular region and fibrinolysis (Figure 4.2)—indicating a down-regulated immune response and upregulation in fibrinolysis. Figure 4.3 further indicates that from a statistical standpoint, it appears that there is an enrichment in the number of BALF proteins that interact with HIV proteins compared to a random sample with the same number of gene products: 1/3 of the differentially expressed proteins in BALF can also interact with HIV proteins.

We further explored the functional relationships among differentially expressed proteins in HIV using gene product interaction network analysis (Figure 4.4). Complex biological networks possess topologic properties that are functionally informative [39]. For example, highly connected nodes—known as hubs, disproportionately affect the stability of networks and may be important orchestrators of biological processes [40]. Figure 4.4 depicts several such hubs, including biological modules identified in Gene Ontology analysis (Figure 4.2) such as enrichment in complement module (C4B, C5, C4BP, C4BPA) and fibrinolysis (PLG, F12, APOH) in the BALF of normal volunteers. A few nodes (i.e., NFkB, LDL, and HDL) were

included to improve the spanning structure of the network which highlighted the potential role of cholesterol in HIV infection.

Although HAART can control viraemia, high levels of HIV return rapidly after its cessation[41]. HIV along with a number of viruses depend on cholesterol for replication which occur at lipid rafts, the sphingolipid and cholesterol enriched domain of the plasma membrane [42, 43]. Inhibition of cholesterol synthesis by lovastatin reduces HIV-1 particle production in infected cells [44] and in the presence of cholesterol removing drug, β-cyclodextrin, the HIV-1 virion is rendered incompetent for cell entry [51, 52]. Furthermore, cholesterol-depleted cells have shown to be unable to form clusters of CD4 and CXCR4 (or CCR5) which reduces their susceptibility for membrane fusion by the HIV surface protein gp120 [43, 50].

We used shotgun proteomics to establish the BALF proteome in HIV patients not on HAART therapy. We showed that integrative statistical analysis showed that even in asymptomatic HIV patients, there is a unique protein profile distinguishable from normal volunteers. Gene ontology and network analysis both indicated fibrinolysis and suppression of the immune response including the complement system. Last of all, a role of cholesterol in HIV virulence in the lung environment was independently validated corresponding with previous published studies.

Table 4.1 Characteristics of persons enrolled in the study

HIV status	CD4	Time Infected	Smoker	Age	Sex	Race
POS	445	~5 yr	N	34	M	Caucasian
POS	405	~4 yr	N	34	M	Caucasian
POS	426	~10 yr	Y	36	M	Caucasian
POS	253	~20 yr	Y	43	M	African American
POS	171	~1.5 yr	N	37	M	Asian American
POS	503	~11 yr	Y	46	M	African American
NEG	N/A	N/A	N	20	M	Caucasian
NEG	N/A	N/A	N	30	M	Caucasian
NEG	N/A	N/A	N	32	F	Caucasian
NEG	N/A	N/A	N	23	F	African American
NEG	N/A	N/A	N	45	M	Caucasian

Table 4.2 Characteristics of control and HIV subjects and their BALF samples

Patient ID	HIV status	HIV-1 RNA plasma copies /ml (screening)	CD4	Time Infected	Therapy at first sample	Smoker	Age	Sex	Race
832713	**POS**	**14, 642** 4/20/05	445	~5 yr	None[1]	N	34	M	Caucasian
570729	**POS**	**<50** 6/01/05	405	~4 yr	None[2]	N	34	M	Caucasian
180508	**POS**	**76,691** 7/19/05	426	~10 yr	None[3]	Y	36	M	Caucasian
21320	**POS**	**40,304** 8/29/05	253	~20 yr	None	Y	43	M	African
344801	**POS**	**8,001** 1/19/07	171	~1.5 yr	None	N	37	M	Asian
160293	**POS**	**299** 1/23/07	503	~11 yr	None[4]	Y	46	M	African
314629	**NEG**	N/A	N/A	N/A	N/A	N	20	M	Caucasian
904871	**NEG**	N/A	N/A	N/A	N/A	N	30	M	Caucasian
390227	**NEG**	N/A	N/A	N/A	N/A	N	32	F	Caucasian
143952	**NEG**	N/A	N/A	N/A	N/A	N	23	F	African
304507	**NEG**	N/A	N/A	N/A	N/A	N	45	M	Caucasian

[1]Viral load dipped to 400 copies during brief period of HAART 5 months before BAL
[2]Patient had history of HAART for 3 years, stopped 1 year before BAL
[3]Patient had history of HAART reducing levels in 1197, 2000, 2001
[4]Patient had history of HAART 6 years before BAL

Table 4.3 List of common proteins identified in BALF of normal volunteers and HIV subjects ranked by spectral counts.

IPI	Protein Name	Entrez Gene ID	Spectral Counts
IPI00745872	albumin	213	1211
IPI00783987	complement component 3	718	1098
IPI00022463	transferrin	7018	487
IPI00004573	polymeric immunoglobulin receptor	5284	368
IPI00855918	mucin 5B, oligomeric mucus/gel-forming	727897	312
IPI00291410	chromosome 20 open reading frame 114	92747	280
IPI00010471	lymphocyte cytosolic protein 1 (L-plastin)	3936	243
IPI00296083	surfactant, pulmonary-associated protein B	6439	238
IPI00478003	alpha-2-macroglobulin	2	235
IPI00418471	vimentin	7431	206
IPI00219365	moesin	4478	187
IPI00220327	keratin 1 (epidermolytic hyperkeratosis)	3848	153
IPI00021841	apolipoprotein A-I	335	150
IPI00022488	hemopexin	3263	138
IPI00553177	serpin peptidase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 1	5265	129
IPI00329801	annexin A5	308	126
IPI00017601	ceruloplasmin (ferroxidase)	1356	120
IPI00550991	serpin peptidase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 3	12	114
IPI00218914	aldehyde dehydrogenase 1 family, member A1	216	111
IPI00382470	heat shock protein 90kDa alpha (cytosolic), class A member 1	3320	106
IPI00009342	IQ motif containing GTPase activating protein 1	8826	101
IPI00073772	fructose-1,6-bisphosphatase 1	2203	96
IPI00019502	myosin, heavy chain 9, non-muscle	4627	94
IPI00219757	glutathione S-transferase pi	2950	93
IPI00242956	Fc fragment of IgG binding protein	8857	93
IPI00009865	keratin 10 (epidermolytic hyperkeratosis; keratosis palmaris et plantaris)	3858	91
IPI00294578	transglutaminase 2 (C polypeptide, protein-glutamine-gamma-glutamyltransferase)	7052	85
IPI00550069	ribonuclease/angiogenin inhibitor 1	6050	78
IPI00032291	complement component 5	727	72
IPI00550363	transgelin 2	8407	72

IPI00019038	lysozyme (renal amyloidosis)	4069	70
IPI00298994	talin 1	7094	70
IPI00027848	mannose receptor, C type 1	4360	65
IPI00022432	transthyretin (prealbumin, amyloidosis type I)	7276	65
IPI00219217	lactate dehydrogenase B	3945	62
IPI00304273	apolipoprotein A-IV	337	62
IPI00019359	keratin 9 (epidermolytic palmoplantar keratoderma)	3857	61
IPI00219525	phosphogluconate dehydrogenase	5226	61
IPI00291878	surfactant, pulmonary-associated protein D	6441	58
IPI00291005	malate dehydrogenase 1, NAD (soluble)	4190	56
IPI00032220	angiotensinogen (serpin peptidase inhibitor, clade A, member 8)	183	56
IPI00291262	clusterin	1191	56
IPI00013895	S100 calcium binding protein A11	6282	56
IPI00026314	gelsolin (amyloidosis, Finnish type)	2934	56
IPI00006114	serpin peptidase inhibitor, clade F (alpha-2 antiplasmin, pigment epithelium derived factor), member 1	5176	56
IPI00216691	profilin 1	5216	54
IPI00169383	phosphoglycerate kinase 1	5230	54
IPI00291866	serpin peptidase inhibitor, clade G (C1 inhibitor), member 1, (angioedema, hereditary)	710	53
IPI00029739	complement factor H	3075	53
IPI00186290	eukaryotic translation elongation factor 2	1938	52
IPI00028091	ARP3 actin-related protein 3 homolog (yeast)	10096	52
IPI00022431	alpha-2-HS-glycoprotein	197	52
IPI00027223	isocitrate dehydrogenase 1 (NADP+), soluble	3417	50
IPI00019591	complement factor B	629	50
IPI00465352	calcyphosine	828	49
IPI00215746	fatty acid binding protein 4, adipocyte	2167	48
IPI00465248	enolase 1, (alpha)	2023	48
IPI00022774	valosin-containing protein	7415	47
IPI00292530	inter-alpha (globulin) inhibitor H1	3697	45
IPI00410714	hemoglobin, alpha 1; hemoglobin, alpha 2	3039; 3040	45
IPI00014055	napsin A aspartic peptidase	9476	45
IPI00218918	annexin A1	301	44
IPI00005118	hexokinase 3 (white cell)	3101	44
IPI00295777	glycerol-3-phosphate dehydrogenase 1 (soluble)	2819	43
IPI00216699	fermitin family homolog 3 (Drosophila)	83706	43

IPI00003590	quiescin Q6 sulfhydryl oxidase 1	5768	42
IPI00292858	thymidine phosphorylase	1890	41
IPI00032179	serpin peptidase inhibitor, clade C (antithrombin), member 1	462	39
IPI00027463	S100 calcium binding protein A6	6277	39
IPI00000874	peroxiredoxin 1	5052	39
IPI00008494	intercellular adhesion molecule 1 (CD54), human rhinovirus receptor	3383	38
IPI00012011	cofilin 1 (non-muscle)	1072	38
IPI00019568	coagulation factor II (thrombin)	2147	37
IPI00645078	ubiquitin-like modifier activating enzyme 1	7317	36
IPI00007047	S100 calcium binding protein A8	6279	34
IPI00178926	immunoglobulin J polypeptide, linker protein for immunoglobulin alpha and mu polypeptides	3512	34
IPI00298971	vitronectin	7448	34
IPI00027462	S100 calcium binding protein A9	6280	33
IPI00156689	vesicle amine transport protein 1 homolog (T. californica)	10493	33
IPI00010896	chloride intracellular channel 1	1192	33
IPI00027444	serpin peptidase inhibitor, clade B (ovalbumin), member 1	1992	33
IPI00022371	histidine-rich glycoprotein	3273	33
IPI00022395	complement component 9	735	32
IPI00451624	cartilage acidic protein 1	55118	32
IPI00006705	secretoglobin, family 1A, member 1 (uteroglobin)	7356	32
IPI00465439	aldolase A, fructose-bisphosphate	226	32
IPI00456969	dynein, cytoplasmic 1, heavy chain 1	1778	32
IPI00793199	annexin A4	307	32
IPI00017704	coactosin-like 1 (Dictyostelium)	23406	32
IPI00246058	programmed cell death 6 interacting protein	10015	31
IPI00295400	tryptophanyl-tRNA synthetase	7453	31
IPI00419237	leucine aminopeptidase 3	51056	31
IPI00292950	serpin peptidase inhibitor, clade D (heparin cofactor), member 1	3053	31
IPI00220301	peroxiredoxin 6	9588	30
IPI00295386	carbonyl reductase 1	873	30
IPI00642211	arginyl aminopeptidase (aminopeptidase B)	6051	29
IPI00019580	plasminogen	5340	28
IPI00066193	secretoglobin, family 3A, member 1	92304	28
IPI00294158	biliverdin reductase A	644	28
IPI00413451	serpin peptidase inhibitor, clade B	5269	28

