

Chapter 17

Transcriptional Networks – Control of Lung Maturation

Yan Xu and Jeffrey A. Whitsett

Abstract This chapter provides examples of the application of tools of bioinformatics and functional genomics to integration of large-scale gene expression data with array independent genomic information to reveal transcriptional regulatory networks. The focus is on networks that control lung maturation and surfactant homeostasis, which can serve as a prototype for study of other complex biological processes. Prenatal maturation of the respiratory system is fundamentally important for the transition to airbreathing at birth. Lung immaturity is a major cause of morbidity and mortality in newborn infants and underlies the pathogenesis of acute respiratory failure (respiratory distress syndrome) and chronic respiratory dysfunction (bronchopulmonary dysplasia), associated with preterm birth. The immaturity of type II alveolar epithelial cells and the lack of pulmonary surfactant lipids and proteins that are needed to reduce surface tension in the alveolar saccules cause atelectasis and respiratory insufficiency after preterm birth. Lung maturation includes diverse structural, cellular, and biochemical changes in lung architecture and function that are precisely coordinated by genetic and environmental factors that synchronize the length of gestation with the process of lung maturation. Analysis of lung specific gene deletion and mutation mouse models, using tools of functional genomics, permits the identification of new genes and pathways controlling surfactant lipid homeostasis and lung maturation. The signaling and

Y. Xu, Ph.D. (✉)

Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, OH, USA

Divisions of Pulmonary Biology and Biomedical Informatics, Cincinnati Children's Hospital Medical Center, 3333 Burnet Avenue, MLC7009, Cincinnati, OH 45229-3039, USA

e-mail: Yan.Xu@cchmc.org

J.A. Whitsett, M.D.

Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, OH, USA

Perinatal Institute, Cincinnati Children's Hospital Medical Center, 3333 Burnet Avenue, MLC7009, Cincinnati, OH 45229-3039, USA

transcriptional mechanisms that influence lung growth and maturation needed to support the abrupt adaptation to airbreathing at birth are of considerable interest, and will provide a rational basis for the design of new treatment strategies for neonatal pulmonary disease.

17.1 Introduction

The lung is a complex organ consisting of more than 40 distinct cell types derived from ectodermal, mesenchymal and endodermal precursors that serve diverse specialized functions related to gas exchange, host defense, and ion transport (Whitsett and Matsuzaki 2006). Perinatal lung “maturation” is necessary for the transition from intrauterine to extra-uterine life. The maturation process includes structural, biochemical and physiologic changes that in the mouse begin approximately at embryonic day 15 and increase dramatically prior to birth. During late gestation, epithelial cells lining the airways and lung saccules differentiate in association with increased production of surfactant, a complex mixture of lipids and proteins that is essential for reducing surface tension created at the air-liquid interface in the alveoli required for air breathing at birth (McMurtry 2002; Burri 1984). Epithelial cell differentiation is accompanied by dilation of peripheral lung saccules, the thinning of mesenchyme, and extensive growth of the pulmonary vasculature. Lung maturation is a carefully timed process. Preterm infants with immature lungs are not able to make sufficient surfactant, causing Respiratory Distress Syndrome (RDS), the major cause of mortality and morbidity in preterm infants (Dubin 1990; Grenache and Gronowski 2006). Since the discovery that RDS is caused by the lack of pulmonary surfactant (Avery and Mead 1959), the structure, function and clinical relevance of surfactant lipid and proteins have been studied extensively in both human and animals (Weaver and Beck 1999; Weaver and Whitsett 1991; Whitsett 2006; Whitsett et al. 1995; Whitsett and Weaver 2002). Although the introduction of surfactant replacement therapy and antenatal glucocorticoid treatment for prevention of RDS have significantly reduced perinatal morbidity and mortality, preterm birth rates and associated pulmonary morbidities have increased progressively during recent decades in North America (Goldenberg et al. 2008; Gravett et al. 2010). New and alternative treatments for the clinical problems related to lung immaturity are needed. Progress in prevention and therapy of lung disease in preterm infants will rest on understanding of genetic, environmental, and hormonal inputs that control lung maturation and surfactant homeostasis.

The post-genomic era has provided unanticipated tools enabling the elucidation of the complex biological processes critical for organ formation and function. The availability of complete sequences of the human, mouse and other genomes and the introduction of mRNA microarray and other high throughput technologies for study of gene and protein expression enable simultaneous quantification of expression levels of thousands of genes/proteins. These technologies bring new perspectives to the study of gene networks and their regulation, providing access to molecular

mechanisms underlying various diseases and phenotypes. While phenotypic outcomes of experiments that are designed to manipulate the expression or function of genes of interest in animal models of disease have been highly successful in the study of disease pathogenesis, the global impact of gene perturbations remains to be fully explored. Systems biology, supported by bioinformatic, proteomic, and genomic data, is providing increasing insight into the pathogenesis of human disease, including those affecting the lung. The density of bioinformatics data relevant to lung biology is increasing exponentially, the thorough analysis of which will provide insights into the processes underlying lung maturation and function. In this chapter, we share experiences applying mRNA microarray analyses and functional genomic applications in conjunction with transgenic mouse models to identify key genes and pathways regulating lung maturation. We provide examples regarding the application of systems biology to integrating lung specific gene expression data with array independent data to generate transcriptional networks regulating surfactant homeostasis during lung development. These approaches can serve as models for study of transcriptional networks that regulate development in other organ systems, especially the use of informatics tools to analyze large data sets.

17.2 Lung Maturation and Surfactant Homeostasis

Lung development can be divided into five morphologically distinct stages that begin near mid-gestation and continue through the early postnatal period. The embryonic stage is distinguished by the formation of the lung buds and division of the trachea and esophagus (E9–11.5) (days 9–11.5 of embryonic development of the mouse). The pseudoglandular stage (E11.5–15.5) is characterized by branching of the conducting airways, formation of the peripheral acinar tubules and buds, and vasculogenesis; the canalicular stage (E15.5–17.5), by expansion of the acinar tubules and buds, angiogenesis, and differentiation of alveolar epithelial type I and II cells; the saccular stage (E17.5 – PN5), by dilation of the terminal respiratory sacs, thinning of the mesenchyme, and deposition of elastin; and the alveolar stage (PN5–30), by maturation of the alveolar-capillary bed, alveolar ducts and alveoli (Perl and Whitsett 1999; Maeda et al. 2007). “Perinatal lung maturation” occurs within the canalicular and saccular stages, at which time differentiation of peripheral distal respiratory epithelial cells mediates increased production of pulmonary surfactant lipids and proteins that are essential for lung function and host defense after birth.

Lung surfactant is a complex mixture of lipid and protein, composed of approximately 80% phospholipids, 10% neutral lipids (particularly cholesterol) and 10% surfactant-associated proteins (SP-A, -B, -C and -D). The most abundant phospholipid is phosphatidylcholine with saturated dipalmitoyl-PC (DPPC) as the predominant form. DPPC is responsible for the surface tension-lowering properties of surfactant (Goerke 1998). The lipid components are synthesized in the

endoplasmic reticulum of the type II cell, modified in Golgi apparatus, transported to the lamellar bodies where surfactant is assembled, stored and secreted by exocytosis into the alveolar airspace (Perez-Gil and Weaver 2010). In late gestation, alveolar type II cells dramatically increase production of surfactant lipid and protein to reduce surface tension created at the air-liquid interface and prepares the lung for extrauterine life (Johansson and Curstedt 1997).

17.3 Gene Expression Profiling and Transcriptional Regulatory Network

Advances in genomic and proteomic technologies have been heralded as the new biological revolution after completion of the human genome project. Gene profiling is considered to be the heart of ‘omics’ because it not only enables a global view regarding expression of genes as cells contribute to specific biological processes, but also infers regulatory relationships between individual genes (Bonner et al. 2003; Bullinger and Valk 2005; Liang et al. 2004; Margalit et al. 2005).

Lung development is a highly coordinated and precisely timed process. During lung maturation, distinct groups of genes are induced or suppressed at specific developmental stages to guide cell proliferation, morphogenesis, and differentiation. An important aspect of gene regulation is mediated by transcription factors (TFs), which bind to cis-elements of the target genes (TGs). Signaling molecules (SM) activate TF responses to biological signals that change the transcription rates of TGs, allowing cells to make needed proteins at the right times and in the right amounts. The interactions between TF and TG can be graphically represented by Transcriptional Regulatory Networks (TRN), in which TF/SMs and TGs are represented as nodes and the interaction relationships are represented as edges. TFs do not function alone in higher organisms, instead, TFs form complexes and bind to *cis* regulatory modules (CRMs), DNA sequences (typically 50–1,000 bp in size) that contain multiple transcription factor binding sites (TFBSs) clustered into modular structures (Jeziorska et al. 2009). CRMs are responsive to specific combinations of TFs, the precise combinatorial interactions of transcription factors providing transcriptional activation appropriate to cell conduct and function (Jeziorska et al. 2009).

A fundamental challenge in the “post genomic era” is to decode transcriptional networks that direct intricate patterns of gene expression typical of complex organs. Although many key regulators have been identified in the lung, how TFs interact with each other and with SMs to regulate groups of target genes mediating perinatal lung development remains unclear. Here, we summarize recent work seeking to decipher transcriptional networks regulating lung maturation and surfactant homeostasis and consider current efforts directed toward the more challenging problem of generating predictive models that account for the dynamic and context dependent TRNs of lung function and disease.

