Aspen Lung Conference 2010

Systems Biology of Lung Diseases—Progress in the Omics Era

Dean Sheppard¹

¹Lung Biology Center, Division of Pulmonary, Critical Care, Allergy, and Sleep, Department of Medicine, University of California, San Francisco, San Francisco, California

The 53rd annual Thomas L. Petty Aspen Lung Conference focused on the dramatic progress that has been made in the past several years in applying large-scale, unbiased data acquisition (“omics”) to the study of lung biology and disease. The conference organizers, Mark Geraci, Ivor Douglas, Stephen Rennard, and David Schwartz, put together a terrific program, and the invited speakers and participants presented data describing the rapid evolution of experimental approaches that should encourage pulmonary scientists to begin to think about a true molecular systems biology of the lung.

Keywords: systems biology; genomics; proteomics

PULMONARY SYSTEMS BIOLOGY

“Systems biology” continues to mean different things to different scientists. The simplest operational definition is the process of integrating and quantifying information from multiple sources into predictive models. By this definition, pulmonary scientists have a long tradition of systems biology, dating back more than 100 years to the integration of anatomy, physics, and *in vivo* biology that has long characterized the study of pulmonary physiology. One recent, elegant example of a more sophisticated physical systems biology of the lung was the careful analysis by Ross Metzger and Mark Krasnow of airway branching patterns that led to a detailed functional map of three subroutines repeatedly used during the process of airway branching morphogenesis (1). Of course, a more grandiose view is that systems biology has as its ultimate goal the integration of everything into a single picture. Obviously, a unifying, quantitative model of how all of the cellular and molecular components of the lung interact to regulate lung health and disease is not likely to be forthcoming in the near future.

TOWARD A MOLECULAR SYSTEMS BIOLOGY OF THE LUNG

A critical first step in thinking about a molecular systems biology of the lung is the systematic identification of the molecular components that are present in health and disease. The development of high-throughput approaches to determine DNA sequences and messenger RNA (mRNA) abundance and to simultaneously analyze large numbers of proteins has facilitated impressive progress in this critical first step. Many of the talks at this meeting focused on progress made using these approaches.

GENETICS OF LUNG DISEASE

The tremendous progress in methods for rapid, inexpensive DNA sequencing has revolutionized the study of human genetics over the past decade. Invited talks by Kathleen Barnes and William Cookson outlined the evolution of studies of the genetics of asthma and other complex airway diseases from analysis of candidate genes to genome-wide linkage studies and, more recently, genome-wide association studies (GWAS). Both Dr. Barnes and Dr. Cookson described the most recent results of metaanalysis of pooled genome-wide association studies in more than 10,000 patients with asthma and a larger number of control subjects. One interesting feature of these results was the failure to identify any of the several dozen candidate genes previously reported to be associated with asthma at accepted levels of genome-wide significance. A small number of polymorphisms did reach genome-wide significance, but these are estimated to explain only approximately 4% of the estimated heritability of asthma. There was robust discussion of the possible explanations for the 96% unidentified by this large GWAS analysis, with possible systems biology explanations including gene-gene interactions, gene-environment interactions, and epigenetic factors. Of course, it remains possible that much of this heritability could be explained by a large number of rare sequence variants, which would not be expected to be detected by GWAS.

James Loyd discussed one of the poster children for the value of unbiased genetic studies: the example of how identification of the importance of genes in transforming growth factor β superfamily signaling pathways in familial primary pulmonary hypertension has opened the door to completely new insights into the biology underlying this group of disorders. Michaela Aldred discussed an interesting follow-up of observations from Akiko Hata's laboratory identifying a novel pathway by which Sma and Mad related protein (SMAD) proteins downstream of transforming growth factor β and bone morphogenetic protein (BMP) signaling interact with the Droscha complex to specifically enhance the processing of a subset of microRNAs (miRNAs) (2). These observations suggest a potential connection between defining mutations in specific BMP signaling proteins and a systemwide modulation of cell behavior mediated by miRNAs.

Christine Garcia discussed the dramatic genetic discoveries of mutations that affect protein folding of surfactant protein C and more recently A and in components of telomerase in families with pulmonary fibrosis. She pointed out how these findings led to discoveries of defects in these pathways in patients with sporadic pulmonary fibrosis and how these insights have fundamentally changed how we think about the underlying mechanisms and potential therapeutic interventions for idiopathic pulmonary fibrosis.

