A systems approach to mapping transcriptional networks controlling surfactant homeostasis

Yan Xu1,2*, Minlu Zhang2,3, Yanhua Wang1, Pooja Kadambi4, Vrushank Dave1, Long J Lu2,3, Jeffrey A Whitsett1

Abstract

Background: Pulmonary surfactant is required for lung function at birth and throughout life. Lung lipid and surfactant homeostasis requires regulation among multi-tiered processes, coordinating the synthesis of surfactant proteins and lipids, their assembly, trafficking, and storage in type II cells of the lung. The mechanisms regulating these interrelated processes are largely unknown.

Results: We integrated mRNA microarray data with array independent knowledge using Gene Ontology (GO) similarity analysis, promoter motif searching, protein interaction and literature mining to elucidate genetic networks regulating lipid related biological processes in lung. A Transcription factor (TF) - target gene (TG) similarity matrix was generated by integrating data from different analytic methods. A scoring function was built to rank the likely TF-TG pairs. Using this strategy, we identified and verified critical components of a transcriptional network directing lipogenesis, lipid trafficking and surfactant homeostasis in the mouse lung.

Conclusions: Within the transcriptional network, SREBP, CEBPA, FOXA2, ETSF, GATA6 and IRF1 were identified as regulatory hubs displaying high connectivity. SREBP, FOXA2 and CEBPA together form a common core regulatory module that controls surfactant lipid homeostasis. The core module cooperates with other factors to regulate lipid metabolism and transport, cell growth and development, cell death and cell mediated immune response. Coordinated interactions of the TFs influence surfactant homeostasis and regulate lung function at birth.

Background

Pulmonary surfactant is a lipid-protein complex that is synthesized by type II epithelial cells in the alveoli. Surfactant is stored in intracellular organelles known as lamellar bodies and is secreted into airspace by exocytosis. Surfactant lipids form monolayer and multilayer that line the alveolar surface, reducing surface tension created at the air-liquid interface. Pulmonary surfactant is essential for the proper inflation and function of the lung [1]. Surfactant deficiency is associated with premature birth, lung infection or injury. Mutations in genes critical for surfactant production or function can cause lung atelectasis and respiratory failure [2]. Surfactant homeostasis is maintained by a balance among multi-tiered processes, including the synthesis assembly, trafficking, storage, secretion recycling and degradation of surfactant proteins and lipids. While the structures and functions of pulmonary surfactant proteins and lipids have been extensively studied, little is known regarding the genetic and cellular mechanisms integrating the complex processes controlling surfactant lipid homeostasis.

Transcriptional regulation of lipogenesis has been extensively studied in the liver and adipocytes. A number of TFs have been identified controlling the expression of lipogenic enzymes and genes in the lipogenic pathways including Sterol Regulatory Element Binding Protein (SREBP) isoforms, CCAAT-enhancer binding protein (C/EBP) isoforms, nuclear hormone receptors (NR1H2 and NR1H3) and peroxisome proliferator activated receptors (PPAR) [3-7]. SREBP has two genes (Srebf1 and 2) encoding for three protein isoforms, SREBP-1a, SREBP-1c and SREBP-2. SREBPs are synthesized as inactive precursors and activated by proteolysis in the Golgi apparatus. SREBP-2 primarily activates cholesterol biosynthetic genes whereas SREBP-1c predominantly activates genes involved in fatty acid production [4]. The C/EBPs belong to the basic-leucine zipper class
of TFs. Six isoforms have been identified; all of which act as homo-or heterodimers via highly conserved bZIP domain [8]. The involvement of C/EBPs in lipogenesis is strongly supported by both in vitro and in vivo data. In adipocytes, C/EBPα, SREBP-1c and PPARγ induce fatty acid biosynthesis, but only C/EBPα is essential [9].

Lung maturation 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 have lead to the identification of several TFs in lung epithelium that are crucial to lung maturation and respiratory adaptation, including TTF-1, FOXA2 and C/EBPα. TTF-1 binds to the promoters of lung specific genes such as Sftpa, Sftpb, Sftpc, Sftpd and Scgb1a1 and increases their expression [10,11]. The deletion of Foxa2 or Cebpα from lung epithelial cells resulted in the lack of surfactant lipids and proteins, lack of appropriate differentiation of type I and II cells and absence of lamellar body formation, indicating delayed peripheral lung maturation [12,13]. Comparative microarray analysis show that although these TFs bind to distinct cis-elements in the promoter region of target genes, they are able to influence the expression of many common targets involved in surfactant proteins 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, CEBPα and Titf1 may share common transcription network regulating perinatal lung maturation and postnatal adaptation [12-15]. The majority of information regarding the role of SREBP has been focused to cholesterol and fatty acid metabolism in tissues such as liver and adipose [4,16,17]. SREBP-1c is expressed in the developing lung, where its expression increases during late gestation, concomitantly with the perinatal increases in surfactant lipid synthesis and the induction of genes critical for surfactant function [18,19]. SREBP activates CTP:phosphocholine cytidylyltransferase, the rate-limiting enzyme for phosphatidylcholine synthesis and increases surfactant phosphatidylcholine synthesis in the mouse lung [20-22]. These data strongly support the notion that in lung, SREBP may play an important role in surfactant and phospholipid homeostasis.

A fundamental challenge in the “post genomic era” is to decode transcriptional networks that direct intricate patterns of gene expression in complex organisms. In the lung, how TFs interact with each other and signaling molecules to regulate groups of gene targets mediating distinct but integrated aspects of cell or organ function, and how lipid homeostasis is integrated with maturation of type II epithelial cells remain unclear. It is highly likely that surfactant lipid homeostasis is controlled by complex interactions among transcriptional networks that integrate distinct but interrelated aspects of alveolar cell biology, e.g., lung maturation, host defense and surfactant function. Several strategies have been devised to decipher regulatory components and networks, each is partially successful and none is without limitations. Microarray analysis reveals mRNAs that change significantly in expression, but fails to assign these changes to biological events. The GO annotation and literature mining enable the association of genes with biological processes and pathways, but are limited to current knowledge. TF-TG correlation takes into account that expression profiles of TFs and their targets are often correlated and genes with highly correlated profiles are likely to be regulated by the same TF(s). In some instances, however, TFs regulate their targets, not by changing their own expression, but by post-transcriptional mechanisms such as transcript stability, binding site accessibility, interaction with tissue-specific co-factors or chromatin structures [23,24]. Promoter analysis seeking conserved or common TFBSs in promoters of co-expressed genes can identify the potential cis-elements, but may not inherently identify the binding TF or its role in transcription; moreover, this analysis is often associated with high numbers of false positive predictions due to the short and degenerate nature of many TFBS motifs. In the present study, we sought biological consistency and comprehensiveness by using a systems approach to integrate analytic results from independent and complementary resources, including gene expression profiling, protein interaction, functional annotation, promoter and literature mining, to develop a map of genetic networks regulating lung lipogenesis and surfactant homeostasis that are critical for lung function, focusing on the roles of key TFs in the network.

Results and Discussion
We retrieved microarray data from a lung specific gene expression database that measures lung mRNA responses to genetic modification of various genes important to lung development and function (see “Data collection, processing and storage”). Total of 194 mRNA microarray samples from 27 distinct mouse models were used in this study (Table 1).

Clustering and functional classification revealed three lipid enriched gene clusters
Cluster analysis grouped 1498 genes that significantly changed in response to various gene perturbations into 29 clusters (Additional file 1). Genes in each cluster were further classified according to GO classification by Biological Processes to test the biological relevance of each cluster. The criteria for an enriched functional class were P < 0.01 in Fisher Exact Test, the functional
term being shared by more than 20% of the genes in the cluster. Most clusters (26/29) had enriched functional classes according to the criteria, indicating their functional coherence within each cluster (Table 2).

In the present study, we sought to identify the transcription networks regulating perinatal surfactant lipid homeostasis. “Lipid biosynthesis/metabolism/transport” was enriched in 4 out of 29 clusters and SREBP was a member in three of the clusters. We chose to focus on three SREBP related lipid clusters (C1, C2 and C28) for compactness and simplicity of the network (C23 was not included since SREBP was not in the cluster and lipid metabolism was not the predominant functional class of this cluster).

In addition to the commonly enriched functions, i.e. “lipid biosynthesis and metabolism”, each cluster has its uniquely enriched functionality (Table 3). Cluster 1 is functionally enriched in “lung” and “vascular” development, with corresponding mouse phenotypes that include “abnormal vascular development, alveolar morphology and respiratory mechanics”. Membrane/Insoluble fraction is the most enriched cellular component in cluster 1. Cluster 2 is the smallest cluster among the three and is enriched for “lipid metabolism and lipid transport”. Mouse phenotypes associated with the cluster 2 include “abnormal respiratory alveolar morphology and abnormal lipid homeostasis”. “Endoplasmic reticulum (ER)” is the most enriched cellular component in this gene cluster. Tissue distribution analysis indicated that the expression of genes in this cluster is most abundantly expressed in the lung. These functional annotations aligned well with the fact that surfactant lipid and proteins are synthesized and assembled in the ER of alveolar type II cells. Cluster 28 is functionally enriched in “lipid metabolism” and “response to external/chemical stimulus”, the corresponding mouse phenotype being “abnormal blood chemistry and alveolar morphology”. Overall, the functional classifications indicate that lung