17.4 Functional Genomics to Study Lung Maturation and Surfactant Homeostasis

Microarray technology has been widely applied to many aspects of pulmonary biology (Mayburd et al. 2006; Minn et al. 2005; Wang et al. 2000; Zuo et al. 2002). Our own studies have contributed to the contemporary body of knowledge applying functional genomics approaches to study the transcriptional regulatory programs controlling lung development, function, and disease. In the lung, distinct sets of signaling molecules and transcription factors interact to implement the structural maturation and cell type specific differentiation of the lung. Through the manipulation of a number of key transcription factors and signaling molecules in transgenic mouse models and the application of genome-wide transcriptional profiling analysis to these models, we have identified target genes, pathways, and physiologic consequences in response to the deletion or mutation of many lung transcription factors (Nkx2-1, Foxa1/a2, C/EBP α , Hif1 α , STAT3, NFAT, SREBP, SPDEF, SOX17, PTEN, NF1, Nkx2-9, KLF5, FOXM1, and CATNB) and signaling molecules (MIA, SHH, CSF2R, FGF, RSPO1) (Bridges et al. 2006; Dave et al. 2006; DeFelice et al. 2003; Lian et al. 2004; Martis et al. 2006; Matsuzaki et al. 2006; Metzger et al. 2007; Miller et al. 2004; Mucenski et al. 2005; Wan et al. 2004, 2005, 2008; Xu et al. 2003, 2006, 2007, 2009; Maeda et al. 2011). Many of the key regulators of lung maturation are also critical for early embryogenesis, their disruption often causing embryonic lethality preceding lung formation that begins at approximately E9 in the mouse. The application of conditional mutagenesis in specific lung compartments has been useful in identifying the role of factors critical to early embryogenesis, lung morphogenesis and differentiation. Through these studies, multiple TFs and signaling pathways have been implicated in the structural and functional adaptation of the lung at birth (Maeda et al. 2007). Factors important for lung development and function are not exclusively lung-specific; lung specificity is derived from the unique combinations and interactions of TFs. An understanding of the individual TFs and their interactions in the context of lung development requires systemic approaches to connect distinct but interrelated components to define the transcription networks governing lung maturation and differentiation.

Substantial gaps remain in causally linking patterns of gene expression to the transcriptional mechanisms regulating cell behavior. Microarray data alone cannot distinguish direct and indirect cellular responses. While using genome expression data to elucidate TRNS is far from transparent, additional layers of information and integrative approaches are needed to link gene expression, cell biology, and lung function.

17.4.1 Clustering

Genes belonging to the same co-expression cluster are likely to be co-regulated by similar TFs or belong to the same transcriptional regulatory network. Clustering

provides important insights into regulatory networks by grouping genes on the basis of similarity of their expression patterns under various experimental conditions. Genes selected from microarray analyses are often grouped into distinct clusters. Genes in each cluster are further classified according to Gene Ontology (GO) and shared transcription factor binding sites in the regulatory regions of genes within the cluster to identify the potential biological themes and common regulatory mechanism represented by these unique gene sets. Classical clustering algorithms including K-means, SOM and Hierarchical clustering generally emphasize clear group separations; any given entity will be assigned to only one cluster. In real biology, however, many proteins have multiple roles in cellular responses to various conditions. Fuzzy Heuristic Partition (Fu and Medico 2007; Gasch and Eisen 2002) considers each gene to be a member of every cluster with a variable degree of membership and enables the assignment of genes to more than one cluster with different degrees of membership. Using stringent membership cutoffs, most of the genes in each cluster are highly correlated across all experimental conditions. As the degree of membership decreases, additional genes join the cluster based on their expression similarity under various experimental conditions, enabling the identification of context-dependent regulation. We prefer to evaluate clustering performance based on its ability to produce biologically meaningful clusters using the GO database as a common reference (Datta and Satten 2008; Pihur et al. 2007) rather than to emphasize cluster separation.

Clustering is most commonly used to identify co-expressed genes. Genes also may be clustered based on their functional annotations, shared promoter/regulatory cis-elements, and biochemical and morphological measurements. Figure 17.1 is an example of using multivariate correlation analysis of mRNA expression with lung physiology, biochemical, and morphological measurements. As depicted in Fig. 17.1a, qPCR data analysis of 53 mRNAs previously associated with lung function and structure revealed three major clusters. Cluster 1 genes (including surfactant, *Abca3* and *Slc34a2*) are induced dramatically before birth; cluster 2 genes (*Nkx2-1*, *Pdgfa*, *Lpcat*, etc.) are moderately induced, while cluster 3 genes do not significantly change during perinatal lung maturation. Using mouse embryonic day 15 (E15) mRNA expression as baseline and 1-day before birth as peak maturation, the rate of gene induction prior to birth can be assessed. Clusters 1 and 2 genes are induced earlier and faster in B6 mice (born after 19.5 days gestation) than in A/J mice (born after 20.5 days gestation), indicating the dynamic mRNA changes in cluster 1 are required for the “shortened” lung maturation process. In Fig. 17.1b, multivariate correlation analysis of mRNA expression with dynamic changes of body weight, lung weight, SatPC and morphometric measurements, we identified a subset of mRNAs, including *Sftpa*, *Sftpb*, *Sftpc*, *Sftpd*, *Slc34a2*, *Scgb1a1*, *Cebpa* and *Aqp5*, that were highly correlated with SatPC, body weight, lung weight, and the fractional area of airspace, biochemical and morphological features of lung maturation. Likewise, mRNAs associated with lipid homeostasis, including *Scd1*, *Abca3*, *Fabp5*, and *Lpcat1*, were correlated with lung weight and fractional area of airspace; while another distinct subset of mRNAs, including *Tubb3*, *Pygb*, and

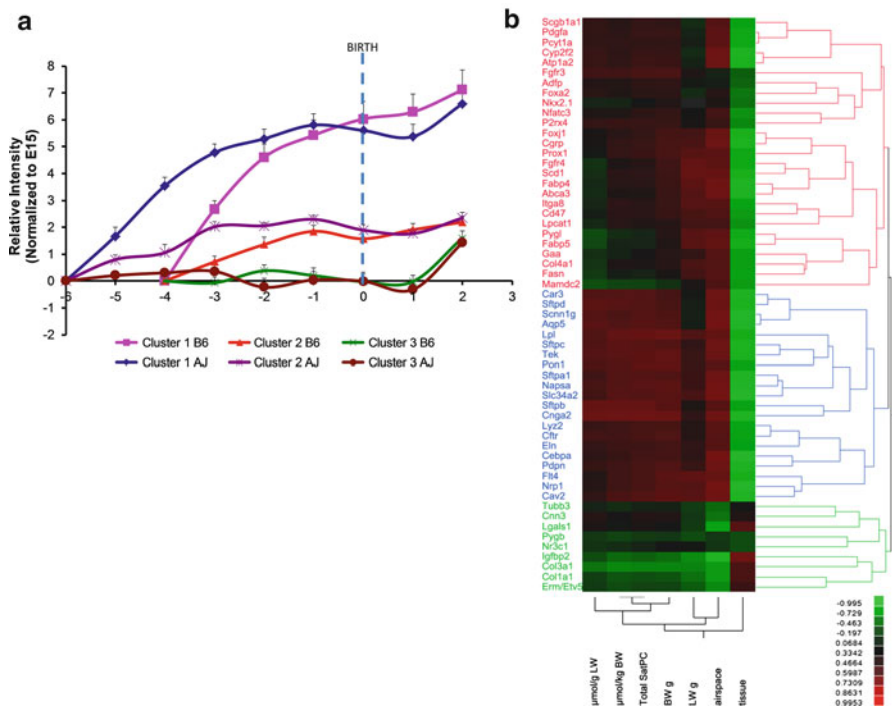


Fig. 17.1 Clustering analysis on selective lung maturation markers: (a) Dynamic mRNA expression levels of 53 lung maturation markers were measured by qPCR in both A/J and B6 mouse strains. qPCR data analysis revealed three major clusters. Profile chart is generated based on the normalized cluster mean and standard error of each gene cluster. The X-axis represents days before and after birth (E15-P2): 0 is the day of birth. The Y-axis represents relative expression normalized to E15. (b) Dynamic mRNA expression profiles of 53 genes at different gestational ages for the A/J mice were correlated with body weight (*BW*), lung weight (*LW*), SatPC ($\mu\text{mol/gLW}$, $\mu\text{mol/gBW}$, and total), and morphometric measurements (airspace, tissue) at corresponding gestational ages using multivariate correction function from JMP 9 (SAS Institute Inc, NC). The heat map was generated based on data from A/J mice using Ward’s minimum variance method to estimate cluster similarity. Gradients in the red and green color range indicate positive and negative correlation, respectively. The levels of mRNAs in red clusters were highly correlated with ontogenic changes in lung SatPC and fractional area of airspace mRNAs in blue clusters were moderately correlated with SatPC, but closely correlated with fractional area of airspace mRNAs in green clusters correlated well with the fractional area of the tissue compartment (Adapted with permission from Figure S5 in Besnard et al. 2011)

Igfbp2, were best correlated with fractional area of the tissue compartment (i.e., mesenchyme). Thus, expression of a subset of mRNAs that encode proteins involved in surfactant homeostasis was highly correlated with increasing SatPC (surfactant lipid), body weight, lung weight, and structural maturation of the lung as gestation proceeded.

17.4.2 Functional Classification

After identifying the major co-expressed gene groups, one general question to address is: “What is unique about this gene set?” There are two common approaches to this question; a reductionism or “cherry-picking” approach and a systems approach. A widely used hypothesis driven approach is gene centric, identifying mRNAs of your interest and choosing them for further study. The systems approach is unbiased, seeking to understand the general themes, trends, and biological meanings buried in the data, rather than to identify a single gene or gene network of interest. Biological knowledge and concepts integration represent an unbiased way to identify the potential biological themes represented by distinct gene data sets. Such processes also help in assigning putative roles to previously uncharacterized genes. As each gene is associated with multiple biological annotations from various resources (Gene Ontology terms, Medical Subject Headings and keywords, pathways, protein–protein interactions, protein functional domains, phenotypes, literature/abstract etc.), enrichment of genes in certain functional categories can be determined using Fisher’s exact test to compare the occurrence of the term in the experimental gene set of interest, with annotations in the rest of the genome as reference. Thus overrepresented functional categories can be identified in your gene list. Multiple pre-compiled web-based functional annotation tools including Onto-Express (Khatri et al. 2002), GoMiner (Zeeberg et al. 2003), DAVID (Dennis et al. 2003), GSEA (Subramanian et al. 2005), and ToppGene (Chen et al. 2009) have been developed that release the user from the burden of compiling and updating the vast and increasing abundance of annotations. Most of these methods are capable of implementing corrections by comparing functional representations within random gene lists and then generating adjusted p-values that represent the probability of observing a given categorical enrichment in experimental data sets. For genes within a cluster, Kappa similarity can be measured to estimate functional similarity between genes based on the number of shared annotation terms (McGinn et al. 2004). Kappa similarity values range from 0 to 1, the higher the value of Kappa, the stronger the overall agreement in annotation terms.

