Systems biology approaches to identify developmental bases for lung diseases

Soumyaroop Bhattacharya[1] and **Thomas J. Mariani**[1]

[1]Division of Neonatology and Center for Pediatric Biomedical Research, University of Rochester Medical Center, 601 Elmwood Ave, Box 850, Rochester, NY 14642

Abstract

A greater understanding of the regulatory processes contributing to lung development could be helpful to identify strategies to ameliorate morbidity and mortality in premature infants and to identify individuals at risk for congenital and/or chronic lung diseases. Over the past decade, genomics technologies have enabled the production of rich gene expression databases providing information for all genes across developmental time or in diseased tissue. These data sets facilitate systems biology approaches for identifying underlying biological modules and programs contributing to the complex processes of normal development, and those that may be associated with disease states. The next decade will undoubtedly see rapid and significant advances in redefining both lung development and disease at the systems level.

Introduction

There is a wealth of biological information available for chronic diseases of the lung (1). However, a critical gap in knowledge that may create challenges in development of therapeutic modalities for lung diseases, is the lack of complete understanding the development mechanisms of the normal pulmonary system, and how alterations therein, contribute to disease pathophysiology. It is becoming increasingly clear that many respiratory diseases have their origin early in life and are influenced by developmental, as well as genetic and environmental factors. A thorough understanding of lung development is needed to understand how the environmental and genetic factors affect the process.

Lung morphogenesis occurs both prenatally and postnatally, and is typically divided into five phases, with the final alveolar phase occurring principally after birth in humans and rodents. The initiation of lung formation as a bud off the lateral foregut endoderm ('embryonic' stage) occurs from 26 days post conception to up to 5 weeks of gestation, and the corresponding duration in mouse is through embryonic day 9.5. Expansion of the bronchial tree, including formation of the bronchi and bronchioles, occurs during the pseudoglandular stage that runs from 5 to 16 weeks of gestation in humans, with corresponding time period in mouse being embryonic days 14.5–16.5. The distal region of

Corresponding author: Thomas J. Mariani, PhD, Division of Neonatology and Center for Pediatric Biomedical Research, University of Rochester Medical Center, 601 Elmwood Ave, Box 850, Rochester, NY 14642.

the lung where gas-exchange will ultimately occur expands exponentially during the canalicular period which spans from 16th to 26th week of gestation in humans, and during embryonic days 16.5–17.5 in mice. In the saccular stage, from 26th to 36th week in humans (and embryonic day 17.5 to post natal day 5 in mice), the air spaces in the respiratory portion of the lung mature to include include surfactant- (type II) and non-surfactant producing (type I) pneumocytes. The final stage of lung development, termed alveolarization or alveogenesis, begins prior to birth in humans and extends through at least the first decade of life, while occurring entirely postnatally in mice. In this stage, there is a vast expansion of the surface area of the lung, such that the adult human lung has roughly the same surface area as a tennis court, and a substantial reorganization of the capillaries to facilitate gas exchange.

The lung is a complex three-dimensional organ whose functions depend upon the formation and maintenance of dynamic interactions between multiple cell and tissue systems. These include the highly branched system of airway tubes and terminal alveolar sacs, a complex hierarchy of respiratory and non-respiratory epithelial cells that lines these tubes and sacs, blood and lymphatic vessels, nerves, smooth muscle cells and fibroblasts, and cells of the immune system. Defects, not only in individual components, but in interactions among them, lead to significant respiratory disorders that affect neonates, infants, juveniles and adults.

Systems biology has been defined as the study of the interactions between the components of biological systems, and how these interactions give rise to the function and behavior of that system(2). However, in practical terms, systems biology still means different things to different people. It can be interpreted as the ability to obtain, integrate and analyze complex data sets using interdisciplinary tools, from multiple platforms namely, genomics, epigenetics, transcriptomics, proteomics, metabolomics among others. The advantage it has over "classical" or traditional approaches is that it considers the organ as a whole, and involves modeling of the entire system through integration of its various components. Once individual components accurately mimic the responses to specific stimuli, they can be integrated into a system that can be used to understand organ physiology including development, disease progression, therapeutic interventions, and to predict the molecular responses to biological perturbations (3). This is an ideal approach for obtaining a greater understanding of the complexity of the lung, its formation and development (Figure 1).

The application of comprehensive and unbiased functional genomics methods, complementing focused approaches that draw upon decades of research, provide a more integrated approach to better understand developmental processes and regulatory networks. Over the last few years, there has been a great expansion of genomic data that has facilitated redefining lung development at the molecular level, provided a better understanding of the mechanisms of lung pathophysiology, and identified putative markers for early disease detection, diagnosis, and treatment. However, when compared to other disciplines, there has been a lag in implementing systems biology approaches in the pulmonary field, and it is more so evident in the area of lung development research. At a recent NIH workshop, participants have recommended, among other things, developing strategies for integration of systems biology approaches to decipher the mechanisms of developmental origins of lung

diseases (4). An expansion of systems-level approaches to integrate multiple levels of molecular and functional information has great potential for discovery of underlying mechanisms occurring during the process of lung development. Since "omics" technologies are completely unsupervised, they have the potential to discover new and unsuspected links between processes and pathways during development.

Global Gene Expression of Lung Development

A greater understanding of the regulatory processes contributing to lung development could help ameliorate morbidity and mortality in premature infants, and identify individuals at risk for congenital and/or chronic lung diseases. The development of high-throughput approaches to determine DNA sequences and mRNA abundance, and to simultaneously analyze large numbers of proteins, has facilitated impressive progress in this direction. Genomics technologies also have provided rich gene expression databases containing information for specific genes across development (Table 1).

Microarrays

The most commonly applied approach for high-throughput genome-wide technologies is gene expression microarrays. For over a decade now, microarrays have been proved useful in gaining novel insights for many human diseases. Microarray technology has revolutionized the search for disease biomarkers by simultaneous comparison of expression changes of thousands of genes. This has led to an increasing number data sets being deposited in public databases, such as Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/), Array Express (http://www.ebi.ac.uk/arrayexpress/), Stanford Microarray Database (SMD; http://smd.stanford.edu/), among others. These data sets enable systems biology approaches for identifying biological modules contributing to complex processes. The rapid emergence of microarray technology in combination with information about sequences and function of genes provides a wealth of data in both human and animal models.