Table 1 Microarray Data Used In This Study

| Array Name | Investigator | Mouse Model | Reference |
|------------|--------------|-------------|-----------|
| CEBPA_KO   | Ikeda        | Cebpaflox/flox mice: (tetO)/CMV-Cre<sup>tg</sup>/SP-C-rtTA<sup>tg</sup>/Cebpaflox<sup>lox/lox</sup>, E18.5 | Martis, et al. 2006 |
| CNB        | Dave         | Cnb<sup>tg/lox</sup> mice: CCSP<sup>-rtTA</sup>(tetO)CMV-Cre/Cnb<sup>lox/lox</sup> | Dave, et al. 2006 |
| CTNNB1_ACT | Mucenski     | Catenb<sup>2</sup><sub>lox/lox</sub> mice: CCSP<sup>-rtTA</sup>(tetO)CMV-Cre<sup>tg/tyg</sup> or tyg<sup>tg/tyg</sup> (tetO)CMV-Cre<sup>tg/tyg</sup> or tyg<sup>tg/tyg</sup>, Catenb<sup>2</sup><sub>lox/lox</sub> (ex3) | Mucenski, et al. 2005 |
| CTNNB1_KO  | Mucenski     | SP-C<sup>-rtTA</sup>+<sup>tg</sup>/tg or tyg<sup>tg/tyg</sup>, <sup>β</sup>-catenin<sup>lox/lox</sup> | Mucenski, et al. 2003 |
| Cyclopamine_ | Shannon     | Lung explant culture treated with Cyclopamine for 1-3 days | Xu, et al. 2006 |
| DS08      | Whitsett     | CFTR<sup>lox/lox</sup> mice: FABP-<sup>HCFCTR</sup>/<sup>−/−</sup>/SP-C-Δ508CFTR<sup>+/−</sup> | Whitsett, et al. 2002 |
| FGF18_OE  | Whitsett     | SP-C<sup>-rtTA</sup> and (tetO)-CMV-FGFR-18 | Whitsett, et al. 2002 |
| Fgf21ib    | Perl         | SP-C<sup>-rtTA</sup> and (tetO)-CMV-Fgf21ib<sup>tm1lkf</sup>/<sup>tm1lkf</sup>, E11.5-13.5 lung | Perl, et al. 2003 |
| Foxa2_KO   | Whitsett     | Foxa2<sup>2</sup><sub>LoxP/LoxP</sub> mice: SP-C<sup>-rtTA</sup>+/<sup>tg</sup>/Cre<sup>lox/lox</sup>/Foxa2<sup>lox/lox</sup>tg<sup>tg/tyg</sup>, E18.5 | Wan, et al. 2004 |
| FoxaDKO    | Whitsett     | Foxa2<sup>2</sup><sub>LoxP/LoxP</sub>, Foxa1<sup>−/−</sup> mice: Foxa1<sup>−/−</sup>/SPC<sup>-rtTA</sup>+/<sup>tg</sup>/Cre<sup>lox/lox</sup>/Foxa2<sup>lox/lox</sup>tg<sup>tg/tyg</sup>, E14.5 | Wan, et al. 2005 |
| FoxM1_KO   | Whitsett     | Foxm1<sup>−/−</sup> mice: E18.5 | Wang, et al. 2004 |
| HIF1KO     | Shannon      | Hif1<sup>2</sup><sub>LoxP/LoxP</sub> mice: SP-C<sup>-rtTA</sup>+/<sup>tg</sup>/tetO<sup>−</sup>/Cre<sup>−</sup>/Hif1<sup>−/−</sup>/Bv<sup>−/−</sup>, PND1 | Xu, et al. 2006 |
| LAL        | Yan          | Lal<sup>−/−</sup> mice, 4month | Lian, et al. 2004 |
| MIA        | Shannon      | tetO<sup>2-Cre</sup>/SPC<sup>-rtTA</sup>/Mia<sup>1</sup>, E18.5 | Lin, et al. 2008 |
| SHH12.5    | Shannon      | Shh<sup>−/−</sup> mice, E12.5 | Liu, et al. 2007 |
| SHH13.5    | Whitsett     | Shh<sup>−/−</sup> mice: SP-C<sup>-rtTA</sup>+/<sup>tg</sup>/tetO<sup>−</sup>/Cre<sup>−</sup>/Shh<sup>−/−</sup>/Bv<sup>−/−</sup>, E13.5 | Miller, et al. 2004 |
| SHH18.5    | Whitsett     | Shh<sup>−/−</sup> mice: SP-C<sup>-rtTA</sup>+/<sup>tg</sup>/tetO<sup>−</sup>/Cre<sup>−</sup>/Shh<sup>−/−</sup>/Bv<sup>−/−</sup>, E18.5 | Miller, et al. 2004 |
| SPA_KO     | Levine       | alveolar macrophage from Sftpa<sup>−/−</sup> mice | Xu, et al. 2005 |
| SPC_2M     | Glasser      | Sftp<sup>c</sup>-/ <sup>c</sup> mice, 2month | Zhang, et al. 2006 |
| SPC_PNDI   | Glasser      | Sftp<sup>c</sup>-/ <sup>c</sup> mice, PND1 | Zhang, et al. 2006 |
| SPC_typedl | Glasser      | Isolated typical cells from Sftp<sup>c</sup>-/ <sup>c</sup> mice | Glasser, et al. 2003 |
| SPD<AM     | Whitsett     | Isolated alveolar macrophage from Sftp<sup>d</sup>-/ <sup>d</sup> mice | Glasser, et al. 2003 |
| SPD_typedl | Ikeda        | Isolated typical cell from Sftp<sup>d</sup>-/ <sup>d</sup> mice | Korfhagen, et al. 1998 |
| Stat3<sub>tyll</sub> | Ikeda | Type II cells from TetO<sup>Cre</sup>/SPC<sup>-rtTA</sup>/Stat3<sup>tm1lkf</sup>/, 7 week | Xu, et al. 2007 |
| SUS5402    | Shannon      | Lung explant culture treated with 0.1% DMSO or SUS5402, E12.5 | Metzger, et al. 2007 |
| TTF1_Lung  | Whitsett     | Tiff<sup>1<sub>LMM</sub></sup> mice:Tiff<sup>1</sup> phosphorylation mutant, E18.5 lung | Defelice, et al. 2003 |
| TTF1_Thyroid | Whitsett    | Tiff<sup>1<sub>LMM</sub></sup> mice:Tiff<sup>1</sup> phosphorylation mutant, E18.5 thyroid | Defelice, et al. 2003 |
lipid metabolism is closely associated with lung development and is required for various stress responses.

Identification of commonly enriched TFBS

In general, transcriptional regulation is mediated by the binding of TF or their partners to specific binding sites (TFBS) in the regulatory regions of the target genes (TG). The TFBSs are often located in close proximity to the transcription start site (TSS), but sometimes can be located at more remote locations [25-27]. It is assumed that functional TFBS are subject to greater selective pressure, and therefore will be evolutionarily conserved across species [28-30]. To identify over-represented TFBSs in a given cluster, we took three approaches. First, we searched 3 kb upstream genomic sequence for TFBS in the evolutionarily conserved regions (ECR) that are over-represented in a gene cluster [28,31]. We then searched proximal promoter regions (1.2 kb) for over-represented TFBS in the cluster [32]. We also determined the over-represented TFBS frequency in the proximal promoter region for each gene in the cluster. The relative importance of a TFBS was determined by the average ranking order of the ECR, promoter and frequency analysis. The results are summarized in Figure 1. TFBS for CEBP (CCAAT/Enhancer Binding Protein Family), HNF3B (FOXA2) and SREBP (SREBF1/2) are common to all three clusters, likely indicating the universal roles of these factors in lung lipid metabolism. TFBS for TTF1, HNF3 (FOXA1/2), TCF4, SOX9 and BARBIE (barbiturate-inducible element) were commonly enriched in cluster 1 and 2 genes. The enrichment of this group of TFBS in cluster 1 and 2 gene promoters may indicate their related roles in lung development and morphogenesis. In addition to commonly enriched TFBS among the clusters, we identified TFBS uniquely enriched for each cluster. For example, CIZ (Cas-associated zinc finger protein), OCT (POU2F1) and ETS2 were unique to C1 genes; HNF1 and EGR1 were unique to C2 genes; NFAT and STAT6 were unique to C28 genes. This was consistent with the finding that the three clusters have shared as well as unique functionalities.

Determination of TF-TG functional similarity and expression correlation

It is assumed that genes share similar annotations are likely to be involved in similar biological processes. We used kappa statistics to quantitatively measure the degree of agreement how TF-TG sharing annotation terms [33]. Kappa result ranges from 0 to 1. The higher the value of Kappa, the stronger the agreement is. The
annotation terms are downloaded from DAVID knowledgebase http://david.abcc.ncifcrf.gov/. We calculated the kappa similarity between the enriched TFs of a given cluster (determined via promoter analysis) and genes in the same cluster. Table 4 lists top ranked genes according to their functional similarity (kappa) to that of Srebf1 and Cebpa.

Expression profiles of transcriptional regulators and their targets are correlated in many cases, and genes regulated by the same regulators are likely to be co-expressed [34-37]. We considered TFs in each cluster as potential regulators of the genes in the same cluster. We determined the TF-TG correlations using Pearson correlation. Srebf1 and Cebpa expression profiles
Table 4 TF-TG functional similarity and expression correlation (CEBPA and SREBP)

| CEBPA          | Known Targets | Rank by Kappa Similarity | Known Targets | Rank by Correlation | SREBP         | known Targets | Rank by Kappa Similarity | Known Targets | Rank by Correlation | Known Targets |
|----------------|---------------|--------------------------|---------------|---------------------|---------------|---------------|--------------------------|---------------|-----------------------|---------------|
| Cebpa          | Kyrmizi et al. 2006 | Cebpa                 | Kyrmizi et al. 2006 | Srebf1          | She et al. 2005 | Srebf1         | She et al. 2005            |               |                       |               |
| Foxf1a         | Kim et al. 2005 | S100 g                 | Mtdh          |                     |                |                |                          |               |                       |               |
| Foxa1          | Lpcat1        | Supt16h                | Wdr81         |                     |                |                |                          |               |                       |               |
| Ets1           | Lefterova et al. 2008 | Sftpb                | Id2           |                     | Id2           | Sftpb          | Id2                       | 633016G13Rik |                       |               |
| Sox7           | Dlk1         | Shimomura et al. 1998 | Ebfs          |                     |                |                |                          |               |                       |               |
| Foxa2          | Martis et al. 2006 | Serpinb6b           | Ef5           |                     | Ef5           | Wars           | Reed et al. 2008          |               |                       |               |
| Wwtr1          | Timp3        | Lefterova et al. 2008 | Ankbl         |                     | Ankbl         | Lyzs           |                          |               |                       |               |
| Ef5            | Edil3        | Fah                    | Serpinf1      |                     |                |                |                          |               |                       |               |
| Smad5          | Abca3        | Fli1                   | Dhar7         |                     | Dhar7         | Reed et al. 2008 |                          |               |                       |               |
| Tbx4           | 3110001120Rik | Soat1                  | Farell, et al. 2005 | Siva1           | Siva1         |                |                          |               |                       |               |
| Fli1           | Bex2         | Ahr                    | Ivano et al. 2005 | Ndst1           | Ndst1         |                |                          |               |                       |               |
| Ahr            | Tspan11      | Cdkn2b                 | Cds2          |                     |                |                |                          |               |                       |               |
| Etv5           | Vsnl1        | Foxo3                  | Bcl6b         |                     |                |                |                          |               |                       |               |
| Id2            | Tavor et al. 2003 | Cd38                 | Sox7          |                     | Sox7          | Matn4          |                          |               |                       |               |
| Runx1t1        | Rochford et al. 2004 | 1190002N15Rik       | Mdi1p1        |                     | Mdi1p1        | Ier3           |                          |               |                       |               |
| MeF2c          | Pard6b       | Cbra2t3                | Scd1          | Horton et al. 2002 |                |                |                          |               |                       |               |
| Ebf1           | Jimenez, et al. 2007 | Emp2                | Met2c         | Hck                 |                |                |                          |               |                       |               |
| Klf7           | Id2          | Tavor et al. 2003      | Myb           |                     | Myb           | Dag1           |                          |               |                       |               |
| Prdm1          | Kit          | Zkx                    | Bcl2a1a       |                     |                |                |                          |               |                       |               |
| Jun            | Rangatia et al. 2002 | Mme                 | Acsi4         | Sox7                | Sox7          |                |                          |               |                       |               |
| Utp111         | B3gnt2       | Cebpa                  | Pedersen et al. 2007 | Ahr             | Ivano et al. 2005 |                |                          |               |                       |               |
| Tcfcp211       | Ndst1        | Dhar7                  | Rab6b         |                     |                |                |                          |               |                       |               |
| Cbra2t3        | Lyz          | Lefterova et al. 2008 | Etv5          |                     | Etv5          | Slic1a5        |                          |               |                       |               |
| Fos            | Cammenga et al. 2003 | Lyz                 | Lefterova et al. 2008 | Foxa1          | Slic3a2     |                |                          |               |                       |               |
| Stat3          | Numata et al. 2005 | Syne2                | Foxa2         | Enpp2                |                |                |                          |               |                       |               |
| Sox2           | Tgoln1       | Rab2                   | Hdc           | Ai et al. 2006       |                |                |                          |               |                       |               |
| Myb            | Verbeek, et al. 1999 | Klf7              | Runx1t1       | Sftpb                |                |                |                          |               |                       |               |
| Srebf1         | Le et al. 2002 | Atp6v1b2               | Tbx4          | Kdr                  |                |                |                          |               |                       |               |
| Klf9           | Me1          | Lefterova et al. 2008 | Tcfcp211      | Tsn                  |                |                |                          |               |                       |               |
| Foxo3a         | Tcfcp911     | Upk3b                  | Rtkn2         |                     |                |                |                          |               |                       |               |
| Cdkn2b         | Rtkn2        | Foxf1a                 | Zdhhc14       |                     |                |                |                          |               |                       |               |
| Ankbl          | Krt19        | Sox2                   | Dtna          |                     |                |                |                          |               |                       |               |
correlated well with many of the genes in the lipid clusters across various experimental conditions, there were 50 genes correlated with Srebf1 and 60 genes correlated with Cebp with a correlation coefficient ≥0.5. Table 4 lists genes whose mRNA expression was strongly correlated with that of Srebf1 and Cebp in the rank order of the Pearson product-moment correlation coefficient. As indicated in the Table 4, regulation of a number of the top ranked genes by Srebf1 and Cebp was experimentally confirmed through literature search, indicating TF-TG functional similarity and expression correlation can be useful features for TF-TG prediction.