	(ovalbumin), member 6		
IPI00018219	transforming growth factor, beta-induced, 68kDa	7045	27
IPI00021302	sushi domain containing 2	56241	27
IPI00297487	cathepsin H	1512	26
IPI00025512	heat shock 27kDa protein 1	3315	26
IPI00022394	complement component 1, q subcomponent, C chain	714	25
IPI00643920	transketolase (Wernicke-Korsakoff syndrome)	7086	25
IPI00296654	bactericidal/permeability-increasing protein-like 1	80341	25
IPI00453473	histone cluster 1, H4i; histone cluster 1, H4a; histone cluster 1, H4d; histone cluster 1, H4f; histone cluster 1, H4k; histone cluster 1, H4j; histone cluster 1, H4c; histone cluster 1, H4h; histone cluster 1, H4b; histone cluster 1, H4e; histone cluster 1, H4l; histone cluster 2, H4a; histone cluster 4, H4; histone cluster 2, H4b	8294; 8359; 8360; 8361; 8362; 8363; 8364; 8365; 8366; 8367; 8368; 8370; 121504; 554313	24
IPI00163207	peptidoglycan recognition protein 2	114770	24
IPI00477992	complement component 1, q subcomponent, B chain	713	24
IPI00021842	apolipoprotein E	348	24
IPI00023673	lectin, galactoside-binding, soluble, 3 binding protein	3959	23
IPI00654755	hemoglobin, beta	3043	23
IPI00018246	hexokinase 1	3098	23
IPI00012503	prosaposin (variant Gaucher disease and variant metachromatic leukodystrophy)	5660	23
IPI00022426	alpha-1-microglobulin/bikunin precursor	259	22
IPI00374315	chromosome 6 open reading frame 58	352999	22
IPI00218192	inter-alpha (globulin) inhibitor H4 (plasma Kallikrein-sensitive glycoprotein)	3700	22
IPI00024095	annexin A3	306	22
IPI00024915	peroxiredoxin 5	25824	22
IPI00297160	CD44 molecule (Indian blood group)	960	22
IPI00744692	transaldolase 1	6888	20

IPI00060800	similar to common salivary protein 1	124220	20
IPI00005161	actin related protein 2/3 complex, subunit 2, 34kDa	10109	20
IPI00022974	prolactin-induced protein	5304	19
IPI00216049	heterogeneous nuclear ribonucleoprotein K	3190	19
IPI00021854	apolipoprotein A-II	336	19
IPI00219682	stomatin	2040	18
IPI00479722	proteasome (prosome, macropain) activator subunit 1 (PA28 alpha)	5720	18
IPI00005160	actin related protein 2/3 complex, subunit 1B, 41kDa	10095	17
IPI00163563	phosphatidylethanolamine-binding protein 4	157310	16
IPI00292946	serpin peptidase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 7	6906	16
IPI00022392	complement component 1, q subcomponent, A chain	712	16
IPI00029658	EGF-containing fibulin-like extracellular matrix protein 1	2202	15
IPI00026272	histone cluster 1, H2ae; histone cluster 1, H2ab	3012; 8335	15
IPI00219219	lectin, galactoside-binding, soluble, 1 (galectin 1)	3956	15
IPI00010180	carboxylesterase 1 (monocyte/macrophage serine esterase 1)	1066	15
IPI00013808	actinin, alpha 4	81	14
IPI00003935	histone cluster 2, H2be	8349	13
IPI00031461	GDP dissociation inhibitor 2	2665	13
IPI00257508	dihydropyrimidinase-like 2	1808	13
IPI00216008	glucose-6-phosphate dehydrogenase	2539	13
IPI00414320	annexin A11	311	13
IPI00297779	chaperonin containing TCP1, subunit 2 (beta)	10576	12
IPI00025084	calpain, small subunit 1	826	12
IPI00013698	N-acylsphingosine amidohydrolase (acid ceramidase) 1	427	12
IPI00296608	complement component 7	730	11
IPI00018873	nicotinamide phosphoribosyltransferase	10135	11
IPI00219077	leukotriene A4 hydrolase	4048	11
IPI00103356	integrin, beta 2 (complement component 3 receptor 3 and 4 subunit)	3689	10
IPI00641950	guanine nucleotide binding protein (G protein), beta polypeptide 2-like 1	10399	10
IPI00021263	tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, zeta polypeptide	7534	9

IPI00215997	CD9 molecule	928	9
IPI00328609	serpin peptidase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 4	5267	8
IPI00022432	transthyretin (prealbumin, amyloidosis type I)	7276	65
IPI00219217	lactate dehydrogenase B	3945	62
IPI00304273	apolipoprotein A-IV	337	62
IPI00019359	keratin 9 (epidermolytic palmoplantar keratoderma)	3857	61
IPI00219525	phosphogluconate dehydrogenase	5226	61
IPI00291878	surfactant, pulmonary-associated protein D	6441	58
IPI00291005	malate dehydrogenase 1, NAD (soluble)	4190	56
IPI00032220	angiotensinogen (serpin peptidase inhibitor, clade A, member 8)	183	56
IPI00291262	clusterin	1191	56
IPI00013895	S100 calcium binding protein A11	6282	56
IPI00026314	gelsolin (amyloidosis, Finnish type)	2934	56
IPI00006114	serpin peptidase inhibitor, clade F (alpha-2 antiplasmin, pigment epithelium derived factor), member 1	5176	56
IPI00216691	profilin 1	5216	54
IPI00169383	phosphoglycerate kinase 1	5230	54
IPI00291866	serpin peptidase inhibitor, clade G (C1 inhibitor), member 1, (angioedema, hereditary)	710	53
IPI00029739	complement factor H	3075	53
IPI00186290	eukaryotic translation elongation factor 2	1938	52
IPI00028091	ARP3 actin-related protein 3 homolog (yeast)	10096	52
IPI00022431	alpha-2-HS-glycoprotein	197	52
IPI00027223	isocitrate dehydrogenase 1 (NADP+), soluble	3417	50
IPI00019591	complement factor B	629	50
IPI00465352	calcyphosine	828	49
IPI00215746	fatty acid binding protein 4, adipocyte	2167	48
IPI00465248	enolase 1, (alpha)	2023	48
IPI00022774	valosin-containing protein	7415	47
IPI00292530	inter-alpha (globulin) inhibitor H1	3697	45
IPI00410714	hemoglobin, alpha 1; hemoglobin, alpha 2	3039; 3040	45
IPI00014055	napsin A aspartic peptidase	9476	45
IPI00218918	annexin A1	301	44
IPI00005118	hexokinase 3 (white cell)	3101	44
IPI00295777	glycerol-3-phosphate dehydrogenase 1	2819	43

	(soluble)		
IPI00216699	fermitin family homolog 3 (Drosophila)	83706	43
IPI00003590	quiescin Q6 sulfhydryl oxidase 1	5768	42
IPI00292858	thymidine phosphorylase	1890	41
IPI00032179	serpin peptidase inhibitor, clade C (antithrombin), member 1	462	39
IPI00027463	S100 calcium binding protein A6	6277	39
IPI00000874	peroxiredoxin 1	5052	39
IPI00008494	intercellular adhesion molecule 1 (CD54), human rhinovirus receptor	3383	38
IPI00012011	cofilin 1 (non-muscle)	1072	38
IPI00019568	coagulation factor II (thrombin)	2147	37
IPI00645078	ubiquitin-like modifier activating enzyme 1	7317	36
IPI00007047	S100 calcium binding protein A8	6279	34
IPI00178926	immunoglobulin J polypeptide, linker protein for immunoglobulin alpha and mu polypeptides	3512	34
IPI00298971	vitronectin	7448	34
IPI00027462	S100 calcium binding protein A9	6280	33
IPI00156689	vesicle amine transport protein 1 homolog (T. californica)	10493	33
IPI00010896	chloride intracellular channel 1	1192	33
IPI00027444	serpin peptidase inhibitor, clade B (ovalbumin), member 1	1992	33
IPI00022371	histidine-rich glycoprotein	3273	33
IPI00022395	complement component 9	735	32
IPI00451624	cartilage acidic protein 1	55118	32
IPI00006705	secretoglobin, family 1A, member 1 (uteroglobin)	7356	32
IPI00465439	aldolase A, fructose-bisphosphate	226	32
IPI00456969	dynein, cytoplasmic 1, heavy chain 1	1778	32
IPI00793199	annexin A4	307	32
IPI00017704	coactosin-like 1 (Dictyostelium)	23406	32
IPI00246058	programmed cell death 6 interacting protein	10015	31
IPI00295400	tryptophanyl-tRNA synthetase	7453	31
IPI00419237	leucine aminopeptidase 3	51056	31
IPI00292950	serpin peptidase inhibitor, clade D (heparin cofactor), member 1	3053	31
IPI00220301	peroxiredoxin 6	9588	30
IPI00295386	carbonyl reductase 1	873	30
IPI00642211	arginyl aminopeptidase (aminopeptidase B)	6051	29
IPI00019580	plasminogen	5340	28
IPI00066193	secretoglobin, family 3A, member 1	92304	28

IPI00294158	biliverdin reductase A	644	28
IPI00413451	serpin peptidase inhibitor, clade B (ovalbumin), member 6	5269	28
IPI00018219	transforming growth factor, beta-induced,	7045	27
IPI00021302	sushi domain containing 2	56241	27
IPI00297487	cathepsin H	1512	26
IPI00025512	heat shock 27kDa protein 1	3315	26
IPI00022394	complement component 1, q subcomponent, C chain	714	25
IPI00643920	transketolase (Wernicke-Korsakoff syndrome)	7086	25
IPI00296654	bactericidal/permeability-increasing protein-like 1	80341	25
IPI00453473	histone cluster 1, H4i; histone cluster 1, H4a; histone cluster 1, H4d; histone cluster 1, H4f; histone cluster 1, H4k; histone cluster 1, H4j; histone cluster 1, H4c; histone cluster 1, H4h; histone cluster 1, H4b; histone cluster 1, H4e; histone cluster 1, H4l; histone cluster 2, H4a; histone cluster 4, H4; histone cluster 2, H4b	8294; 8359; 8360; 8361; 8362; 8363; 8364; 8365; 8366; 8367; 8368; 8370; 121504; 554313	24
IPI00163207	peptidoglycan recognition protein 2	114770	24
IPI00477992	complement component 1, q subcomponent, B chain	713	24
IPI00021842	apolipoprotein E	348	24
IPI00023673	lectin, galactoside-binding, soluble, 3 binding protein	3959	23
IPI00654755	hemoglobin, beta	3043	23
IPI00018246	hexokinase 1	3098	23
IPI00012503	prosaposin (variant Gaucher disease and variant metachromatic leukodystrophy)	5660	23
IPI00022426	alpha-1-microglobulin/bikunin precursor	259	22
IPI00374315	chromosome 6 open reading frame 58	352999	22
IPI00218192	inter-alpha (globulin) inhibitor H4 (plasma Kallikrein-sensitive glycoprotein)	3700	22
IPI00024095	annexin A3	306	22
IPI00024915	peroxiredoxin 5	25824	22
IPI00297160	CD44 molecule (Indian blood group)	960	22