In the field of lung biology, most of the gene expression datasets have been two-class (e.g. case-control or treatment-control) comparison studies looking for markers of lung diseases. However, the application of expression profiling methods to study lung development has lagged behind. The first applications of high throughput analysis methods to study global gene expression of the developing lung were published at the beginning of the new millennium. Golpon and colleagues in 2001 looked at the expression of Hox genes in normal and diseased (emphysema and pulmonary hypertension) human adult lungs and fetal (at 12 weeks of gestation) and adult mouse lung tissue using Affymetrix human and murine microarrays (5). Using basic analytical approaches they characterized differences in the pattern of HOX gene expression among fetal, adult, and diseased lung specimens. Similarly, Kaplan used cDNA arrays to identify developmentally regulated genes in the lungs of wild-type and transgenic mice with targeted hypomorphic disruption in the Glucocorticoid Receptor (GRhypo) gene at embryonic day 18 and postnatal day 1(6). They identified changes in expression of 31 genes in GRhypo mice when compared to wild-type mice. Lin & Shannon used oligonucleotide arrays to compare the expression of 38,018 known genes and ESTs in the embryonic mouse (E13.5) lung and trachea (7). They identified 204 genes,

including novel and known lung-specific genes, as differentially expressed in lung tissue. One novel gene, melanoma inhibitory activity (MIA), was suggested to be a potential marker of lung epithelial differentiation. Liu and Hogan carried out genome-wide expression profiling in an effort to understand mechanisms of branching morphogenesis. This study compared gene expression in epithelial tissue from the tips of branching tubes, to those of more proximal region, in embryonic day 11.5 lung buds, identifying 20 genes as differentially expressed (8).

In all these studies, expression profiling served only as a supporting component of a larger study, focusing on only one individual time point. None of these studies had generated a comprehensive microarray data spanning across stages of development. Our group undertook a large-scale gene expression analysis of murine lung development in which we included all stages of lung development beginning at embryonic day 12 and continuing to adulthood (9). We applied various clustering approaches to segregate genes into groups in accordance their developmental expression patterns. We identified genes encoding regulatory proteins with highly correlated expression patterns to those of extracellular matrix genes. This analysis revealed previously unknown associations among the expression patterns of genes that may have functional significance in this complex process. In a similar study, Bonner and colleagues used oligonucleotide microarrays to study developmental expression patterns across different lung development stages, and showed various patterns associated with lung development and the temporal regulation of key regulatory pathways (10). They analyzed RNA from four samples at each stage of lung development using Affymetrix U74Av2 microarrays representing over 12,000 genes and ESTs. They identified 1346 genes and ESTs as significantly different in at least one stage. Lu and colleagues used microarrays to characterize the transcriptomic profiles of proximal and distal regions of the mouse respiratory tract at embryonic day 11.5, when branching morphogenesis is initiating, to gain insights into the pathways that potentially distinguish proximal non-branching from the distal branching region. This study identified 83 genes up-regulated in branching region while 128 up-regulated in non-branching regions (11).

One of the major limitations of all these studies exists in their experimental designs, as most suffer from limited samples size or lack of replicates, which hampers the ability to accurately identify true changes. Another significant limitation has been the lack of corollary spatial information within the lung. Spatial information is essential for hypothesis development as localized gene expression is critical in development, morphogenesis and differentiation. As most of these studies have used lung tissues, and not isolated cell types, and the lung being a complex organ with a heterogeneous mixture of cell types, they may not capture cell specific data. The complex anatomy of the lung generally makes identifying cell type specific gene expression changes difficult and hence requires additional methods such as immunohistochemistry or laser capture microdissection. An interesting, possible exception is Okubo and Hogan, who examined differential gene expression from RNA isolated from the caudal lobe (endoderm and mesoderm) of SftpC-CatCLef1 transgenic and wild-type embryo (e18.5) lungs using the Affymetrix MOE430 microarray (12). Statistical analysis of the array data identified a mixture of cells expressing marker genes characteristic of different cell types indicating a shift in cell lineage commitments. Another exception is the study by O'Reilly and colleagues, who used genome-wide expression analysis a

genetically-labeled sub-set of type II pneumocytes to identify novel epithelial mechanisms of innate immune responses to respiratory viral infection (13).

RNA-Seq

With the advent of RNA-Seq technology, a massive parallel sequencing approach to measure gene expression at the RNA level, researchers are now capable of building on microarray data to provide additional insights into the transcriptome of normal and disease processes occurring in the lung. Hackett and colleagues have recently used RNA-Seq to characterize the transcriptome of small airway epithelial cells. They observed that in addition to genes previously known to be expressed by Clara cells, which are markers of the predominant airway epithelial cell type, genes characteristic of neuroendocrine cells were highly expressed as well (14).

miRNA

In addition to studies investigating changes in expression at the messenger RNA (mRNA) level, changes in expression of microRNAs (miRNA), protein and their metabolites have been studied to a lesser extent. miRNAs are a family of small noncoding RNAs (21–25 nucleotides in length) found in almost all mammalian genomes. miRNA registry databases such as the Sanger Institute miRBase (http://www.mirbase.org/) contain annotations for all published miRNAs that were either experimentally validated for mature miRNA expression or computationally predicted for the corresponding hairpin structures (15). microRNAs play an important roles in cell proliferation, differentiation, tumorigenesis and organ development. miRNAs are estimated to be responsible for regulating the expression of at least a quarter of the human genome. Very few miRNA studies have assessed miRNA expression during lung development. Bhaskaran and colleagues used a custom miRNA microarray platform to profile the expression of miRNAs at different stages in rat lung development and identified 21 miRNAs that were significantly changed during this process (16). Yang and colleagues used a microarray that covered probes for more than 1891 miRNAs to profile expression at three separate time points during rat lung development at embryonic days 16, 19, and 21. They identified 167 miRNAs as differentially expressed (including 81 upregulated and 86 downregulated) during rat lung development (17). Dong and colleagues used a cross-platform approach to study the regulation of miRNAs in mouse lung organogenesis (18). They used both miRNA and mRNA expression profiling across all recognized stages of lung development beginning at embryonic day 12 and continuing to adulthood. They analyzed the expression patterns of dynamically regulated miRNAs and mRNAs and further correlated those with protein levels from an existing mass-spectrometry derived protein database for lung development.