TF-TG functional similarity measure is limited by known annotations for a given gene. Likewise, correlation does not always hold true. For example, previous studies using lung selective deletion of Foxa2 in the mouse demonstrated that Foxa2 is critical for lung maturation and is involved in the expression regulation of genes in surfactant lipid synthesis [13]. The promoter and gene ontology analysis also indicate that Foxa2 is an important regulator in the mouse lung lipid network. Foxa2 mRNA levels were poorly correlated with genes in the lipid clusters, there were only 5 genes that correlated with Foxa2 with a correlation coefficient ≥0.5. We confirmed by qRT-PCR that Foxa2 mRNA expression levels do not substantially change during lung maturation (data not shown). TFs can regulate their targets independently of their own levels of expression, for example by changing chromatin structure, histone-modification states, nucleosome positions in vivo, phosphorylation status, and binding site accessibility with other partners [23,24]. In other words, a positive correlation between TF and TG provides useful evidence for a regulatory relationship; a poor expression correlation does not necessarily indicate there is no relationship between a given TF-TG pair. Our predication is based on the combined evidence from mRNA expression correlation with promoter profiles and gene ontology similarity; the latter two methods do not require expression correlation.

Prediction of Gene Regulatory Interactions via Data Integration

We then predicted TF-TG interaction based on the integration of evidence from TF-TG correlations, promoter TFBS information, TF-TG kappa similarity and TF-TG interaction evidence. An integrative scoring function was developed to rank the possibility of TF-TG relationship, and to prioritize and associate each target with its potential regulators (detail see METHODS section). Based on these regulatory relationships, we constructed a lung lipid regulatory network. Using the cut off confidence score of 0.5, the overall connectivity of each TF was calculated and summarized in Table 5. HNF3, ETSF, SREBP, CEBP, GATA and IRFF were the most common TFBSs across the three lipid clusters with the highest connectivity in the network. Using this method, we linked the TFs to their potential target genes in three lipid clusters in the order of confidence score (Additional files 2, 3, 4). The TFBS of SREBP, HNF3 and CEBP are commonly enriched in all three lipid clusters and share many downstream targets. Additional files 5, 6, 7 listed top ranked potential targets for SREBP, CEBP and HNF3 according to the confidence score from the integrative analysis of three lipid related clusters. Within the top 100 predicted targets for CEBP, SREBP and HNF3, 49 were common between SREBP and CEBP, 44 were common between CEBP and HNF3, and 35 were common between SREBP and HNF3; suggesting complex crosstalk and interactions

Table 4 TF-TG functional similarity and expression correlation (CEBPA and SREBP) (Continued)

| Fah   | Slc34a2 | Stat3 | Lphn3 |
|-------|---------|-------|-------|
| Mapk14 | Kumar et al. 2003 | Prdx6 | Aytl2 | Lpcat1 |
| Cyp4v3 | Fabp5 | Ets1 | Scd2 | Tabor et al. 1999 |
| Elovl1 | Ier3 | Exosc7 | Emp2 |
| Qk | Lefterova et al. 2008 | Scd1 | Christy et al. 1989 | Elovl1 |
| Rcan1 | Cds5 | Fos | Cyp4v3 |
| Exocs7 | Exocs7 | Ggcx | Mid1p1 |
| Gadd45g | Kdr | Klf9 | Lyz |

Correlation: the expression profile similarities between TF and genes in the same cluster were calculated using Pearson Correlation and ranked accordingly from high to low based on the correlation coefficient. The top 40 genes with the highest correlation with Cebpa and Srebf1 are listed in Table 4.

Kappa similarity was calculated to estimate functional similarity between TF and genes based on the shared annotation terms. TF-TG functional association were ranked from high to low based on the Kappa value. The top 40 genes sharing the highest functional annotations with Cebpa and Srebf1 are listed in Table 4.

We collected the positive TF-TG relationships from Ingenuity knowledge base (Ingenuity), Transfac 11.3 (Biobase), Eldorado (Genomatix) and PubMed. References for the known TF-TG relationships are listed in the table.
among CEBP, SREBP and HNF3 in the proposed lung lipid network.

This method enables identification of genes of interest and their regulators in rank order of their confidence level (Table 6). For example, Abca3 is predicted to be regulated by TFs in the order of SREBP1, HNF3 (FOXA1/2), TTF1, EGR (EGR1), E2F (multiple family members) and CEBPA. ABCA3 is a known phosphatidylcholine transporter and plays an essential role in pulmonary surfactant lipid metabolism and lamellar body biogenesis [38,39]. Abca3 mutations are associated with surfactant deficiency and fatal respiratory distress syndrome [40-42]. Our previous studies showed that Abca3 gene expression was regulated by SREBP, CEBPA and FOXA2 [12,13,43]. Abca3 promoter activity was regulated by both lung selective TFs including TTF1, CEBPA and FOXA2 and the lipogenic TF SREBP1. The direct binding of SREBP1c to Abca3 promoter was confirmed by ChIP assay [44]. Thus Abca3 expression is regulated by both cis-acting cassettes, providing a mechanism by which surfactant and lipid homeostasis may be integrated at the transcriptional level [44]. In addition to known regulators, our model predicts EGR and E2F as potential important regulators for Abca3 expression. ELOVL1 encodes a microsomal enzyme involved in tissue-specific synthesis of very long chain fatty acids and sphingolipids [45,46]. Little is known about Elov11 expression regulation other than that its mRNA expression is correlated with SREBP1 in brown adipocytes [47]. Elov11 was grouped in Clusters 1 and 2, together with Abca3 and our analysis predicts its control by SREBP1, CEBP, HNF3, TTF1 and TCF4, sharing similar regulation with Abca3. Scl34a2 encodes Na(+)/Pi cotransporter 2B (NPT2B), is expressed most strongly in lung and only in apical membrane of alveolar type II

Table 5 Summary of TF connectivity

| TFBS | Total Connectivity | C1 | C2 | C28 | TF in Lung |
|------|-------------------|----|----|-----|------------|
| CEBP | 447               | 238| 51 | 158 | Cebpa, Cebpb, Cebpd, Cebpog |
| IRFF | 404               | 239| 0  | 165 | Irf1, Irf2, Irf3, Irf5, Irf7 |
| HNF3 | 359               | 228| 51 | 80  | Foxa1, Foxa2 |
| GATA | 358               | 218| 44 | 96  | Gata6, Gata1 |
| ETSF | 344               | 172| 14 | 158 | Ets1, Ets2, Etv5, Nfe221, Elf2 |
| SREB | 312               | 162| 43 | 107 | Srebf1 |
| FOXO | 268               | 151| 0  | 117 | Foxo1, Foxo4, Foxo3a |
| FKHD | 213               | 95 | 25 | 93  | Foxc2, Foxc1 |
| HAND | 201               | 94 | 0  | 107 | Lmo2 |
| STAT | 182               | 0  | 0  | 182 | Stat6, Stat3 |
| MEF2 | 176               | 110| 0  | 66  | Mef2a |
| NFAT | 169               | 0  | 0  | 169 | Ilf3, Nfatc3 |
| CP2F | 168               | 83 | 0  | 85  | Atf4, Tcfcp2, Atf3, Atf1 |
| NFBX | 166               | 78 | 0  | 88  | Nkb1 |
| EREF | 165               | 113| 0  | 52  | Emsa |
| LEFF | 150               | 97 | 0  | 53  | Lef1 |
| HFF | 134               | 75 | 19 | 40  | Foxa1a, Foxa1 |
| PARF | 134               | 134| 0  | 0   | Tef, Tead1 |
| AP1R | 129               | 60 | 0  | 69  | Nfe2 |
| LEFF | 121               | 87 | 34 | 0   | Tcf4 |
| CI2F | 117               | 117| 0  | 0   | Znf384 |
| HAND | 113               | 43 | 0  | 70  | Tcf12 |
| BARBIE | 111             | 98 | 13 | 0   | Unknown |
| NKXH | 106               | 68 | 38 | 0   | Nkox2-1 |
| SORY | 104               | 93 | 11 | 0   | Sox5, Sox9 |
| NR2F | 98                | 37 | 20 | 41  | Hnf4a, Nrf2f1, Nrf2f2 |
| OCT | 92                | 92 | 0  | 0   | Pou2f1, Pou6f2 |
| CREB | 79                | 0  | 0  | 69  | Creb1 |
| MYOD | 74                | 0  | 0  | 74  | Myog |
| NKOX7 | 62              | 39 | 0  | 23  | Nkox6-2 |
| EBOX | 59                | 21 | 38 | 0   | Tcf4, Max |
| PS3F | 57                | 0  | 57 | 0   | Trp53 |
| RORA | 57                | 0  | 57 | 0   | Rora |
| HAML | 54                | 0  | 54 | 0   | Runx2, Pebp1 |
| RXRF | 54                | 0  | 54 | 0   | Nrlh2 |
| GREF | 48                | 48 | 0  | 0   | Nrf3c1 |
| BRN5 | 46                | 25 | 21 | 0   | Pou6f1 |
| HESF | 46                | 20 | 26 | 0   | Hes1 |
| EGRF | 45                | 45 | 0  | 0   | Egr1, Wt1 |
| HDXH | 44                | 0  | 44 | 0   | Meis1 |
| SPIF | 39                | 39 | 0  | 0   | Klf11 |
| HNF1 | 38                | 0  | 38 | 0   | Hnf1a, Hnf1b, Hmbox1 |
| E2FF | 33                | 33 | 0  | 0   | E2f1, E2f2, E2f3, E2f4, E2f5, E2f7 |
| SMAD | 23                | 23 | 0  | 0   | Smad4 |

