Published in final edited form as:

Am J Physiol Lung Cell Mol Physiol. 2007 July ; 293(1): L45–L51. doi:10.1152/ajplung.00139.2007.

Genomics and proteomics of lung disease: conference summary

J. Usha Raj¹, Constantin Aliferis², Richard M. Caprioli³, Allen W. Cowley Jr.⁴, Peter F. Davies⁵, Mark W. Duncan⁶, David J. Erle⁷, Serpil C. Erzurum⁸, Patricia W. Finn⁹, Harry Ischiropoulos¹⁰, Naftali Kaminski¹¹, Steven R. Kleeberger¹², George D. Leikauf¹³, James E. Loyd¹⁴, Thomas R. Martin¹⁵, Sadis Matalon¹⁶, Jason H. Moore¹⁷, John Quackenbush¹⁸, Tara Sabo-Attwood¹⁹, Steve D. Shapiro²⁰, Jan E. Schnitzer²¹, David A. Schwartz²², Lisa M. Schwiebert²³, Dean Sheppard⁷, Lorraine B. Ware¹⁴, Scott T. Weiss²⁴, Jeff A. Whitsett²⁵, Mark M. Wurfel²⁶, and Michael A. Matthay²⁷

¹Division of Neonatology, Harbor-UCLA Medical Center, Geffen School of Medicine at University of California, and Los Angeles Biomedical Research Institute, Los Angeles, California

²Department of Biomedical Informatics, Vanderbilt University, Nashville, Tennessee

³Mass Spectrometry Research Center and Department of Biochemistry, Vanderbilt University, Nashville, Tennessee

⁴Department of Physiology, Medical College of Wisconsin, Milwaukee, Wisconsin

⁵Institute for Medicine and Engineering, University of Pennsylvania, Philadelphia, Pennsylvania

⁶Department of Pediatrics, Section of Pulmonary Medicine, University of Colorado at Denver and Health Sciences Center, Aurora, Colorado

⁷Lung Biology Center, Department of Medicine, University of California, San Francisco, California

⁸Department of Pathobiology, Lerner Research Institute, Cleveland Clinic Foundation, Cleveland, Ohio

⁹Pulmonary and Critical Care Medicine, Department of Medicine, School of Medicine, University of California, San Diego, California

¹⁰Stokes Research Institute, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania

¹¹Dorothy P. and Richard P. Simmons Center for Interstitial Lung Disease, Pulmonary, Allergy, and Critical Care Medicine, University of Pittsburgh Medical Center, Pittsburgh, Pennsylvania

¹²Laboratory of Respiratory Biology, National Institute of Environmental Health Sciences, National Institutes of Health, Research Triangle Park, North Carolina

¹³Department of Environmental Health, University of Cincinnati, Cincinnati, Ohio

¹⁴Division of Allergy, Pulmonary, and Critical Care Medicine, Vanderbilt University School of Medicine, Nashville, Tennessee

- ¹⁵Department of Medicine, Division of Pulmonary and Critical Care Medicine, University of Washington, Seattle, and Medical Research Service of the Veterans Affairs Puget Sound Healthcare System, Seattle, Washington
- ¹⁶Department of Anesthesiology, University of Alabama at Birmingham, Birmingham, Alabama
- ¹⁷Computational Genetics Laboratory, Norris-Cotton Cancer Center, Dartmouth Medical School, Lebanon, New Hampshire
- ¹⁸Department of Biostatistics and Computational Biology and Department of Cancer Biology, Dana-Farber Cancer Institute, Boston, and Department of Biostatistics, Harvard School of Public Health, Boston, Massachusetts
- ¹⁹Department of Pathology, College of Medicine, University of Vermont College of Medicine, Burlington, Vermont
- ²⁰Department of Medicine, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania
- ²¹Sidney Kimmel Cancer Center, San Diego, California
- ²²National Institute of Environmental Health Sciences, National Toxicology Program, Research Triangle Park, North Carolina
- ²³Department of Physiology and Biophysics, University of Alabama at Birmingham, Alabama
- ²⁴Channing Laboratory, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts
- ²⁵Division of Pulmonary Biology, University of Cincinnati College of Medicine, Cincinnati, Ohio
- ²⁶Division of Pulmonary and Critical Care Medicine, Harborview Medical Center, University of Washington, Seattle, Washington
- ²⁷University of California, Cardiovascular Research Institute, San Francisco, California

Genomics refers to the analysis of genomes. A genome is the complete set of DNA sequences that code for the hereditary material that is passed on from generation to generation. Thus genomics is the sequencing and analysis of all genes and transcripts in an organism. Proteomics refers to the analysis of the complete set of proteins or proteome. Bioinformatics involves the integration of computers, software tools, and databases in an effort to address biological questions. In addition to genomics and proteomics, there are many more areas of biology where bioinformatics is being applied (i.e., metabolomics, transcriptomics). Systems biology involves the integration of genomics, proteomics, and bioinformatics information to create a whole system view of a biological entity.

In this postgenomic era, we face formidable challenges as scientists and physicians to make sense of the huge amount of sequence information generated by the various genome projects in improving our understanding of human disease. Although our understanding of the pathology of lung diseases has grown rapidly in recent decades, the underlying mechanisms of many diseases remain obscure. The complexity of gene regulation and protein expression has made it difficult to globally map regulatory pathways that control basic cellular processes both in lung development and also in lung injury and repair. The rapid increase in

the development of new technologies in genomic and proteomic research that enables systems-wide analysis of cellular processes has enabled some small measure of progress in understanding common human diseases. These technological advances depend on the tools of many disciplines, including computational biology, chemistry, protein biochemistry, and mass spectrometry. This conference (Genomics and Proteomics of Lung Disease, November, 2006) brought together several leading investigators who discussed progress that has been made in the utilization of these powerful tools of genomics and proteomics in understanding lung disease and outlined directions for future work.