Proteomics

There are very few studies that have assessed genome-wide expression at the protein or metabolomics level. Cox and colleagues studied protein expression during mouse lung organogenesis from embryonic day 13.5 until adulthood using gel-free two-dimensional liquid chromatography coupled to shotgun tandem mass spectrometry (MudPIT)(19). They correlated protein expression patters with gene expression profiles obtained from a previously published microarray dataset (9). Computational modeling of the proteomic

profiles in conjunction with DNA microarray data identified groups of genes with statistically significant correlation in expression levels of proteins and transcripts during lung development.

Phosphoproteomics is a branch of proteomics that identifies and characterizes proteins containing a phosphate group as a post-translational modification. Compared to expression analysis, phosphoproteomics provides information on change in phosphorylation status, which reflects a change in protein activity. While phosphoproteomics expands the current knowledge about the numbers and types of phosphoproteins, its greatest promise is the rapid analysis of entire phosphorylation based signaling networks (20). Application of phosphoproteomics in the pulmonary field is still in the developing phase, and has primarily been focused on in-vitro studies of lung cancer (21–23). Giorgianni and colleagues have recently generated the phosphoproteomic profile of human Bronchoalveolar Lavage (BAL) fluid which characterized phosphorylation of several proteins involved in lung function and disease mechanisms (24).

Metabolomics

Metabolomics is a global approach to understanding regulation of metabolic pathways and networks of a biologic system (25). Feihn and colleagues used GC-TOF chromatograms on lungs of smoke-exposed pregnant rats and their fetuses to study altered metabolic phenotypes in developing lungs affected by cigarette smoke (26).

Using multivariate statistics they identified 46 metabolites that were differentially regulated in fetal lungs that were exposed prenatally to environmental tobacco smoke, indicating alterations in metabolic phenotypes of developing lungs due to cigarette smoke. Fetal lungs showed major down regulation for free fatty acids, but only a few up-regulations of metabolites like ketone bodies and sugar phosphates.

High throughput transcriptome profiling technologies have enabled recent comprehensive studies of the genome–wide expression patterns and global biological processes underlying organogenesis in animal models. However, corresponding developmental studies in human are generally lacking. We have recently generated genome-wide expression profiling from human fetal lung tissue specimens to identify global transcriptomic features of the developing human lung (27). Our dataset comprised of expression profiles from 38 samples, representing 29 distinct time-points, spanning the pseudoglandular and early canalicular stages of human lung development, represents an encyclopedia of gene expression data. Analysis of this brief developmental time interval using principal component analysis (PCA) identified a set of 3,223 characteristic genes contributing to the changes in the developing lung transcriptome, including both known and novel markers, capable of defining the features of human lung development.

Development as Predictor of Disease

The 'developmental origins of adult disease' hypothesis, often referred to as the 'Barker hypothesis' named after its leading proponent David Barker, states that adverse influences early in development, and particularly during intrauterine life, can result in long-term

changes in physiology and metabolism, which result in increased disease risk in adulthood (28). Current genomic data supports the concept that developmentally predominant genes and pathways are commonly associated with disease pathogenesis.

The idea of shared molecular mechanisms between tumorigenesis and organogenesis has been discussed since the latter part of the last century (29). Kho and colleagues looked at genome-wide expression data generated from human cerebellar brain tumors (medulloblastomas), and normal mouse cerebellar tissues collected from postnatal days 1–60 (30). They used Principal Component Analysis (PCA) to project the profiles of human medulloblastomas onto a normal mouse cerebellar development temporal series. They found that the human medulloblastomas had genomic profiles most similar to early stage mouse cerebella, and normal human cerebella were more similar to adult mouse cerebella. This approach proved informative in the pulmonary system as well. Liu and colleagues compared genome-wide expression in lung cancer samples to gene expression in normal mice during development, and found similarities between expression patterns in the lung cancer subtypes and the developing mouse lung (31). They observed that cancer prognosis in humans was correlated to lung maturity in mouse. When expression in the human cancer was more similar to mature mouse lung cells, the prognosis was better, and when it was similar to that of very immature mouse lung cells, prognosis was poor.

One of the initial applications of "systems biology" was integrated genomics, such as combining genome-wide expression with genetic association data. We have reported the identification of multiple COPD susceptibility genes through the integration of human genetics with gene expression profiling of normal lung development and diseased lung tissue (32). Specifically, we used an integrative approach involving genetic linkage data, genome-wide association data, and gene expression profiling data to identify Serine Protease Inhibitor E2 (SERPINE2) as a novel candidate susceptibility gene for COPD (33) (see Figure 2). SERPINE2, an inhibitor of thrombin and plasmin, was known to promote extracellular matrix production and inhibit apoptosis; however its role in pulmonary system had not been previously explored. Analysis of microarray data exploring gene expression changes during embryonic lung development (9, 10) showed SERPINE2 had highest expression during airspace morphogenesis. SERPINE2 is located within a region of chromosome 2 that had previously been defined as a linkage region for early-onset COPD. Analysis of genome-wide association data from a family-based population identified multiple SERPINE2 polymorphisms to be associated with COPD, an observation that was further replicated in a case-control population (34). Using genome-wide expression profile data from two independent populations (32, 35), we observed that SERPINE2 expression was significantly correlated with measures of pulmonary function in human COPD patients (36). We subsequently applied this approach of integrated genomics to identify IREB2 (37), SOX5 (38) and FGF7 (39) as additional COPD susceptibility genes. Further, we have looked at the developmental expression patterns of expression of COPD marker genes (32) and found 27 of those to be changing during early embryonic development (27), suggesting this disease process involves dysregulation of developmental pathways.

Other studies have used genome-wide expression data from both human and murine lung development in combination with genetic association data to identify asthma susceptibility

genes. Although there was no significant over-representation of the asthma genes among genes differentially expressed during lung development, differential expression in more than 10 asthma candidate genes during development was observed (40). Similarly, *Wnt* signaling genes that were differentially expressed during fetal lung development were associated also with impaired lung function in asthmatic children (41).