17.4.3 Identification of Common TFBS Motifs and Modules

It is reasonable to predict that co-expressed genes are likely to be co-regulated through common regulatory mechanisms, via the presence and function of common TFBS and binding modules at regulatory sites within the genome. Because “motif” searches are associated with many false positive predictions due to the short and degenerate nature of many TFBS motifs, several approaches can be used to reduce false positives and improve predictive accuracy.

- **Apply comparative genomics:** Programs such as Genome RVista (<http://genome.lbl.gov/vista/>) and DiRE (<http://dire.dcode.org>) are used to identify evolutionarily conserved regulatory elements in co-expressed gene clusters. These programs define precompiled evolutionary conserved regions (ECR) via human and mouse whole genome alignment. The locations of putative TFBSs are precomputed for each genome using vertebrate position weighted matrices from TRANSFAC matrix library. Evolutionarily conserved TFBSs are identified at specific genome locations (promoter, intron, UTR or intergenic regions, etc.) at defined strengths.
- **Search for over-represented TFBSs in proximal promoter regions:** cis-Element over-representation (Clover) (Frith et al. 2004) can be used to identify conserved TFBSs that are over- or under-represented in the given promoter set.
- **Search for motif cluster and CRM:** Since it is known that TFBS are not evenly distributed, finding motif peaks within the promoter region is likely to indicate functional regulatory regions. Cluster-Buster, a Hidden Markov Model based method (Frith et al. 2003) can be used to identify clusters of motifs in a given gene sequence. Matbase (Genomatix) contain well documented, experimentally confirmed promoter modules with synergistic, antagonistic or additive functions. Comparison of predicted CRMs with the known TF modules is used to identify and cross validate meaningful TFBS combinations.

17.4.4 Concept Integration

The combination of approaches that include unsupervised clustering analysis, gene set enrichment analysis, promoter and literature mining with microarray analysis is useful in identifying general modes of action and in forming initial hypotheses regarding the potential targets and regulatory mechanisms underlying experimental data. Bioinformatics data mining and wet-bench confirmation is then integrated to study selected genes and associated pathways. Figure 17.2 provides an example of an integrative analysis of a gene cluster containing 45 genes derived from microarray analysis of 194 samples under 27 conditions. GO classification indicates a significant enrichment of genes involved in lipid metabolism ($P = 0.000472$) and phosphate transport ($P = 0.0055$). Motif and module search were carried out to search for phylogenetically conserved common regulatory elements. EBOX, SP1F, SREB and EGRF are the most significantly enriched TFBS for this gene cluster. Among these, EBOX was detected in 43 out of 45 gene promoters. SREB shares high matrix similarity with EBOX. SREBP binds to both EBOX and SREB *in vivo* (Bennett et al. 1995). Further, SREB/SP1F and EGRF/ZBPF likely form a regulatory module in the promoters of genes in this cluster. SREBP-SP1 may play important role in controlling lung lipid homeostasis, a concept supported by the demonstration that SREB regulates a number of genes involved in lipogenesis in the lung (Plantier et al. 2012; Besnard et al. 2009; Mason et al. 2003).

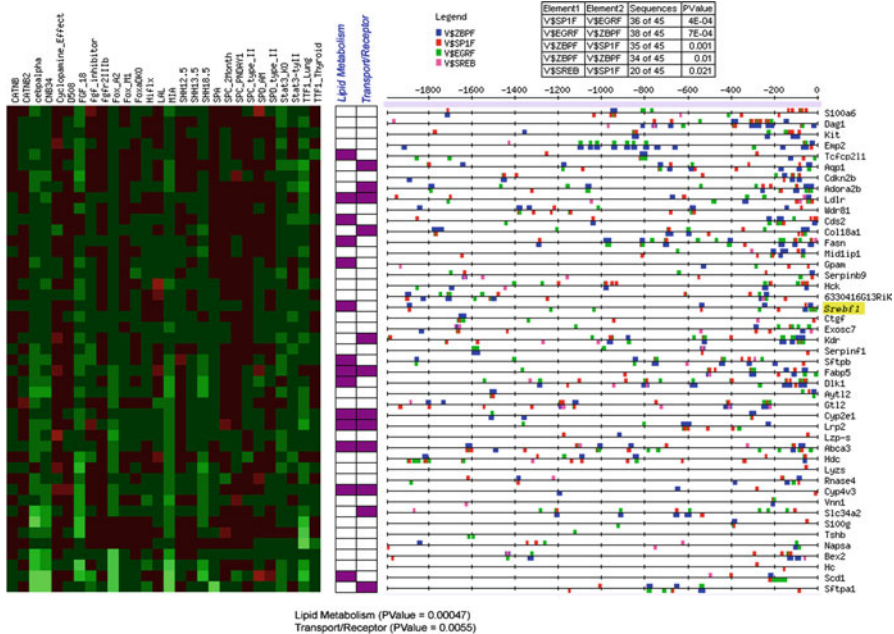


Fig. 17.2 Integrative analysis of a “Lipid” enriched cluster: (a) Heatmap represents gene expression profiles across 27 conditions. Intensity in the red and green color range indicates up-regulated and down-regulated mRNAs, respectively. Each row represents a single gene. Each column represents a particular experimental condition. Each box represents the normalized RNA intensity value. Similarity measures were assessed utilizing Euclidean distance. (b) Significant functional classifications. Colored boxes indicate genes with the respective classification. The calculated P value for the enriched classifications is shown at the bottom. (c) Promoter analysis. SREB (pink), SPIF (red), EGRF (green) and ZFPF (blue) were indicated in the 2 kb upstream promoter sequences of the cluster genes. Those TFBSs are overly represented in this cluster, conserved across human and mouse genomes and tend to form composite modules as shown in the top Table

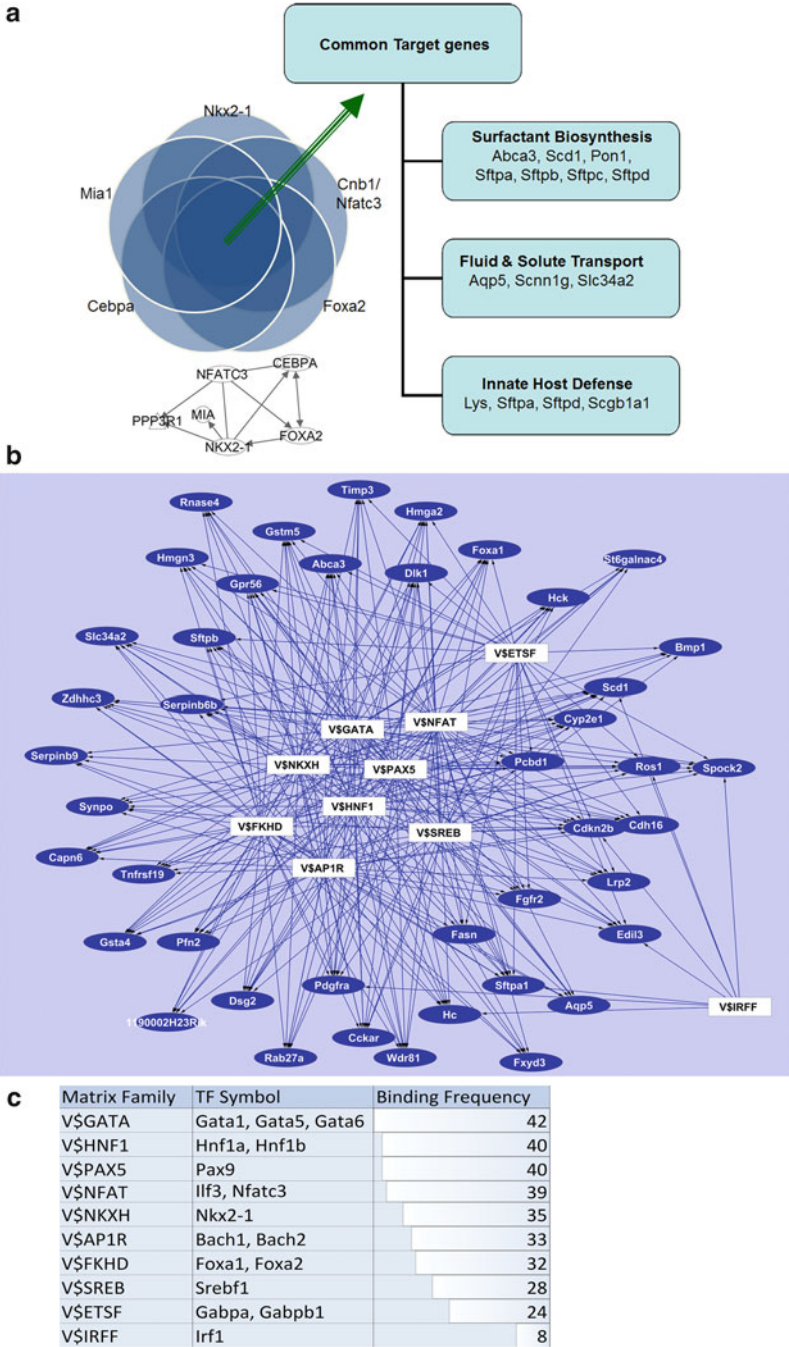
17.4.5 Meta-Analyses of Microarray Data

As a result of the wide application of mRNA microarray technology, there are rapidly growing collections of available microarray data sets that can be used for subsequent analysis. “Meta-analysis” of microarray data uses statistical tools to combine and synthesize results from several related, but independent, microarray experiments (Hong and Breitling 2008). The objectives include increasing power to detect an overall treatment effect, assessment of the variability among studies, and maximizing the use of data available. We have carefully studied genome wide responses to experimental perturbations in more than 80 distinct mRNA microarray experiments. These data are a useful resource for mapping transcriptional networks regulating various aspects of lung development, function, and disease. Applying

meta-analysis to combinations of existing lung microarray datasets has been useful in testing hypotheses and revealing insights into potential regulatory mechanisms. Using this strategy, we have compared mRNA microarray experiments from distinct mouse models that share delayed lung maturation phenotypes (details see below).