Table 5 Summary of TF connectivity (Continued)

| ZBPF | 23  | 0   | 23  | 0   | Zfp219 |
| NKK6 | 22  | 22  | 0   | 0   | Nkox6-1 |
| LXHF | 21  | 21  | 0   | 0   | Lhx3 |
| AP2F | 19  | 0   | 19  | 0   | Tcfap2c |
| PTBP | 19  | 0   | 19  | 0   | Top |
| GLIF | 16  | 0   | 16  | 0   | Zic2 |
| BCDF | 15  | 0   | 15  | 0   | Ctx |
| SPZ1 | 11  | 0   | 11  | 0   | Spz1 |
| FAX2 | 10  | 0   | 10  | 0   | Pox2 |
| MTF1 | 9   | 0   | 9   | 0   | Mtf1 |
| ZSF5 | 8   | 0   | 8   | 0   | Zfp161 |

We calculated the confidence score based on the integrative evidence of TF-TG relationship. Using the cut off confidence score of 0.5, the overall connectivity of each TF to its potential TGs within three clusters were calculated and summarized in Table 5. The corresponding TFs expressed in lung were also listed.
cells, the cells that produce and secrete surfactant. Because of this localization, it was proposed that the function of the gene product is to take up phosphate from the alveolar fluid [48,49]. Mutations in SLC34A2 cause pulmonary alveolar microlithiasis [48,50]. We utilized transient transfection promoter assays and confirmed the activation of Elovl1 and Slc34a2 transcription by both SREBP1 and CEBPA (see data validation section). DLK1 encodes an EGF like homeotic transmembrane protein that acts as a negative regulator of Notch1 and adipocyte differentiation [51,52].

Our analysis predicts its control by CEBP, HNF3, SREBP1, EGR1, HNF1 and GATA1. Both Elovl1 and Dlk1 are highly enriched in alveolar type II cells. Based on the present model, we hypothesize that genes such as Slc34a2, Dlk1 and Elovl1 may share similar transcription regulation with Abca3 in the lung where they are likely to influence surfactant metabolism.

**Construction of a lipid gene regulatory network in the mouse lung**

A lung “Lipid gene regulatory network” was generated by combining the predicted TF-TG relationships from the three clusters. In Figure 2, 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, ETSE, CEBP, GATA and IRF1 are clear regulatory hubs in this network, these TFs are likely to be key regulators controlling surfactant lipid homeostasis in the lung via the regulation of genes within the three lipid-related clusters. The roles of several key TFs in the proposed network have been partially confirmed by previous studies from our group and others, including SREBP1, FOXA2, CEBP, GATA and IRF1. The finding that previously experimentally validated transcriptional regulators of surfactant homeostasis were identified as key hubs in present unbiased network, strongly support the reliability of our proposed model.

Due to the complexity and modularity of the biological networks, we have focused on several important sub-networks. Figure 3 depicts the CEBPA-SREBP centered sub-network. 3A represents top ranked common gene targets for CEBP and SREBP and 3B represents top ranked unique gene targets for CEBP and SREBP.

---

**Table 6** Selected Genes and their potential regulators in rank order

| Gene  | TF Rank | Gene  | TF Rank |
|-------|---------|-------|---------|
| ABCA3 | 1       | PRDX6 | 5       |
| SFTPC | 2       | SLC34A2 | 3      |
| SOAT1 | 1       | DLK1  | 1       |
| ZDHHC3 | 2     | SREBP1 | 1     |
| LPCAT1 | 3      | SREBP1 | 1     |
| ELOVL1 | 4     | SREBP1 | 1     |
| SREBP1 | 1     | SREBP1 | 1     |

Genes were selected based on their functional relevance to surfactant biosynthesis/transport. For each gene of interest, its potential TF regulators were predicted in rank order of TF-TG confidence score and TF-TG relationships were supported by promoter assay. Top 10 TFBS and their corresponding TF in lung are listed under the selected genes.
SREBP and CEBPA control lung maturation. A subset of predicted targets whose regulation was unknown previously was experimentally confirmed by *in vitro* promoter reporter assays (Figure 4).

**Experimental validation of predicted TF-TG relationships**

Network prediction was validated through promoter reporter assays, transgenic animal models and literature confirmation. Since the integrative analysis predicted SREBP, CEBP and HNF3 as key regulators in the lipid related transcription network in lung, we focused on these three TFs to validate the network predictions derived from the bioinformatics analysis.

Gene promoter assays were carried out on selective TF-TG pairs utilizing the following selection criteria: 1) confidence score, prioritizing top ranked gene targets of SREBP and CEBP, 2) tissue and cell specificity i.e. lung epithelial type II cell enrichment and subcellular location in endoplasmic reticulum or Golgi, 3) functional annotation that is lipid related and 4) originality, seeking novel targets with new function. Applying these criteria, we selected the first set of candidate genes likely modulating lipid homeostasis in the lung epithelial cells, including *Elov1*, *Slc34a2* and *Zdhhc3*, their functionality, expression and subcellular location as listed in Table 7. Figure 4 shows the promoter-reporter assays using C/EBPα and SREBP1c cotransfected with ~1 kb *Elov1*, *Slc34a2* and *Zdhhc3* promoter-luciferase constructs in mouse lung epithelial cells (MLE-15)[56,57]. Consistent with our prediction (Table 6 and Additional files 2, 3, 4), CEBPA and SREBP1c activated *Elov1* and *Slc34a2* promoters; *Zdhhc3* was only regulated by SREBP1c but
not by CEBPA. The functions of Elovl1 and Zdhhc3 in lung biology are unknown whereas Scl34a2 has recently been linked to alveolar microlithiasis.[48,50].

Transgenic mice were used in conjunction with mRNA microarray to identify genes and processes regulated by TFs and signaling molecules. The correlation between the genomic response of selective TF perturbation using transgenic mouse models and the integrative prediction derived from the present study provide in vivo evidence for the predicted TF-TG regulatory relationships. We compared predicted SREBP, HNF3 and CEBP targets with the genes differentially expressed in the lung after selective deletion of Scap (SREBP cleavage-activating protein), Foxa2 and Cebpa from respiratory epithelial cells.[43,58]. These three arrays were not included in previously array analysis and network development, therefore can be used as independent data for validation. Genes with high confidence score (score > 0.55) were used as positive prediction, genes with low confidence score (score < 0.45) were used as negative control. Based on the binomial probability calculation, predicted gene targets showed significant overlap with genes responsive to the deletion of the respective TFs in vivo (p-value for SREBP: 1.03E-08, p-value for FOXA2: 0.0037, p-value for CEBPA: 1.61E-05). Within the top 100 ranked candidate targets for CEBP

---

**Figure 3** Graphic representation of a CEBPA-SREBP centered sub-network. The graphic representation of a CEBPA-SREBP centered sub-network, showing the potential connections between SREBP, CEBPA and their predicted gene targets. 3A represents top ranked common gene targets for CEBP and SREBP and 3B represents top ranked unique gene targets for CEBP or SREBP. Solid line represented literature-validated relationships and dotted lines represent predicted relationships. Known markers of lung maturation and function are highlighted in purple.

---

**Figure 4** Promoter reporter assay of predicted C/EBPA and SREBP targets in transient transfection of MLE-15 cells. Schematic representation of the ≥1 kb Scl34a2, Elovl1 and Zdhhc3 promoter-luciferase constructs made in pGL3 reporter plasmids are depicted above the graphs. C/EBPa (green) and SREBP1c (red) represent consensus motifs on each mouse gene promoter. Transcription start sites are shown at +1 bp. The dose response effects of C/EBPa and SREBP1c expression after co-transfection with fixed amounts of the promoter-reporter constructs were assessed in MLE-15 cells, an immortalized mouse lung epithelial cell line, as measured by luciferase activity in 6-well plates. Values represent two independent experiments carried out in duplicate with means ± S.D. (n = 6).
Table 7 Functionality and subcellular location of selected genes

| Symbol | Description | Expression & Subcellular Location | Function | Disease |
|--------|-------------|-----------------------------------|----------|---------|
| Elov1  | Elongation of very long chain fatty acid protein1 | Expressed in lung type II cells. Endoplasmic Reticulum Membrane | Tissue-specific synthesis of very long fatty acids and sphingolipids. May catalyze the conversion of beta-ketoacyl CoA to beta-hydroxyacyl CoA or Reduction of trans-2-enoyl CoA to the saturated acyl CoA derivative. | Parkinson’s disease |
| Scl34a2| Solute carrier family 34 (sodium phosphate), Member 2 | Apical Membrane of Type II cells | Actively transporting phosphate into cells via Na+ cotransport. May have a role in the synthesis of surfactant in lungs’ alveoli. | Pulmonary alveolar microlithiasis, ovarian cancer |
| Soat1  | Sterol O-acyltransferase 1 | Expressed in lung type II cells. Endoplasmic Reticulum Membrane | Catalyzes the formation of fatty acid-cholesterol esters. Plays a role in lipoprotein assembly and dietary cholesterol absorption. | Atherosclerosis |
| Zdhhc3| Zinc finger DHHC domain-containing protein 3 | Expressed in lung type II cells. Golgi apparatus | Palmitoyltransferase with broad specificity; membrane protein trafficking | |
| Lpcat1| Acyltransferase-like 2 (Atyl2) | Expressed in lung type II cells. Endoplasmic Reticulum and Golgi Apparatus Membrane | Mediates the conversion of 1-acyl-sn-glycero-3-phosphocholine (LPC) into phosphatidylcholine (PC). May synthesize phosphatidylcholine in pulmonary surfactant, play a pivotal role in respiratory physiology. | Migraines |
| Stard4| START domain-containing protein 4 | Expressed in lung type II cells. Mitochondria. | May be involved in the intracellular transport of sterols or other lipids. May bind cholesterol or other sterols. | Huntington’s disease |
| Dlk1  | Protein delta homolog 1 (DLK1) | Expressed in lung type II cells. Membrane | May function in adipocyte differentiation and in neuroendocrine differentiation | Lung cancer, bronchiolo-alveolar adenocarcinoma, blepharophimosis, obesity, neoplasia, hypertiglycemia |
| Prdx6 | Peroxiredoxin 6 | Expressed in lung type II cells. Cytoplasm, Lysosome, lung secretory organelles. | Involved in redox regulation of the cell. Can reduce H2O2 and short chain organic, fatty acid, and phospholipid hydroperoxides. May play a role in the regulation of phospholipid turnover as well as in protection against oxidative injury | Acute allergic pulmonary eosinophilia, asthma, follicular adenoma, Huntington’s disease, neoplasia |
| Abca3  | ATP-binding cassette, sub-family A (ABC1), member 3 | Expressed in lung type II cells. Membrane | Plays an important role in the formation of pulmonary surfactant, probably by transporting lipids such as cholesterol | Surfactant metabolism dysfunction, inflammation, respiratory failure, atelectasis |