We have recently studied genome-wide expression changes in lungs of patients with Bronchopulmonary Dysplasia (BPD), a chronic lung disease of the newborn with a strong developmental contribution. We identified genes and pathways that are involved in disease pathogenesis (42). One of the pathways found to be associated with gene expression changes in BPD is sonic hedgehog signaling, which has previously been linked to both lung development and COPD, a chronic lung disease of the aged (32, 43). In all, thirty-one of 159 genes identified as dysregulated (~20%) could be linked through a single network with IGF1 as the central node (Figure 3A). However, when BPD gene expression analysis was limited to genes also involved in early embryonic lung development as identified by expression profiling (27), we observed that CDK1 became the central node, instead of IGF1 (Figure 3B). These data demonstrate that consideration of developmental processes can have a significant impact on disease gene discovery, both helping to identify novel genes and pathways, and modifying interpretations of disease-associated pathophysiological processes. These studies further indicate that regulatory mechanisms governing development of the mammalian lung consist of a complex set of discrete, yet overlapping pathways, that when altered, may potentially lead to the onset of chronic diseases either perinatally (BPD) or in the aged (COPD).

'Hubs and Spokes' model of Genomes and Systems – State of Science

Tremendous efforts, primarily focused over the past two decades, have led to the identification of numerous critical transcription factors, secreted growth factors and their receptors that play an essential role in proper lung formation. However, nearly all studies to date have involved linear dissection of a single molecule or pathway. A full understanding of lung development and its pathophysiological perturbations will require integration of the complex mechanisms driving morphogenesis and cellular differentiation.

In an early application of systems biology to understand the mechanisms of asthma, Novershtern and colleagues compiled a gene expression database from five publicly available mouse microarray datasets consisting of 4,305 gene sets (44). Using this collection of genome-wide expression datasets, they generated a network of functional groups for asthma, dominated primarily by immune response classes. Whitsett and Matsuzaki used a combination of mouse genetics and functional genomics to integrate individual molecules involved in cellular differentiation and surfactant production into a 'circuit' necessary to prepare the lung for the 'transition to air breathing' (45). Similarly, Xu and colleagues employed a systems biology approach to generate a transcriptional network describing regulation of surfactant homeostasis in the lung. Instead of focusing on individual genes, they identified gene sets as regulatory hubs in networks of TFs (46). Even though this approach will not identify epigenetic, post-transcriptional and gene-environmental interactions critical to gene regulation, it provides a systematic view and working model of a

transcriptional network regulating the formation and metabolism of the pulmonary surfactant system.

Classical descriptions of mammalian lung development have focused upon the transition through histo-morphological stages. Although these stages, and their morphological correlates, are highly conserved across species, significant differences exist in their relative length and timing. As an example, birth occurs in the saccular stage in rodents, but in the alveolar stage in humans. As the rodent lung is not biochemically immature at birth (e.g., with respect to surfactant), this example highlights the differences in molecular and histological development. In fact, it has long been appreciated that discrete molecular processes occurring during lung development, such as branching morphogenesis or respiratory epithelial cell differentiation, are not bounded by histological stages. Therefore, it is rational to anticipate that lung development could be defined in terms of discrete molecular transitions analogous to histological stages (see Figure 4).

In previous work, we used unsupervised principal component analysis of genome–wide expression data to identify global transcriptomic features of mouse lung development (47). Gene expression variation was found to be associated with macroscopic biological features such as age and alveolar formation. In particular, we observed an overlying biological program, which accounted for much of the genome-wide variation in expression, which defined the distance in age of the lung from the day of birth. We termed this program the "time-to-birth" signature. We also identified groups of genes with expression patterns corresponding to the time-to-birth molecular signature. These analyses suggested the possibility of characterizing lung development in molecular, in addition to histological terms. We have also subsequently applied the same approach to genome-wide expression patterns of developing human lung from samples spanning the pseudoglandular and canalicular stages (estimated 53–154 days post conception, dpc) (27). We observed that global shifts in gene expression (e.g., molecular phases) during human lung development sometimes parallel histologically–defined stages. However, as in the mouse, we identified novel, distinct molecular phases that did not correlate with histological stages. We conclude that molecular phases of lung development may correlate with appreciated histo-morphological stages, novel sub-stages or may cross boundaries of stages. These phases, in turn, are composed of overlapping (rather than unique) sets of genes (see Figure 4). Reconsidering normal developmental processes at this level of resolution should improve our understanding of essential molecular events during both normal development, and pathological derangements of the lung, and may provide further insights into critical windows of development wherein environmental exposures may lead to subsequent disease.

Summary

A greater understanding of the regulatory pathways controlling lung development are essential for attempts to identify individuals at increased risk for chronic lung disease (in the newborn, juvenile or adult periods), to better define pathogenic mechanisms of disease, and to identify targets for therapeutic intervention. Likewise, attempts to identify lung disease biomarkers will benefit from a greater knowledge of physiological context, including dynamic gene expression levels during normal states, such as organ development. It is

rational to hypothesize that a majority of disease-related genes and pathways are congruent with a physiological set that also contributes to organ formation. Current data supports the concept that developmentally predominant genes and pathways are commonly associated with disease pathogenesis. In order to gain a deeper understanding of the functional and regulatory pathways that play critical role during complex mechanisms of pulmonary organogenesis, integrative systems biology approaches can be applied to combine experimental techniques with genome-level information, and computational methods (modeling and simulation) (48). Given the advances and availability of high-throughput technologies to define biological processes in incredible detail, significant advances in our understanding of lung development are feasible and achievable. Systems biology approaches promise to provide a much more complete map of the functioning and interaction of the signaling networks and groups of co-expressed genes are involved in lung formation, which may give insights to explain critical aspects of disease pathogenesis. We have presented here a snapshot of applications of different systems level approaches in exploring molecular mechanisms of lung development. Similar approaches are already being applied in pulmonary medicine to discover the bases of complex lung diseases and to overcome the limitations faced in development diagnostic markers and therapeutic targets (49).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Bhattacharya S, Mariani TJ. Array of hope: expression profiling identifies disease biomarkers and mechanism. Biochem Soc Trans. 2009; 37:855–862. [PubMed: 19614607]

2. Snoep JL, Bruggeman F, Olivier BG, Westerhoff HV. Towards building the silicon cell: a modular approach. Biosystems. 2006; 83:207–216. [PubMed: 16242236]

3. Hood L, Rowen L, Galas DJ, Aitchison JD. Systems biology at the Institute for Systems Biology. Brief Funct Genomic Proteomic. 2008; 7:239–248. [PubMed: 18579616]

4. Morrisey EE, Cardoso WV, Lane RH, et al. Molecular Determinants of Lung Development: NHLBI Workshop. Proc Am Thorac Soc. In Press.