17.4.6 Analysis of Genetic Models Influencing Lung Maturation

Lung function at birth is highly dependent on the differentiation and function of the respiratory epithelium that, in turn, produces pulmonary surfactant lipids and proteins. Studies from the conditional deletion or mutation of specific genes led to the identification of several transcription factors and signaling molecules that serve cell-autonomous roles in the respiratory epithelium that are critical for respiratory adaptation at birth, including NKX2-1, FOXA2, C/EBP α , MIA1 and CNB1. Thyroid transcription factor gene 1 (TTF-1/NKX2-1) belongs to the NK2 class of homeobox transcription factors. *Nkx2-1(-/-)* embryos die at birth from respiratory failure due to a profound failure of lung formation (Kimura et al. 1996); absence of *Nkx2-1* activity leads to inhibition of distal lung morphogenesis and epithelial cell differentiation (Yuan et al. 2000). Mice bearing a mutation in *Nkx2-1* (i.e., serine phosphorylation sites are mutated to alanine), substantially rescued lung formation, but impaired lung maturation, causing respiratory failure at birth (DeFelice et al. 2003). Deletion of *Foxa2* (forkhead box protein A2, a winged-helix transcription factor), *Cebpa* (CCAAT enhancer binding protein α) and *Cnb1* (Calcineurin b1) from lung epithelial cells or misexpression of *Mia1* (Melanoma inhibitory activity protein) in lung epithelial cells caused respiratory distress at birth with phenotypic and biochemical changes similar to those observed in NKX2-1 mutant mice, namely: decreased expression of surfactant mRNA and proteins, lack of appropriate differentiation of type I and II cells, and absence of lamellar body formation, indicating delayed peripheral lung maturation (Dave et al. 2006; DeFelice et al. 2003; Martis et al. 2006; Wan et al. 2004; Lin et al. 2008). To better understand the mechanisms underlying the similarity of the perinatal lung maturational defects, we employed genome-wide transcriptional profiling to study genomic responses to phosphorylation mutation of *Nkx2-1*, lung epithelial cell specific deletion of *Foxa2*, *Cebpa*, *Cnb1* and misexpression of *Mia1*. Meta-analysis of these microarray datasets showed that although these transcription factors and signaling molecules act through different signaling pathways and bind to distinct cis-elements, they influence the expression of many common targets involved in surfactant protein and lipid biosynthesis (e.g., *Abca3*, *Scd1*, *Pon1*, *Sftpa*, *Sftpb*, *Sftpc* and *Sftpd*), fluid and solute transport (e.g., *Aqp5*, *Scnn1g*, *Slc34a2*) and innate host defense (e.g., *Lys*, *Sftpa*, *Sftpd* and *Scgb1a1*), suggesting that *Foxa2*, CEBP α , *Cnb1*, *Mia1* and TTF-1 may share a common transcriptional network regulating perinatal lung maturation and postnatal respiratory adaptation (Fig. 17.3a).



Based on the assumption that co-expressed genes sharing similar phenotypes are likely controlled by the same sets of TFs, the altered genes from the “phenocluster” analysis were subjected to promoter common TFBS search to identify statistically over-represented TFBSs and CRMs as compared to their general occurrence in promoter regions of the mouse genome. Using a p-value of 0.05, TFBSs were ranked based on the total binding frequency in the promoter region of genes in the cluster presence (Fig. 17.3c). Cytoscape v2.8.2 was used to generate transcriptional regulatory networks for visualizing molecular interaction networks between this group of genes and predicted TFs. *Nfatc3*, *Nkx2-1* and *Foxa1/a2* were among the most over-represented TFs in the lung with high connectivity to this group of genes in the network (Fig. 17.3b). We linked the TFs to their potential target genes, generating data that further supports the concept that these TFs work in concert to control the transcription of genes involved in lung maturation.

17.4.7 mRNA Microarray and ChIP-seq Integration

Chromatin immunoprecipitation (ChIP) followed by genomic tiling microarray (ChIP-chip) or massive parallel sequencing (ChIP-seq) are two of the most widely used approaches for genome-wide identification of physical interactions between TFs and the regulatory DNA sequences to which they bind. Such studies provide direct evidence of regulatory relationships (Ho et al. 2011). The integration of ChIP-seq and mRNA microarray results enables further dissection of the direct vs. indirect effects of selected TFs *in vivo*. Differentially expressed genes identified from mRNA microarray analysis with positive binding sites from ChIP-seq analysis are considered to be direct transcriptional targets. Genes with altered expression in mRNA microarray analysis, but not in ChIP-seq, are likely indirect targets. Promising peaks that are not associated with changes in mRNA level are likely to be nonfunctional binding sites. Because expression data and ChIP-seq data provide complementary information, predictions of TF-TG relationships based on integration of their data are more accurate than predictions based on single data sources. For instance, to more fully understand the NKX2-1 transcriptional activities



Fig. 17.3 Meta-analysis of microarrays from mouse models sharing common perinatal respiratory distress phenotypes at birth: (a) Comparative microarray analysis of lung selective deletions of *Foxa2*, *Cebpa*, *Cnbl*, mutation of *Nkx2-1* and misexpression of *Mial* revealed common targets involved in surfactant biosynthesis, fluid and solute transport and innate defense. (b) Over-represented TFBS were identified using Clover and TRANSFAC software in the 2 kb promoter regions of the genes altered in all five microarrays. Cytoscape v2.8.2 was used to generate TRN via mapping TF matrix to predicted target genes. Each *white rectangle* represents a TF matrix family and each *blue oval* represents a target gene. (c) The summary of binding frequency of each TF to its potential TGs within 2 kb promoter regions. If multiple TFs are associated with the same TF matrix family, only TFs with abundantly expressed in lung were selected as representative TFs in the table

and direct downstream targets in lung, we preformed integrative analysis of NKX2-1 ChIP-seq and NKX2-1 over-expression mRNA microarray data (Secondary analysis of data from (Maeda et al. 2012)) in following steps: (1) identify differentially expressed genes in response to the NKX2-1 overexpression through mRNA microarray analysis, (2) determine high confidence peaks containing a NKX2-1 binding sites through peak calling of ChIP-seq data, (3) annotate the peak regions and find closest gene in the region, (4) map differentially expressed genes to high confidence peak regions, (5) discover novel motifs and mapping known TFBSs by scanning the known TFBS library using position weighted matrix, (6) build an NKX2-1 consensus binding site, and (7) identify NKX2-1 containing cis-regulatory modules by searching for the presence of additional TFBS within NKX2-1 containing peak sequences that represent potential NKX2-1 interacting partners. We show that the AP1, GATA, RXR, and PAX family of TFs are ranked highest to form cis-regulatory modules with NKX2-1 (Fig. 17.4). Based on our previous studies, *Nkx2-1*, *Cnbl*, *Cebpa* and *Foxa2* belong to same “phenoclusters” (i.e., when perturbed, exhibit phenotypes with similar morphological features) and shared a group of common downstream genes that are known to play important roles in lung maturation and function. Taking advantage of next generation sequencing, we assessed the probability that these TFs form cis-regulatory modules. We scanned the presence of NFAT, CEBPA and FOXA2 binding sites within the peak regions containing NKX2-1 binding site as well as their association with changes in expression from corresponding mRNA microarray analysis. We used random sequences of fragments as reference to calculate the binomial probability of their association. Interestingly, the co-binding probability and frequency of NKX2-1/CEBPA, NKX2-1/FOXA2 and NKX2-1/NFAT were significantly enriched in the positive peak regions compared with random DNA fragments, as indicated in Fig. 17.4. These data support the concept that CEBPA, FOXA2 and NFAT act as interaction partners with NKX2-1 to regulate gene expression during lung development and maturation.

Through comparative mRNA microarray data from experiments with related phenotypes, clustering, promoter analysis, gene set functional enrichment analysis, literature mining and correlation with experimental data (physiology, biochemistry and morphometric measurements), we have made substantial progress in mapping the transcription regulation of lung “maturation.” A similar approach could be taken to identification of transcriptional regulatory networks underlying the development and maturation of other tissues and organs.

17.5 Toward a Systems Level Understanding of Surfactant Homeostasis

While microarray data can be used for unbiased gene and pathway discovery, identifying direct versus indirect genomic responses remain challenging. Functional enrichment analysis and literature mining enable the association of genes with

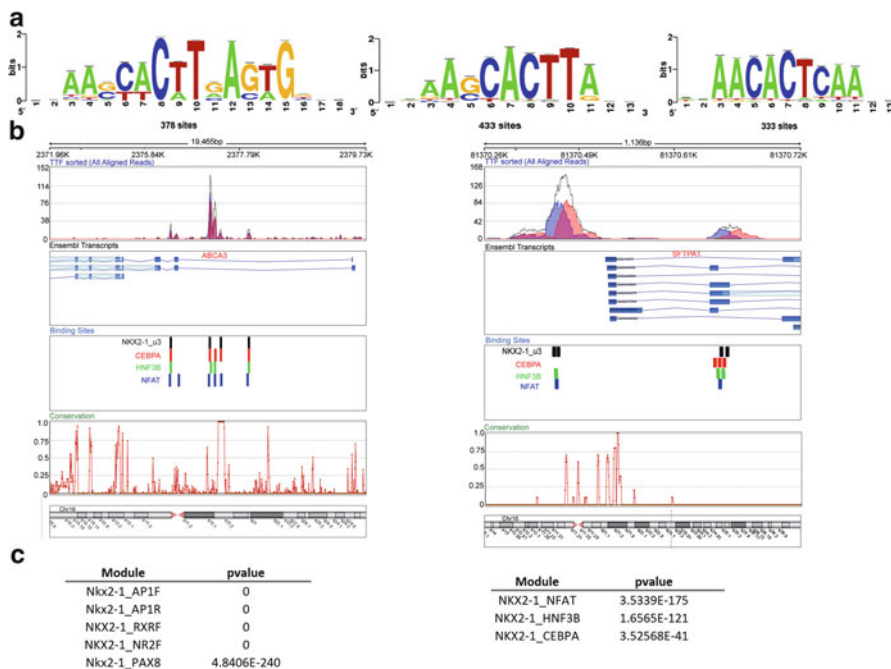


Fig. 17.4 Nkx2-1 Chip-Seq analysis: (a) Logos of the motifs discovered by peak-motifs for Nkx2-1. Peaks were selected using MACS peak detection program in Galaxy package with p -value $< e^{-40}$. Next, we applied ‘peak-motifs’ in RSAT package to discover binding motifs in the detected peak regions and compare discovered motifs to TRANSFAC and JASPAR databases to predict associated transcription factors. (b) Examples of peaks and TFBSs (Nkx2-1: dark blue, CEBPA: red, FOXA2: green and NFAT: baby blue) detected in the human SFTPC and ABCA3 gene locus. (c) The top enriched TFBS modules that co-occurred within 5–100 bp of NKX2-1-binding sequences were determined by RegionMiner (Genomatix) within all peaks identified from ChIP-seq analysis and listed at left panel. The probability of NKX2-1 forming modules with NFAT, FOXA2 and CEBPA were determined within peaks associating with genes altered expression from corresponding NKX2-1 mRNA microarray analysis. The results are compared with random picked sequence fragments with the length similar to the NKX2-1 containing peak regions to determine the binomial probability (right panel)