Type II cell expression information is obtained from PBGE DB. Subcellular location is based on Gene Ontology http://amigo.geneontology.org/ DB and GeneCard http://www.genecards.org/. Protein function is based on Uniprot Knowledgebase http://www.uniprot.org/uniprot/. Disease information is based on the Ingenuity knowledgebase (Ingenuity) and GeneCard http://www.genecards.org/.

(Cebpa/b/g), 35 mRNA were decreased in response to the Cebpa deletion in the lung in vivo. Likewise, within the top 100 ranked candidate targets for SREBP (Srebf1/2), 25 mRNAs decreased in response to the Scap deletion in vivo; and within the top 100 ranked candidate targets for HNF3 (Foxa1/2), 21 mRNAs were decreased in response to the Foxa2 deletion (Additional files 5, 6, 7). These results provide evidence that SREBP, HNF3 and CEBPA regulate the predicted gene targets expression in lung in vivo.

Literature mining provides another resource to validate computational predictions for the enriched TFs and their potential target genes in the three lipid clusters identified in the present study. We used MedScan Reader, a Natural Language Processing (NLP) text-mining tool [59] (Ariadne Genomics) to search the entire PubMed database. For each TF - TG pair, this algorithm extracts various types of regulatory mechanisms and the effects of regulation by recognizing different domain-specific named entities in the input text and extracting functional relationships among them. As indicated in Additional files 5, 6, 7, all experimentally confirmed SREBP targets were ranked in the top 5% of our predictions; all confirmed HNF3 targets were ranked in the top 10% of our predictions. In the case of CEBPA, all of the experimentally confirmed CEBPA targets were ranked within the top 30% of our prediction with the score >0.5, 86% of them were ranked in the top 10% of our prediction. CCAAT/enhancer binding proteins (C/EBP) include multiple family members that bind to CEBP binding sites with different affinities; that may influence the precision of the present prediction.

Taken together, the consistency of results from in vitro reporter assays, transgenic mice and literatures...
support the validity of the present approach and its potential utility for predicting important TF-TG relationships in the proposed transcription regulatory network.

Biological implication of the lung lipid transcription networks

In the present study, we identified both general and context dependent regulators of lung lipid homeostasis related to pulmonary surfactant. The TFBS of SREBP, HNF3B and CEBP are commonly enriched in all three lipid related clusters and share many downstream targets. We hypothesize that SREBP, CEBP and HNF3 family of TFs form core regulatory modules to maintain surfactant production. Consistent with our model, previous studies demonstrated that the deficiency of hepatic C/EBP in leptin-deficient mouse leads to impaired SREBP signaling [60], C/EBPβ and SREBP-1 form complexes in hepatocytes and synergistically regulate the transcription of lipogenesis associated genes such as Acly and Acss2 [6]. Recent work from Payne et al. [17] demonstrated that SREBP-1c is directly regulated by C/EBP factors during adipocyte differentiation (α, β and δ) and C/EBPα plays a critical role in regulating SREBP-1c in the later stages of adipogenesis (adipocyte maturation). In the lung, C/EBPα and SREBP-1c play important roles in alveolar type II cells lipogenesis [19]. FOXA2 interacts with C/EBPα in mouse liver [61], FOXA2 is necessary for normal expression of C/EBPα in embryonic mouse lung epithelial cells [12].

Core TFs may cooperate with other factors in a context dependent manner. In addition to “lipid metabolism”, SREBP is associated with target genes involved in other related biological processes in cooperation with other TFs. TTF-1 (gene symbol: Nkx2-1) plays a central role at various stages of lung development, essential for lung cell differentiation, maturation and proliferation, and for the production of surfactant proteins. TTF-1 binds to the promoters of lung specific genes such as Sftpa, Sftpb, Sftpc, Sftpd and Sgb1a1 and increases their expression [10,15,53,55,62,63]. The effects of TTF-1 are likely mediated by its interactions with other TFs and co-activators, including WWTR1 (also known as TAZ [10]), GATA6 [55], RAR [64], NFATC3 [57] and NFI [65]. In the present study, TTF-1 is enriched in Clusters 1 and 2, sharing many targets with SREBP to control lipid and surfactant biosynthesis and transport (Abca3, Prdx6, Sftpa1, Sftpb, Sftpc, Dlk1 and Elov11), Apoptosis (Ahr, Bex2, Flt1, Id2, Mef2c and Runx1t1), transcription regulation (Ahr, Bcl6b, Cebpα, Elf5, Etv5, Foxa2, Jun, Sox7 and Wwtr1) and respiratory disease (Abca3, Aqp5, Cfr, Dlk1, Kdr, Prdx6, Sftpa, Sftpb, Sftpc and Slc34a2). Among these, predicted targets such as CEBPA, FOXA2, WWTR1, JUN, ABCA3, SFTPA and SFTPB have been identified as interaction partners or transcriptional targets of TTF-1[12,44,63,66-68]; targets like AHR, CEBPA, ID2 and DLK1 have the same relationships with SREBP[69-73], but little information is available regarding combinatorial regulation of targets by multiple transcription factors.

EGR is uniquely enriched in Cluster 2 genes (lipid cluster). EGR-1 belongs to C2H2-type zinc-finger protein family and activates genes required for differentiation and mitogenesis. In lung, EGR-1 plays a key role in the pathogenesis of IL-13-induced inflammatory responses [74]. The role of EGR-1 in lipid metabolism is unknown. Present study identified a number of EGR and SREBP shared common targets that associated with lung disease or function (Abca3, Aqp5, Foxa2, Cebpα, Kdr and Sftpb), lipid metabolism (Abca3, Soat1, Dlk1, Scd1, Scd2, Lpcat1 and Fapbp5), cell growth and proliferation (Btg3, Dlk1, Emp2 and Pdia5). Among these, Scd1 and 2 are known target of SREBF1 [75], their mRNA expression are also dependent on EGR2 [76]. SCD and FABP5 are known to play important roles in lung specific phospholipids/surfactant biosynthesis [19,77]. LPCAT1 encodes lysophosphatidylcholine acyl-transferase catalyzing the conversion of lysophosphatidylcholine to phosphatidylcholine in the remodeling pathway of phatidylcholine biosynthesis [78]. LPCAT1 is highly expressed in lung type II cells and plays a critical role in regulating surfactant phospholipid/surfactant biosynthesis [79].

Known disease associated genes were identified through the present network analysis. As predicted in Figure 3, ABCA3, DLK1, VEGFA, AGER, SLC34A2 and surfactant proteins are co-regulated by SREBP and CEBPA. Deficiency or mutation of surfactant proteins and ABCA3 cause interstitial lung disease and respiratory failure [40,41,80], PRDX6 is associated with allergic pulmonary eosinophilia and asthma [81,82], DLK1 is associated with bronchiolo-alveolar adenocarcinoma and lung cancer [83], VEGFA (vascular endothelial growth factor A) and KDR (VEGFR, a member of VEGF receptor) play important roles in lung maturation [84] and pulmonary hypertension [85], AGER (advanced glycosylation end product-specific receptor, also known as RAGE) is associated with acute allergic pulmonary eosinophilia [81], and mutations of SLC34A2 cause pulmonary alveolar microlithiasis [50]. The finding that the present approach identified genes and processes associated with human lung disease indicates its potential utility for the discovery of new genes and biomarkers that may be useful in understanding the pathogenesis of lung disorders.

Conclusions

We employed a systems biology approach to begin mapping a transcriptional network regulating surfactant
homeostasis in the lung. We identified novel and known TFs, signaling molecules and potential target genes within the network. SREBP, CEBP, HNF3, ETS, GATA and IRF were identified as regulatory hubs with high connectivity. We propose that SREBP, HNF3B and CEBP form a common core regulatory module mediating surfactant lipid homeostasis. These key TFs likely interact with other TF partners to regulate lung growth (OCT and NFKB), differentiation and maturation (TTF1 and EGR1), pulmonary host defense and inflammatory responses (IRF, NFAT and STAT). The present study provides a systematic view and working model of a transcriptional network regulating the formation and metabolism of the pulmonary surfactant system.

The current approach also has several important limitations. The approach is unlikely to identify epigenetic, post-transcriptional and gene-environmental interactions that may play important roles in gene regulation [23,24]. Likewise, we have not emphasized the study of enzymatic transport activities of the many enzymes and proteins identified in the network. All these will be important for our long-term understanding of lung lipid homeostasis, but are beyond the scope of the present study.

Methods

Data Collection, processing and storage
We have developed a relational database to store, manage and maximally utilize gene expression profile data collected from multiple investigators in Cincinnati Children’s Hospital Medical Center, Division of Pulmonary Biology. We analyzed 194 microarray samples from 27 independent microarray experiments in this study (Table 1). Data was normalized using the Robust Multi-chip Average model [86] from R/Bioconductor package. The detection of differential expression was preformed using unpaired two-group Student’s t-test for mutant and control at the P value ≤ 0.05. Additional filters for positive candidate selection include a minimum of 1.5 fold change in absolute ratio and a minimum of 67% Present call by Affymetrix algorithm. We identified 1498 genes that significantly changed in response to the gene perturbations in at least 5 experimental conditions. The full gene set derived from mRNA profiling is listed in Additional file 1.