5. Golpon HA, Geraci MW, Moore MD, et al. HOX genes in human lung: altered expression in primary pulmonary hypertension and emphysema. Am J Pathol. 2001; 158:955–966. [PubMed: 11238043]

6. Kaplan F, MacRae T, Comber J, et al. Application of expression microarrays to the investigation of fetal lung development in a glucocorticoid receptor knockout mouse model. Chest. 2002; 121:90S. [PubMed: 11893717]

7. Lin S, Shannon JM. Microarray analysis of gene expression in the embryonic lung. Chest. 2002; 121:80S–81S. [PubMed: 11893703]

8. Liu Y, Hogan BL. Differential gene expression in the distal tip endoderm of the embryonic mouse lung. Gene Expr Patterns. 2002; 2:229–233. [PubMed: 12617806]

9. Mariani TJ, Reed JJ, Shapiro SD. Expression profiling of the developing mouse lung: insights into the establishment of the extracellular matrix. Am J Respir Cell Mol Biol. 2002; 26:541–548. [PubMed: 11970905]

10. Bonner AE, Lemon WJ, You M. Gene expression signatures identify novel regulatory pathways during murine lung development: implications for lung tumorigenesis. J Med Genet. 2003; 40:408–417. [PubMed: 12807961]

11. Lu J, Qian J, Izvolsky KI, Cardoso WV. Global analysis of genes differentially expressed in branching and non-branching regions of the mouse embryonic lung. Dev Biol. 2004; 273:418–435. [PubMed: 15328023]

12. Okubo T, Hogan BL. Hyperactive Wnt signaling changes the developmental potential of embryonic lung endoderm. J Biol. 2004; 3:11. [PubMed: 15186480]

13. O'Reilly MA, Yee M, Buczynski BW, et al. Neonatal oxygen increases sensitivity to influenza A virus infection in adult mice by suppressing epithelial expression of Ear1. Am J Pathol. 2012; 181:441–451. [PubMed: 22677423]

14. Hackett NR, Butler MW, Shaykhiev R, et al. RNA-Seq quantification of the human small airway epithelium transcriptome. BMC Genomics. 2012; 13:82. [PubMed: 22375630]

15. Zhou T, Garcia JG, Zhang W. Integrating microRNAs into a system biology approach to acute lung injury. Transl Res. 2011; 157:180–190. [PubMed: 21420028]

16. Bhaskaran M, Wang Y, Zhang H, et al. MicroRNA-127 modulates fetal lung development. Physiol Genomics. 2009; 37:268–278. [PubMed: 19439715]

17. Yang Y, Kai G, Pu XD, Qing K, Guo XR, Zhou XY. Expression profile of microRNAs in fetal lung development of Sprague-Dawley rats. Int J Mol Med. 2012; 29:393–402. [PubMed: 22160159]

18. Dong J, Jiang G, Asmann YW, et al. MicroRNA networks in mouse lung organogenesis. PLoS One. 2010; 5:e10854. [PubMed: 20520778]

19. Cox B, Kislinger T, Wigle DA, et al. Integrated proteomic and transcriptomic profiling of mouse lung development and Nmyc target genes. Mol Syst Biol. 2007; 3:109. [PubMed: 17486137]

20. Lim YP. Mining the tumor phosphoproteome for cancer markers. Clin Cancer Res. 2005; 11:3163–3169. [PubMed: 15867208]

21. Wang YT, Tsai CF, Hong TC, et al. An informatics-assisted label-free quantitation strategy that depicts phosphoproteomic profiles in lung cancer cell invasion. J Proteome Res. 2010; 9:5582–5597. [PubMed: 20815410]

22. López E, Cho WCS. Phosphoproteomics and Lung Cancer Research. Int J Mol Sci. 2012; 13:12287–12314. [PubMed: 23202899]

23. Sudhir PR, Hsu CL, Wang MJ, et al. Phosphoproteomics identifies oncogenic Ras signaling targets and their involvement in lung adenocarcinomas. PLoS One. 2011; 6:e20199. [PubMed: 21637843]

24. Giorgianni F, Mileo V, Desiderio DM, Catinella S, Beranova-Giorgianni S. Characterization of the phosphoproteome in human bronchoalveolar lavage fluid. Int J Proteomics. 2012:460261. [PubMed: 22997577]

25. Nicholson JK, Connelly J, Lindon JC, Holmes E. Metabonomics: a platform for studying drug toxicity and gene function. Nat Rev Drug Discov. 2002; 1:153–161. [PubMed: 12120097]

26. Fiehn, O.; Kumar, D.; Wohlgemuth, G., et al. Biochemical Mapping of Metabolic Alterations in Lungs of Rat Embryos. 56th ASMS Conference on Mass Spectrometry and Allied Topics; Denver, CO. 2008.

27. Kho AT, Bhattacharya S, Tantisira KG, et al. Transcriptomic analysis of human lung development. Am J Respir Crit Care Med. 2010; 181:54–63. [PubMed: 19815808]

28. Barker DJ, Osmond C. Infant mortality, childhood nutrition, and ischaemic heart disease in England and Wales. Lancet. 1986; 1:1077–1081. [PubMed: 2871345]

29. Rather LJ. Langenbeck on the mechanism of tumor metastasis and the transmission of cancer from man to animal. Clio Med. 1975; 10:213–225. [PubMed: 50912]

30. Kho AT, Zhao Q, Cai Z, et al. Conserved mechanisms across development and tumorigenesis revealed by a mouse development perspective of human cancers. Genes Dev. 2004; 18:629–640. [PubMed: 15075291]

31. Liu H, Kho AT, Kohane IS, Sun Y. Predicting survival within the lung cancer histopathological hierarchy using a multi-scale genomic model of development. PLoS Med. 2006; 3:e232. [PubMed: 16800721]