biological processes and pathways, but are limited to current annotated knowledge. Correlations between regulators and potential targets are largely based on their shared expression patterns, taking into account the likelihood that expression patterns of TF and their targets are often correlated and groups of genes sharing highly correlated expression profiles are likely to share TF(s). This approach, however, will miss TFs that regulate their targets via processes that are independent of their own levels of expression, for example, their activity may be mediated primarily by post-transcriptional mechanisms. Analyses seeking conserved or common TFBSs in promoters of co-expressed genes can identify the potential cis-elements, but does not prove the actual binding of the TF; moreover, TFBS prediction is often associated with high numbers of false positive predictions because of the short

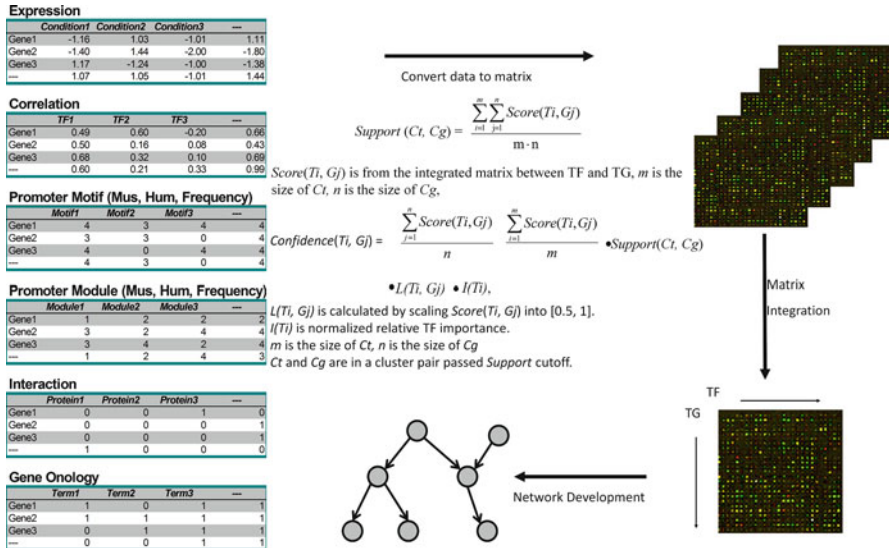


Fig. 17.5 Work flow for construction of a TRN regulating surfactant lipid homeostasis: In order to predict TF-TG interactions using combined evidence, we developed an algorithm to integrate expression profiling with expression-independent data (protein interactions, functional annotation, promoter analyses, and literature mining). TFs were further clustered according to an integrated matrix compiled from data sources including TF-TG functional similarity, TF-TG expression correlation matrix and TF-TG interaction matrix. Each value in the four matrices was scaled from 0 to 1 and summed into the integrated TF-TG matrix. TGs were grouped into sub-clusters based on an integrated matrix, combining and capturing information from four data sources: gene expression, TF-TG correlation, promoter TFBS prediction and GO functional similarity. We calculated the relative confidence score of TF-TG associations by combining the data obtained

(generally encompassing 5–15 base pairs), and degenerate nature of many TFBS motifs. Likewise, the important regulatory roles of non-coding RNAs are not readily identified in the RNA microarray platforms.

To generate a model by which surfactant homeostasis is controlled, we retrieved 194 mRNA microarray samples from 27 distinct mouse models in which TF/SMs modifications were made in mouse models of lung disease. We utilized a systems approach to integrate expression profiling with evidence from other resources, including TF-TG correlation, protein interactions, functional annotation, promoter analyses, and literature mining. Figure 17.5 illustrates the workflow used to build our network model. Detailed methodologies can be found in (Xu et al. 2010). Briefly, we identified 1,498 genes that significantly changed in mRNA microarray analyses in response to the gene perturbations under at least 5 out of 27 experimental conditions. We further clustered differentially expressed genes using Fuzzy clustering by a local approximation of membership algorithm (Fu and Medico 2007). Three co-expressed gene clusters that were highly enriched in genes influencing lipid synthesis and transport were identified. For each cluster, over-represented

TF binding sites, TF-TG expression correlations, TF-TG functional similarity and TF-TG protein-protein interaction were determined. TFs were further clustered according to an integrated matrix compiled from four types of data sources: TFBS-TG scoring, TF-TG functional similarity, TF-TG expression correlations, and TF-TG interaction matrices. Each value in the four matrices was scaled from 0 to 1 and summed into the integrated TF-TG matrix. TGs were grouped into sub-clusters based on an integrated matrix, combining and capturing information from four data sources: gene expression, TF-TG correlation, promoter TFBS prediction and GO functional similarity. In the integrated matrix, each row represents a gene, and each column represents a feature from one of the four matrices. We calculated the relative confidence score of TF-TG associations by combining the resulting data. *Confidence* describes the possibility of a true positive TF-TG relationship according to the integrated information available. The TFBS-TG pairs with the highest *Confidence* scores represent candidates for experimental validation.

Based on predicted, ranked TF-TG relationships, we constructed a “lung lipid regulatory network.” In Fig. 17.6, we show a sub-network consisting of the TFs with the highest connectivity (score ≥ 0.6 , top 4.5%) among three gene clusters. SREBP, HNF3, ETSF, CEBP, GATA and IRFF are clear regulatory hubs in this network, and are TFs likely to be key regulators controlling surfactant lipid homeostasis via control of genes within three lipid-related clusters. The roles of several key TFs in the proposed network have been partially confirmed by previous studies, including SREBP1, FOXA2, CEBPA, ETV5 and GATA6 (Martis et al. 2006; Wan et al. 2004; Besnard et al. 2009; Bruno et al. 2000; Lin et al. 2006; Liu et al. 2002). IRF1 encodes interferon regulatory factor 1, a member of the interferon regulatory transcription factor family. The finding that IRF may serve as an important regulator in lung lipid homeostasis merits further experimental validation. Transcriptional regulators of surfactant homeostasis that had previously been experimentally validated were identified as key hubs in this unbiased network, supporting the reliability of the proposed model. The TFBS of SREBP, HNF3 and CEBP are commonly enriched in all three lipid related clusters and share many downstream targets, suggesting complex interactions among CEBP, SREBP and HNF3 in the proposed lung lipid network. Many of the network predictions for the targets of *Cebpa*, *Srebfl* and *Foxa2* were validated through the combination of promoter reporter assays, transgenic animal models, and literature mining (Xu et al. 2010).

17.6 Transcriptional Programs Controlling Perinatal Lung Maturation

Complex genetic programs that are influenced by multiple environmental and temporal dependent factors likely control the timing of lung maturation. Cross-sectional integrative gene expression profiling analysis does not take into account the dynamic nature of the transcriptional programs accompanying lung maturation. We have recently combined genetic, genomic and bioinformatics methods to

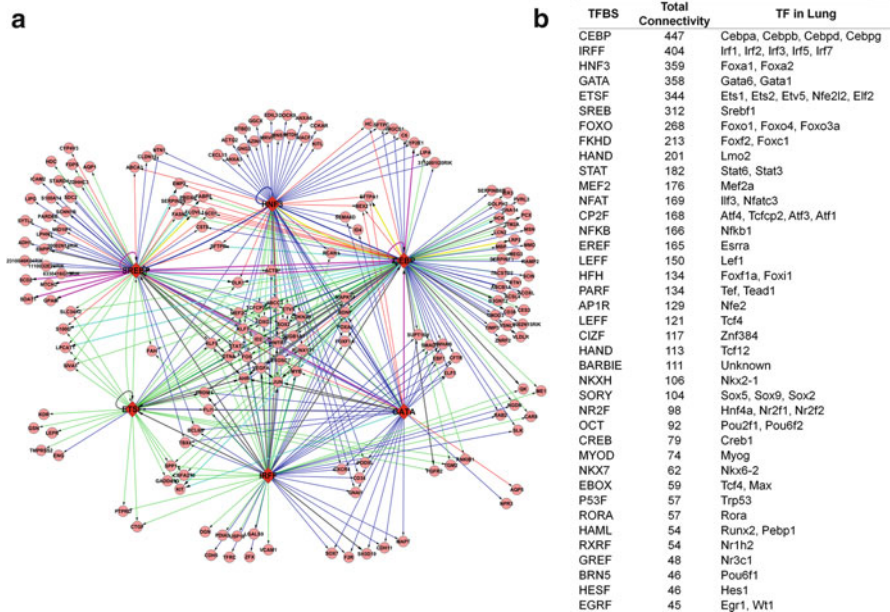


Fig. 17.6 TRN composed of predicted TF-TG pairs with the highest connectivity: (a) The graphic representation of a sub-network consisting of predicted TF-TG pairs with confidence cutoff as 0.60 (top 4.5%) and the top 6 TFs with the highest connectivity. The network has 183 nodes and 386 links. Round nodes represent TGs, red diamond nodes represent TFs. Blue edges indicate the TF-TG predictions from C1, red edges for C2, green for C28, yellow for both C1 and C2, brown for both C1 and C28, light blue for both C2 and C28, and pink edges for TF-TG prediction from C1, C2, and C28. The thickness of the edge corresponds to the frequency of the TF-TG prediction from all three clusters. (b) Confidence score was calculated based on the integrative evidence of TF-TG relationship. The overall connectivity of each TF to its potential TGs within three clusters were calculated and summarized in b. The corresponding TFs expressed in lung were also listed (Adapted with permission from Figure 2 in Xu et al. 2010)

elucidate the relationship between the length of gestation and lung function at birth in two inbred mouse strains (C57BL/6 J and A/J), whose gestational length differed by 1 day (Besnard et al. 2011; Murray et al. 2010; Xu et al. 2012). Lung maturation, as indicated by SatPC (surfactant lipid), lung histology, and the expression of surfactant genes, occurs earlier in B6 than A/J mice. Shorter gestation in B6 mice was associated with advanced morphological and biochemical pulmonary development and better perinatal survival when compared to A/J pups born prematurely (Besnard et al. 2011).