IPI00744692	transaldolase 1	6888	20
IPI00060800	similar to common salivary protein 1	124220	20
IPI00005161	actin related protein 2/3 complex, subunit 2, 34kDa	10109	20
IPI00022974	prolactin-induced protein	5304	19
IPI00216049	heterogeneous nuclear ribonucleoprotein K	3190	19
IPI00021854	apolipoprotein A-II	336	19
IPI00219682	stomatin	2040	18
IPI00479722	proteasome (prosome, macropain) activator subunit 1 (PA28 alpha)	5720	18
IPI00005160	actin related protein 2/3 complex, subunit 1B, 41kDa	10095	17
IPI00163563	phosphatidylethanolamine-binding protein 4	157310	16
IPI00292946	serpin peptidase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 7	6906	16
IPI00022392	complement component 1, q subcomponent, A chain	712	16
IPI00029658	EGF-containing fibulin-like extracellular matrix protein 1	2202	15
IPI00026272	histone cluster 1, H2ae; histone cluster 1, H2ab	3012; 8335	15
IPI00219219	lectin, galactoside-binding, soluble, 1 (galectin 1)	3956	15
IPI00010180	carboxylesterase 1 (monocyte/macrophage serine esterase 1)	1066	15
IPI00013808	actinin, alpha 4	81	14
IPI00003935	histone cluster 2, H2be	8349	13
IPI00031461	GDP dissociation inhibitor 2	2665	13
IPI00257508	dihydropyrimidinase-like 2	1808	13
IPI00216008	glucose-6-phosphate dehydrogenase	2539	13
IPI00414320	annexin A11	311	13
IPI00297779	chaperonin containing TCP1, subunit 2 (beta)	10576	12
IPI00025084	calpain, small subunit 1	826	12
IPI00013698	N-acylsphingosine amidohydrolase (acid ceramidase) 1	427	12
IPI00296608	complement component 7	730	11
IPI00018873	nicotinamide phosphoribosyltransferase	10135	11
IPI00219077	leukotriene A4 hydrolase	4048	11
IPI00103356	integrin, beta 2 (complement component 3 receptor 3 and 4 subunit)	3689	10
IPI00641950	guanine nucleotide binding protein (G protein), beta polypeptide 2-like 1	10399	10
IPI00021263	tyrosine 3-monooxygenase/tryptophan 5-	7534	9

	monooxygenase activation protein, zeta polypeptide		
IPI00215997	CD9 molecule	928	9
IPI00328609	serpin peptidase inhibitor, clade A (alpha)	5267	8

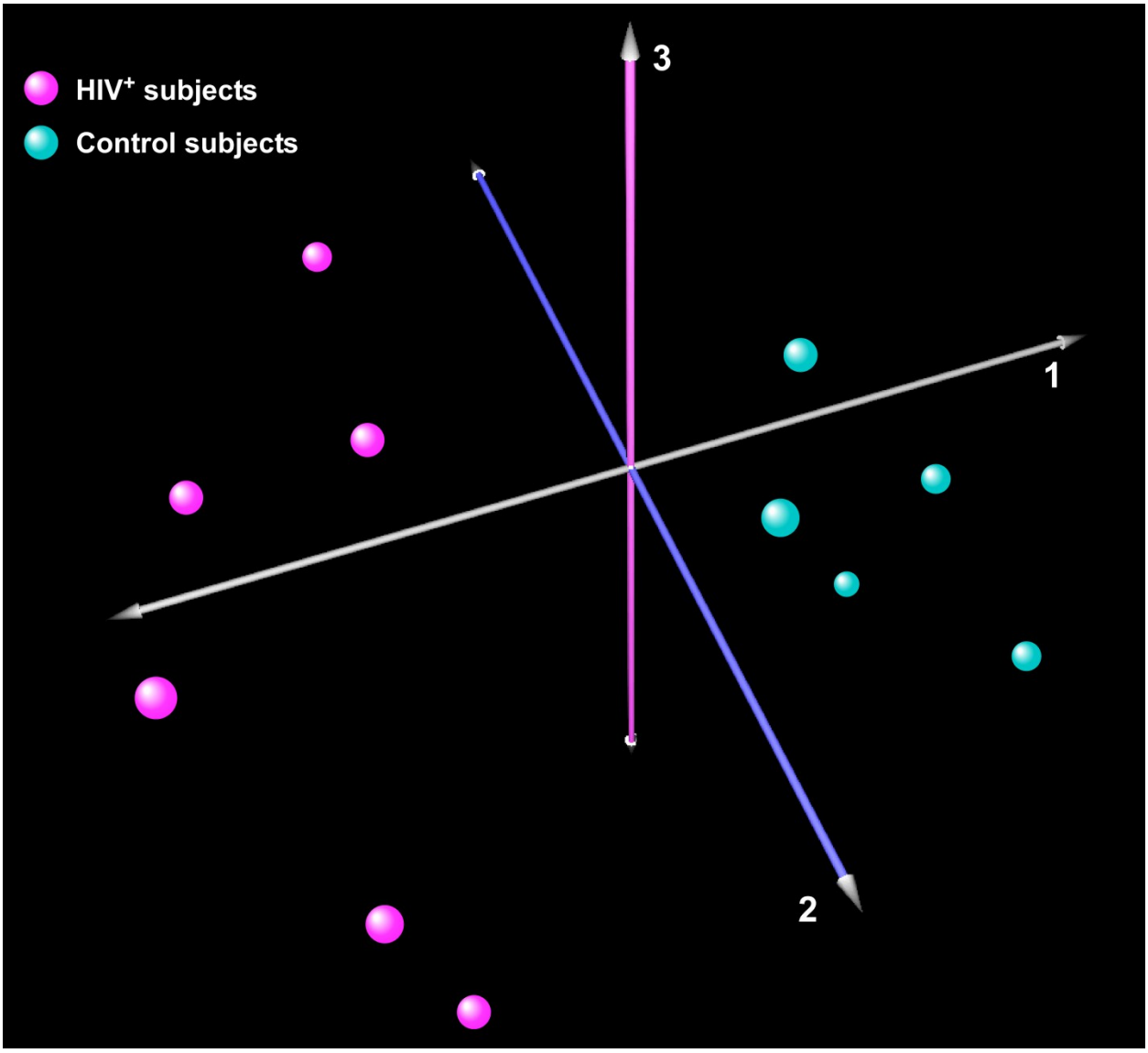

Figure 4.1 Principle component analysis on proteins identified by shotgun proteomics in the normal BALF of five volunteers and 6 HIV patients. Grey spheres represent individual normal volunteers and red for HIV patients

Table 4.4 List of 88 differential proteins found in BALF samples of 5 normal volunteers to 6 HIV patients. Orange highlighted IPI identifiers represent proteins enriched in BALF of HIV subjects.

IPI	Protein Name	Entrez Gene ID	Spectral Index
IPI00011302	CD59 molecule, complement regulatory protein	966	-1
IPI00152418	CD55 molecule, decay accelerating factor for complement (Cromer blood group)	1604	-1
IPI00221221	arachidonate 15-lipoxygenase	246	-0.96377
IPI00016786	cell division cycle 42 (GTP binding protein, 25kDa)	998	-0.94971
IPI00007910	solute carrier family 34 (sodium phosphate), member 2	10568	-0.94548
IPI00328350	family with sequence similarity 129, member A	116496	-0.9386
IPI00018451	calcium and integrin binding 1 (calmyrin)	10519	-0.92708
IPI00418163	complement component 4B (Childo blood group)	721	-0.90591
IPI00000190	CD81 molecule	975	-0.90113
IPI00022462	transferrin receptor (p90, CD71)	7037	-0.89744
IPI00217906	guanine nucleotide binding protein (G protein), alpha inhibiting activity polypeptide 2	2771	-0.88095
IPI00021891	fibrinogen gamma chain	2266	-0.84202
IPI00021727	complement component 4 binding protein, alpha	722	-0.84127
IPI00657682	glutathione S-transferase A1	2938	-0.8374
IPI00294739	SAM domain and HD domain 1	25939	-0.83333
IPI00289758	calpain 2, (m/II) large subunit	824	-0.83122
IPI00219682	stomatin	2040	-0.81367
IPI00027482	serpin peptidase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 6	866	-0.8
IPI00215767	UDP-Gal:betaGlcNAc beta 1,4- galactosyltransferase, polypeptide 1	2683	-0.8
IPI00295851	coatomer protein complex, subunit beta 1	1315	-0.8
IPI00456969	dynein, cytoplasmic 1, heavy chain 1	1778	-0.77191
IPI00027497	glucose phosphate isomerase	2821	-0.77157
IPI00024067	clathrin, heavy chain (Hc)	1213	-0.73856
IPI00000581	OTU domain, ubiquitin aldehyde binding 1	55611	-0.72892
IPI00026216	aminopeptidase puromycin sensitive	9520	-0.72892
IPI00413451	serpin peptidase inhibitor, clade B (ovalbumin), member 6	5269	-0.72338
IPI00009104	RuvB-like 2 (E. coli)	10856	-0.72204
IPI00022977	creatine kinase, brain	1152	-0.71808
IPI00021302	sushi domain containing 2	56241	-0.71519
IPI00302925	chaperonin containing TCP1, subunit 8 (theta)	10694	-0.71369

IPI00032313	S100 calcium binding protein A4	6275	-0.71015
IPI00011285	calpain 1, (mu/I) large subunit	823	-0.68763
IPI00298497	fibrinogen beta chain	2244	-0.68739
IPI00001639	karyopherin (importin) beta 1	3837	-0.67607
IPI00302927	chaperonin containing TCP1, subunit 4 (delta)	10575	-0.67546
IPI00329331	UDP-glucose pyrophosphorylase 2	7360	-0.65937
IPI00012303	selenium binding protein 1	8991	-0.65762
IPI00289499	5-aminoimidazole-4-carboxamide ribonucleotide formyltransferase/IMP cyclohydrolase	471	-0.64896
IPI00847989	pyruvate kinase, muscle	5315	-0.64691
IPI00306960	asparaginyl-tRNA synthetase	4677	-0.64153
IPI00645078	ubiquitin-like modifier activating enzyme 1	7317	-0.63415
IPI00010720	chaperonin containing TCP1, subunit 5 (epsilon)	22948	-0.63333
IPI00032291	complement component 5	727	-0.61497
IPI00027848	mannose receptor, C type 1	4360	-0.60993
IPI00296654	bactericidal/permeability-increasing protein-like 1	80341	-0.60901
IPI00297779	chaperonin containing TCP1, subunit 2 (beta)	10576	-0.60901
IPI00008164	prolyl endopeptidase	5550	-0.6
IPI00009856	palate, lung and nasal epithelium associated	51297	-0.6
IPI00016342	RAB7A, member RAS oncogene family	7879	-0.6
IPI00296353	Rho GTPase activating protein 18	93663	-0.6
IPI00304557	chromosome 20 open reading frame 70	140683	-0.6
IPI00418495	CD36 molecule (thrombospondin receptor)	948	-0.6
IPI00456750	family with sequence similarity 129, member B	64855	-0.6
IPI00646877	surfactant, pulmonary-associated protein A1B; surfactant, pulmonary-associated protein A2B; surfactant, pulmonary-associated protein A1; surfactant, pulmonary-associated protein A2	6435; 6436; 653509; 729238	-0.6
IPI00903112	lactotransferrin	4057	-0.6
IPI00294004	protein S (alpha)	5627	-0.6
IPI00020436	RAB11B, member RAS oncogene family	9230	-0.6
IPI00217766	scavenger receptor class B, member 2	950	-0.6
IPI00029012	eukaryotic translation initiation factor 3, subunit A	8661	-0.6
IPI00030023	histamine N-methyltransferase	3176	-0.6
IPI00305978	aldo-keto reductase family 7, member A2 (aflatoxin aldehyde reductase)	8574	-0.6
IPI00019580	plasminogen	5340	0.597178
IPI00006705	secretoglobin, family 1A, member 1 (uteroglobin)	7356	0.623577
IPI00022895	alpha-1-B glycoprotein	1	0.632415
IPI00555812	group-specific component (vitamin D binding protein)	2638	0.652437
IPI00004656	beta-2-microglobulin	567	0.666667
IPI00019581	coagulation factor XII (Hageman factor)	2161	0.666667