32. Bhattacharya S, Srisuma S, Demeo DL, et al. Molecular biomarkers for quantitative and discrete COPD phenotypes. Am J Respir Cell Mol Biol. 2009; 40:359–367. [PubMed: 18849563]

33. Demeo DL, Mariani TJ, Lange C, et al. The SERPINE2 gene is associated with chronic obstructive pulmonary disease. Am J Hum Genet. 2006; 78:253–264. [PubMed: 16358219]

34. Zhu G, Warren L, Aponte J, et al. The SERPINE2 gene is associated with chronic obstructive pulmonary disease in two large populations. Am J Respir Crit Care Med. 2007; 176:167–173. [PubMed: 17446335]

35. Spira A, Beane J, Pinto-Plata V, et al. Gene expression profiling of human lung tissue from smokers with severe emphysema. Am J Respir Cell Mol Biol. 2004; 31:601–610. [PubMed: 15374838]

36. Bhattacharya S, Srisuma S, Demeo DL, et al. Microarray data-based prioritization of chronic obstructive pulmonary disease susceptibility genes. Proc Am Thorac Soc. 2006; 3:472. [PubMed: 16921109]

37. DeMeo DL, Mariani T, Bhattacharya S, et al. Integration of genomic and genetic approaches implicates IREB2 as a COPD susceptibility gene. Am J Hum Genet. 2009; 85:493–502. [PubMed: 19800047]

38. Hersh CP, Silverman EK, Gascon J, et al. SOX5 is a candidate gene for chronic obstructive pulmonary disease susceptibility and is necessary for lung development. Am J Respir Crit Care Med. 2011; 183:1482–1489. [PubMed: 21330457]

39. Brehm JM, Hagiwara K, Tesfaigzi Y, et al. Identification of FGF7 as a novel susceptibility locus for chronic obstructive pulmonary disease. Thorax. 2011; 66:1085–1090. [PubMed: 21921092]

40. Melen E, Kho AT, Sharma S, et al. Expression analysis of asthma candidate genes during human and murine lung development. Respir Res. 2011; 12:86. [PubMed: 21699702]

41. Sharma S, Tantisira K, Carey V, et al. A role for Wnt signaling genes in the pathogenesis of impaired lung function in asthma. Am J Respir Crit Care Med. 2010; 181:328–336. [PubMed: 19926868]

42. Bhattacharya S, Go D, Krenitsky DL, et al. Genome-wide transcriptional profiling reveals connective tissue mast cell accumulation in bronchopulmonary dysplasia. Am J Respir Crit Care Med. 2012; 186:349–358. [PubMed: 22723293]

43. Shi W, Chen F, Cardoso WV. Mechanisms of lung development: contribution to adult lung disease and relevance to chronic obstructive pulmonary disease. Proc Am Thorac Soc. 2009; 6:558–563. [PubMed: 19934349]

44. Novershtern N, Itzhaki Z, Manor O, Friedman N, Kaminski N. A functional and regulatory map of asthma. Am J Respir Cell Mol Biol. 2008; 38:324–336. [PubMed: 17921359]

45. Whitsett JA, Matsuzaki Y. Transcriptional regulation of perinatal lung maturation. Pediatr Clin North Am. 2006; 53:873–887. viii. [PubMed: 17027615]

46. Xu Y, Zhang M, Wang Y, et al. A systems approach to mapping transcriptional networks controlling surfactant homeostasis. BMC Genomics. 2010; 11:451. [PubMed: 20659319]

47. Kho AT, Bhattacharya S, Mecham BH, Hong J, Kohane IS, Mariani TJ. Expression profiles of the mouse lung identify a molecular signature of time-to-birth. Am J Respir Cell Mol Biol. 2009; 40:47–57. [PubMed: 18664640]

48. Auffray C, Imbeaud S, Roux-Rouquie M, Hood L. From functional genomics to systems biology: concepts and practices. C R Biol. 2003; 326:879–892. [PubMed: 14744094]