17.6.1 Dynamic Profiling

Taking into account the dynamic nature of the transcriptional programs accompanying lung maturation, we designed genome-wide, time-course microarray

studies to systemically explore the dynamic regulation of lung maturation in both B6 and A/J mouse strains in order to discover genes, pathways, and associated transcriptional networks underlying lung maturation in the two strains of mice that differ in gestational length (Xu et al. 2012). Briefly, lung samples from each mouse strain were collected daily from E15.5 to P0 at precise gestational ages (Murray et al. 2010). Lung RNAs isolated from the two mouse strains at different time points in development were hybridized to Mouse Gene 1.0 ST Array ($n = 3/\text{strain}/\text{time}$). Dynamic lung mRNA expression profiling from B6 and A/J mice were compared to identify: (1) genes and bioprocesses commonly altered in both strains during lung maturation, (2) transcription factors and signaling molecules (TF/SMs) that changed at different stages of lung maturation, (3) pathways and transcriptional networks controlling lung maturation, and (4) strain dependent effects on lung maturation.

To identify temporal-dependent gene expression changes during lung maturation, a functional Bayesian approach (Angelini et al. 2008) was used to analyze time dependent changes in lung mRNAs from both mouse strains. Next, we identified temporal dependent expression patterns and matched dynamic profiles of transcription factors and targets during lung maturation using STEM (Short Time-series Expression Miner), a clustering algorithm specifically designed for the analysis of short time series gene expression datasets (5–8 time points) (Ernst and Bar-Joseph 2006). Comprehensive knowledge integration was employed to identify temporal dependent bioprocesses, key TF/SMs and associated TRNs and to reveal the potential biological interrelationships among the matched TF/SMs and their proposed target genes within each cluster.

To identify strain dependent gene expression during lung maturation, we selected genes that were differentially expressed in A/J and B6 mice at E18.5, but unchanged when comparing A/J at E19.5 vs. B6 at E18.5. The selection is based on two sets of evidence: (1) sample correlation analysis suggested that B6 mRNAs at E18.5 were most similar to A/J mRNAs at E19.5 and least similar to A/J mRNAs at E18.5. (2) A/J fetuses delivered 2 days prematurely (at E18.5) failed to expand their lungs and died of respiratory failure soon after birth, while A/J fetuses delivered 1 day prematurely (at E19.5) survived (87.4%), a survival rate similar to that for B6 mice born 1 day prematurely at E18.5 (82.5%) (Besnard et al. 2011). Together these data support the concept that lung mRNAs differ most between the two strains at E18.5 and that mRNAs modulated during “catch up” in A/J mice at E19.5, are likely important for lung maturation and function.

Through comprehensive bioinformatics data mining, both temporal and strain dependent gene expression patterns were identified during lung maturation. As illustrated in Fig. 17.7, bioprocesses and key regulators associated with different stages of lung development were identified. Lung development, cell adhesion and movement, lipid metabolism, and proliferation were induced early in lung maturation (E15–16 in the pseudoglandular stage). *Hopx*, *Cebpa*, *Tcf21* and *Klf5* were predicted to be important transcriptional regulators at this stage, a finding consistent with gene deletion studies that support their important roles in prenatal lung maturation (Martis et al. 2006; Wan et al. 2008; Quaggin et al. 1999; Yin et al. 2006). TF/SMs regulating vasculature development and apoptosis were induced at E16–17

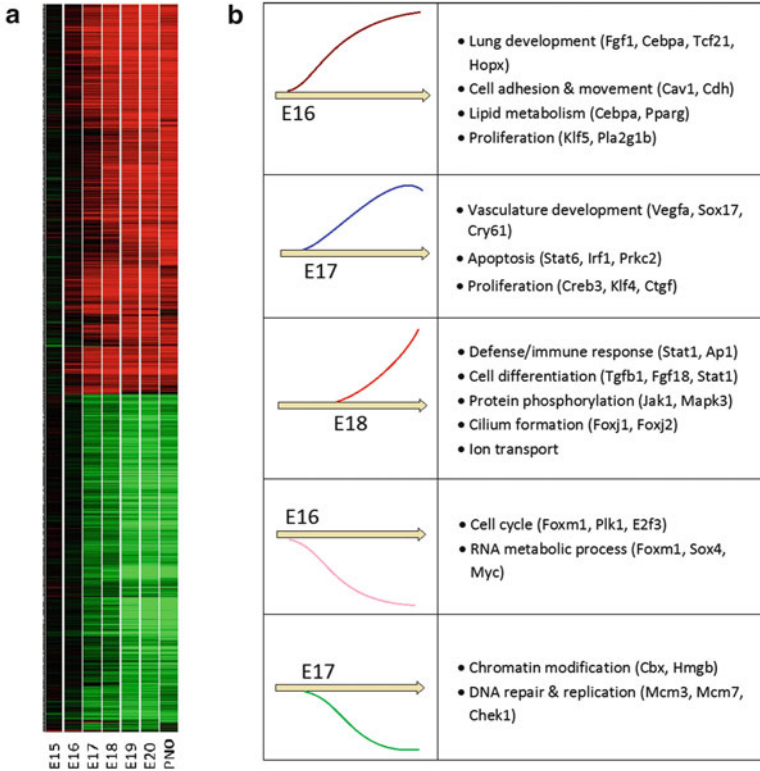


Fig. 17.7 Dynamic regulation of lung maturation: (a) Heatmap of temporal dependent gene expression changes during lung maturation (E15-PN0). (b) Schematized depiction of bioprocesses and predicted key regulators changed dynamically with advancing gestation

(canalicular stage), *Vegfa*, *Sox17* and *Stat3/6* representing important regulators at this stage. Innate defense/immune responses, cell differentiation, protein phosphorylation, ion transport, and cilium formation were induced at later gestational ages (E18–20, in the saccular stage), *Stat1*, *Tgfb1* and *Foxj1* being important regulators associated with this stage of maturation. Cell cycle and chromatin assembly were repressed during lung maturation. FOXM1, PLK1, chromobox, SWI/SNF and high mobility group families of transcription factors were predicted to play important roles in the negative regulation of lung cell proliferation that occurs in late gestation.

Prior to birth, innate immune responses and surfactant production are critical and connected processes that positively influence lung maturation necessary for respiration and survival after birth. In contrast, epigenetic regulators are likely to play a repressive role by altering chromatin structure and controlling the cell cycle. We hypothesize that precise regulation and balance among the positive and negative gene networks are likely critical determinants coordinating the timing of lung maturation with gestational length that differs in the B6 and A/J mouse strains.

17.6.2 Sub-networks Control Distinct Biological Processes During Lung Maturation

Transcriptional regulatory networks can be divided into sub-networks of interconnected genes, each of which represents a functional unit of the entire network. Each unit is driven by tissue or cell type specific TF/SM hubs and acts at specific times and in specific cell types. All units work coordinately to control spatiotemporal processes of lung development and maturation. Due to the complexity and modularity of the TRNs, it is more desirable and experimentally feasible to focus on sub-networks. For example, in our recent genome-wide time-course mRNA microarray study in two strains of mice, we identified multiple bioprocesses induced during the saccular stage of lung development, at E16.5–E17.5, including cell adhesion, lipid metabolism/transport and vasculature development. These major bioprocesses were controlled by CAV1/CDH1, CEBPA/PPARG and VEGFA centered sub-networks, respectively. Innate defense/immune responses were induced at later gestational ages (E19.5–20.5). STAT1, AP1, and EGFR are important regulators of these responses in the sub-network. Expression of RNAs associated with the cell cycle was repressed during prenatal lung maturation and was associated with a FOXM1/PLK1 centered sub-network. These sub-networks consist of a small group of effector genes, usually centered around one or several interrelated hubs. Effector genes in the sub-network tend to be co-expressed, transcriptionally co-regulated, and perform a similar cellular function or work in concert to influence a specific developmental process. Perturbations and experimental validation of hubs and effector genes in the sub-networks will help further delineate the complicated biological processes involved in lung maturation.

17.7 Conclusions and Future Directions

In this chapter, we have summarized functional genomic and systems biology approaches that can be applied to the study of transcriptional regulation of tissue and organ development. The example chosen to illustrate uses of tools and databases is perinatal lung maturation and surfactant homeostasis. We have emphasized the importance of using integrative approaches to achieve comprehensive understanding of regulatory mechanisms controlling lung maturation at systems level. To date, none of the transcription factors participating in lung development are exclusively lung-specific. The unique combinations and interactions among TF/SMs are likely to provide the basis for emergence of lung structural and functional specificities that drive lung maturation. A systems level identification of the individual TF/SMs and their unique interactions in the context of different stages of lung development will be needed to further understand the temporal regulation of structural and functional changes involved in lung maturation prior to birth. A thorough understanding of the molecular mechanisms controlling the timing of normal lung maturation will

promote the understanding of the pathogenesis of lung diseases associated with preterm birth, providing new therapeutic and diagnostic tools to treat pulmonary diseases in infants. This approach takes full advantage of genomic resources to provide an accurate prediction of lung transcriptional networks. The systems biology strategy discussed in this review is powerful and will be highly relevant to studies of other diseases and developmental processes. Like RDS, most lung diseases are complex, and a single gene or protein approach is less likely to identify the cause of acute and chronic lung diseases. Systems biology strategies will facilitate our understanding the pathogenesis and treatment of diseases of the lung and other organs in the future.