Cluster analysis
Clustering is a powerful way to explore complex gene expression data by grouping them on the basis of similarity of their expression patterns. We compared methods among K-means, QT clustering and Fuzzy Heuristic Partition [87,88] in this study. Only Fuzzy Heuristic Partition allows genes to be assigned to more than one cluster with different degrees of membership. At a very stringent membership cutoff, most of the genes in each cluster were highly correlated across all experimental conditions. As the membership cut-off decreases, additional genes were assigned to the cluster based on their expression similarity on a subset of experimental conditions. This enables the identification of context-dependent regulation. We further clustered differentially expressed genes using Fuzzy clustering by local approximation of membership algorithm [87] with parameter setting -KNN: 7; Max App: 500; Membership Range: 35%. We evaluated the clustering performance based on its ability to produce biologically meaningful clusters using the Gene Ontology database as a common reference [89,90].

Functional classification
After identifying co-expressed gene groups, we sought to identify the potential biological themes represented by these distinct gene sets. Such processes are helpful in assigning the functional linkage to gene groups and the evaluation of clustering quality. Genes in each cluster were uploaded to DAVID, a pre-compiled web-based functional annotation tool [91] for gene ontology analysis. For each GO term, a Fisher’s exact test was used to compare the occurrence of the term in the list of interest and the rest of the genome as a reference list to identify over-represented functional categories in each gene list. For genes within a cluster, Kappa similarity was measured to estimate functional similarity between genes based on the number of shared annotation terms [33]. A TF-TG Kappa similarity matrix was created with each value ranging from 0 to 1, the higher the value of Kappa, the stronger the overall agreement in annotation terms.

TF-TG Correlation
We consider TFs in a given cluster as “candidate regulators” of that cluster. The expression profile similarity between TF and genes in each cluster were calculated using Pearson Correlation and a TF-TG correlation matrix was generated with each value ranging from +1 to -1, indicating the perfect positive and negative correlation, respectively.

Identification of common TFBS motif and module
Motif search is often associated with a large number of false positive predictions due to the short and degenerate nature of many TFBS motifs. Several approaches were used to reduce false positives and improve the prediction accuracy. 1) Apply comparative genomics: Genome RVista http://genome.lbl.gov/vista/ and DiRE http://dire.dcode.org were used to identify evolutionarily conserved regulatory elements that were over-represented in our co-expressed gene clusters [28,31,92].
Both use precompiled evolutionary conserved regions (ECR) via human and mouse whole genome alignment. The locations of putative TFBSs were precomputed for each genome using vertebrate position weighted matrices from TRANSFAC matrix library version 10.2. For Genome RVista, we chose conserved TFBSs located 3 kb upstream of transcription start site with the p-value cutoff at 0.005. For DiRE, we chose conserved TFBSs from the top three conserved ECRs (which can be located in intron, UTR or intergenic regions) and the promoter ECRs. Over-represented TFBSs from both programs were combined for further analysis. 2) Search for over-represented TFBSs in proximal promoter region: since the majority of functional TFBSs are found in the promoter region of eukaryotic genomes, cis-element over-representation (Clover) [93] was used to determine the conserved TFBSs that were over-or under-represented in the given promoter set. 3) Search for Cluster and composite of TFBSs: 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 [93] was used to identify clusters of pre-specified motifs in a given gene cluster. Perl scripts were used to extract common composite sites from the motif clusters identified by Cluster-Buster algorithm. For approaches in 2) and 3), we used proximal promoter sequences of genes in the cluster of interest (1 kb up stream and 200 bp downstream of TSS, Ensembl release, version 52). We used MousePromoters_v19 from Ensembl release 19.32 as the background set, which contains 20,028 mouse promoters of the same region. 4) Both TRANSCompel database [94] and Matbase (Genomatix) contain well documented, experimentally confirmed promoter modules with synergistic, antagonistic, or additive functions. Comparison with these prior known TF modules can be used to identify and verify meaningful TFBS combinations.

The relative importance of a TFBS is determined by the average ranking order of ECR, prompter and frequency analysis. A TFBS-TG matrix was derived from promoter mining. The score between a TFBS, Ti, and a gene, Gj, is defined as \( \text{TFBS}(Ti, Gj) \equiv <0,1,2,3> \) The higher the score, the more certainty the TF-TG relationship: 0 means no evidence, 1 indicates the evidence from high throughput screen or computational prediction or gene co-citation from databases \( \leq 10 \). 2 means supporting evidence is from more than one resources and gene co-citation \( \geq 10 \). 3 means direct experimental evidence or evidence from multiple resources.

**Data Integration**

We calculated the relative confidence score of TF-TG associations by combining the data obtained. One key assumption of our integrative approach is that TGs sharing expression and functional similarity are likely to be regulated by the same TF(s), and TFs sharing expression and functional similarity are likely to form functional modules to regulate the same group of TG(s). We grouped TF using hierarchical clustering, according to an integrated matrix compiled from four types of data sources: a TFBS-TG scoring matrix, a TF-TG functional similarity matrix, a TF-TG expression correlation matrix and a TF-TG interaction matrix. Each value in the four matrices was scaled from 0 to 1 and summed into the integrated TF-TG matrix. The TF-TG matrix was further normalized and scaled between 0 and 1, denoted as \( \text{Score}(Ti, Gj) \). We grouped TGs into sub-clusters using hierarchical clustering, 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 define \( \text{Support} \) between each TF cluster \( Ct \) and each TG cluster \( Cg \) as

\[
\text{Support}(Ct, Cg) = \frac{\sum_{i=1}^{m} \sum_{j=1}^{n} \text{Score}(Ti, Gj)}{m \cdot n},
\]

where \( \text{Score}(Ti, Gj) \) is from the integrated matrix between TF and TG, \( m \) is the size of \( Ct \), \( n \) is the size of \( CgTi \) or \( Ct, Cg \).

\( \text{Support} \) describes the connectivity between a TF cluster and a gene cluster. The value of \( \text{Support} \) ranges from 0 to 1. Given a threshold of \( \text{Support} \), for instance, 0.25, satisfying TF-TG cluster pairs are extracted as correlated cluster pairs. Given a correlated cluster pair, we further define \( \text{Confidence} \) between TF-TG pairs within this cluster pair as

\[
\text{Confidence}(Ti, Gj) = \frac{\sum_{i=1}^{m} \text{Score}(Ti, Gj)}{\text{Support}(Ct, Cg) \cdot \sum_{m} \text{Score}(Ti, Gj) - \text{Score}(Ti, Gj) \cdot \sum_{m} \text{Support}(Ct, Cg) \cdot \sum_{m} \text{Score}(Ti, Gj)}.\]

**Knowledge Base and Interaction Search**

We collected the positive TF-TG relationships from: Ingenuity knowledge base (Ingenuity), Transfac 11.3 (Biobase) [94], Eldorado (Genomatix), PreMod [95], protein interaction databases HPRD [96] and BioGRID [97]. A TF-TG interaction matrix was formed from the combined resources. Interaction score is defined as \( \text{Interaction}(Ti, Gj) \equiv <0,1,2,3> \).
where $L(T_i, G_j)$ is calculated by scaling $Score(T_i, G_j)$ into $[0.5, 1]$. $l(T_i)$ is normalized relative TF importance ranging from $[0.8, 1.2]$. $m$ is the size of $C_T$, $n$ is the size of $C_g$, $T_i \subseteq C_T$ and $G_j \subseteq C_g$, and $C_T$ and $C_g$ are in a cluster pair passed Support cutoff. All factors are equally weighted in the equation.

Confidence describes the possibility of a true positive TF-TG relationship according to the integrated information. The first factor of confidence $(T_i, G_j)$ denotes the connectivity between a $T_i$ from a cluster $C_T$ and all genes in a cluster $C_g$, the second factor measures the connectivity between a $G_j$ from cluster $C_g$ and all TFs in cluster $C_T$, the fourth factor implies the connectivity between $T_i$ and $G_j$, and the fifth factor $l(T_i)$ denotes the relative importance of $T_i$ in our analysis. We rank TFBS-TG pairs based on the normalized confidence score for each TF-TG pair. The TFBS-TG pairs with the highest confidence scores will be selected for experimental validation. For each cluster, we generated a TF-TG association table ranked according to the confidence score. A network graph linking TFs and their TGs was generated using Cytoscape 2.6. http://www.cytoscape.org/.

Cell Culture, Transfection, and Reporter Gene Assays

The MLE-15 cell is an immortalized mouse lung epithelial cell line that maintains some of the morphological and functional characteristics of type II epithelial cells. MLE-15 cells were cultured in HITES medium [56] for functional characterization of mouse Elovl1, Slc34a2 and Zdhhc3 promoters. Approximately 1 Kb of the $5'$-upstream regulatory regions comprising the proximal promoter were PCR amplified, including the transcription start site (TSS) and a part of the $5'$-untranslated (5'-UT) region as depicted in Figure 4. The promoter fragments were confirmed by sequencing from both ends and cloned to generate promoter-luciferase vectors in pGL3-basic plasmid (Promega) and used in transient transfection assays using Fugene 6 at a DNA/Fugene ratio of 1:3 according to the manufacturer's instructions (Roche Applied Science). Briefly, 6-well plates at 30-50% confluence were transfected with a fixed amount of each promoter-luciferase plasmid and various amounts of CMV-based cDNA expressing transactivator plasmids mouse C/EBPα (kind gift from Dr. McKnight, University of Texas Southwestern Medical Center at Dallas) or human SREBP1c [98]. Total DNA was normalized with corresponding CMV-empty vectors, and transfection efficiency was normalized to β-galactosidase activity using 100 ng/well of pCMV β-galactosidase. Two days after transfection, luciferase and β-galactosidase assays were performed using 20 μl of the supernatant according to a previous protocol [55]. The light units were assayed by luminometry (Berthold Technologies GmbH & Co., Calmbach, Germany). Data obtained represent the average of three transfection experiments, each carried out in duplicate ($n = 6$) and depicted as means ± S.D. unless stated otherwise.

Availability

All published microarrays and mouse models we used in this study are listed in Table 1 with references. Unpublished microarray data used in this study are available upon request. Perl scripts for extracting results from Cluster-Buster and confident score calculation can be freely downloaded from http://research.cchmc.org/pbge/jsp/links_v2.jsp.