Genomic approaches to understanding complex human diseases

Traditional genomic analyses, such as studies of gene mutations, polymorphisms, and linkage mapping, have been used widely to delineate genetic variation in complex disorders. Genomic and proteomic approaches may improve our understanding of disorders such as systemic hypertension, asthma, interstitial lung disease or interstitial pulmonary fibrosis, primary pulmonary hypertension, infant respiratory distress syndrome, chronic obstructive pulmonary disease, acute lung injury, occupational lung diseases, and cystic fibrosis.

The polygenic nature of complex diseases, gene-environmental interactions, ethnic diversity, and the general heterogeneity of human populations has made the identification of causal genes by traditional linkage studies difficult. Systemic hypertension is a perfect example of a complex clinical disorder with polygenic and environmental determinants. Human studies have highlighted the complexity of both the genetic and environmental determinants. Blood pressure regulation is polygenic with many different genetic loci (pleiotropy and epistasis of common variants). There is a low probability that a random individual of a given genotype will display the trait (low penetrance). The severity of the disease is generally low (low expressivity); gender, lifestyle, and ecogenetic context is of great importance, and there are few replications among different populations.

Allen Cowley, in his keynote talk, described his research group's approach to understanding the genetics of systemic hypertension (8). Their research has therefore emphasized the utility of using rat models in which natural genetic variants (and combinations of these variants) that influence blood pressure have been captured and inbred. The goal has been to determine the genetic and physiological basis of the protection from salt-induced hypertension and related phenotypes that occur with substitution of BN alleles (a salt-insensitive strain) into the genomic background of the Dahl salt-sensitive rat. Although they have shown that five different rat chromosomes contribute importantly to salt-sensitive hypertension and renal dysfunction, they are currently focused on chromosome 13. They have developed and phenotyped 23 overlapping congenic strains spanning chromosome 13 to narrow the specific region(s) containing genes responsible for the trait of salt sensitivity and renal and vascular dysfunction. Studies are underway to determine how genes within four narrow congenic regions contribute to genome-wide responses and mechanistic pathways to improve sodium excretory function and protect the organism from hypertension and associated phenotypes. Strategies for identification and prioritization of candidate genes in narrow congenic regions that are responsible for the protection from hypertension and for final validation of the candidate gene using transgenic approaches were presented (19).

Jim Loyd reported that familial pulmonary artery hypertension (FPAH) exhibits features of vertical transmission, incomplete penetrance, and earlier age of onset in subsequent generations. By collecting DNA specimens from families with PAH, the group was able to localize FPAH to chromosome 2q32 in 1997. The responsible gene therein was identified to be bone morphogenetic protein receptor II (BMPRII) in 2000 (20). Mutations in BMPRII are now known to be the basis of FPAH in 80% of families (21). Nearly 150 different BMPRII mutations have been described, including mutation in each of the 13 exons of the four functional domains. Mutations in other transforming growth factor (TGF)- β family members, including ALK-1 and endoglin, have also been described in association with PAH (10, 23). BMPRII mutation is clinically expressed in only 20% of mutation carriers, so it is believed that modifier genes determine the clinical expression. It is not known how many modifiers participate, or what functions they exhibit, such as enhancement vs. suppression, or both, resulting in the clinically expressed phenotype. A recently developed mouse strain with the BMPRII mutation holds promise as a model of PAH, which should allow better understanding of pathogenesis and screening of novel treatments.

Innate immune inflammatory responses represent an important quantitative trait that shows considerable interindividual variability in the normal population (45). Mark Wurfel described studies that have identified common genetic variation in the Toll-like receptor (TLR) pathway that account for a portion of the variability in this quantitative trait and showed that this genetic variation can also predict poor clinical outcomes in the complex clinical phenotypes of sepsis and septic shock. They screened a panel of single nucleotide polymorphisms (SNPs) from 45 genes in the TLR pathway and identified associations between 20 SNPs and TLR agonist-induced responses in peripheral leukocytes from healthy volunteers ex vivo. These validated previously published SNPs in TLR4 and TLR5 and also identified new functional SNPs in other TLRs. In a cohort of critically ill patients with sepsis and septic shock, they found that these SNPs in specific TLRs predicted increased mortality, greater organ failure, and higher prevalence of infection with gram-positive organisms. These studies demonstrated that understanding how common genetic variation influences innate immune responses in the normal population can lead to the identification of new markers of susceptibility to outcomes in sepsis and septic shock. Thus understanding interindividual variation in TLR-mediated innate immune responses in the normal population provides a relatively unbiased way to identify individuals who are relatively susceptible or resistant to sepsis, septic shock, and related organ failure (44, 45).

Dean Sheppard described a general strategy for combining the advantages of knockout mice and microarray technology both to generate hypotheses about the molecular mechanisms underlying observed knockout phenotypes and to predict additional phenotypes. An attractive feature of this approach is that hypotheses and predictions can then be tested using additional knockout or transgenic lines. One example of this approach has been the continuing evaluation of mice generated more than a decade ago that lack a single integrin subunit (β_6) and are thus deficient in the epithelially restricted integrin heterodimer, $\alpha_v\beta_6$. These mice were initially noted to have exaggerated inflammatory responses in epithelial organs (16) but later were found to be protected in models of tissue fibrosis, including pulmonary fibrosis. Based on these phenotypes, the integrin was suspected to have a role in regulating the function of the anti-inflammatory and profibrotic growth factor, TGF- β (30).

Microarray studies comparing pulmonary gene expression in wild-type and knockout mice at baseline and after treatment with the fibrogenic drug bleomycin identified a cluster of TGF-B-inducible genes that were clearly expressed and induced at higher levels in wild-type animals (18), providing strong support for this hypothesis. Data from these microarrays also demonstrated a surprisingly early induction of TGF-B-inducible genes, well before the onset of pulmonary fibrosis. These results suggested that this pathway might play a role in modulating early responses to bleomycin, such as the development of acute lung injury. This possibility turned out to be correct because the β_6 knockout mice were completely protected from bleomycin-induced pulmonary edema, and this response was also inhibited by blocking TGF-B (36).