IPI00299150	cathepsin S	1520	0.666667
IPI00217465	histone cluster 1, H1c	3006	0.666667
IPI00219025	glutaredoxin (thioltransferase)	2745	0.666667
IPI00000861	LIM and SH3 protein 1	3927	0.666667
IPI00022431	alpha-2-HS-glycoprotein	197	0.675594
IPI00375676	ferritin, light polypeptide	2512	0.688021
IPI00291867	complement factor I	3426	0.742157
IPI00298828	apolipoprotein H (beta-2-glycoprotein I)	350	0.775779
IPI00166729	alpha-2-glycoprotein 1, zinc-binding	563	0.799702
IPI00302592	filamin A, alpha (actin binding protein 280)	2316	0.812269
IPI00008580	secretory leukocyte peptidase inhibitor	6590	0.833333
IPI00301579	Niemann-Pick disease, type C2	10577	0.833333
IPI00179589	myotrophin	136319	0.833333
IPI00001699	PYD and CARD domain containing	29108	0.833333
IPI00025318	SH3 domain binding glutamic acid-rich protein like	6451	0.833333
IPI00075248	calmodulin 1 (phosphorylase kinase, delta); calmodulin 2 (phosphorylase kinase, delta); calmodulin 3 (phosphorylase kinase, delta)	801; 805; 808	0.833333
IPI00019038	lysozyme (renal amyloidosis)	4069	0.882596
IPI00020091	orosomucoid 2	5005	0.888803
IPI00019943	afamin	173	1
IPI00022429	orosomucoid 1	5004	1

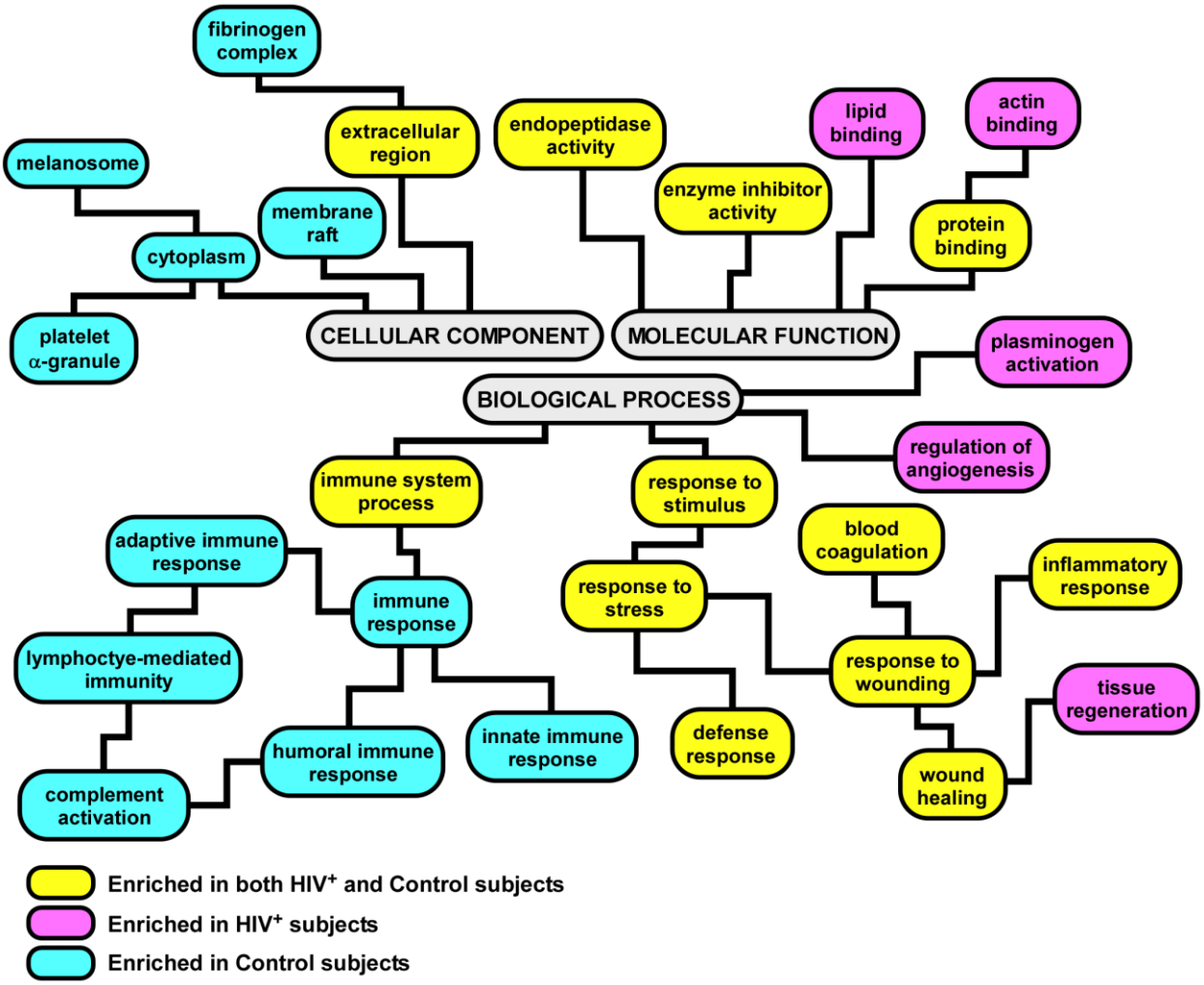

Figure 4.2 Functional analysis of differentially expressed proteins in the BALF of patients with HIV and Controls

Table 4.5 Functional analysis of differentially expressed proteins in the BALF of patients with HIV and Controls.

Enriched in BALF of HIV patients

Gene Ontology Category	Number of Proteins	Enrichment *P*-value
Extracellular Region	19	1.7×10^{-10}
Response to wounding	9	6.8×10^{-7}
Acute inflammatory response	5	1.4×10^{-5}
Regulation of fibrinolysis	3	8.2×10^{-5}
Inflammatory response	6	1.2×10^{-4}

Enriched in BALF of Control subjects

Gene Ontology Category	Number of Proteins	Enrichment *P*-value
Response to wounding	12	1.1×10^{-5}
Humoral immune response	6	2.0×10^{-5}
Acute inflammatory response	6	5.6×10^{-5}
Complement activation, classical	4	2.4×10^{-5}
Innate immune response	6	2.8×10^{-4}
Defense response	10	9.9×10^{-4}

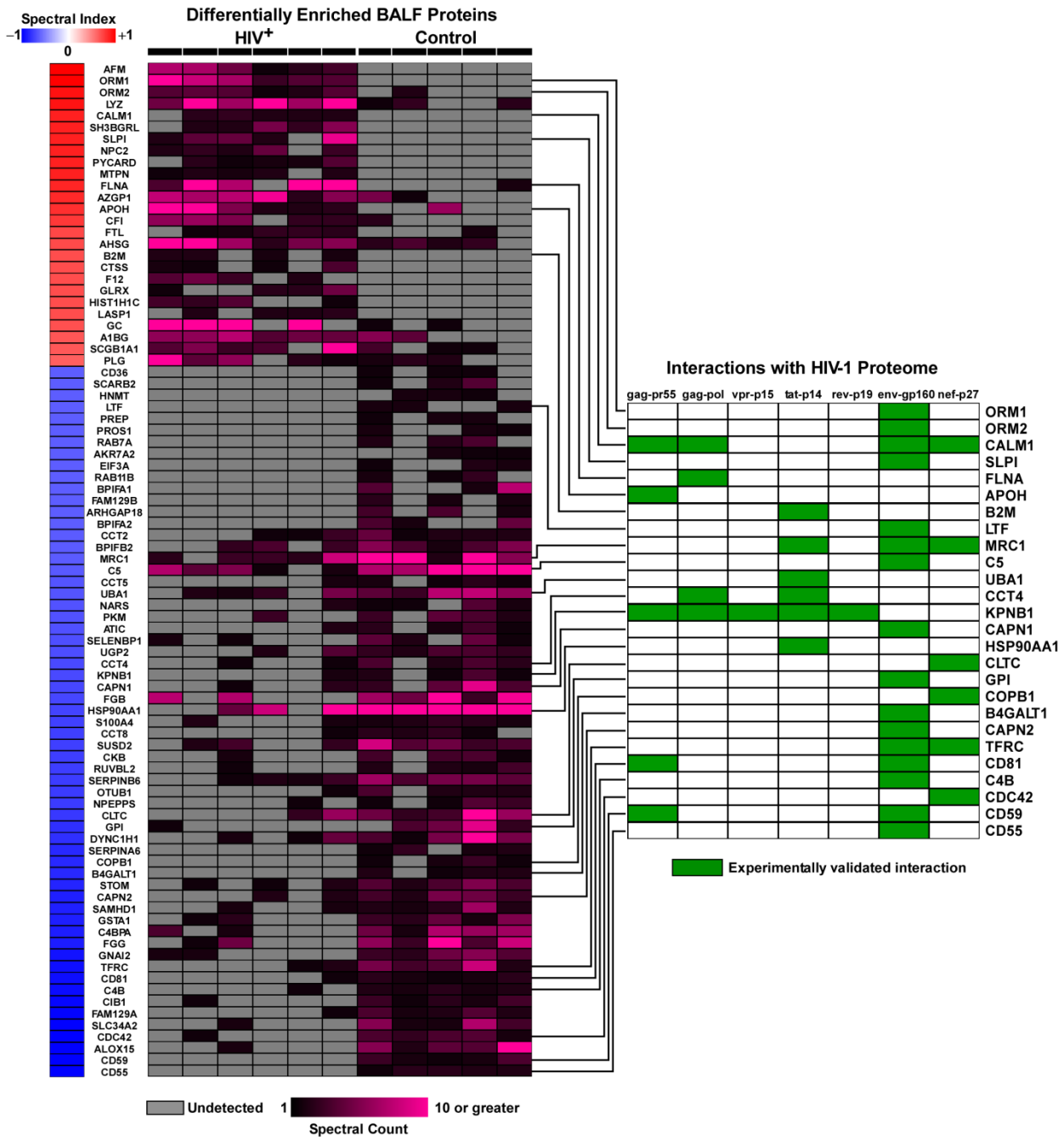

Figure 4.3 Hierarchical clustering of differentially expressed proteins in BALF of HIV subjects and volunteers. Protein abundance is represented on a gradient from black to pink, with proteins showing high spectral counts represented by pink, proteins showing low spectral counts represented by black, and undetected proteins represented by gray. Each column represents a patient sample. The y-axis also displays the spectral index for each protein with a value close to +1 represented by red (enriched in the HIV subjects) and a value close to -1 represented by blue (enriched in normal volunteers). The green-colored heatmap represents validation interactions with HIV.