INSIGHTS FROM GLOBAL ANALYSES OF mRNA

The first description of use of microarrays to analyze global patterns of gene expression in the lung was published only a decade ago. Since that time there has been an explosion of

(Received in original form October 21, 2010; accepted in final form November 13, 2010)

Correspondence and requests for reprints should be addressed to Dean Sheppard, M.D., Lung Biology Center, University of California, San Francisco, Box 2922, San Francisco, CA 94143-2922. E-mail: dean.sheppard@ucsf.edu

Proc Am Thorac Soc Vol 8, pp 199–202, 2011

DOI: 10.1513/pats.201010-065MS

Internet address: www.atsjournals.org

papers using this approach. A simple Medline search using the search terms “expression microarray” and “lung” yielded a list of 1,400 papers, which I am sure is a gross underestimate of the numbers of published papers using this approach in the past decade. These studies have led to a number of stunning new insights into the molecular mechanisms underlying normal lung development and virtually every significant lung disease. Access to expression microarrays has become so widespread that semiquantitative exploration of changes in the concentration of individual mRNAs, for example by Northern blotting or standard polymerase chain reaction, is rapidly becoming a scientific historical relic. It is thus not surprising that there were a number of talks about new biological insights based on microarray analysis. Joseph Nevins and Avrum Spira discussed new bioinformatic approaches that can generate reliable fingerprints of multiple signaling pathways and allow inferences, at a systems level, of the pathways that might have been activated in cells and tissues prior to sampling. Both speakers described how this approach can be used to classify patients with heterogeneous diseases, such as lung cancer, and could be used to improve and individualize treatment of malignancies, including lung cancer.

On first analysis, this vast and rapidly accumulating data set should provide the pulmonary community with wonderful opportunities to approach lung development and disease at a systems level. Examples include the ability to identify coordinate changes in groups of functionally related genes, coordinate changes in multiple genes within a well-established signaling pathway, and relationships between DNA sequence variants and coordination of expression of multiple downstream genes. Efforts of pulmonary scientists to take advantage of this potential have also highlighted the limitations of tools currently available. Analysis of functional groups depends on accurately compiling such groups, and the precision and completeness of current databases have serious limitations. Similarly, analysis of coordinate changes within a signaling pathway or network depends on accurate maps of these pathways and networks, another area that is very much a work in progress. Open source sites where investigators can build and post their own versions of such networks have been helpful, but also chaotic, and it is difficult to establish accuracy and quality control in such a rapidly evolving area.

APPROACHES TO DEAL WITH CELLULAR COMPLEXITY AND DISTINCT MICROENVIRONMENTS IN THE LUNG

Many of the published studies of pulmonary microarray analysis have been performed on RNA from whole lung tissue. These analyses have been surprisingly informative, but can obviously be confounded by changes in cellular composition rather than changes in gene expression in specific cell types. Analysis of whole lung gene expression can also be insensitive to dramatic changes in gene expression in relatively rare cellular populations and in specific lung microenvironments (e.g., within the conducting airway epithelium or at bronchoalveolar duct junctions). Several groups have attempted to get around these problems using laser microdissection, with some success. In humans, the airway epithelial microenvironment can be sampled by brushing, an approach that was highlighted in talks by Prescott Woodruff and Avrum Spira. We have recently adapted this approach to the mouse. In mice, the development of an increasing array of cell-specific promoters should allow investigators to sort individual cell types from mice expressing cell-specific fluorescent reporters. Doug Kuperman and David Erle described a clever variation on this approach by genetically restricting signaling to the cytokine IL-13 to cells within the

conducting airways (3). With this approach they were able to identify a small number of genes that are specifically induced *in vivo* in epithelial cells in response to IL-13 (4). Many of the genes have subsequently been shown to be functionally important in models of allergic asthma.

GLOBAL ANALYSES OF mRNA ABUNDANCE HAVE BEEN A MAJOR BOON TO REDUCTIONIST BIOLOGY

Somewhat paradoxically, the broad availability of tools for global analysis of gene expression and large data sets of gene expression in health and disease has been even more helpful to classic reductionist biologists than to emerging systems biologists. For example, a classic reductionist approach is to knock-down or overexpress a single gene of interest to explore the roles of that gene in altering cell or *in vivo* functions and to identify the simplest linear pathways involving the gene product of interest. Many important functional effects either rapidly or eventually result in significant changes in gene expression. In the dark ages of the 20th century, investigators had to select and painstakingly evaluate such effects one gene at a time, a slow process that was generally limited by what was already known about a pathway of interest. Thus, even single unknown steps in a linear pathway often went unidentified. In contrast, now that we can simultaneously identify changes in expression of every gene it has been possible to rapidly fill in some of these missing gaps. In this context, the more reductionists already know about the details of the pathway they have been studying the easier it is to use global gene expression data to infer the nature of the missing pieces. Such successes in reductionist biology are obviously of great benefit to systems biologists, who will have increasingly accurate pathways and networks to include in evolving bioinformatic analysis tools. Skip Garcia's talk described the identification of several genes that play critical roles in the development of acute lung injury. His talk demonstrated the value of combining global analyses of gene expression, targeted analysis of DNA sequence, and detailed information (from previous reductionist approaches from the same laboratory) about pathways controlling relevant biology (in this case actin reorganization and endothelial cell contraction).