Additional material

Additional file 1: Data collecting and Clustering
Additional file 2: Support & Confidence Calculation For C1 Genes
Additional file 3: Support & Confidence Calculation For C2 Genes
Additional file 4: Support & Confidence Calculation For C28 Genes
Additional file 5: Top Ranked CEBP Targets According To The Integrative Score
Additional file 6: Top Ranked SREBP Targets According To The Integrative Score
Additional file 7: Top Ranked HNF3 Targets According To The Integrative Score

Abbreviations

ABC3: ATP-binding cassette sub-family A member 3; ACRYL: ATP citrate lyase; ACOX1: acyl-CoA oxidase-like; ACS52: acyl-CoA synthetase short-chain family member 2; ADORA2B: adenosine A2b receptor; AHR: aryl hydrocarbon receptor; APP: approximation steps; AQP5: aquaporin 5; BARBIE: barbiturate-inducible element; BCL6: B-cell CLL/lymphoma 6, member B; BEX2: brain expressed X-linked 2; BIOGRID: Biological General Repository for Interaction Datasets; BTG3: B-cell translocation gene 3; CEBPA: CCAAT-enhancer-binding protein alpha; CFTR: cystic fibrosis transmembrane conductance regulator; CHIP: Chromatin immunoprecipitation; CIZ: Car-associated zinc finger protein; CLOVER: cis-element over-representation; CMV: Cytomegalovirus; DAVID: Database for Annotation, Visualization and Integrated Discovery; DIRE: Distant Regulatory Elements of co-regulated genes; DLK1: delta-like 1 homolog; ECR: Evolutionarily Conserved Regions; EGF: epidermal growth factor; EGR: Early growth response; ELFS: EFα-like factor-S; ELOVL1: elongation of very long chain fatty acids-like 1; EMP2: epithelial membrane protein 2; ENAC: epithelial sodium channel; ENPP2: ectonucleotide pyrophosphatase/phosphodiesterase 2; ER: Endoplasmic Reticulum; ERMD: eTS-related molecule; ERR1: estrogen receptor related 1; ETS: erythroleukemia virus E26 oncogene homolog; ETF5: ETFα-like factor-S; ETV5: ETS variant gene 5; FABP5: fatty acid binding protein 5; FLI1: Friend leukemia integration 1; FOXA2: forkhead box A2; GATA6: GATA binding protein 6; GO: Gene Ontology; GPM6: glycerol-3-phosphate acyltransferase, mitochondrial; HES1: hairy and enhancer of split 1; HITES: hydrocortisone, insulin, transferrin, estrogen, and selenium; HNF3: Hepatocyte Nuclear Factor 3; HPRD: Human Protein Reference Database; ID2: inhibitor of DNA binding 2; IFI1: interferon regulatory factor 1; IFI2: interferon regulatory factors; JUN: v-jun sarcoma virus 17 oncogene homolog; KDR: kinase insert domain protein receptor; KNN: k-Nearest-neighbours; LEPF1: lymphoid enhancer binding factor 1; LIPG: lipase, endothelial; LMO2: LIM domain only 2 complex; LPICAT1: lysophosphatidylcholine acyltransferase 1; MEF2C: myocyte enhancer factor 2C; MLE-15: Murine lung epithelial cells; MITC2: mitochondrial carrier homolog 2; NFI: nuclear factor I; NFAT1: Nuclear factor of activated T-cells; NFATC3: nuclear factor of activated T-cells, calcineurin-dependent 3, NFE2;
nuclear factor, erythroid derived 2; NFkB: nuclear factor of kappa light polypeptide gene enhancer in B-cells; NK2X1-2: NKX2 box 1; NLP: Natural Language Processing; NOTCH1: neurogenic locus notch homolog protein 1; NPT2B: Na(+)/Pi co-transporter 2B; NR1H2/3: nuclear receptor subfamily 1 group, member 2/3; OCT1: organic cation transporter 1; PDXA: protein disulfide isomerase associated 5; POU2F1: POU domain, class 2, transcription factor 1; PPAR: peroxisome proliferator-activated receptor; PRDX6: peroxiredoxin 6; PREMO: predicted transcriptional regulatory modules; QT: Quality Threshold; RAR: retinoic acid receptor; RUNKLT1: runt-related transcription factor 1; translocated to, 1; RVISTA: Rank Vista; S.D.: Standard Deviation; SCAP: SREBP cleavage-activating protein; SCBD: stearyl-CoA-Coenzyme A desaturase; SGC81A1: secretoglobin, family 1A, member 1; SLC34A2: solute carrier family 34 (sodium phosphate), member 2; SOAT1: sterol O-acyltransferase 1; SOX9: Sex determining region Y-Box 9; SP1: Sp1 transcription factor (specificity protein 1); SPP1: secreted phosphoprotein 1; SREBP: Sterol Regulatory Element Binding Proteins; SREPN9: serotonin peptidase inhibitor, clade B (ovabumin), member 9; STATs: signal transducer and activator of transcription 6; STFPA: surfactant, pulmonary-associated protein A; STFPB: surfactant, pulmonary-associated protein B; STFPC: surfactant, pulmonary-associated protein D; STFPD: surfactant, pulmonary-associated protein D; TCF4: transcription factor 4; TF: Transcription Factor; TFB5: Transcription Factor Binding Site; TG: Target Gene; TRANFCAC: Transcriptional Factor Database; TSS: Transcription start site; TTF-1: thyroid transcription factor 1; VEGFA: vascular endothelial growth factor A; WARS: tryptophanyl-tRNA synthetase, WWTR1: WW domain containing transcription regulator 1, ZDHHC3: zinc finger, DHHC domain containing 3.

Acknowledgements

This work was supported by NHLBI Grants HL61646 (Y.X) and HL085610 (J.A.) and AHA Grant SDG-0830101N (D.V) and ACS Grant IRG-92-026-12 (LJL). We thank Dr. Raj Bhatnagar for manuscript review.

Author details

1Division of Molecular Biology, Department of Pediatrics, Cincinnati Children’s Hospital Medical Center, University of Cincinnati College of Medicine, Cincinnati, OH, USA. 2Division of Biomedical Informatics, Department of Pediatrics, Cincinnati Children's Hospital Medical Center, University of Cincinnati College of Medicine, Cincinnati, OH, USA. 3Department of Computer Science, University of Cincinnati College of Medicine, Cincinnati, OH, USA. 4Department of Biomedical Engineering, University of Cincinnati, Cincinnati, OH, USA.

Authors’ contributions

YX designed and coordinated the overall project, participated in the statistical analysis and drafted the manuscript. MZ and LJL participated in the design, drafting and computational analysis of the data integration section. YW carried out multiple data analysis and assisted manuscript preparation. PK assisted in the data analysis and manuscript preparation. VO carried out promoter reporter assays and wrote that part of the manuscript. JAW provided miRNA data, contributed to study design and to the writing and revising of the manuscript. All authors read and approved the final manuscript.

Received: 5 October 2009 Accepted: 26 July 2010

Published: 26 July 2010

References

1. Johansson J, Curstedt T: Molecular structures and interactions of pulmonary surfactant components. Eur J Biochem 1997, 244(3):675-693.
2. Whitsett JA, Weaver TE: Hydrophobic surfactant proteins in lung function and disease. N Engl J Med 2002, 347(26):2141-2148.
3. Dieder C, Krey G, Keller H, Givel F, Heftmann S, Wahli W: Control of the peroxisomal beta-oxidation pathway by a novel family of nuclear hormone receptors. Cell 1992, 68(5):879-887.
4. Horton JD, Shah NA, Warrington JA, Anderson NN, Park SW, Brown MS, Goldstein JL: Combined analysis of oligonucleotide microarray data from transgenic and knockout mice identifies direct SREBP target genes. Proc Natl Acad Sci USA 2003, 100(21):12007-12012.
5. Kliewer SA, Forman BM, Blumberg B, Ong ES, Borgmeyer U, Mangelsdorf DJ, Umesono K, Evans RM: Differential expression and activation of a family of murine peroxisome proliferator-activated receptors. Proc Natl Acad Sci USA 1996, 91(15):7355-7359.
6. Pedersen TA, Bresheechenko O, Garcia-Silva S, Ermarkova O, Kurz E, Mandrup S, Porre ET, Nerlov C: Distinct C/EBPalpha motifs regulate lipogenic and gluconeogenic gene expression in vivo. Endo J 2007, 26(4):1081-1093.
7. Rosen ED, Walkey CJ, Puigserver P, Spiegelman BM: Transcriptional regulation of adipogenesis. Genes Dev 2000, 14(11):1293-1307.
8. Lekstrom-Himes J, Xanthopoulos KG: Biological role of the CCAAT/ enhancer-binding protein family of transcription factors. J Biol Chem 1998, 273(44):26545-26548.
9. Lane MD, Lin FT, MacDougald OA, Vasseur-Cognet M: Control of adipocyte differentiation by CCAAT/enhancer binding protein alpha (C/EBP alpha). Int J Obes Relat Metab Disord 1996, 20(Suppl 3):S91-96.
10. Ray MK, Chen CY, Schwartz RJ, DeMayo FJ: Transcriptional regulation of a mouse Clara cell-specific protein (mCC10) gene by the NKx2 family members thyroid transcription factor 1 and cardiac muscle-specific homebox protein (CSX). Mol Cell Biol 1996, 16(12):2056-2064.
11. Whitsett JA, Glasser SW: Regulation of surfactant protein gene expression. Biochimica et biophysica acta 1998, 1408(2-3):303-311.
12. Martis PC, Whitsett JA, Xu Y, Peri AK, Wan H, Ikegami M: C/EBPalpha is required for lung maturation at birth. Development 2006, 133(16):1155-1164.
13. Wan H, Xu Y, Ikegami M, Stahlan MT, Kaestner KH, Ang SL, Whitsett JA: Foxa2 is required for transition to air breathing at birth. Proceedings of the National Academy of Sciences of the United States of America 2004, 101(40):14449-14454.
14. Defelice 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):35574-35583.
15. Minoo P, Hu L, Xing Y, Zhu NL, Chen H, Borok Z, Li C: Physical and functional interactions between homeodomain NK2.1 and winged helix/foreskin FOXA1 in lung epithelial cells. Mol Cell Biol 2007, 27(6):2155-2165.
16. Horton JD, Goldstein JL, Brown MS: SREBP: activators of the complete program of cholesterol and fatty acid synthesis in the liver. J Clin Invest 2002, 109(9):1125-1131.
17. Payne VA, Au WS, Lowe CE, Rahman SM, Friedman JE, O'Reilly S, Rockford JJ: C/EBP transcription factors regulate SREBP1c gene expression during adipogenesis. The Biochemical journal 2009, 425:215-223.
18. Mason RJ, Pan T, Edeen KE, Nielsen LD, Zhang F, Longmire 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):244-255.
19. Zhang F, Pan T, Nielsen LD, Mason RJ: Lipogenesis in fetal rat lung: Importance of C/EBPalpha, SREBP-1c, and stearyl-CoA desaturase. Am J Respir Cell Mol Biol 2004, 30(2):174-183.
20. Mallampalli RK, Ryan AJ, Carroll JL, Osborne TF, Thomas CP: Lipid deprivation increases surfactant phosphatidylcholine synthesis via a sterol-sensitive regulatory element within the CTP:phosphocholine cytidylyltransferase promoter. The Biochemical journal 2002, 362(Pt 1):81-92.
21. Kast HR, Nguyen CM, Anisfeld AM, Ericsson J, Edwards PA: CTP: phosphocholine cytidylyltransferase, a new sterol and SREBP-responsive gene. Journal of lipid research 2001, 42(8):1266-1272.
22. Ryan AJ, McCoy DM, Mathur SN, Field FJ, Mallampalli RK: Lipoprotein deprivation stimulates transcription of the CTP:phosphocholine cytidylyltransferase gene. Journal of lipid research 2001, 41(8):1268-1277.
23. Kapdan N, Moore IK, Fondufe-Mittendorf Y, Gosssett AJ, Tillo D, Field Y, LeProust EM, Hughes TR, Lieb JD, Widom J, Segal E: The DNA-encoded nucleosome organization of a eukaryotic genome. Nature 2007, 458(7238):362-366.
24. Mikkelsen TS, Ku M, Jaffe DB, Bachurski CJ, Clark JC, Whitsett JA: Functional interactions between homeodomain Foxa2 and winged helix/foreskin FoxA1 in lung epithelial cells. Mol Cell Biol 2007, 27(6):2155-2165.
25. Boulika T: A compilation and classification of DNA binding sites for protein transcription factors from vertebrates. Crit Rev Eukaryot Gene Expr 1994, 4(2-3):117-321.