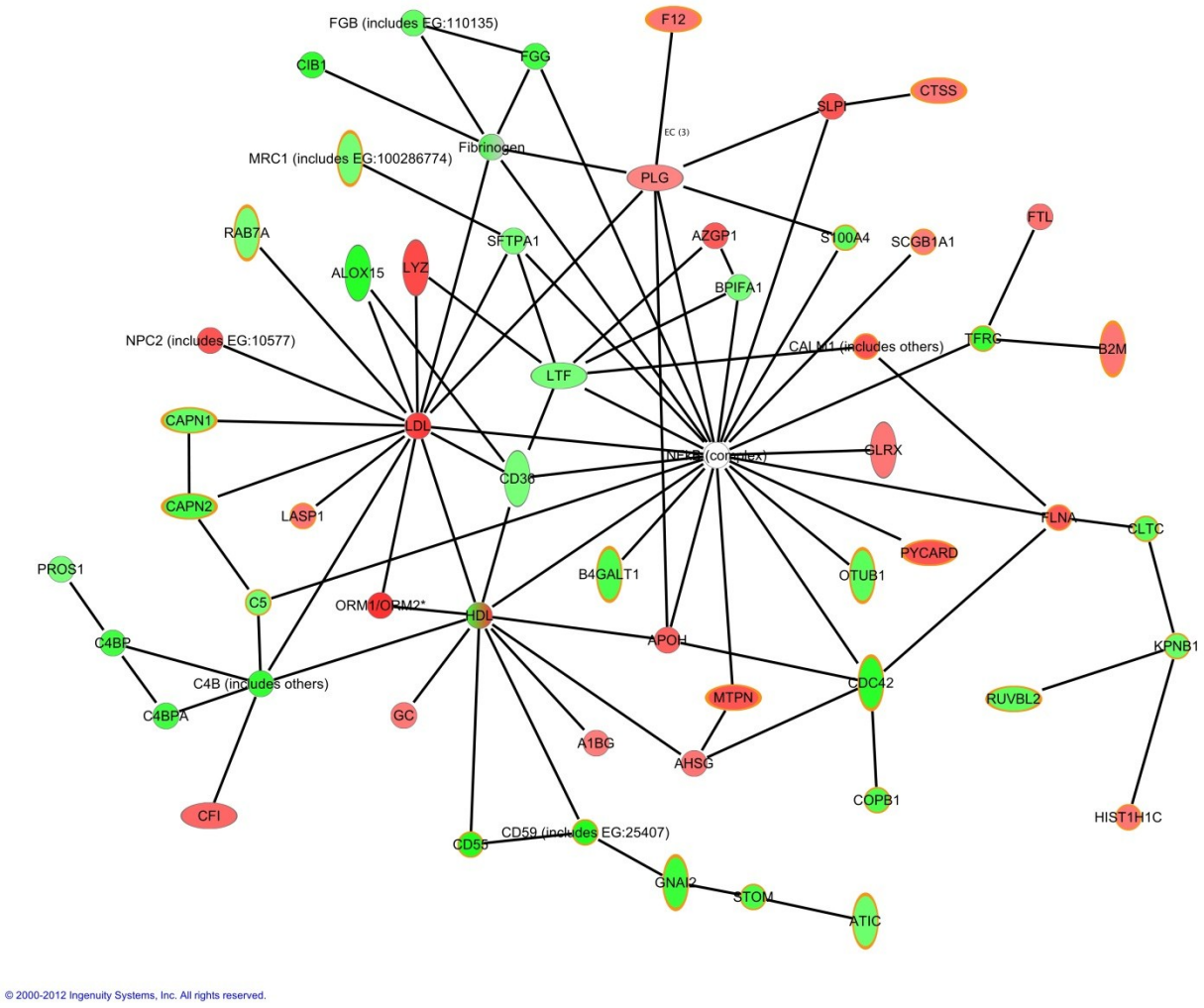

Figure 4.4 Protein interactome in BALF of HIV. This network was constructed by limiting the nodes to differentially expressed proteins (red, enriched in HIV; green, depleted in HIV). Complex interactions among these proteins highlight the relationships between the biological modules identified in Gene Ontology analysis (Table 4.5) such as complement module (C4B, C5, C4BP, C4BPA) and Fibrinolysis (PLG, F12, APOH)

4.6 Notes to Chapter 4

1. White, N.C., et al., The growth and the control of human immunodeficiency virus in the lung: implications for highly active antiretroviral therapy. Eur J Clin Invest, 1999. 29(11): p. 964-72.

2. Pearce, T.E., et al., Uniform detection of HIV-1 in alveolar macrophages of pediatric but not adult AIDS patients. J Leukoc Biol, 1993. 53(6): p. 722-6.

3. de Leon, F.C. and E.J. Britt, The noninfectious respiratory complications of infection with HIV. Curr Opin Pulm Med, 1995. 1(3): p. 223-33.

4. Hanna, Z., et al., Transgenic mice expressing human immunodeficiency virus type 1 in immune cells develop a severe AIDS-like disease. J Virol, 1998. 72(1): p. 121-32.

5. Fitzpatrick, E.A., et al., Role of virus replication in a murine model of AIDS-associated interstitial pneumonitis. Exp Lung Res, 1999. 25(8): p. 647-61.

6. Agostini, C., et al., CXC chemokines IP-10 and mig expression and direct migration of pulmonary CD8+/CXCR3+ T cells in the lungs of patients with HIV infection and T-cell alveolitis. Am J Respir Crit Care Med, 2000. 162(4 Pt 1): p. 1466-73.

7. Teruya-Feldstein, J., et al., Chemokine gene expression and clonal analysis of B cells in tissues involved by lymphoid interstitial pneumonitis from HIV-infected pediatric patients. Mod Pathol, 2001. 14(10): p. 929-36.

8. Steinberg, K.P., et al., Evolution of bronchoalveolar cell populations in the adult respiratory distress syndrome. Am J Respir Crit Care Med, 1994. 150(1): p. 113-22.

9. Matute-Bello, G., et al., Neutrophil apoptosis in the acute respiratory distress syndrome. Am J Respir Crit Care Med, 1997. 156(6): p. 1969-77.

10. Greene, K.E., et al., Serial changes in surfactant-associated proteins in lung and serum before and after onset of ARDS. Am J Respir Crit Care Med, 1999. 160(6): p. 1843-50.

11. Link, A.J., et al., Direct analysis of protein complexes using mass spectrometry. Nat Biotechnol, 1999. 17(7): p. 676-82.

12. Dennis, G., Jr., et al., DAVID: Database for Annotation, Visualization, and Integrated Discovery. Genome Biol, 2003. 4(5): p. P3.

13. Calvano, S.E., et al., A network-based analysis of systemic inflammation in humans. Nature, 2005. 437(7061): p. 1032-7.

14. Jensen, L.J., et al., STRING 8--a global view on proteins and their functional interactions in 630 organisms. Nucleic Acids Res, 2009. 37(Database issue): p. D412-6.

15. Saphire, A.C., P.A. Gallay, and S.J. Bark, Proteomic analysis of human immunodeficiency virus using liquid chromatography/tandem mass spectrometry effectively distinguishes specific incorporated host proteins. J Proteome Res, 2006. 5(3): p. 530-8.

16. Thali, M., et al., Functional association of cyclophilin A with HIV-1 virions. Nature, 1994. 372(6504): p. 363-5.

17. Cimarelli, A. and J. Luban, Translation elongation factor 1-alpha interacts specifically with the human immunodeficiency virus type 1 Gag polyprotein. J Virol, 1999. 73(7): p. 5388-401.

18. Gurer, C., A. Cimarelli, and J. Luban, Specific incorporation of heat shock protein 70 family members into primate lentiviral virions. J Virol, 2002. 76(9): p. 4666-70.

19. Kalsdorf, B., et al., HIV-1 infection impairs the bronchoalveolar T-cell response to mycobacteria. Am J Respir Crit Care Med, 2009. 180(12): p. 1262-70.

20. Pantaleo, G., et al., Role of lymphoid organs in the pathogenesis of human immunodeficiency virus (HIV) infection. Immunol Rev, 1994. 140: p. 105-30.

21. de Villartay, J.P., et al., HLA antigens on peripheral red blood cells: analysis by flow cytofluorometry using monoclonal antibodies. Tissue Antigens, 1985. 26(1): p. 12-9.

22. Swann, S.A., et al., HIV-1 Nef blocks transport of MHC class I molecules to the cell surface via a PI 3-kinase-dependent pathway. Virology, 2001. 282(2): p. 267-77.

23. Greenberg, M.E., A.J. Iafrate, and J. Skowronski, The SH3 domain-binding surface and an acidic motif in HIV-1 Nef regulate trafficking of class I MHC complexes. EMBO J, 1998. 17(10): p. 2777-89.

24. Le Gall, S., et al., Distinct trafficking pathways mediate Nef-induced and clathrin-dependent major histocompatibility complex class I down-regulation. J Virol, 2000. 74(19): p. 9256-66.

25. Andrade, A.C. and B.R. Cotter, Endothelial function and cardiovascular diseases in HIV infected patient. Braz J Infect Dis, 2006. 10(2): p. 139-45.

26. Dube, M.P., et al., Relationship of body composition, metabolic status, antiretroviral use, and HIV disease factors to endothelial dysfunction in HIV-infected subjects. AIDS Res Hum Retroviruses, 2010. 26(8): p. 847-54.

27. Calza, L., et al., Prevalence of diabetes mellitus, hyperinsulinaemia and metabolic syndrome among 755 adult patients with HIV-1 infection. Int J STD AIDS, 2011. 22(1): p. 43-5.

28. Capili, B., J.K. Anastasi, and O. Ogedegbe, HIV and general cardiovascular risk. J Assoc Nurses AIDS Care, 2011. 22(5): p. 362-75.

29. Funderburg, N.T., et al., Increased tissue factor expression on circulating monocytes in chronic HIV infection: relationship to in vivo coagulation and immune activation. Blood. 115(2): p. 161-7.

30. Kuller, L.H., et al., Inflammatory and coagulation biomarkers and mortality in patients with HIV infection. PLoS Med, 2008. 5(10): p. e203.

31. Luscombe, N.M., et al., Genomic analysis of regulatory network dynamics reveals large topological changes. Nature, 2004. 431(7006): p. 308-12.

32. Vidal, M., M.E. Cusick, and A.L. Barabasi, Interactome networks and human disease. Cell. 144(6): p. 986-98.

33. Ho, D.D., et al., Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection. Nature, 1995. 373(6510): p. 123-6.

34. Wei, X., et al., Viral dynamics in human immunodeficiency virus type 1 infection. Nature, 1995. 373(6510): p. 117-22.

35. Cohen, O.J. and A.S. Fauci, Current strategies in the treatment of HIV infection. Adv Intern Med, 2001. 46: p. 207-46.

36. Campbell, S.M., S.M. Crowe, and J. Mak, Lipid rafts and HIV-1: from viral entry to assembly of progeny virions. J Clin Virol, 2001. 22(3): p. 217-27.

37. Manes, S., et al., Membrane raft microdomains mediate lateral assemblies required for HIV-1 infection. EMBO Rep, 2000. 1(2): p. 190-6.

38. Maziere, J.C., et al., Lovastatin inhibits HIV-1 expression in H9 human T lymphocytes cultured in cholesterol-poor medium. Biomed Pharmacother, 1994. 48(2): p. 63-7.

39. Jain, M.K. and P.M. Ridker, Anti-inflammatory effects of statins: clinical evidence and basic mechanisms. Nat Rev Drug Discov, 2005. 4(12): p. 977-87.

40. Nishibori, M., H.K. Takahashi, and S. Mori, The regulation of ICAM-1 and LFA-1 interaction by autacoids and statins: a novel strategy for controlling inflammation and immune responses. J Pharmacol Sci, 2003. 92(1): p. 7-12.

41. Giguere, J.F. and M.J. Tremblay, Statin compounds reduce human immunodeficiency virus type 1 replication by preventing the interaction between virion-associated host intercellular adhesion molecule 1 and its natural cell surface ligand LFA-1. J Virol, 2004. 78(21): p. 12062-5.

42. Chan, D.C. and P.S. Kim, HIV entry and its inhibition. Cell, 1998. 93(5): p. 681-4.

43. Nguyen, D.H., et al., Dynamic reorganization of chemokine receptors, cholesterol, lipid rafts, and adhesion molecules to sites of CD4 engagement. Exp Cell Res, 2005. 304(2): p. 559-69.

44. Viard, M., et al., Role of cholesterol in human immunodeficiency virus type 1 envelope protein-mediated fusion with host cells. J Virol, 2002. 76(22): p. 11584-95.

45. Campbell, S.M., S.M. Crowe, and J. Mak, Virion-associated cholesterol is critical for the maintenance of HIV-1 structure and infectivity. AIDS, 2002. 16(17): p. 2253-61.

46. Guyader, M., et al., Role for human immunodeficiency virus type 1 membrane cholesterol in viral internalization. J Virol, 2002. 76(20): p. 10356-64.

47. Zheng, Y.H., et al., Nef increases the synthesis of and transports cholesterol to lipid rafts and HIV-1 progeny virions. Proc Natl Acad Sci U S A, 2003. 100(14): p. 8460-5.