49. Auffray C, Adcock IM, Chung KF, Djukanovic R, Pison C, Sterk PJ. An integrative systems biology approach to understanding pulmonary diseases. Chest. 2010; 137:1410–1416. [PubMed: 20525651]
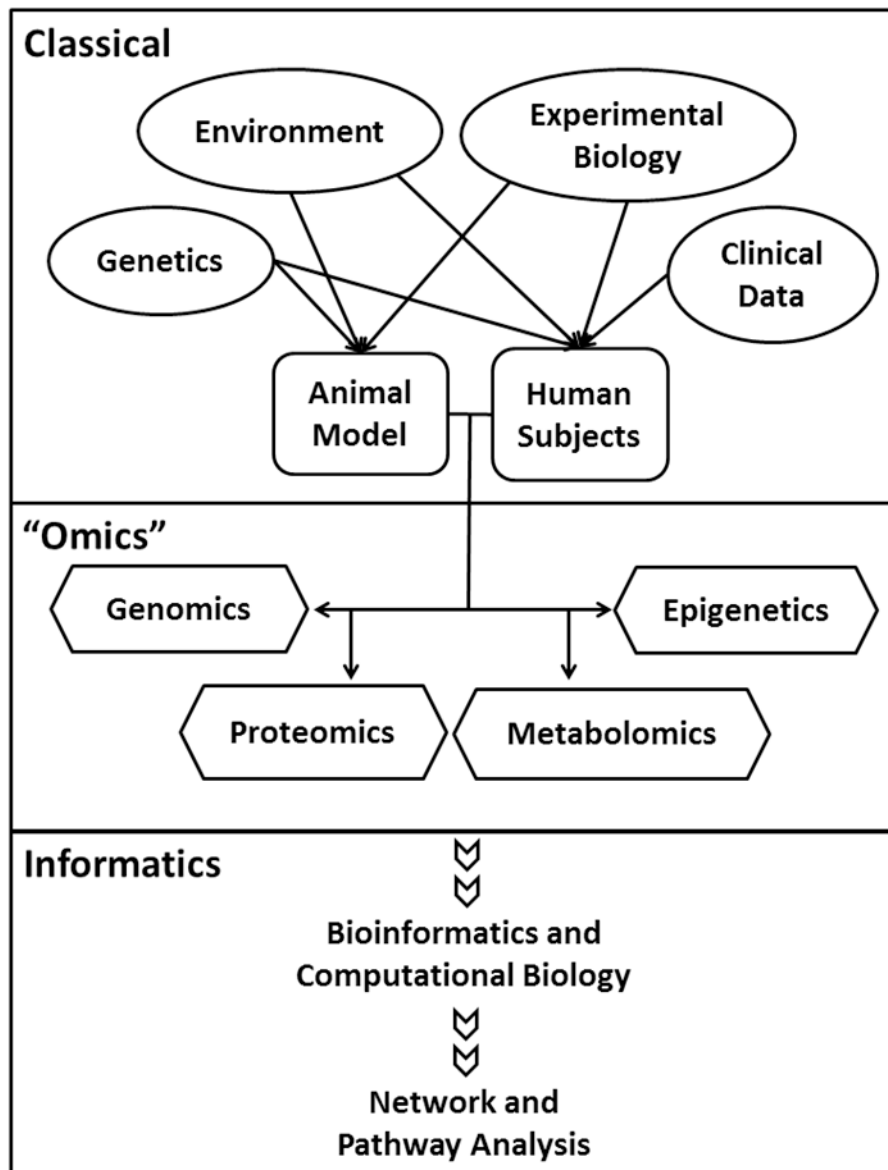
Figure 1.
Systems biology approaches in lung development. Systems biology involves integration of "classical" data collection methods, including genetic, environmental, experimental, and clinical data, with various "omics" approaches through both human subjects and animal models. The "omics" data from different sources can be integrated through various computational informatics approaches. The ultimate goal of systems biology is to understand the underlying mechanisms and pathways involved in complex biological processes including development and disease.
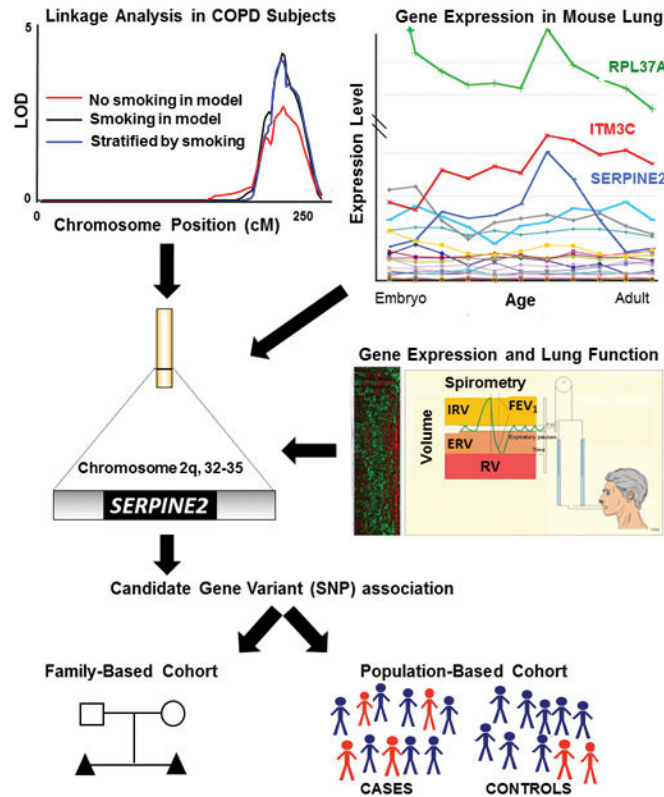
Figure 2.
Discovery of SERPINE2 as a candidate gene for COPD susceptibility. "Systems biology" using a combination of genetic linkage, microarray gene expression and genetic association studies were used to identify SERPINE2 as a candidate susceptibility gene for COPD. LOD, logarithm of odds score. FEV_1, Forced Expiratory Volume in 1 second. RV, Residual Volume. ERV, Expiratory Reserve Volume. IRV, Inspiratory Reserve Volume. In the pedigree chart, square represents male, circle represents female, and triangles represent offspring. White indicates unaffected and black indicates diseased subjects. In the lower right of the figure, red indicates females and blue indicates male.
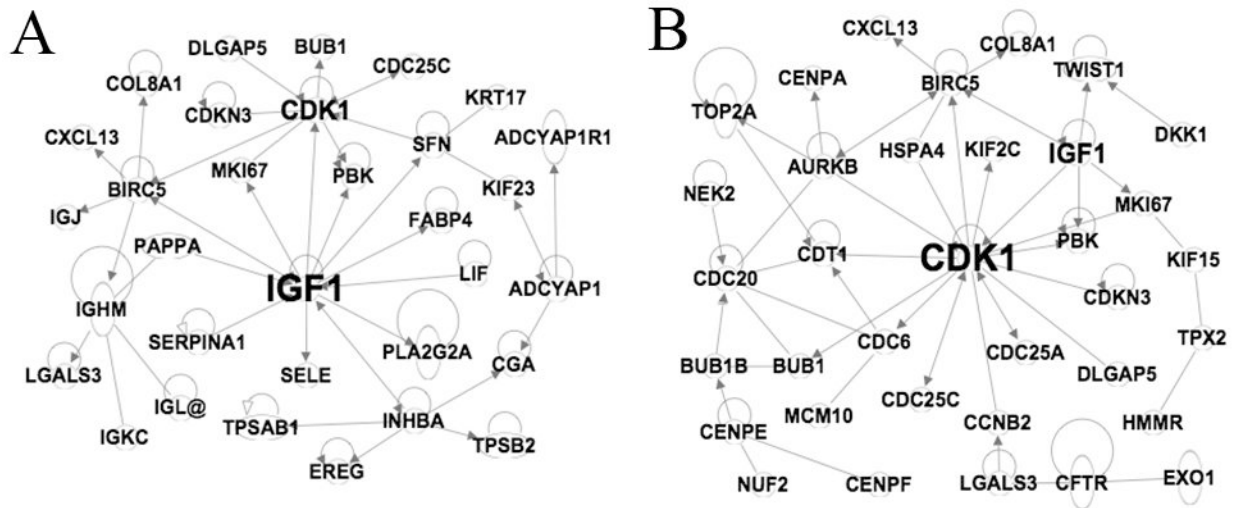
Figure 3.
Developmental data alters disease mechanism prediction. (A) Pathway analysis identified a cluster of 31 genes, from among 159 genes significantly affected in BPD, which were related to IGF1. (B) Integrating normal human lung developmental data into the analysis, shifts CDK1 from a peripheral to a central node in this BPD-related pathway.
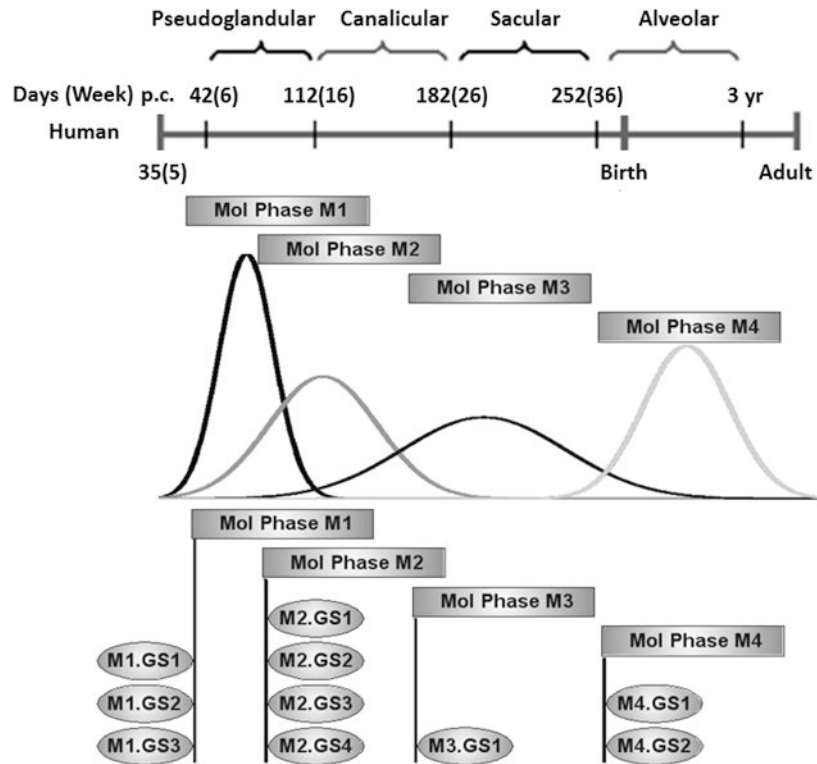
Figure 4.
Defining lung development as molecular phases. Lung development is divided into five distinct histological stages that occur during specific time-periods. Alternately, lung development may be defined as a set of non-distinct molecular phases (Mol Phase MX), each occurring in waves and composed of overlapping gene sets (GS). For instance, in this example, the pseudoglandular stage includes 2 molecular phases (M1, M2). These phases are composed of 3 common gene sets (GS1-3), but differ by the contributions of gene set (GS4).