Comparison of global pulmonary gene expression between unchallenged wild-type and β_6 knockout mice also suggested previously unexpected roles for this integrin. The most differentially expressed gene encoded matrix metalloprotease 12 (MMP-12) is also called macrophage metalloelastase. MMP-12 knockout mice had been previously shown to be protected from pulmonary emphysema in response to chronic exposure to tobacco smoke (18). This result thus suggested that the $\alpha_v\beta_6$ integrin and TGF-B might normally play a role in suppressing MMP-12 expression and that loss of this pathway might lead to the development of emphysema. Before the microarray analysis, there had been no suggestion of emphysema in β_6 knockout mice, but based on these results, cohorts of mice were allowed to age, and lungs from these animals were evaluated morphometrically. By 6 mo of age, a small but statistically significant increase in alveolar size was observed, and this effect progressively increased with age up to 14 mo (29). The effect was clearly due to loss of the integrin β_6 subunit in alveolar epithelial cells because it could be rescued by transgenic expression of this subunit only in a subset of alveolar epithelial cells using the surfactant protein C promoter.

Steven Shapiro presented data that illustrated how animal models can play an important role in the understanding of the pathogenesis of chronic obstructive pulmonary disease (COPD) (40). The applicability of findings to human COPD depends on several factors, including the disease model and similarities in mouse structure and function between species. There are many examples in the literature of transgenic mice that have contributed to the understanding of COPD. Several studies demonstrate the complexity of inflammatory networks and how unexpected findings in animal models have led to the search for new potential mediators in human disease. Gene-targeting studies of α_1 -antitrypsin (α_1 -AT) and emphysema in mice have demonstrated that the genetic locus for α_1 -AT in mice is complex and that the loss of one gene is lethal in embryo lung development. This underlines the differences between mice and humans that limit the ability to translate between systems in some instances. Gene targeting has also highlighted complex roles for TGF-B in COPD and has been used to determine important molecules and pathways in COPD. Both transgenic and gene-targeted models suffer limitations, and their applicability to human COPD may be dependant on several factors, some of which are still being appreciated. The more that is known about similarities and differences, the better the knowledge will be that is gained for COPD.

George Leikauf and Tara Sabo-Attwood presented data to illustrate the use of knockout mice and/or microarray technology to study the pathogenesis of lung injury. Tara Sabo-Attwood presented her work on asbestos-induced fibrogenesis (37), and George Leikauf presented work on acute lung injury (ALI). His research group utilized nickel-induced lung injury in mice as a model of ALI (26) and have found metallothionein to be one of the greatest noted in transcriptome-wide analyses of gene expression. For example, their previous studies have shown that mice deficient in the tyrosine kinase domain ($TK^{-/-}$) of the receptor *Mst1r* have an increased susceptibility to nickel (Ni)-induced ALI. Using microarray analyses, they found that a total of 343 transcripts were significantly changed, either by Ni treatment or between genotypes (43). Dr. Leikauf also discussed the use of whole genome association analysis in the identification of critical genes involved in ALI.

Jeff Whitsett reported on their identification of complex networks of genes and transcriptional factors that regulate lung development. Mammalian lung is lined by a diversity of epithelial cell types that vary during development, regionally along the cephalo-caudal and dorsal-ventral axes of the respiratory tract and among species. Formation and differentiation of the respiratory epithelium is strongly influenced by a number of transcription factors including TTF-1, GATA-6, NF-1, Foxa1, Foxa2, β -catenin, Foxj1, Ets family members, C/EBP α , Sox family members, p63, and others. The proteins are expressed in various pulmonary cell types to influence respiratory epithelial cell differentiation and gene expression. Gene deletion and addition studies in the mouse demonstrate the importance of these transcription factors in the formation and differentiation of the lung before birth and their roles in various aspects of lung function and homeostasis after birth. Many of the transcription factors interact at multiple levels, coregulating each other, interacting directly via protein-protein interactions, and by both distinct and cooperative interactions at binding sites on specific transcriptional target genes. RNA microarray analyses of lung RNAs were utilized to identify shared and distinct transcriptional targets and participants in the networks. Many of these transcription factors are expressed postnatally and are regulated during repair of the lung following injury or during regrowth following unilateral pneumonectomy. Dynamic changes in the expression of TTF-1, β -catenin, Sox, Ets, Fox family members, and Stat-3 accompany the repair of the respiratory epithelium. Together, these studies support the concept that transcriptional programs that mediate epithelial cell differentiation during lung morphogenesis also play important roles during injury and repair (24).

Independently, David Erle and Scott Weiss have utilized microarray analyses to dissect the genomics of asthma. Dr. Erle's research group applied this technique to the study of IL-13-induced gene expression changes in airway epithelial cells; IL-13 is an important regulatory cytokine in asthma pathogenesis (49). Scott Weiss' team integrated microarray analysis with traditional gene linkage and association mapping to identify key pathways or candidate genes that are involved in the asthma response (4). Naftali Kaminski contrasted two approaches to analysis of microarray experiments, a reductionist "cherry picking approach" and a global "systems biology" approach (41). He presented examples of the application of the two approaches to the understanding of human pulmonary fibrosis including the use of microarrays to classify lung diseases (39) and to identify biomarkers for lung fibrosis (33). Elizabeta Kovkarova from his team presented the use of laser capture microscopy with

microarray analysis to profile gene expression in idiopathic pulmonary fibrosis microenvironments, and Sean Studer presented the use of microarray analysis to evaluate changes in gene expression profiles that may predict the likelihood of lung transplant rejection. Patricia Finn presented a summary of her team's work on interaction networks in allergic asthmatic responses. She presented data to support their thesis that a combination of differential gene expression (46) plus topological characteristics of the interaction network provides enhanced understanding of allergic responses and may have important implications for interpreting gene expression data in relationship to biological responses.