48. van 't Wout, A.B., et al., Nef induces multiple genes involved in cholesterol synthesis and uptake in human immunodeficiency virus type 1-infected T cells. J Virol, 2005. 79(15): p. 10053-8.

49. Cui, H.L., et al., HIV-1 Nef mobilizes lipid rafts in macrophages through a pathway that competes with ABCA1-dependent cholesterol efflux. J Lipid Res, 2012. 53(4): p. 696-708.

50. Mujawar, Z., et al., Human immunodeficiency virus impairs reverse cholesterol transport from macrophages. PLoS Biol, 2006. 4(11): p. e365.

Chapter 5

Conclusions

Bronchoalveolar lavage (BAL) samples the components of the alveolar epithelial lining fluid (ELF). The protein composition of ELF reveals the effects of external factors and/or diseases affecting the lung. Various external factors and lung diseases induce detectable biochemical modifications in ELF which may aid in early diagnosis or understanding of lung pathogenesis. Proteomic analysis of BALF from normal volunteers defined the expressed proteome of lung airspace and provided a proteomic standard from which investigators can compare to differing diseased states, e.g., human immunodeficiency virus (HIV) and ventilator-associated pneumonia (VAP). Results of proteomic methodology involving protein separation by 2-D gel electrophoresis and mass spectrometry (MS)-based proteomic analysis of the normal lung proteome, defined by BAL, were summarized. Our methodology, shotgun proteomics involving techniques such as rational gas phase fractionation (GPF) and analysis with a hybrid LTQ Velos mass spectrometer, is the most comprehensive BALF proteome in normal volunteers to date.

For the first time, the BALF proteome from patients diagnosed with HIV not on highly active antiretroviral therapy (HAART) was described using the shotgun proteomic methodology. Principal component analysis shows a differentiation between the proteome of normal and HIV infected subjects, indicating that HIV affects the BALF environment. In comparison with the normal BALF proteome, proteins enriched in the BALF proteome of HIV patients were mapped to their functional pathways using gene ontology (GO) analysis. The results indicated immune response suppression. HIV infects t-cells of the innate immune response which later trigger the

adaptive immune response. Both immune systems were shown to be under-enriched in the BALF proteome of HIV subjects.

Network analysis was performed on the differentially expressed proteins in the BALF of normal and HIV subjects, displaying the interactome of proteins in functional pathways. A hub in the protein interactome of HIV subjects further supports the possibility of a pivotal role of cholesterol in HIV, further supporting other studies done on the investigation of cholesterol in HIV. These studies were able to independently validate results with proteomics of BALF in HIV infected patients already known and published in the literature about HIV infectivity. Proteomic analysis necessitates a larger cohort of HIV subjects enrolled to identify specific protein biomarkers to gain a more in-depth insight on how HIV leads to lung pathogenesis.

Clinical diagnosis of ventilator-associated pneumonia (VAP) is challenging. Delayed diagnosis may lead to incorrect treatment and subsequent complications and morbidity. Discovering protein profiles specific to VAP will aid in future diagnostic endeavors and may unravel mechanisms of disease progression. Bronchoalveolar lavage fluid (BALF) from acute lung injury patients (ALI), either later confirmed or not to have pneumonia based on bacterial cell cultures, were collected and analyzed by shotgun proteomics. Our study represents the largest investigation of the airspace proteome in patients with ALI and VAP. When compared to the normal BALF proteome, principle component analysis distinguished normal subjects from acute lung injury patients. Gene ontology analysis indicated the BALF proteome of ALI patients was highly enriched in proteins involved in inflammation, defense response and immunity. However, those later confirmed for VAP (VAP^{+}) and those without it (VAP^{-}) did not segregate in the ALI population.

Within the ALI group, we identified differentially expressed proteins between VAP^+ compared to VAP^- patients. Hierarchical clustering was performed that distinguished VAP^+ from VAP^- based on the differentially expressed proteins. Functional analysis of these proteins suggested activation of pro-inflammatory pathways in VAP^+ while reparative processes were induced in the VAP^- patients. We further explored functional relationships using network analysis among differentially expressed proteins. The results indicated key critical mediators of immune response such as integrin beta 2 (ITGB2), integrin alpha M (ITGAM), and myeloperoxidase (MPO)—all of which were up-regulated in VAP^+ patients and directly interacted with each other. In contrast, fibronectin 1 (FN1) was downregulated in VAP^+ subjects. FN1, a principal component of the extracellular matrix, is a key regulator of epithelial wound repair. These findings suggest that VAP is associated with an imbalance between overabundance of pro-inflammatory mediators and reduced expression of reparative proteins.

A limited proteomic signature was identified using robust statistical methods that accurately separated VAP^+ and VAP^- in ALI patients. Lactotransferrin (LTF), a key component of the respiratory tract antimicrobial defense system and innate immunity, and S100 calcium binding protein A8 (S100A8), an antimicrobial protein expressed by leukocytes and epithelial cells, was tested in a group of ALI patients undergoing bronchoscopy for suspicion of VAP. This group was not in the original ALI population recruited for the study. We confirmed that the expression patterns of LTF and S100A8 in the validation cohort mimicked those of the original 30 ALI subjects, with both proteins being significantly more abundant in BALF of VAP^+ patients. This observation implies that these proteomic signatures independently validated VAP in ALI patients.

BIBLIOGRAPHY

Aderka, D., J.M. Le, and J. Vilcek, *IL-6 inhibits lipopolysaccharide-induced tumor necrosis factor production in cultured human monocytes, U937 cells, and in mice.* J Immunol, 1989. **143**(11): p. 3517-23.

Agerberth, B. and G.H. Gudmundsson, *Host antimicrobial defence peptides in human disease.* Curr Top Microbiol Immunol, 2006. **306**: p. 67-90.

Alberts, B.A.J., Julian Lewis, Martin Raff, Keither Roberts, Peter Walters, *Molecular Biology of the Cell*, ed. Garland. 2002, New York, NY: Garland Science. 1463.

Anderson, N.L. and N.G. Anderson, *The human plasma proteome: history, character, and diagnostic prospects.* Mol Cell Proteomics, 2002. **1**(11): p. 845-67.

Ashburner, M., et al., *Gene ontology: tool for the unification of biology. The Gene Ontology Consortium.* Nat Genet, 2000. **25**(1): p. 25-9.

Ashutosh, K., *Nitric oxide and asthma: a review.* Curr Opin Pulm Med, 2000. **6**(1): p. 21-5.

Barnes, P.J., *Reactive oxygen species and airway inflammation.* Free Radic Biol Med, 1990. **9**(3): p. 235-43.

Barnouin, K., *Two-dimensional gel electrophoresis for analysis of protein complexes.* Methods Mol Biol, 2004. **261**: p. 479-98.

Baughman, R.P. and M. Drent, *Role of bronchoalveolar lavage in interstitial lung disease.* Clin Chest Med, 2001. **22**(2): p. 331-41.

Beck, G. and G.S. Habicht, *Immunity and the invertebrates.* Sci Am, 1996. **275**(5): p. 60-3, 66.

Beier, J., et al., *Induced sputum methodology: Validity and reproducibility of total glutathione measurement in supernatant of healthy and asthmatic individuals.* J Lab Clin Med, 2004. **144**(1): p. 38-44.

Bell, D.Y., et al., *Plasma proteins of the bronchoalveolar surface of the lungs of smokers and nonsmokers.* Am Rev Respir Dis, 1981. **124**(1): p. 72-9.

Bell, D.Y. and G.E. Hook, *Pulmonary alveolar proteinosis: analysis of airway and alveolar proteins.* Am Rev Respir Dis, 1979. **119**(6): p. 979-90.

Bernard, G.R., et al., *The American-European Consensus Conference on ARDS. Definitions, mechanisms, relevant outcomes, and clinical trial coordination.* Am J Respir Crit Care Med, 1994. **149**(3 Pt 1): p. 818-24.

Bissonnette, E.Y., et al., *Potentiation of tumor necrosis factor-alpha-mediated cytotoxicity of mast cells by their production of nitric oxide.* J Immunol, 1991. **147**(9): p. 3060-5.

Bowler, R.P., et al., *Proteomic analysis of pulmonary edema fluid and plasma in patients with acute lung injury.* Am J Physiol Lung Cell Mol Physiol, 2004. **286**(6): p. L1095-104.

Boyton, R.J. and P.J. Openshaw, *Pulmonary defences to acute respiratory infection.* Br Med Bull, 2002. **61**: p. 1-12.

Calvano, S.E., et al., *A network-based analysis of systemic inflammation in humans.* Nature, 2005. **437**(7061): p. 1032-7.

Capron, F., *[Bronchoalveolar lavage].* Arch Anat Cytol Pathol, 1997. **45**(5): p. 255-60.

Chastre, J. and J.Y. Fagon, *Ventilator-associated pneumonia.* Am J Respir Crit Care Med, 2002. **165**(7): p. 867-903.

Chen, J., et al., *Exploration of the normal human bronchoalveolar lavage fluid proteome.* Proteomics Clin Appl, 2008. **2**(4): p. 585-95.

Cohen, J., *The immunopathogenesis of sepsis.* Nature, 2002. **420**(6917): p. 885-91.

Colinge, J., et al., *OLAV: towards high-throughput tandem mass spectrometry data identification.* Proteomics, 2003. **3**(8): p. 1454-63.

Cotran, K., *Robbins Patholgic Basis of Disease*. 1998, Philadelphi: W.B Saundes Company.

Dennis, G., Jr., et al., *DAVID: Database for Annotation, Visualization, and Integrated Discovery.* Genome Biol, 2003. **4**(5): p. P3.

Determann, R.M., et al., *Serial changes in soluble triggering receptor expressed on myeloid cells in the lung during development of ventilator-associated pneumonia.* Intensive Care Med, 2005. **31**(11): p. 1495-500.

Downey, G.P., et al., *Neutrophil sequestration and migration in localized pulmonary inflammation. Capillary localization and migration across the interalveolar septum.* Am Rev Respir Dis, 1993. **147**(1): p. 168-76.

El Kebir, D., et al., *Myeloperoxidase delays neutrophil apoptosis through CD11b/CD18 integrins and prolongs inflammation.* Circ Res, 2008. **103**(4): p. 352-9.

Elder, A., et al., *Lung inflammation induced by endotoxin is enhanced in rats depleted of alveolar macrophages with aerosolized clodronate.* Exp Lung Res, 2005. **31**(6): p. 527-46.

Farkas, A.M., T.M. Kilgore, and M.T. Lotze, *Detecting DNA: getting and begetting cancer.* Curr Opin Investig Drugs, 2007. **8**(12): p. 981-6.

Finlay, B.B. and G. McFadden, *Anti-immunology: evasion of the host immune system by bacterial and viral pathogens.* Cell, 2006. **124**(4): p. 767-82.

Fiorentino, D.F., et al., *IL-10 acts on the antigen-presenting cell to inhibit cytokine production by Th1 cells.* J Immunol, 1991. **146**(10): p. 3444-51.

Fu, X., et al., *Spectral index for assessment of differential protein expression in shotgun proteomics.* J Proteome Res, 2008. **7**(3): p. 845-54.

Gaston, B., et al., *The biology of nitrogen oxides in the airways.* Am J Respir Crit Care Med, 1994. **149**(2 Pt 1): p. 538-51.

Geijtenbeek, T.B. and S.I. Gringhuis, *Signalling through C-type lectin receptors: shaping immune responses.* Nat Rev Immunol, 2009. **9**(7): p. 465-79.

Gharib, S.A., et al., *Of mice and men: comparative proteomics of bronchoalveolar fluid.* Eur Respir J, 2010. **35**(6): p. 1388-95.

Gharib, S.A., et al., *Mapping the lung proteome in cystic fibrosis.* J Proteome Res, 2009. **8**(6): p. 3020-8.