THE RISE OF miRNAs

The past few years have seen a meteoric rise in the appreciation of important roles for small noncoding RNAs in regulating broad patterns of gene expression relevant to development and disease. MiRNAs are particularly attractive to nascent systems biologists because of their roles in quantitatively regulating large numbers of targets that operate cooperatively in networks. In Joe Loscalzo's beautiful talk on systems approaches to biology and disease he pointed out the role that miRNAs can play as supernodes in scale-free networks. Talks by Avi Spira, Prescott Woodruff, Michaela Aldred, and Jadranka Milosevic described potential roles for miRNAs in multiple lung diseases. Although no talks at this meeting discussed the importance of the miRNAs that circulate in the blood in high concentration in exosomes, it is anticipated that reports about circulating miRNAs will emerge soon. These miRNAs might be especially useful as informative biomarkers but could also provide insight into disease pathogenesis once more is known about their cellular origins and biological effects.

Many small RNAs modulate protein concentrations through interactions with the 3' untranslated regions (UTRs) of mRNAs. 3' UTRs are also the targets for a number of RNA-binding proteins that can dramatically modify mRNA stability. David Erle outlined an exciting new method to systematically identify

critical regions of 3' UTRs and to ultimately accurately determine the rules governing interactions of these regions with multiple pathways for regulating mRNA stability and translation.

Although microarrays are currently the method of choice to analyze gene expression at a near genome-wide scale, Avi Spira described dramatic results demonstrating that next generation sequencing can provide much deeper coverage, identifying many low-abundance species with differential expression undetectable by standard microarrays. Once the cost, accessibility, and analytic tools for this approach are improved, next-generation sequencing seems likely to replace expression microarrays and provide a series of new insights into the functions of noncoding RNAs.

ANALYSIS OF PROTEIN EXPRESSION, STRUCTURE, AND FUNCTION

Rapid success over the past decade in comprehensively evaluating DNA sequence and RNA expression has whetted the scientific appetite for a similar comprehensive approach to analyzing proteins. Although enormous progress has been made, whole proteome analysis has proved to be a harder nut to crack. One important problem is that none of the currently available methods for detecting proteins approaches the sensitivity of methods for analyzing RNA. Bead-based multiplexing assays and arrays of monoclonal antibodies are improvements over Western blotting and individual ELISA assays, but antibodies are not yet available for even most human proteins. Mass spectrometry has made major advances but is still limited in sampling low-abundance proteins in complex pools, and quantitative comparisons remain challenging in many experimental circumstances. Pierre Massion discussed the technical underpinnings, strengths, and limitations of current protein detection methods. He showed how shotgun proteomics could identify a large number of proteins undetectable by matrix-assisted laser desorption/ionization time of flight spectrometry (MALDI-TOF) MALDI-TOF. One exciting development discussed in two talks at this meeting was the development of DNA aptamers as protein detection tools. If this approach can be extended to a larger fraction of the proteome it might make simultaneous, quantitative detection of proteins more rapid and accessible.

It should be immediately apparent that even precise and comprehensive quantitation of protein expression is only the beginning of what will be needed for a systems approach to protein function. Protein function is largely determined not by amino acid sequence but by structure, and structure is in turn dramatically affected by alternative splice variants and by post-translational modification, many of which are quite transient. Thus, many of the most important advances in systems biology over the next decade are likely to depend on improved methods to rapidly solve the structures of most proteins, to comprehensively analyze a wide range of protein modifications (e.g., phosphorylation, acetylation, glycosylation, sulfation, ubiquitination, sumoylation, and lipidation).