In addition to the identification of causal genes, research must also determine how these genes interact with environmental stimuli to either preserve health or cause disease. The application of genetics and genomics to problems in environmental health represents a potentially effective strategy to substantially impact morbidity and mortality. David Schwartz, Director of the National Institute of Environmental Health Sciences, was a featured speaker at the meeting and stated how recent advances in human and molecular genetics provide an unparalleled opportunity to understand how genes and genetic changes interact with environmental stimuli to either preserve health or cause disease (38). The fields of environmental genetics and environmental genomics have enormous potential to develop our ability to accurately assess the risk of developing disease, identify and understand basic pathogenic mechanisms that are critical to disease progression, and to more precisely phenotype disease subtypes (48). Collaborative approaches that team together environmental scientists with molecular biologists, geneticists, physiologists, and physician scientists are critical to the investigation of environmental aspects of human health. Moreover, exploiting eukaryotic model systems (yeast, *Caenorhabditis elegans*, zebrafish, *Drosophila*, and rodents) will accelerate our understanding of environmental exposures on human health. Schwartz presented compelling data on how environmental insults can result in DNA methylation and epigenetic phenomenon (15), a process that may be involved in predisposing susceptible humans to asthma.

Steven Kleeberger highlighted the fact that considerable effort is ongoing to understand the genetic basis of susceptibility to the pulmonary effects of environmental exposures. Association studies in populations exposed to environmental stimuli (22) and subjects exposed under controlled conditions (47) have tested the importance of candidate susceptibility genes in pulmonary responses to the stimuli. Positional cloning studies using inbred mouse models have also identified genes that determine differential susceptibility to environmental stimuli including ozone and particulate matter (7, 26). Translational investigations that test in human populations candidate genes identified in mouse models should yield mechanistic insight to differential susceptibility (21). Continued investigation into the mechanisms of interaction between genetic background and environmental exposures should yield novel intervention strategies and means to identify susceptible individuals.

Together, these investigators illustrated the utility of genomic-based approaches in the identification and assessment of causal genes that promote disease development and progression. Collaborative approaches that team together molecular biologists, geneticists,

physiologists, and physician scientists are critical to the further investigation and improved treatment of complex genetic disorders.

Data mining and bioinformatics in genomics

Several challenges still remain. The gene association studies in complex human diseases to identify causal pathways remain difficult as they present biostatistical challenges in the analysis of large data sets in humans. Jason Moore, John Quackenbush, and Constantin Aliferis presented different approaches to analysis of large data sets. Dr. Moore discussed the important role of biological knowledge in genome-wide studies of epistasis (28). Constantin Aliferis presented experiments with microarray gene expression and mass spectrometry data using novel biomarker selection algorithms that reverse-engineer local regions of gene and protein regulatory networks. These results suggest that the new algorithms discover accurate, compact, statistically reproducible, and pathway-localized biomarkers and signatures. He also cautioned that predictive ability does not imply biological importance, and most predictive biomarkers are not necessarily causative and should not be interpreted as such unless specialized discovery methods are used. In general, because of rich network connectivity, it is conceivable that in diseases that affect multiple pathways, hundreds if not thousands of genes are affected and are thus predictive of phenotype. Furthermore, he explained that understanding complex networks requires highly specialized algorithms and larger sample sizes than differential expression or predictive modeling. New methods of development call for extensive design and testing. In bioinformatics, it is not unusual for a new method to be invented and introduced in a short period of time just to analyze a specific data set. When little is known about the formal properties of the new method (in terms of correctness, completeness, sample and time complexity), results should not be easily trusted, however. Another pitfall is the overinterpretation of gene clustering. If two genes cluster together, this does not imply that they are always in the same pathways or share the same functional roles. He concluded that the opportunities for revolutionary discoveries using gene expression microarrays (as well as other high-throughput technologies) are immense but so are the challenges for their proper analysis and interpretation (1).

Use of proteomics in understanding lung diseases

The field of proteomics has come to encompass not only the identification and quantification of proteins but also the determination of protein localization, modifications, interactions, activities, and functions (9). Investigators in lung disease are beginning to take full advantage of the growing array of proteomic methodologies to improve our understanding of both the normal and the diseased lung.

Several presenters focused on posttranslational modification of proteins that alter protein structure and function. Harry Ischiropoulos presented work on the *S*-nitrosylation of proteins in human aortic smooth muscle cells. *S*-nitrosylation, the formal transfer of nitrosonium to a reduced cysteine, is a reversible and selective posttranslational modification, regulating protein activity, localization, and stability, while also functioning as a general sensor for cellular redox balance. Selective *S*-nitrosylation of cysteine residues in proteins to form *S*-

nitrosocysteine is a major emerging mechanism by which nitric oxide acts as a signaling molecule (14, 17). A proteomic approach using selective peptide capturing and site-specific adduct mapping was employed to identify the targets of *S*-nitrosylation in human aortic smooth muscle cells upon exposure to *S*-nitrosocysteine and propylamine propylamine NONOate (12). This strategy identified 20 unique *S*-nitrosocysteine-containing peptides belonging to 818 proteins including cytoskeletal proteins, chaperones, proteins of the translational machinery, vesicular transport, and signaling. Further refinements of the proteomic approaches can be taken to identify endogenous *S*-nitrosylated proteins and their function. Serpil Erzurum presented a summary of her work on the proteomics of oxidant lung injury in asthma. The identification of nitrated proteins and effects of posttranslational modifications on protein function has allowed greater understanding of the pathophysiological consequences of ROS and RNS in asthma (11). Sadis Matalon discussed the functional implications of nitration of surfactant protein A (SP-A). Exposure of human SP-A results in nitration of all three tyrosines, oxidation of side chains, and significant aggregation, which interfered with its ability to serve as a ligand for *Pneumocystis carinii* adherence to alveolar macrophages (50). Mass spectrometry was used to identify specific nitration sites. The major nitrated peptide was the tryptic fragment Tyr161-Arg179 located in the carbohydrate recognition domain. Sequencing of this nitrated peptide with electrospray ionization liquid chromatography-tandem mass spectrometry demonstrated that the nitration was equally distributed on Tyr164 and Tyr166 (13). These studies indicate that nitration of a single tyrosine in the SP-A carbohydrate recognition domain decreases its ability to aggregate lipids and bind mannose. These investigators provided evidence that overproduction of reactive oxygen and nitrogen intermediates causes oxidative modifications of several important lung proteins, and these modifications are associated with loss of function.