Looking forward, we anticipate that transcriptome analysis, ChIP-seq, expression profiling, and assessment of the important regulatory roles of non-coding RNAs will provide essential building blocks for the construction of TRNs regulating signaling and metabolic pathways underlying the complexity of organ formation and function. The emerging view is that many developmental sub-networks are largely operative at the level of individual cells, each of which expresses a unique combination of proteins. Studies of the transcriptome at the level of single cells level pose major technical challenges that have not yet been solved. Currently, the use of mRNA microarray technology has been widely applied at the organ level. The examples described in this chapter are largely based on studies using whole lung samples. Nevertheless, the lung is a complex organ consisting of more than 40 distinct cell types that may be regulated by cell type-specific TF/SMs combinations that activate additional sub-networks of genes to determine the cell fate and function, for example, those involved in Type I and II cell differentiation and surfactant production. Isolating cell-type-specific mRNA, using Laser Capture Microdissection (LCM) and Fluorescence Activated Cell Sorting (FACS) will be useful in measuring RNA expression patterns at a single cell level (Okaty et al. 2011; Chung et al. 2005; Arlotta et al. 2005). These approaches will significantly improve the sensitivity and resolution of mRNA analyses and enable the construction of TRNs that directly reflect the regulation of expression in a cell autonomous manner.

The validation and reconstruction of TRNs based on experimental data presents a major scientific challenge. Currently, TRNs can be experimentally validated via TF-centered (Protein to DNA) or gene-centered (DNA to protein) methods (Arda and Walhout 2010). TF-centered methods such as chromatin immunoprecipitation and gene-centered methods such as reporter gene assays are commonly used to delineate TF-TG relationships. However, these methods are difficult to apply to larger sets of genes, complex network predictions, and within the context of intact chromatin. The recent introduction of ChIP-seq and high throughput siRNA/shRNA based loss-of-function studies provide powerful tools for functional validation of predictions made by computational analyses. How to translate cell culture findings to *in vivo* responses also remains a major issue. New technologies and experimental approaches may be required to bring the process of experimental validation to the next level that are compatible with the complex computational network predictions and high-throughput, genome-wide technologies.

References

- Angelini C, Cutillo L, De Canditiis D, Mutarelli M, Pensky M. BATS: a Bayesian user-friendly software for analyzing time series microarray experiments. *BMC Bioinform.* 2008;9:415.
- Arda HE, Walhout AJ. Gene-centered regulatory networks. *Brief Funct Genomics.* 2010;9(1):4–12.
- Arlotta P, Molyneaux BJ, Chen J, Inoue J, Kominami R, Macklis JD. Neuronal subtype-specific genes that control corticospinal motor neuron development in vivo. *Neuron.* 2005;45(2):207–21.
- Avery ME, Mead J. Surface properties in relation to atelectasis and hyaline membrane disease. *AMA J Dis Child.* 1959;97(5, Part 1):517–23.
- Bennett MK, Lopez JM, Sanchez HB, Osborne TF. Sterol regulation of fatty acid synthase promoter. Coordinate feedback regulation of two major lipid pathways. *J Biol Chem.* 1995;270(43):25578–83.
- Besnard V, Wert SE, Stahlman MT, Postle AD, Xu Y, Ikegami M, Whitsett JA. Deletion of scap in alveolar type II cells influences lung lipid homeostasis and identifies a compensatory role for pulmonary lipofibroblasts. *J Biol Chem.* 2009;284(6):4018–30.
- Besnard V, Wert SE, Ikegami M, Xu Y, Heffner C, Murray SA, Donahue LR, Whitsett JA. Maternal synchronization of gestational length and lung maturation. *PLoS One.* 2011;6(11):e26682.
- 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(6):408–17.
- Bridges JP, Xu Y, Na CL, Wong HR, Weaver TE. Adaptation and increased susceptibility to infection associated with constitutive expression of misfolded SP-C. *J Cell Biol.* 2006;172(3):395–407.
- Bruno MD, Korfhagen TR, Liu C, Morrisey EE, Whitsett JA. GATA-6 activates transcription of surfactant protein A. *J Biol Chem.* 2000;275(2):1043–9.
- Bullinger L, Valk PJ. Gene expression profiling in acute myeloid leukemia. *J Clin Oncol.* 2005;23(26):6296–305.
- Burri PH. Fetal and postnatal development of the lung. *Annu Rev Physiol.* 1984;46:617–28.
- Chen J, Bardes EE, Aronow BJ, Jegga AG. ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. *Nucleic Acids Res.* 2009;37(Web Server issue):W305–11.
- Chung CY, Seo H, Sonntag KC, Brooks A, Lin L, Isacson O. Cell type-specific gene expression of midbrain dopaminergic neurons reveals molecules involved in their vulnerability and protection. *Hum Mol Genet.* 2005;14(13):1709–25.
- Datta S, Satten GA. A signed-rank test for clustered data. *Biometrics.* 2008;64(2):501–7.
- Dave V, Childs T, Xu Y, Ikegami M, Besnard V, Maeda Y, Wert SE, Neilson JR, Crabtree GR, Whitsett JA. Calcineurin/Nfat signaling is required for perinatal lung maturation and function. *J Clin Invest.* 2006;116(10):2597–609.
- DeFelicis M, Silberschmidt D, DiLauro R, Xu Y, Wert SE, Weaver TE, Bachurski CJ, Clark JC, Whitsett JA. TTF-1 phosphorylation is required for peripheral lung morphogenesis, perinatal survival, and tissue-specific gene expression. *J Biol Chem.* 2003;278(37):35574–83.
- Dennis Jr G, Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, Lempicki RA. DAVID: database for annotation, visualization, and integrated discovery. *Genome Biol.* 2003;4(5):P3.
- Dubin SB. Assessment of fetal lung maturity: in search of the Holy Grail. *Clin Chem.* 1990;36(11):1867–9.
- Ernst J, Bar-Joseph Z. STEM: a tool for the analysis of short time series gene expression data. *BMC Bioinform.* 2006;7:191.
- Frith MC, Li MC, Weng Z. Cluster-buster: finding dense clusters of motifs in DNA sequences. *Nucleic Acids Res.* 2003;31(13):3666–8.
- Frith MC, Fu Y, Yu L, Chen JF, Hansen U, Weng Z. Detection of functional DNA motifs via statistical over-representation. *Nucleic Acids Res.* 2004;32(4):1372–81.
- Fu L, Medico E. FLAME, a novel fuzzy clustering method for the analysis of DNA microarray data. *BMC Bioinform.* 2007;8:3.

- Gasch AP, Eisen MB. Exploring the conditional coregulation of yeast gene expression through fuzzy k-means clustering. *Genome Biol.* 2002;3(11):RESEARCH0059.
- Goerke J. Pulmonary surfactant: functions and molecular composition. *Biochim Biophys Acta.* 1998;1408(2-3):79-89.
- Goldenberg RL, Culhane JF, Iams JD, Romero R. Epidemiology and causes of preterm birth. *Lancet.* 2008;371(9606):75-84.
- Gravett MG, Rubens CE, Nunes TM. Global report on preterm birth and stillbirth (2 of 7): discovery science. *BMC Pregnancy Childbirth.* 2010;10 Suppl 1:S2.
- Grenache DG, Gronowski AM. Fetal lung maturity. *Clin Biochem.* 2006;39(1):1-10.
- Ho JW, Bishop E, Karchenko PV, Negre N, White KP, Park PJ. ChIP-chip versus ChIP-seq: lessons for experimental design and data analysis. *BMC Genomics.* 2011;12:134.
- Hong F, Breitling R. A comparison of meta-analysis methods for detecting differentially expressed genes in microarray experiments. *Bioinformatics.* 2008;24(3):374-82.
- Jeziorska DM, Jordan KW, Vance KW. A systems biology approach to understanding cis-regulatory module function. *Semin Cell Dev Biol.* 2009;20(7):856-62.
- Johansson J, Curstedt T. Molecular structures and interactions of pulmonary surfactant components. *Eur J Biochem.* 1997;244(3):675-93.
- Khatri P, Draghici S, Ostermeier GC, Krawetz SA. Profiling gene expression using onto-express. *Genomics.* 2002;79(2):266-70.
- Kimura S, Hara Y, Pineau T, Fernandez-Salguero P, Fox CH, Ward JM, Gonzalez FJ. The T/ebp null mouse: thyroid-specific enhancer-binding protein is essential for the organogenesis of the thyroid, lung, ventral forebrain, and pituitary. *Genes Dev.* 1996;10(1):60-9.
- Lian X, Yan C, Yang L, Xu Y, Du H. Lysosomal acid lipase deficiency causes respiratory inflammation and destruction in the lung. *Am J Physiol Lung Cell Mol Physiol.* 2004;286(4):L801-7.
- Liang M, Cowley AW, Greene AS. High throughput gene expression profiling: a molecular approach to integrative physiology. *J Physiol.* 2004;554(Pt 1):22-30.
- Lin S, Perl AK, Shannon JM. Erm/thyroid transcription factor 1 interactions modulate surfactant protein C transcription. *J Biol Chem.* 2006;281(24):16716-26.
- Lin S, Ikegami M, Xu Y, Bosserhoff AK, Malkinson AM, Shannon JM. Misexpression of MIA disrupts lung morphogenesis and causes neonatal death. *Dev Biol.* 2008;316(2):441-55.
- Liu C, Glasser SW, Wan H, Whitsett JA. GATA-6 and thyroid transcription factor-1 directly interact and regulate surfactant protein-C gene expression. *J Biol Chem.* 2002;277(6):4519-25.
- Maeda Y, Dave V, Whitsett JA. Transcriptional control of lung morphogenesis. *Physiol Rev.* 2007;87(1):219-44.
- Maeda Y, Chen G, Xu Y, Haitchi HM, Du L, Keiser AR, Howarth PH, Davies DE, Holgate ST, Whitsett JA. Airway epithelial transcription factor NK2 homeobox 1 inhibits mucous cell metaplasia and Th2 inflammation. *Am J Respir Crit Care Med.* 2011;184:421-29.
- Maeda Y, Tsuchiya T, Hao H, Tompkins DH, Xu Y, Mucenski M, Du L, Keiser AR, Fukazawa T, Naomoto Y, Nagayasu T, Whitsett JA. KRAS^{G12D} and Nkx2-1 haploinsufficiency induce invasive mucinous adenocarcinoma of the lung (submitted to *J Clin Invest*, 2012).
- Margalit O, Somech R, Amariglio N, Rechavi G. Microarray-based gene expression profiling of hematologic malignancies: basic concepts and clinical applications. *Blood Rev.* 2005;19(4):223-34.
- Martis PC, Whitsett JA, Xu Y, Perl AK, Wan H, Ikegami M. C/EBP{alpha} is required for lung maturation at birth. *Development.* 2006;133(6):1155-64.
- Mason RJ, Pan T, Edeen KE, Nielsen LD, Zhang F, Longphre M, Eckart MR, Neben S. Keratinocyte growth factor and the transcription factors C/EBP alpha, C/EBP delta, and SREBP-1c regulate fatty acid synthesis in alveolar type II cells. *J Clin Invest.* 2003;112(2):244-55.
- Matsuzaki Y, Xu Y, Ikegami M, Besnard V, Park KS, Hull WM, Wert SE, Whitsett JA. Stat3 is required for cytoprotection of the respiratory epithelium during adenoviral infection. *J Immunol.* 2006;177(1):527-37.