PROTEIN-PROTEIN INTERACTIONS

Proteins generally exert their biological effects by directly contacting other proteins and either enzymatically modifying these targets or otherwise inducing changes in their conformation. Emerging systems biology is thus highly dependent on accurate and comprehensive descriptions of protein interaction networks, and building databases of protein interaction networks has thus been a major focus of many systems biologists. The most stunning successes have occurred in unicellular organisms, such as budding yeast, including a recent comprehensive mass spectrometry analysis of more than 1,000 yeast proteins

that form complexes with kinases and/or phosphatases (5). William Balch discussed a beautiful example of systems-level analysis of the pathways that regulate protein folding, the maintenance and proper targeting of folded proteins, and the appropriate removal of misfolded proteins, a process he calls proteostasis. He described how this analysis has informed the identification of novel treatments to restore molecular and cellular homeostasis in pulmonary disorders of proteostasis, including cystic fibrosis and α_1 -antitrypsin deficiency. For multicellular organisms, and multicellular organs like the lung, the enormous variability of such networks in cells differentiated to perform unique tasks will no doubt keep many of us busy for the next several years. As with analysis of gene expression, progress in this area is likely to be highly dependent on both improved methods for systematic evaluations of protein interactions and one-by-one interactions identified and/or confirmed through reductionist approaches.

INTEGRATION OF GLOBAL RNA AND PROTEIN ANALYSIS INTO DIAGNOSIS AND TREATMENT

For scientists interested in disease-related biology, a major promise of systems biology is the development of tools to improve diagnostic acumen, better characterize clinical complexity, and improve treatment outcomes with reduced toxicity through personalized medicine. Beautiful talks by Joseph Nemins and Joseph Loscalzo highlighted the promise of such approaches for cancer and cardiopulmonary disease and helped to lay out a roadmap for how improvements in systems biology could ultimately revolutionize the practice of medicine.

TOWARD A UNIFIED MOLECULAR SYSTEMS BIOLOGY OF THE LUNG

This meeting highlighted the substantial progress that pulmonary scientists have made over the past decade in systematically evaluating the genetic bases for lung disease, in the expression patterns of mRNA and noncoding RNAs in lung development and disease, and in developing the tools to evaluate protein expression in lung cancer, normal and diseased lung tissue, and bronchoalveolar lavage fluid. This rapidly expanding database has already provided reductionist pulmonary biologists with a number of novel mechanistic insights that would have been inconceivable in the absence of these new approaches. Many of these insights have already led to the identification and partial validation of new targets for the treatment of currently untreatable or difficult-to-treat diseases, such as pulmonary fibrosis, acute lung injury, pulmonary hypertension, and lung cancer.

The past decade has also produced stunning achievements in refining a physical systems biology of the lung, perhaps best exemplified by a new detailed model of the steps regulating lung branching morphogenesis. The next decade will very likely see even more rapid progress in developing a molecular atlas of lung development and disease, including new insights into the roles of epigenetics and of noncoding RNAs as master regulators of disease-related programs. We can also expect comprehensive data about protein splicing, post-translational modification, protein structure, and protein-protein interactions in lung biology and disease.

The challenge of the next decade will be to extend this excellent start toward developing more comprehensive insights into beginning to develop a true molecular systems biology of the lung. In addition to identifying all of the component parts, we will need to know more about the structure of each component, how the components are physically connected (physical networks), and how they act on one another (functional net-

works). We will also need to develop methods to determine the precise quantitative relationships among components and how the components and their interactions change in real time. Such progress will depend on the continued use of reductionist approaches to accurately determine the actual interactions between components and how they differ with cell type, differentiation state, and local microenvironments. Thus progress in pulmonary biology in the “omics era” will continue to depend on synergistic and iterative interactions between systems thinking and reductionist experimentation, requiring all of us to think globally and act locally.

Author Disclosure: D.S. was a consultant, received grant support, and owns a patent through Stromedix. He receives royalties from Chemicon and he received grant support from the NIH.

References

1. Metzger RJ, Klein OD, Martin GR, Krasnow MA. The branching programme of mouse lung development. *Nature* 2008;453:745–750.
2. Davis BN, Hilyard AC, Lagna G, Hata A. SMAD proteins control DROSHA-mediated microRNA maturation. *Nature* 2008;454:56–61.
3. Kuperman DA, Huang X, Koth LL, Chang GH, Dolganov GM, Zhu Z, Elias JA, Sheppard D, Erle DJ. Direct effects of interleukin-13 on epithelial cells cause airway hyperreactivity and mucus overproduction in asthma. *Nat Med* 2002;8:885–889.
4. Kuperman DA, Lewis CC, Woodruff PG, Rodriguez MW, Yang YH, Dolganov GM, Fahy JV, Erle DJ. Dissecting asthma using focused transgenic modeling and functional genomics. *J Allergy Clin Immunol* 2005;116:305–311.
5. Breitkreutz A, Choi H, Sharom JR, Boucher L, Neduva V, Larsen B, Lin ZY, Breitkreutz BJ, Stark C, Liu G, *et al.* A global protein kinase and phosphatase interaction network in yeast. *Science* 2010;328:1043–1046.