Ghio, A.J., et al., *Iron and iron-related proteins in the lower respiratory tract of patients with acute respiratory distress syndrome.* Crit Care Med, 2003. **31**(2): p. 395-400.

Giorgianni, F., D.M. Desiderio, and S. Beranova-Giorgianni, *Proteome analysis using isoelectric focusing in immobilized pH gradient gels followed by mass spectrometry.* Electrophoresis, 2003. **24**(1-2): p. 253-9.

Goretski, J. and T.C. Hollocher, *Trapping of nitric oxide produced during denitrification by extracellular hemoglobin.* J Biol Chem, 1988. **263**(5): p. 2316-23.

Griese, M., *Pulmonary surfactant in health and human lung diseases: state of the art.* Eur Respir J, 1999. **13**(6): p. 1455-76.

Griese, M., J. Noss, and C. von Bredow, *Protein pattern of exhaled breath condensate and saliva.* Proteomics, 2002. **2**(6): p. 690-6.

Griffin, T.J. and R. Aebersold, *Advances in proteome analysis by mass spectrometry.* J Biol Chem, 2001. **276**(49): p. 45497-500.

Grommes, J. and O. Soehnlein, *Contribution of neutrophils to acute lung injury.* Mol Med, 2011. **17**(3-4): p. 293-307.

Guo, R.F. and P.A. Ward, *Mediators and regulation of neutrophil accumulation in inflammatory responses in lung: insights from the IgG immune complex model.* Free Radic Biol Med, 2002. **33**(3): p. 303-10.

Han, D.K., et al., *Quantitative profiling of differentiation-induced microsomal proteins using isotope-coded affinity tags and mass spectrometry.* Nat Biotechnol, 2001. **19**(10): p. 946-51.

Haslam, P.L. and R.P. Baughman, *Report of ERS Task Force: guidelines for measurement of acellular components and standardization of BAL.* Eur Respir J, 1999. **14**(2): p. 245-8.

Henzel, W.J., et al., *Identifying proteins from two-dimensional gels by molecular mass searching of peptide fragments in protein sequence databases.* Proc Natl Acad Sci U S A, 1993. **90**(11): p. 5011-5.

Herard, A.L., et al., *Fibronectin and its alpha 5 beta 1-integrin receptor are involved in the wound-repair process of airway epithelium.* Am J Physiol, 1996. **271**(5 Pt 1): p. L726-33.

Hermans, C. and A. Bernard, *Lung epithelium-specific proteins: characteristics and potential applications as markers.* Am J Respir Crit Care Med, 1999. **159**(2): p. 646-78.

Hermjakob, H., *The HUPO proteomics standards initiative--overcoming the fragmentation of proteomics data.* Proteomics, 2006. **6 Suppl 2**: p. 34-8.

Hirsch, J., et al., *Proteomics: current techniques and potential applications to lung disease.* Am J Physiol Lung Cell Mol Physiol, 2004. **287**(1): p. L1-23.

Huang da, W., B.T. Sherman, and R.A. Lempicki, *Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources.* Nat Protoc, 2009. **4**(1): p. 44-57.

Imhof, B.A. and D. Dunon, *Leukocyte migration and adhesion.* Adv Immunol, 1995. **58**: p. 345-416.

Issaq, H.J., *The role of separation science in proteomics research.* Electrophoresis, 2001. **22**(17): p. 3629-38.

James, P., et al., *Protein identification by mass profile fingerprinting.* Biochem Biophys Res Commun, 1993. **195**(1): p. 58-64.

Janeway, C.A., Jr., *The immune system evolved to discriminate infectious nonself from noninfectious self.* Immunol Today, 1992. **13**(1): p. 11-6.

Janeway, C.A., Jr. and R. Medzhitov, *Innate immune recognition.* Annu Rev Immunol, 2002. **20**: p. 197-216.

Jensen, L.J., et al., *STRING 8--a global view on proteins and their functional interactions in 630 organisms.* Nucleic Acids Res, 2009. **37**(Database issue): p. D412-6.

Jorens, P.G., et al., *L-arginine-dependent production of nitrogen oxides by rat pulmonary macrophages.* Eur J Pharmacol, 1991. **200**(2-3): p. 205-9.

Keller, A., et al., *Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search.* Anal Chem, 2002. **74**(20): p. 5383-92.

Kicic, A., et al., *Decreased fibronectin production significantly contributes to dysregulated repair of asthmatic epithelium.* Am J Respir Crit Care Med, 2010. **181**(9): p. 889-98.

Kindlen, S., *Physiology for Health Care and Nursing*. 2003, UK: Elsevier Health Sciences.

Kindlen, S., *Physiology for Health Care and Nursing*, S. Kindlen, Editor 2003, Elsevier Health Sciences: UK.

Lawson, P.R. and K.B. Reid, *The roles of surfactant proteins A and D in innate immunity.* Immunol Rev, 2000. **173**: p. 66-78.

Lazarus, S.C., *Role of inflammation and inflammatory mediators in airways disease.* Am J Med, 1986. **81**(5A): p. 2-7.

Lenz, A.G., et al., *Bronchoalveolar lavage fluid proteins in human lung disease: analysis by two-dimensional electrophoresis.* Electrophoresis, 1993. **14**(3): p. 242-4.

Lenz, A.G., et al., *Two-dimensional electrophoresis of dog bronchoalveolar lavage fluid proteins.* Electrophoresis, 1990. **11**(6): p. 510-3.

LeVine, A.M., et al., *Distinct effects of surfactant protein A or D deficiency during bacterial infection on the lung.* J Immunol, 2000. **165**(7): p. 3934-40.

LeVine, A.M., et al., *Surfactant protein D enhances clearance of influenza A virus from the lung in vivo.* J Immunol, 2001. **167**(10): p. 5868-73.

Lindahl, M., B. Stahlbom, and C. Tagesson, *Newly identified proteins in human nasal and bronchoalveolar lavage fluids: potential biomedical and clinical applications.* Electrophoresis, 1999. **20**(18): p. 3670-6.

Link, A.J., et al., *Direct analysis of protein complexes using mass spectrometry.* Nat Biotechnol, 1999. **17**(7): p. 676-82. java/Propub/biotech/nbt0799_676.fulltext java/Propub/biotech/nbt0799_676.abstract.

Liszewski, M.K., et al., *Control of the complement system.* Adv Immunol, 1996. **61**: p. 201-83.

Lo, C.J., M. Fu, and H.G. Cryer, *Interleukin 10 inhibits alveolar macrophage production of inflammatory mediators involved in adult respiratory distress syndrome.* J Surg Res, 1998. **79**(2): p. 179-84.

Luscombe, N.M., et al., *Genomic analysis of regulatory network dynamics reveals large topological changes.* Nature, 2004. **431**(7006): p. 308-12.

Makarov, A., et al., *Dynamic range of mass accuracy in LTQ Orbitrap hybrid mass spectrometer.* J Am Soc Mass Spectrom, 2006. **17**(7): p. 977-82.

Mann, M., P. Hojrup, and P. Roepstorff, *Use of mass spectrometric molecular weight information to identify proteins in sequence databases.* Biol Mass Spectrom, 1993. **22**(6): p. 338-45.

Marieb EN, H.K., *Human anatomy and physiology*, 2010, Benjamin Cummings: San Francisco.

Martinon, F., A. Mayor, and J. Tschopp, *The inflammasomes: guardians of the body.* Annu Rev Immunol, 2009. **27**: p. 229-65.

Medzhitov, R. and C.A. Janeway, Jr., *Decoding the patterns of self and nonself by the innate immune system.* Science, 2002. **296**(5566): p. 298-300.

Meyer, K.C., *The role of bronchoalveolar lavage in interstitial lung disease.* Clin Chest Med, 2004. **25**(4): p. 637-49, v.

Morsey, M.A., et al., *Cytokine profiles following interaction between bovine alveolar macrophages and Pasteurella haemolytica.* Microb Pathog, 1999. **26**(6): p. 325-31.

Moyron-Quiroz, J.E., et al., *Role of inducible bronchus associated lymphoid tissue (iBALT) in respiratory immunity.* Nat Med, 2004. **10**(9): p. 927-34.

Mulligan, M.S., et al., *Tissue injury caused by deposition of immune complexes is L-arginine dependent.* Proc Natl Acad Sci U S A, 1991. **88**(14): p. 6338-42.

Murray J.F., N.J.A., *Murray and Nadel's textbook of respiratory medicine*. 4th ed. 2005, Philadelphia, Pennslyvania: Saunders.

Nathan, C., *Nitric oxide as a secretory product of mammalian cells.* FASEB J, 1992. **6**(12): p. 3051-64.

Nelson DL, C.M., *Lehninger Principles of Biochemistry*. Third ed. 2000, New York, NY: Worth Publishers. 1152.

Nelson, P.J. and A.M. Krensky, *Chemokines, lymphocytes and viruses: what goes around, comes around.* Curr Opin Immunol, 1998. **10**(3): p. 265-70.

Nesvizhskii, A.I., et al., *A statistical model for identifying proteins by tandem mass spectrometry.* Anal Chem, 2003. **75**(17): p. 4646-58.

Noel-Georis, I., et al., *Database of bronchoalveolar lavage fluid proteins.* J Chromatogr B Analyt Technol Biomed Life Sci, 2002. **771**(1-2): p. 221-36.

Noel-Georis, I., et al., *Proteomics as the tool to search for lung disease markers in bronchoalveolar lavage.* Dis Markers, 2001. **17**(4): p. 271-84.

Oppert, M., et al., *Elevations in procalcitonin but not C-reactive protein are associated with pneumonia after cardiopulmonary resuscitation.* Resuscitation, 2002. **53**(2): p. 167-70.

Orchard, S., H. Hermjakob, and R. Apweiler, *The proteomics standards initiative.* Proteomics, 2003. **3**(7): p. 1374-6.

Oudhuis, G.J., et al., *Soluble Triggering Receptor Expressed on Myeloid cells-1 in bronchoalveolar lavage fluid is not predictive for ventilator-associated pneumonia.* Intensive Care Med, 2009. **35**(7): p. 1265-70.

Oury, T.D., B.J. Day, and J.D. Crapo, *Extracellular superoxide dismutase in vessels and airways of humans and baboons.* Free Radic Biol Med, 1996. **20**(7): p. 957-65.

Oury, T.D., B.J. Day, and J.D. Crapo, *Extracellular superoxide dismutase: a regulator of nitric oxide bioavailability.* Lab Invest, 1996. **75**(5): p. 617-36.

Pabst, R. and T. Tschernig, *Lymphocytes in the lung: an often neglected cell. Numbers, characterization and compartmentalization.* Anat Embryol (Berl), 1995. **192**(4): p. 293-9.

Pappin, D.J., P. Hojrup, and A.J. Bleasby, *Rapid identification of proteins by peptide-mass fingerprinting.* Curr Biol, 1993. **3**(6): p. 327-32.

Parbhakar, O.P., et al., *Depletion of pulmonary intravascular macrophages partially inhibits lipopolysaccharide-induced lung inflammation in horses.* Vet Res, 2005. **36**(4): p. 557-69.

Parker, D. and A. Prince, *Innate immunity in the respiratory epithelium.* Am J Respir Cell Mol Biol, 2011. **45**(2): p. 189-201.

Perkins, D.N., et al., *Probability-based protein identification by searching sequence databases using mass spectrometry data.* Electrophoresis, 1999. **20**(18): p. 3551-67.

Pittet, J.F., et al., *Biological markers of acute lung injury: prognostic and pathogenetic significance.* Am J Respir Crit Care Med, 1997. **155**(4): p. 1187-205.

Plymoth, A., et al., *Human bronchoalveolar lavage: biofluid analysis with special emphasis on sample preparation.* Proteomics, 2003. **3**(6): p. 962-72.