Jan Schnitzer presented a novel approach to proteomic mapping of the endothelial cell surfaces and their caveolae isolated from organs and solid tumors for identifying new targets to enhance tissue-specific imaging and therapy (27, 32). A potential clinically useful finding was that caveolae can function like active pumps moving rat lung-specific targeting antibodies across endothelium into the tissue interstitium within seconds of intravenous injection. Peter Davies outlined studies of spatial profiling of endothelial gene and protein expression in the large arteries of swine. Comparisons of endothelial phenotypes in regions that are susceptible to, or protected from, atherosclerosis revealed significant differential expression of genes associated with important pro- and anti-pathological pathways and differential posttranslational modifications of the important PKC- ζ isoenzyme (25, 34). Athero-susceptible sites are closely associated with regions of disturbed blood flow; flow characteristics are proposed to play a major role in the regulation of endothelial phenotype. Multiple arterial regions from disturbed flow (DF) or undisturbed flow (UF) sites were analyzed in swine as a function of imposed risk factors including diet and gender to attempt to modify the susceptible endothelial phenotype. A surprising finding following an unbiased interrogation of the top 25% of differentially expressed genes was a hierarchical clustering in which regional location (UF or DF) completely dominated gender and dietary treatments. The study implicates localized hemodynamics as an important mechanistic determinant of regional endothelial phenotype (35).

Richard Caprioli presented his work on in situ molecular imaging and profiling of proteins in tissues. Profiling and imaging matrix-assisted laser desorption/ionization (MALDI) MS can be used to assess the spatial distribution of peptides and proteins in biological samples and is especially effective in its application to tissue sections. The applications range from low-resolution images of peptides and proteins in selected areas of tissue to high-resolution images of tissue cross sections (5, 6). The Caprioli group has employed this technology in studies of a variety of diseases, comparing proteins differentially expressed in diseased tissue with those in the corresponding normal tissue, and finally correlating these differences with patient outcomes.

Investigators from two of the National Heart, Lung, and Blood Institute-funded Clinical Proteomics Programs presented work on biomarker discovery in ALI. Mark Duncan described proteomic methods to study clinical ALI with the aim of identifying biomarkers of the disorder and its clinical course (2). Duncan and colleagues employed a protein-centric approach based on two-dimensional gel separation followed by mass spectrometry. Plasma and pulmonary edema fluid (EF) from ALI subjects were compared with plasma and bronchoalveolar lavage fluid collected from controls. The changes observed were consistent with a loss of size selectivity of the alveolar-capillary barrier, impaired alveolar type II cell function in ALI, enhanced proteolytic activity in the EF of ALI patients, and an increase in acute-phase proteins. However, many of the changes are common to other lung diseases. Nevertheless, studies such as these are of great value because they offer a powerful alternative first-step to hypothesis generation that is less dependent on insight, instinct, and experience (3).

Lorraine Ware presented a summary of work from the Vanderbilt Clinical Proteomics Program on using MALDI time-of-flight mass spectrometry (MALDI-TOF MS) for discovery of novel biomarkers of adverse clinical outcomes in plasma from patients with ALI/ARDS (2). Working with Caprioli, Ware and colleagues have developed a robotic reverse phase fractionation procedure to reduce high abundance proteins in plasma. As part of the characterization and validation of this fractionation procedure, they discovered that significant protein degradation was ongoing in plasma samples that were collected with EDTA and allowed to sit on ice for 7–8 h. Furthermore, archival plasma samples that had been collected as part of a randomized clinical trial of two ventilator strategies in patients with ARDS also showed significant evidence of protein degradation, suggesting that routine sample collection procedures may not have been adequate for discovery proteomic studies. These findings lead to the conclusion that archival samples may potentially be problematic for mass spectrometry-based discovery proteomics (42). Prospective sample collection for clinical studies may be preferable, and work is ongoing to characterize the role of protein stabilizers and protease inhibitors in sample preservation for discovery proteomics.

Summary

The progress made in understanding the most common lung diseases has been varied. For example, in the case of infant respiratory distress syndrome, the biochemical defect is clear (i.e., immature type II cells unable to make sufficient surface active material), but the contribution of other factors, including genetic and environmental factors, is not clear,

especially in the development of bronchopulmonary dysplasia. In addition, although considerable progress in understanding key genes determining lung development has been made, the link between lung development and disease, injury, and repair is still not established yet. In the case of systemic hypertension, a complex clinical disorder with polygenic and environmental determinants, human studies have highlighted the complexity of both the genetic and environmental determinants. Some progress has been made utilizing animal models, but the studies are labor intensive and expensive and require enormous resources. A major advance in our understanding of the pathogenesis of FPAH has been the identification of a mutation in BMPR2, but pathways by which BMPR2 causes PAH is not known. The low penetrance of this gene suggests the effect of modifier genes that include those of estrogen, androgens, NOS-1, angiotensin-1, PAI-1, and/or the effect of environmental factors. The lack of animal models relevant to the human disease and the lack of early stage biological markers of the disease have impeded progress.