26. Lemmon B, Tjian R: Orchestrated response: a symphony of transcription factors for gene control. Genes Dev 2000, 14(20):2551-2564.

27. Meintjes S, Tjian R: Transcriptional selectivity of viral genes in mammalian cells. Cell 1986, 46(7):795-805.

28. Loos GG, Ozvarehno I, Pachter L, Dubchak I, Rubin EM: Vista for comparative sequence-based discovery of functional transcription factor binding sites. Genome Res 2002, 12(5):832-839.

29. Oettel JM, Wollberg TM, Muzmi DM, Miller W, Gibas RA, Belmont JW: Large-scale comparative sequence analysis of the human and murine Bruton's tyrosine kinase loci reveals conserved regulatory domains. Genome Res 1997, 7(4):315-329.

30. Wasserman WW, Palumbo M, Thompson W, Fickett JW, Lawrence CE: Human-mouse gene comparisons to locate regulatory sites. Nucleic Acids Res 2000, 28(2):225-228.

31. Couronne O, Poliakov A, Bray N, Ishkhavan T, Rybov D, Rubin E, Pachter L, Dubchak I: Strategies and tools for whole-genome alignments. Genome Res 2003, 13(1):73-80.

32. Frith MC, Fu Y, Yu L, Chen JP, Hansen U, Weng Z: Detection of functional DNA motifs via statistical over-representation. Nucleic Acids Res 2004, 32(4):1291-1381.

33. Xu Y, Mayor JA, Gremse D, Kaplan RS: High-yield bacterial expression, purification, and functional reconstitution of the tricarboxylate transport protein from rat liver mitochondria. Biochem Biophys Res Commun 1995, 207(2):783-789.

34. Allocco DJ, Kohane IS, Butte AJ: Quantifying the relationship between co-expression, co-regulation and gene function. BMC Bioinformatics 2004, 5(18):225-228.

35. Ideker T, Thorsson V, Ranish JA, Christmas R, Buhler J, Eng JK, Bumgarner R, Dubchak I: Comparative sequence-based discovery of functional transcription factor binding sites. Genome Res 2000, 13(1):73-80.

36. Fleishman J, Hasty P: Thyroid hormone strongly activates the human ABCA3 promoter in concert with thyroid transcription factor 1. J Biol Chem 2002, 277(11):929-934.

37. Xu Y, Mayor JA, Gremse D, Kaplan RS: High-yield bacterial expression, purification, and functional reconstitution of the tricarboxylate transport protein from rat liver mitochondria. Biochem Biophys Res Commun 1995, 207(2):783-789.

38. Allocco DJ, Kohane IS, Butte AJ: Quantifying the relationship between co-expression, co-regulation and gene function. BMC Bioinformatics 2004, 5(18):225-228.

39. Ideker T, Thorsson V, Ranish JA, Christmas R, Buhler J, Eng JK, Bumgarner R, Dubchak I: Comparative sequence-based discovery of functional transcription factor binding sites. Genome Res 2000, 13(1):73-80.

40. Fleischman J, Hasty P: Thyroid hormone strongly activates the human ABCA3 promoter in concert with thyroid transcription factor 1. J Biol Chem 2002, 277(11):929-934.

41. Xu Y, Mayor JA, Gremse D, Kaplan RS: High-yield bacterial expression, purification, and functional reconstitution of the tricarboxylate transport protein from rat liver mitochondria. Biochem Biophys Res Commun 1995, 207(2):783-789.

42. Allocco DJ, Kohane IS, Butte AJ: Quantifying the relationship between co-expression, co-regulation and gene function. BMC Bioinformatics 2004, 5(18):225-228.

43. Ideker T, Thorsson V, Ranish JA, Christmas R, Buhler J, Eng JK, Bumgarner R, Dubchak I: Comparative sequence-based discovery of functional transcription factor binding sites. Genome Res 2000, 13(1):73-80.

44. Fleischman J, Hasty P: Thyroid hormone strongly activates the human ABCA3 promoter in concert with thyroid transcription factor 1. J Biol Chem 2002, 277(11):929-934.

45. Xu Y, Mayor JA, Gremse D, Kaplan RS: High-yield bacterial expression, purification, and functional reconstitution of the tricarboxylate transport protein from rat liver mitochondria. Biochem Biophys Res Commun 1995, 207(2):783-789.
67. Park KS, Whittsett JA, Di Palma T, Hong JH, Yaffe MB, Zannini M. TAZ interacts with TTF-1 and regulates expression of surfactant protein-C. J Biol Chem 2004, 279(17):17384-17390.

68. Seven-Chronoziou Z, Bachinski CJ, Yan C, Whittsett JA. Regulation of mouse SP-B gene promoter by AP-1 family members. The American journal of physiology 1999, 277(1 Pt 1):L97-98.

69. Gronning LM, Tingsabadh R, Hardy K, Dalen KT, Jat PS, Grudi L, Shepherd PR. Glucose induces increases in levels of the transcriptional repressor li2α via the hexosamine pathway. Am J Physiol Endocrinol Metab 2006, 290(4):E599-606.

70. Iwano S, Nakaya M, Saito T, Asamura F, Karamata T. A possible mechanism for atherosclerosis induced by polycyclic aromatic hydrocarbons. Biochem Biophys Res Commun 2005, 335(1):220-226.

71. Moldes M, Boizard M, Liepvre XL, Feve B, Gaill T, J. Functional antagonism between inhibitor of DNA binding (Iid) and adipocyte determination and differentiation factor 1[sterol regulatory element-binding protein-1c (ADDESREBP-1c) trans-factors for the regulation of fatty acid synthase promoter in adipocytes. Biochem J 1999, 344(Pt 3):873-880.

72. She H, Xiong S, Haza S, Tsukamoto H. Adipogenic transcriptional regulation of hepatic stellate cells. J Biol Chem 2005, 280(36):4959-4967.

73. Shimomura I, Hammer RE, Richardson JA, Ikemoto S, Bashmakov Y, Goldstein JL. brown fat. Insulin resistance and diabetes mellitus in transgenic mice expressing nuclear SREBP-1c in adipose tissue: model for congenital generalized lipodystrophy. Genes Dev 1998, 12(20):3182-3194.

74. Cho SJ, Kang MJ, Horner RJ, Kang HP, Zhang X, Lee PJ, Elias JA, Lee CG. Role of early growth response-1 (Egr-1) in interleukin-13-induced inflammation and remodeling. J Biol Chem 2006, 281(12):1161-1168.

75. Tabor DE, Kim JT, Spiegelman BM, Edwards PA. Identification of conserved cis-elements and transcription factors required for sterol-regulated transcription of stearyl-CoA desaturase 1 and 2. J Biol Chem 1999, 274(29):20603-20610.

76. Nagananj R, Saren J, Le N, Araki T, Watson M, Milbrandt J. EGR2 mutations in inherited neuropathies dominantly-negatively inhibit myelin gene expression. Neuron 2001, 30(2):353-368.

77. Guthmann F, Hohl P, Fechner H, Humbert P, Borchers T, Spener F, Rustov B. Expression of fatty-acid-binding proteins in cells involved in lung-specific lipid metabolism. Eur J Biochem 1998, 253(2):430-436.

78. Nakano H, Shindou H, Hishikawa D, Hayarama T, Ogawara S, Suwabe A, Taguchi R, Shimizu T. Cloning and characterization of mouse lung-type acyl-CoA:lysophosphatidylcholine acyltransferase 1 (LPCAT1). Expression in alveolar type II cells and possible involvement in surfactant lung-specific lipid metabolism. J Biol Chem 2005, 280(36):20140-20147.

79. Chen X, Hyatt BA, Mucenski ML, Mason RJ, Shannon JM: A signed-rank test for clustered data. Biometrics 2008, 64(2):500-507.

80. Whitsett JA, Wert SE, Xu Y: Expression of fatty-acid-binding proteins in cells involved in increased pulmonary blood flow and pulmonary hypertension. American journal of physiology 2003, 285(1):222-231.

81. Dor Y, Keshet E, Lupu F, Nemery B, Dewerchin M, Van Veldhoven P, Plate K, Pathak M, Vande Werve P, Giletti MN, Menon S, Liu F, Chen L, Tait L, Whitsett JA: Functional and phenotypical comparison of myofibroblasts derived from biopsies and bronchoalveolar lavage in mild asthma and scleroderma. Respiratory research 2006, 7:11.

82. Larsen K, Malmstrom J, Lindt M, Dahlqvist R, Larsson H, Nordenberg D, Persson E, Persson JA, Menon S, Liu F, Chen L, Tait L, Whitsett JA: Functional and phenotypical comparison of myofibroblasts derived from biopsies and bronchoalveolar lavage in mild asthma and scleroderma. Respiratory research 2006, 7:11.

83. Laborda J, Sausville EA, Hoffmann T, Notario V: dkl, a putative mammalian hemeocytic gene differentially expressed in small cell lung carcinoma and neuroendocrine tumor cell line. The Journal of biological chemistry 1993, 268(36):38170-38200.

84. Compernelle V, Brussels K, Acker T, Hoet P, Twja M, Beck H, Plaisance S, Dor Y, Keshet E, Lupu F, Nemery B, Deverchin M, Van Veldhoven P, Plate K, Menon S, Linnell J, Cremolet P, Loss of HIF-2alpha and inhibition of VEGF impair fetal lung maturation, whereas treatment with VEGF prevents fetal respiratory distress in premature mice. Nature medicine 2002, 8(7):702-710.

85. Mata-Greenwood E, Meehly B, Sofer SJ, Fineman JR, Black SM: Expression of VEGF and its receptors Flt-1 and Flk-1/KDR is altered in lambs with increased pulmonary blood flow and pulmonary hypertension. American journal of physiology 2003, 285(1):222-231.