Povoa, P., et al., *C-reactive protein as a marker of ventilator-associated pneumonia resolution: a pilot study.* Eur Respir J, 2005. **25**(5): p. 804-12.

Raquil, M.A., et al., *Blockade of antimicrobial proteins S100A8 and S100A9 inhibits phagocyte migration to the alveoli in streptococcal pneumonia.* J Immunol, 2008. **180**(5): p. 3366-74.

Rasband, W.S., *Image J*, 1997-2011, US National Institutes of Health, Bethesda, Maryland USA, http://imagej.nih.gov/ij/.

Raychaudhuri, B., et al., *Interleukin 10 (IL-10)-mediated inhibition of inflammatory cytokine production by human alveolar macrophages.* Cytokine, 2000. **12**(9): p. 1348-55.

Rea-Neto, A., et al., *Diagnosis of ventilator-associated pneumonia: a systematic review of the literature.* Crit Care, 2008. **12**(2): p. R56.

Reynolds, H.Y., *Use of bronchoalveolar lavage in humans--past necessity and future imperative.* Lung, 2000. **178**(5): p. 271-93.

Rimele, T.J., et al., *Interaction of neutrophils with vascular smooth muscle: identification of a neutrophil-derived relaxing factor.* J Pharmacol Exp Ther, 1988. **245**(1): p. 102-11.

Rottoli, P. and E. Bargagli, *Is bronchoalveolar lavage obsolete in the diagnosis of interstitial lung disease?* Curr Opin Pulm Med, 2003. **9**(5): p. 418-25.

Rubartelli, A. and M.T. Lotze, *Inside, outside, upside down: damage-associated molecular-pattern molecules (DAMPs) and redox.* Trends Immunol, 2007. **28**(10): p. 429-36.

Rus, H., C. Cudrici, and F. Niculescu, *The role of the complement system in innate immunity.* Immunol Res, 2005. **33**(2): p. 103-12.

Sabounchi-Schutt, F., et al., *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**(9): p. 1851-60.

Sabounchi-Schutt, F., et al., *Changes in bronchoalveolar lavage fluid proteins in sarcoidosis: a proteomics approach.* Eur Respir J, 2003. **21**(3): p. 414-20.

Saeed, A.I., et al., *TM4: a free, open-source system for microarray data management and analysis.* Biotechniques, 2003. **34**(2): p. 374-8.

Sagel, S.D., M.K. Sontag, and F.J. Accurso, *Relationship between antimicrobial proteins and airway inflammation and infection in cystic fibrosis.* Pediatr Pulmonol, 2009. **44**(4): p. 402-9.

Scherl, A., et al., *Genome-specific gas-phase fractionation strategy for improved shotgun proteomic profiling of proteotypic peptides.* Anal Chem, 2008. **80**(4): p. 1182-91.

Schindler, R., et al., *Correlations and interactions in the production of interleukin-6 (IL-6), IL-1, and tumor necrosis factor (TNF) in human blood mononuclear cells: IL-6 suppresses IL-1 and TNF.* Blood, 1990. **75**(1): p. 40-7.

Schnapp, L.M., et al., *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**(1): p. 86-95.

Schoenborn, J.R. and C.B. Wilson, *Regulation of interferon-gamma during innate and adaptive immune responses.* Adv Immunol, 2007. **96**: p. 41-101.

Second, T.P., et al., *Dual-pressure linear ion trap mass spectrometer improving the analysis of complex protein mixtures.* Anal Chem, 2009. **81**(18): p. 7757-65.

Shull, M.M., et al., *Targeted disruption of the mouse transforming growth factor-beta 1 gene results in multifocal inflammatory disease.* Nature, 1992. **359**(6397): p. 693-9.

Simms, H.H. and R. D'Amico, *Regulation of intracellular polymorphonuclear leukocyte Fc receptors by lipopolysaccharide.* Cell Immunol, 1994. **157**(2): p. 525-41.

Sleigh, M.A., J.R. Blake, and N. Liron, *The propulsion of mucus by cilia.* Am Rev Respir Dis, 1988. **137**(3): p. 726-41.

Sokol, C.L., et al., *A mechanism for the initiation of allergen-induced T helper type 2 responses.* Nat Immunol, 2008. **9**(3): p. 310-8.

Soler, N., et al., *Bronchial microbial patterns in severe exacerbations of chronic obstructive pulmonary disease (COPD) requiring mechanical ventilation.* Am J Respir Crit Care Med, 1998. **157**(5 Pt 1): p. 1498-505.

Song, W.Y., et al., *A receptor kinase-like protein encoded by the rice disease resistance gene, Xa21.* Science, 1995. **270**(5243): p. 1804-6.

Spahr, C.S., et al., *Towards defining the urinary proteome using liquid chromatography-tandem mass spectrometry. I. Profiling an unfractionated tryptic digest.* Proteomics, 2001. **1**(1): p. 93-107.

Steinberg, K.P., et al., *Evolution of bronchoalveolar cell populations in the adult respiratory distress syndrome.* Am J Respir Crit Care Med, 1994. **150**(1): p. 113-22.

Stylianou, E. and J. Saklatvala, *Interleukin-1.* Int J Biochem Cell Biol, 1998. **30**(10): p. 1075-9.

Sumpter, R., Jr., et al., *Regulating intracellular antiviral defense and permissiveness to hepatitis C virus RNA replication through a cellular RNA helicase, RIG-I.* J Virol, 2005. **79**(5): p. 2689-99.

Takeda, K. and S. Akira, *TLR signaling pathways.* Semin Immunol, 2004. **16**(1): p. 3-9.

Taylor, C.F., et al., *The minimum information about a proteomics experiment (MIAPE).* Nat Biotechnol, 2007. **25**(8): p. 887-93.

Thompson, A.B., et al., *Lower respiratory tract lactoferrin and lysozyme arise primarily in the airways and are elevated in association with chronic bronchitis.* J Lab Clin Med, 1990. **115**(2): p. 148-58.

Tibshirani, R., et al., *Diagnosis of multiple cancer types by shrunken centroids of gene expression.* Proc Natl Acad Sci U S A, 2002. **99**(10): p. 6567-72.

Tracey, K.J., et al., *Shock and tissue injury induced by recombinant human cachectin.* Science, 1986. **234**(4775): p. 470-4.

Tschernig, T., W.J. Kleemann, and R. Pabst, *Bronchus-associated lymphoid tissue (BALT) in the lungs of children who had died from sudden infant death syndrome and other causes.* Thorax, 1995. **50**(6): p. 658-60.

Vaandrager, A.B. and L.M. van Golde, *Lung surfactant proteins A and D in innate immune defense.* Biol Neonate, 2000. **77 Suppl 1**: p. 9-13.

Vandal, K., et al., *Blockade of S100A8 and S100A9 suppresses neutrophil migration in response to lipopolysaccharide.* J Immunol, 2003. **171**(5): p. 2602-9.

Vidal, M., M.E. Cusick, and A.L. Barabasi, *Interactome networks and human disease.* Cell, 2011. **144**(6): p. 986-98.

von Bredow, C., P. Birrer, and M. Griese, *Surfactant protein A and other bronchoalveolar lavage fluid proteins are altered in cystic fibrosis.* Eur Respir J, 2001. **17**(4): p. 716-22.

Ward, P.A. and M.S. Mulligan, *New insights into mechanisms of oxyradical and neutrophil mediated lung injury.* Klin Wochenschr, 1991. **69**(21-23): p. 1009-11.

Washburn, M.P., D. Wolters, and J.R. Yates, 3rd, *Large-scale analysis of the yeast proteome by multidimensional protein identification technology.* Nat Biotechnol, 2001. **19**(3): p. 242-7.

Wattiez, R. and P. Falmagne, *Proteomics of bronchoalveolar lavage fluid.* J Chromatogr B Analyt Technol Biomed Life Sci, 2005. **815**(1-2): p. 169-78.

Wattiez, R., et al., *Human bronchoalveolar lavage fluid: two-dimensional gel electrophoresis, amino acid microsequencing and identification of major proteins.* Electrophoresis, 1999. **20**(7): p. 1634-45.

White, N.C., et al., *The growth and the control of human immunodeficiency virus in the lung: implications for highly active antiretroviral therapy.* Eur J Clin Invest, 1999. **29**(11): p. 964-72.

Wu, H.Y. and M.W. Russell, *Nasal lymphoid tissue, intranasal immunization, and compartmentalization of the common mucosal immune system.* Immunol Res, 1997. **16**(2): p. 187-201.

Wu, J., et al., *Differential proteomic analysis of bronchoalveolar lavage fluid in asthmatics following segmental antigen challenge.* Mol Cell Proteomics, 2005. **4**(9): p. 1251-64.

Xie, Q.W., et al., *Cloning and characterization of inducible nitric oxide synthase from mouse macrophages.* Science, 1992. **256**(5054): p. 225-8.

Yates, J.R., 3rd, et al., *Method to correlate tandem mass spectra of modified peptides to amino acid sequences in the protein database.* Anal Chem, 1995. **67**(8): p. 1426-36.

Yates, J.R., 3rd, et al., *Peptide mass maps: a highly informative approach to protein identification.* Anal Biochem, 1993. **214**(2): p. 397-408.

Yokomise, H., et al., *Importance of intrapulmonary lymph nodes in the differential diagnosis of small pulmonary nodular shadows.* Chest, 1998. **113**(3): p. 703-6.

Zou, G.M. and Y.K. Tam, *Cytokines in the generation and maturation of dendritic cells: recent advances.* Eur Cytokine Netw, 2002. **13**(2): p. 186-99.

VITAE

Elizabeth Vi Nguyen was born to Vietnamese immigrants, Vinh The Nguyen and Tu-anh Nguyen, on a warm spring day in 1981. She is more popularly known as Mimi, a name given from a Hmong tradition to ensure evil spirits will never find her on a paper trail. She was born in St. Vincent hospital in the county of Multnomah in Oregon, a hospital she would revisit 14 years later after one of her many near death experiences. She passionately lives and cherishes every moment. She has one older brother, Anthony, who is the first child born in America from her parent's generation of 20 siblings.

Mimi graduated from Mountain View High School (Home of the Thunder) in 2000 with two years of college credit in the suburbs of Vancouver, WA. She studied biochemistry, analytical chemistry, and worked in a physical chemistry laboratory under the supervision of Dr. J. Michael Schurr at the University of Washington. She earned a Bachelor of Science degree in 2004. Her "post college/pre graduate" life entails four months of working full time as both a Coca Cola quality assurance technician and a caregiver, over a year as a researcher at the Howard Hughes Medical Institute characterizing the myelin sheath of macaque monkeys in Dr. John Glomset's laboratory, and six months of back packing through SE Asia.

The excitement of following an uncertain path--paved by the outcome of experimentation--kindled her desire to pursue a life as a scientist. In 2006, she began her graduate work at the University of Washington for Professor David Goodlett. Her research has involved defining the proteome, utilizing a standard "bottom-up" shotgun approach, of bronchoalveolar lavage fluid (BALF) in various respiratory diseases including ventilator-associated pneumonia (VAP) and human immunodeficiency virus (HIV). The work was conducted in conjunction with physicians at Harborview Hospital Trauma Center in Seattle. Her goal is to identify subsets of proteins specific to disease states in order to gain insight into pathogenesis.

Her graduate experience has gained her exposure to the international science scene where she has given two international lectures in Croatia and Scotland. In 2012, she graduated with a Doctor of Philosophy in Medicinal Chemistry from the University of Washington. She has secured a postdoctoral scientist position in Professor Garry Corthals laboratory in Turku, Finland through the Biocenter Finland International Visitor Program. She will extend her training by using new droplet interfaced surface capture affinity mass spectrometry (DISCAMS) technique, recently developed by Deurion Ltd, to study protein phosphorylation.