Asthma is another excellent example of a complex clinical disorder with polygenic and environmental determinants. Although progress has been slow in understanding the pathogenesis of this disease, some selected animal models, such as the mouse model where IL-13 expression is genetically manipulated, have provided new insights. In addition, the study of environmental factors resulting in epigenetic phenomenon that may predispose humans to asthma has been informative. Studies related to ALI and ARDS have included genetic association studies where some associations have been identified; however, validation will be needed including close analysis of the phenotype and contributing causes of lung injury. For example, the candidate genes that indicate susceptibility to sepsis may not identify susceptibility to ARDS. Discovery proteomics, specially utilizing bronchoalveolar lavage fluid from patients with ARDS, is not promising as yet, largely because of methodological issues, although targeted proteomics, perhaps in association with gene association studies, may be of value.

As Michael Matthay explained, the challenge that lies ahead of us in utilizing the modern tools of genomics and proteomics to treat lung disease is best exemplified by the status of a genetically inherited disorder, cystic fibrosis. Although we have known since 1989 that this disease is associated with a genetic mutation in the CFTR gene, which results in the chloride channel not being delivered to the plasma membrane normally, the link between that abnormality and the human lung disease remains poorly understood to date. The impact of gene modification on the phenotype has been shown for polymorphisms of TGF- β , but the therapy is still only supportive. Ironically, the best new therapy is aerosolized hypertonic saline, which increases cough and mucous expectoration.

Future research

Most lung diseases are complex, and no single gene or protein will help us either identify the pathogenesis or lead to new approaches of treating it. The clinical syndromes of lung diseases such as COPD and interstitial pulmonary fibrosis and others mentioned above are heterogeneous. Understanding complex interaction among many genes that results in the disease will require a systematic undertaking aimed at identifying not only single gene pathways but also the network of genes that interact to produce the disease. Perhaps most

importantly, we learned from this conference that it will require coordinated research and creative input from many disciplines to begin to “translate” the genome and unlock the great potential this information has for understanding human development and disease. Targeted proteomics will likely assist us in identifying specific genotypes and phenotypes of these multifactorial and complex disorders.

Acknowledgments

GRANTS

This conference was supported by the American Physiological Society and by grants from National Institutes of Health, National Institute of Environmental Health Sciences, and Sepracor, Inc.

REFERENCES

1. Aliferis CF, Statnikov A, Tsamardinos I. Challenges in the analysis of mass-throughput data: a technical commentary from the statistical machine learning perspective. *Cancer Informatics*. 2006; 2:133–162. [PubMed: 19458765]
2. Bowler RP, Duda B, Chan ED, Enghild JJ, Ware LB, Matthay MA, Duncan MW. Proteomic analysis of pulmonary edema fluid and plasma in patients with acute lung injury. *Am J Physiol Lung Cell Mol Physiol*. 2004; 286:L1095–L1104. [PubMed: 14742308]
3. Brown LM, Helmke SM, Hunsucker SW, Netea-Maier RT, Chiang SA, Heinz DE, Shroyer KR, Duncan MW, Haugen BR. Quantitative and qualitative differences in protein expression between papillary thyroid carcinoma and normal thyroid tissue. *Mol Carcinog*. 2006; 45:613–626. [PubMed: 16788983]
4. Celedon JC, Soto-Quiros ME, Avila L, Lake SL, Liang C, Fournier E, Spesny M, Hersh CP, Sylvia JS, Hudson TJ, Verner A, Klanderma B, Freimer NB, Silverman EK, Weiss ST. Significant linkage to airway responsiveness on chromosome 12q24 in families of children with asthma in Costa Rica. *Hum Genet*. 2007; 120:691–699. [PubMed: 17024367]
5. Chaurand P, Schwartz SA, Billheimer D, Xu BJ, Crecelius A, Caprioli RM. Integrating histology and imaging mass spectrometry. *Anal Chem*. 2004; 76:1145–1155. [PubMed: 14961749]
6. Chaurand P, Schwartz SA, Caprioli RM. Profiling and imaging proteins in tissue sections by MS. *Anal Chem*. 2004; 76:87A–93A.
7. Cho HY, Kleeberger SR. Genetic mechanisms of susceptibility to oxidative lung injury in mice. *Free Radic Biol Med*. 2007; 42:433–445. [PubMed: 17275675]
8. Cowley AW Jr. The genetic dissection of essential hypertension. *Nat Rev Genet*. 2006; 7:829–840. [PubMed: 17033627]
9. Fields S. Proteomics. *Proteomics in genomeland*. *Science*. 2001; 291:1221–1224. [PubMed: 11233445]
10. Geraci MW, Moore M, Gesell T, Yeager ME, Alger L, Golpon H, Gao B, Loyd JE, Tudor RM, Voelkel NF. Gene expression patterns in the lungs of patients with primary pulmonary hypertension: a gene microarray analysis. *Circ Res*. 2001; 88:555–562. [PubMed: 11282888]
11. Ghosh S, Janocha AJ, Aronica MA, Swaidani S, Comhair SA, Xu W, Zheng L, Kaveti S, Kinter M, Hazen SL, Erzurum SC. Nitrotyrosine proteome survey in asthma identifies oxidative mechanism of catalase inactivation. *J Immunol*. 2006; 176:5587–5597. [PubMed: 16622028]
12. Greco TM, Hodara R, Parastatidis I, Heijnen HF, Dennehy MK, Liebler DC, Ischiropoulos H. Identification of *S*-nitrosylation motifs by site-specific mapping of the *S*-nitrosocysteine proteome in human vascular smooth muscle cells. *Proc Natl Acad Sci USA*. 2006; 103:7420–7425. [PubMed: 16648260]
13. Greis KD, Zhu S, Matalon S. Identification of nitration sites on surfactant protein A by tandem electrospray mass spectrometry. *Arch Biochem Biophys*. 1996; 335:396–402. [PubMed: 8914938]
14. Hess DT, Matsumoto A, Kim SO, Marshall HE, Stamler JS. Protein *S*-nitrosylation: purview and parameters. *Nat Rev Mol Cell Biol*. 2005; 6:150–166. [PubMed: 15688001]