- Mayburd AL, Martlinez A, Sackett D, Liu H, Shih J, Tauler J, Avis I, Mulshine JL. Ingenuity network-assisted transcription profiling: identification of a new pharmacologic mechanism for MK886. *Clin Cancer Res.* 2006;12(6):1820–7.
- McGinn T, Wyer PC, Newman TB, Keitz S, Leipzig R, For GG. Tips for learners of evidence-based medicine: 3. Measures of observer variability (kappa statistic). *CMAJ.* 2004;171(11):1369–73.
- McMurtry IF. Introduction: pre- and postnatal lung development, maturation, and plasticity. *Am J Physiol Lung Cell Mol Physiol.* 2002;282(3):L341–4.
- Metzger DE, Xu Y, Shannon JM. Elf5 is an epithelium-specific, fibroblast growth factor-sensitive transcription factor in the embryonic lung. *Dev Dyn.* 2007;236(5):1175–92.
- Miller LA, Wert SE, Clark JC, Xu Y, Perl AK, Whitsett JA. Role of Sonic hedgehog in patterning of tracheal-bronchial cartilage and the peripheral lung. *Dev Dyn.* 2004;231(1):57–71.
- Minn AJ, Gupta GP, Siegel PM, Bos PD, Shu W, Giri DD, Viale A, Olshen AB, Gerald WL, Massague J. Genes that mediate breast cancer metastasis to lung. *Nature.* 2005;436(7050):518–24.
- Mucenski ML, Nation JM, Thitoff AR, Besnard V, Xu Y, Wert SE, Harada N, Taketo MM, Stahlman MT, Whitsett JA. Beta-catenin regulates differentiation of respiratory epithelial cells in vivo. *Am J Physiol Lung Cell Mol Physiol.* 2005;289(6):L971–9.
- Murray SA, Morgan JL, Kane C, Sharma Y, Heffner CS, Lake J, Donahue LR. Mouse gestation length is genetically determined. *PLoS One.* 2010;5(8):e12418.
- Okaty BW, Sugino K, Nelson SB. A quantitative comparison of cell-type-specific microarray gene expression profiling methods in the mouse brain. *PLoS One.* 2011;6(1):e16493.
- Perez-Gil J, Weaver TE. Pulmonary surfactant pathophysiology: current models and open questions. *Physiology (Bethesda).* 2010;25(3):132–41.
- Perl AK, Whitsett JA. Molecular mechanisms controlling lung morphogenesis. *Clin Genet.* 1999;56(1):14–27.
- Pihur V, Datta S, Datta S. Weighted rank aggregation of cluster validation measures: a Monte Carlo cross-entropy approach. *Bioinformatics.* 2007;23(13):1607–15.
- Plantier L, Besnard V, Xu Y, Ikegami M, Wert SE, Hunt AN, Postle AD, Whitsett JA. Activation of sterol response element binding proteins (SREBP) in alveolar type II cells enhances lipogenesis causing pulmonary lipotoxicity. *J Biol Chem.* 2012;287(13):10099–114.
- Quaggin SE, Schwartz L, Cui S, Igarashi P, Deimling J, Post M, Rossant J. The basic-helix-loop-helix protein pod1 is critically important for kidney and lung organogenesis. *Development.* 1999;126(24):5771–83.
- Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A.* 2005;102(43):15545–50.
- Wan H, Xu Y, Ikegami M, Stahlman MT, Kaestner KH, Ang SL, Whitsett JA. Foxa2 is required for transition to air breathing at birth. *Proc Natl Acad Sci U S A.* 2004;101(40):14449–54.
- Wan H, Dingle S, Xu Y, Besnard V, Kaestner KH, Ang SL, Wert S, Stahlman MT, Whitsett JA. Compensatory roles of Foxa1 and Foxa2 during lung morphogenesis. *J Biol Chem.* 2005;280(14):13809–16.
- Wan H, Luo F, Wert SE, Zhang L, Xu Y, Ikegami M, Maeda Y, Bell SM, Whitsett JA. Kruppel-like factor 5 is required for perinatal lung morphogenesis and function. *Development.* 2008;135(15):2563–72.
- Wang T, Hopkins D, Schmidt C, Silva S, Houghton R, Takita H, Repasky E, Reed SG. Identification of genes differentially over-expressed in lung squamous cell carcinoma using combination of cDNA subtraction and microarray analysis. *Oncogene.* 2000;19(12):1519–28.
- Weaver TE, Beck DC. Use of knockout mice to study surfactant protein structure and function. *Biol Neonate.* 1999;76 Suppl 1:15–8.
- Weaver TE, Whitsett JA. Function and regulation of expression of pulmonary surfactant-associated proteins. *Biochem J.* 1991;273(Pt 2):249–64.
- Whitsett JA. Genetic disorders of surfactant homeostasis. *Paediatr Respir Rev.* 2006;7 Suppl 1:S240–2.

- Whitsett JA, Matsuzaki Y. Transcriptional regulation of perinatal lung maturation. *Pediatr Clin North Am.* 2006;53(5):873–87, viii.
- Whitsett JA, Weaver TE. Hydrophobic surfactant proteins in lung function and disease. *N Engl J Med.* 2002;347(26):2141–8.
- Whitsett JA, Noguee LM, Weaver TE, Horowitz AD. Human surfactant protein B: structure, function, regulation, and genetic disease. *Physiol Rev.* 1995;75(4):749–57.
- Xu Y, Clark JC, Aronow BJ, Dey CR, Liu C, Wooldridge JL, Whitsett JA. Transcriptional adaptation to cystic fibrosis transmembrane conductance regulator deficiency. *J Biol Chem.* 2003;278(9):7674–82.
- Xu Y, Liu C, Clark JC, Whitsett JA. Functional genomic responses to cystic fibrosis transmembrane conductance regulator (CFTR) and CFTR(delta508) in the lung. *J Biol Chem.* 2006;281(16):11279–91.
- Xu Y, Ikegami M, Wang Y, Matsuzaki Y, Whitsett JA. Gene expression and biological processes influenced by deletion of Stat3 in pulmonary type II epithelial cells. *BMC Genomics.* 2007;8(1):455.
- Xu Y, Saegusa C, Schehr A, Grant S, Whitsett JA, Ikegami M. C/EBP {alpha} is required for pulmonary cytoprotection during hyperoxia. *Am J Physiol Lung Cell Mol Physiol.* 2009;297(2):L286–98.
- Xu Y, Zhang M, Wang Y, Kadambi P, Dave V, Lu LJ, Whitsett JA. A systems approach to mapping transcriptional networks controlling surfactant homeostasis. *BMC Genomics.* 2010;11:451.
- Xu Y, Wang Y, Besnard V, Ikegami M, Wert SE, Heffner C, Murray SA, Donahue LR, Whitsett JA. Transcriptional programs controlling perinatal lung maturation. *PLoS One.* 2012;7(8):e37046.
- Yin Z, Gonzales L, Kolla V, Rath N, Zhang Y, Lu MM, Kimura S, Ballard PL, Beers MF, Epstein JA, Morrissey EE. Hop functions downstream of Nkx2.1 and GATA6 to mediate HDAC-dependent negative regulation of pulmonary gene expression. *Am J Physiol Lung Cell Mol Physiol.* 2006;291(2):L191–9.
- Yuan B, Li C, Kimura S, Engelhardt RT, Smith BR, Minoo P. Inhibition of distal lung morphogenesis in Nkx2.1(-/-) embryos. *Dev Dyn.* 2000;217(2):180–90.
- Zeeberg BR, Feng W, Wang G, Wang MD, Fojo AT, Sunshine M, Narasimhan S, Kane DW, Reinhold WC, Lababidi S, Bussey KJ, Riss J, Barrett JC, Weinstein JN. GoMiner: a resource for biological interpretation of genomic and proteomic data. *Genome Biol.* 2003;4(4):R28.
- Zuo F, Kaminski N, Eugui E, Allard J, Yakhini Z, Ben-Dor A, Lollini L, Morris D, Kim Y, DeLustro B, Sheppard D, Pardo A, Selman M, Heller RA. Gene expression analysis reveals matrilysin as a key regulator of pulmonary fibrosis in mice and humans. *Proc Natl Acad Sci U S A.* 2002;99(9):6292–7.