Table 1

Studies involving gene expression profiling using different technologies to analyze lung development.

Reference	Study	Year	Organism	Age	Material	Summary
Microarrays						
5	Golpon	2001	Human, Mouse	Adult, Fetal	Lung Tissue	Characterized differences in the pattern of HOX gene expression among fetal, adult, and diseased lung specimens.
6	Kaplan	2002	Mouse	Embryonic, Newborn	Lung Tissue	Identified changes in expression of 31 genes in GRhypo mice when compared to wild-type mice.
7	Lin & Shannon	2002	Mouse	Embryonic	Lung Tissue	Identified 204 genes, including novel and known lung-specific genes, as differentially expressed in the lung.
8	Liu & Hogan	2002	Mouse	Embryonic	Lung Tissue	Identified 20 genes as differentially expressed between epithelial tissue from the tips of branching tubes, to those of more proximal region.
9	Mariani	2002	Mouse	Embryonic, Newborn	Lung Tissue	Identified genes encoding regulatory proteins with highly correlated expression patterns to those of extracellular matrix genes.
10	Bonner	2003	Mouse	Embryonic	Lung Tissue	Identified 1346 genes and ESTs as significantly different in at least one stage of lung development.
11	Lu	2004	Mouse	Embryonic	Lung Tissue	Identified 83 genes up-regulated in branching region while 128 up-regulated in non-branching regions
12	Okubo & Hogan	2004	Mouse	Embryonic	Lung Tissue	Identified a mixture of cells expressing marker genes characteristic of different cell types indicating a shift in cell lineage commitments.
27	Kho	2010	Human	Embryonic	Lung Tissue	Identified a set of 3,223 characteristic genes contributing to the changes in the developing lung transcriptome
13	Oreilly	2012	Mouse	Embryonic	Pneumocyte	Identified novel epithelial mechanisms of innate immune responses to respiratory viral infection
RNA-Seq						
14	Hacket	2012	Human	Adult	Small airway epithelial cells	Observed that in addition to genes related clara cells, which are the markers of airway epithelial cells, genes characteristic of minor cell types such as neuroendocrine cells were highly expressed as well
miRNA						
16	Bhaskaran	2009	Rat	Embryonic, Adult	Lung Tissue	Identified 21 miRNAs that were significantly changed during this process of lung development
17	Yang	2012	Rat	Embryonic	Lung Tissue	Identified 167 miRNAs as differentially expressed during rat lung development.
18	Dong	2010	Mouse	Embryonic, Adult	Lung Tissue	Analyzed the expression patterns of dynamically regulated miRNAs and mRNAs and further correlated those with protein levels from an existing mass-spectrometry derived protein database for lung development
Proteomics						
19	Cox	2007	Mouse	Embryonic, Adult	Lung Tissue	Identified groups of genes with statistically significant correlation in expression levels of proteins and transcripts during lung development.

Pediatr Res. Author manuscript; available in PMC 2013 October 01.

Reference	Study	Year	Organism	Age	Material	Summary
24	Giorgianni	2012	Human	Adult	Broncho-alveolar Lavage (BAL)	Characterized phosphorylation of several proteins involved in lung function and disease mechanisms
Metabolomics						
26	Feihn	2008	Rat	Embryonic	Lung Tissue	Identified 46 metabolites that were differentially regulated in embryonic lungs exposed prenatally to environmental tobacco smoke.