15. Hu M, Yao J, Cai L, Bachman KE, van den Brule F, Velculescu V, Polyak K. Distinct epigenetic changes in the stromal cells of breast cancers. *Nat Genet.* 2005; 37:899–905. [PubMed: 16007089]
16. Huang XZ, Wu JF, Cass D, Erle DJ, Corry D, Young SG, Farese RV Jr, Sheppard D. Inactivation of the integrin beta 6 subunit gene reveals a role of epithelial integrins in regulating inflammation in the lung and skin. *J Cell Biol.* 1996; 133:921–928. [PubMed: 8666675]
17. Jaffrey SR, Erdjument-Bromage H, Ferris CD, Tempst P, Snyder SH. Protein S-nitrosylation: a physiological signal for neuronal nitric oxide. *Nat Cell Biol.* 2001; 3:193–197. [PubMed: 11175752]
18. Kaminski N, Allard JD, Pittet JF, Zuo F, Griffiths MJ, Morris D, Huang X, Sheppard D, Heller RA. Global analysis of gene expression in pulmonary fibrosis reveals distinct programs regulating lung inflammation and fibrosis. *Proc Natl Acad Sci USA.* 2000; 97:1778–1783. [PubMed: 10677534]
19. Kunert MP, Drenjancevic-Peric I, Dwinell MR, Lombard JH, Cowley AW Jr, Greene AS, Kwitek AE, Jacob HJ. Consomic strategies to localize genomic regions related to vascular reactivity in the Dahl salt-sensitive rat. *Physiol Genomics.* 2006; 26:218–225. [PubMed: 16772359]
20. Lane KB, Machado RD, Pauciulo MW, Thomson JR, Phillips JA 3rd, Loyd JE, Nichols WC, Trembath RC. Heterozygous germline mutations in BMPR2, encoding a TGF-beta receptor, cause familial primary pulmonary hypertension. The International PPH Consortium. *Nat Genet.* 2000; 26:81–84. [PubMed: 10973254]
21. Leikauf GD, McDowell SA, Wesselkamper SC, Hardie WD, Leikauf JE, Korfhagen TR, Prows DR. Acute lung injury: functional genomics and genetic susceptibility. *Chest.* 2002; 121(Suppl 3): 70S–75S. [PubMed: 11893692]
22. Li H, Romieu I, Wu H, Sienra-Monge JJ, Ramirez-Aguilar M, Del Rio-avarro E, Del Lara-Sanchez IC, Kistner EO, Gjessing HK, London SJ. Genetic polymorphisms in transforming growth factor beta-1 (TGFβ1) and childhood asthma and atopy. *Hum Genet.* 2007; 21:529–538. [PubMed: 17333284]
23. Machado RD, Pauciulo MW, Thomson JR, Lane KB, Morgan V, Wheeler L, Phillips JA 3rd, Newman J, Williams D, Galie N, Manes A, McNeil K, Yacoub M, Mikhail G, Rogers P, Corris P, Humbert M, Donnai Martensson G, Tranebjaerg L, Loyd JE, Trembath RC, Nichols WC. BMPR2 haploinsufficiency as the inherited molecular mechanism for primary pulmonary hypertension. *Am J Hum Genet.* 2001; 68:92–102. [PubMed: 11115378]
24. Maeda Y, Dave V, Whitsett JA. Transcriptional control of lung morphogenesis. *Physiol Rev.* 2007; 87:219–244. [PubMed: 17237346]
25. Magid R, Davies PF. Endothelial protein kinase C isoform identity and differential activity of PKCζ in an athero-susceptible region of porcine aorta. *Circ Res.* 2005; 97:443–449. [PubMed: 16051884]
26. Mallakin A, Kutcher LW, McDowell SA, Kong S, Schuster R, Lentsch AB, Aronow BJ, Leikauf GD, Waltz SE. Gene expression profiles of Mst1r-deficient mice during nickel-induced acute lung injury. *Am J Respir Cell Mol Biol.* 2006; 34:15–27. [PubMed: 16166746]
27. McIntosh DP, Tan XY, Oh P, Schnitzer JE. Targeting endothelium and its dynamic caveolae for tissue-specific transcytosis in vivo: a pathway to overcome cell barriers to drug and gene delivery. *Proc Natl Acad Sci USA.* 2002; 99:1996–2001. [PubMed: 11854497]
28. Moore, JH. Genome-wide analysis of epistasis using multifactor dimensionality reduction: feature selection and construction in the domain of human genetics. In: Zhu, X.; Davidson, I., editors. *Knowledge Discovery and Data Mining: Challenges and Realities with Real World Data.* New York: IGI Press; 2007.
29. Morris DG, Huang X, Kaminski N, Wang Y, Shapiro SD, Dolganov G, Glick A, Sheppard D. Loss of integrin alpha (v) beta6-mediated TGF-beta activation causes p12-dependent emphysema. *Nature.* 2003; 422:169–173. [PubMed: 12634787]
30. Munger JS, Huang X, Kawakatsu H, Griffiths MJ, Dalton SL, Wu J, Pittet JF, Kaminski N, Garat C, Matthay MA, Rifkin DB, Sheppard D. The integrin alpha v beta 6 binds and activates latent TGF beta-1: a mechanism for regulating pulmonary inflammation and fibrosis. *Cell.* 1999; 96:319–328. [PubMed: 10025398]

31. Newman JH, Wheeler L, Lane KB, Loyd E, Gaddipati R, Phillips JA 3rd. Mutation in the gene for bone morphogenetic protein receptor II as a cause of primary pulmonary hypertension in a large kindred. *N Engl J Med*. 2001; 345:367–371. [PubMed: 11484696]
32. Oh P, Li Y, Yu J, Durr E, Krasinska KM, Carver LA, Testa JE, Schnitzer JE. Subtractive proteomic mapping of the endothelial surface in lung and solid tumors for tissue-specific therapy. *Nature*. 2004; 429:629–635. [PubMed: 15190345]
33. Pardo A, Gibson K, Cisneros J, Richards TJ, Yang Y, Becerril C, Yousem S, Herrera I, Ruiz V, Selman M, Kaminski N. Up-regulation and profibrotic role of osteopontin in human idiopathic pulmonary fibrosis. *PLoS Med*. 2005; 2:e251. [PubMed: 16128620]
34. Passerini AG, Polacek DC, Shi C, Francesco NM, Manduchi E, Grant GR, Pritchard WF, Powell S, Chang GY, Stoeckert CJ Jr, Davies PF. Coexisting proinflammatory and antioxidative endothelial transcription profiles in a disturbed flow region of the adult porcine aorta. *Proc Natl Acad Sci USA*. 2004; 101:2482–2487. [PubMed: 14983035]
35. Passerini AG, Shi C, Francesco NM, Chuan P, Manduchi E, Grant GR, Stoeckert CJ Jr, Karanian JW, Wray-Cahen D, Pritchard WF, Davies PF. Regional determinants of arterial endothelial phenotype dominate the impact of gender or short-term exposure to a high-fat diet. *Biochem Biophys Res Commun*. 2005; 332:142–148. [PubMed: 15896310]
36. Pittet JF, Griffiths MJ, Geiser T, Kaminski N, Dalton SL, Huang X, Brown LA, Gotwals PJ, Kotliansky VE, Matthay MA, Sheppard D. TGF-beta is a critical mediator of acute lung injury. *J Clin Invest*. 2001; 107:1537–1544. [PubMed: 11413161]
37. Sabo-Attwood T, Ramos-Nino M, Bond J, Butnor KJ, Heintz N, Gruber AD, Steele C, Taatjes DJ, Vacek P, Mossman BT. Gene expression profiles reveal increased mClca3 (Gob5) expression and mucin production in a murine model of asbestos-induced fibrogenesis. *Am J Pathol*. 2005; 167:1243–1256. [PubMed: 16251409]
38. Schwartz DA. The importance of gene-environment interactions and exposure assessment in understanding human diseases. *J Expo Sci Environ Epidemiol*. 2006; 16:474–476. [PubMed: 17109021]
39. Selman M, Pardo A, Barerra L, Estrada A, Watson SR, Wilson K, Aziz N, Kaminski N, Zlotnik A. Gene expression profiles distinguish idiopathic pulmonary fibrosis from hypersensitivity Pneumonitis. *Am J Respir Crit Care Med*. 2006; 173:188–198. [PubMed: 16166619]
40. Shapiro SD. Transgenic and gene-targeted mice as models for chronic obstructive pulmonary disease. *Eur Respir J*. 2007; 29:375–378. [PubMed: 17264324]
41. Studer S, Kaminski N. Towards systems biology of human pulmonary fibrosis. *Proc Am Thorac Soc*. 2007; 4:85–91. [PubMed: 17202296]
42. Villanueva J, Philip J, Chaparro CA, Li Y, Toledo-Crow R, DeNoyer L, Fleisher M, Robbins RJ, Tempst P. Correcting common errors in identifying cancer-specific serum peptide signatures. *J Proteome Res*. 2005; 4:1060–1072. [PubMed: 16083255]
43. Wesselkamper SC, McDowell SA, Medvedovic M, Dalton TP, Deshmukh HS, Sartor MA, Case LM, Henning LN, Borchers MT, Tomlinson CR, Prows DR, Leikauf GD. The role of metallothionein in the pathogenesis of acute lung injury. *Am J Respir Cell Mol Biol*. 2006; 34:73–82. [PubMed: 16166738]
44. Wurfel MM. Microarray-based analysis of ventilator-induced lung injury. *Proc Am Thorac Soc*. 2007; 4:77–84. [PubMed: 17202295]
45. Wurfel MM, Park WY, Radella F, Ruzinski J, Sandstrom A, Strout J, Bumgarner RE, Martin TR. Identification of high and low responders to lipopolysaccharide in normal subjects: an unbiased approach to identify modulators of innate immunity. *J Immunol*. 2005; 175:2570–2578. [PubMed: 16081831]
46. Xin Lu Jain VV, Finn PW, Perkins DL. Hubs in gene network exhibit low changes in expression in allergic responses in experimental asthma. *Mol Syst Biol*. 2007; 3:98. [PubMed: 17437023]
47. Yang IA, Savarimuthu S, Kim ST, Holloway JW, Bell SC, Fong KM. Gene-environmental interaction in asthma. *Curr Opin Allergy Clin Immunol*. 2007; 7:75–82. [PubMed: 17218815]
48. Zaas D, Schwartz DA. Genetics of environmental asthma. *Semin Respir Crit Care Med*. 2003; 24:185–196. [PubMed: 16088538]

49. Zhen G, Park SW, Nguyenvu LT, Rodriguez MW, Barbeau R, Paquet AC, Erle DJ. IL-13 and epidermal growth factor receptor have critical but distinct roles in epithelial cell mucin production. *Am J Respir Cell Mol Biol.* 2007; 36:244–253. [PubMed: 16980555]
50. Zhu S, Kachel DL, Martin WJ 2nd, Matalon S. Nitrated SP-A does not enhance adherence of *Pneumocystis carinii* to alveolar macrophages. *Am J Physiol Lung Cell Mol Physiol.* 1998; 275:L1031–L1039.