



# Gene Regulation Network Based Analysis Associated with TGF-beta Stimulation in Lung Adenocarcinoma Cells

Lin Hua<sup>1,2\*,#</sup>, Hong Xia<sup>1,2\*,#</sup>, Wei-ying Zheng<sup>1,2</sup>, Li An<sup>3</sup>

<sup>1</sup>School of Biomedical Engineering, Capital Medical University, Beijing, 100069, China

<sup>2</sup>Beijing Key Laboratory of Fundamental Research on Biomechanics in Clinical Application, Capital Medical University, Beijing, 100069, China

<sup>3</sup>Beijing Institute of Respiratory Medicine, Beijing Chao-Yang Hospital, Capital Medical University, Beijing, 100028, China

\*Corresponding author: Lin Hua, School of Biomedical Engineering, Capital Medical University, Beijing, 100069, China.

Tel: +86-10-83911567; Fax: +86-10-83911552; E-mail: hualin7750@ccmu.edu.cn

# These authors contributed to this work equally

Received: 30 July 2015; Revised: 10 Sep. 2016; Accepted: 8 March 2017; Published online: 28 May 2017

**Background:** Transforming growth factor (TGF)- $\beta$  is over-expressed in a wide variety of cancers such as lung adenocarcinoma. TGF- $\beta$  plays a major role in cancer progression through regulating cancer cell proliferation and remodeling of the tumor micro-environment. However, it is still a great challenge to explain the phenotypic effects caused by TGF- $\beta$  stimulation and the effect of TGF- $\beta$  stimulation on tumor micro-environment.

**Objectives:** To address this issue, in the present study we used two time-course microarray data in human lung adenocarcinoma cells and applied bioinformatics methods to explore the gene regulation network responding to TGF- $\beta$  stimulation in lung adenocarcinoma cells.

**Materials and Methods:** The time-dependent reverse-engineering method, protein-protein interaction network analyses, and calculation of the similarity measures between the links were used to construct gene regulatory network and to extract gene clusters.

**Results:** Utilizing the constructed gene regulation network, we predicted NEFL and LUC7A show the opposite and the same change with C21orf90 if HAND2 is knocked-out after treatment with TGF- $\beta_1$  for 4 hours and for 12 hours respectively. FGG and HSPC009 are predicted to display the opposite change with NEFL if CSMD1 is knocked out after treatment with TGF- $\beta_1$  for 12 hours. Additionally, by integrating two datasets, we specially identified several nested clusters which included those genes regulated by TGF- $\beta$  stimulation in lung adenocarcinoma cells.

**Conclusions:** Our analysis can help a better understanding regarding how TGF- $\beta$  stimulation causes the expression change of a number of the genes and provide a novel insight into TGF- $\beta$  stimulation effect on lung adenocarcinoma cells.

**Keywords:** Gene regulation network; TGF- $\beta$ ; Lung adenocarcinoma cell; Differentially expressed genes.

## 1. Background

Transforming growth factor (TGF)- $\beta$  plays a major role in the initiation and progression of cancers. In the earlier stages, TGF- $\beta$  works as a tumor suppressor via inhibiting cell growth and apoptotic induction in the epithelial cells (1). On the contrary, in the later stages, the epithelial cells will become refractory to the growth inhibitory effect of TGF- $\beta$  which acts as a tumor promoter, increasing the tumor-promoting activity, cause the invasiveness, and metastasis (1, 2). Several previous observations have approved that the normal epithelial cells show differential response to TGF- $\beta$  stimulation as compared

to the tumor cells (3). In response to TGF- $\beta$  stimulation, tumor cells display an increased production of proteases and down-regulation of the inhibitors of the proteases, whereas this is not observed in the normal cells (4). Recent studies have found some novel regulation relationships between TGF- $\beta$  and genes in lung tumor cells. For example, Wang *et al.* have found that TGF- $\beta$  regulates the proliferation of lung adenocarcinoma cells by inhibiting PIK3K3 expression (5). Yu *et al.* unveiled a novel link between TGF- $\beta$  and Rac1. They considered the atypical Rac1 activator DOCK4 as a key component of the TGF- $\beta$ /Smad pathway that promotes

lung adenocarcinoma cell extravasation and metastasis (6). Risolino *et al.* have found that the transcription factor PREP1 induces Epithelial–mesenchymal transition (EMT) and metastasis by controlling the TGF- $\beta$ –SMAD3 pathway in non-small cell lung adenocarcinoma (7). However, it is not clear if various actions of the TGF- $\beta$  on the normal and lung tumor cells are due to differential gene regulations.

We know that gene regulatory networks have an important role in each life processes including cell differentiation, cell metabolism, cell cycle, and signal transduction (8). For example, De and Berx have claimed that the EMT-associated reprogramming of the cells not only suggests that fundamental changes in several regulatory networks might occur but also that an intimate interplay exists between such networks. Disturbance of the controlled epithelial cells' reproduction balance is triggered by altering several layers of regulation (9). Meng *et al.* constructed the lung adenocarcinoma related regulatory network using microarray data and found that FLI and TAL1 promote TGFBR and KDR expression respectively; the result of which is activation of the TGF- $\beta$  signaling pathway (10). Genovese *et al.* uncovered a novel regulation of TGF- $\beta$  signaling via a Smad4 transcriptomic network by miR-34a through constructing a network model based on the complex multidimensional cancer genomic data (11). Vilar *et al.* developed a concise computational model of the TGF- $\beta$  pathway and showed that the first layer of communication with the environment, the ligand-receptor network, is not merely a passive transducer of the signals but rather embeds properties that make it a signal processing unit (12). Therefore, these evidence support the identification of the intricate interplay between genes responsible for the observed phenotypes based on the gene regulation network, and will help to understand how TGF- $\beta$  stimulation affects the biological change of normal cells or tumor cells.

## 2. Objectives

In the present study, we used two time-course microarray data in human normal lung epithelial cell and lung adenocarcinoma cell and applied novel bioinformatics methods to explore the gene regulation networks associated with TGF- $\beta$  stimulation in two different cell lines. We predicted the network change when several genes regulated by TGF- $\beta$  are knocked-out. Our analysis can help to understand better how TGF- $\beta$  causes the expression change of other genes and gives an insight into TGF- $\beta$  effect on lung adenocarcinoma cells, as well as the development of the more effective lung adenocarcinoma treatment strategies.

## 3. Materials and Methods

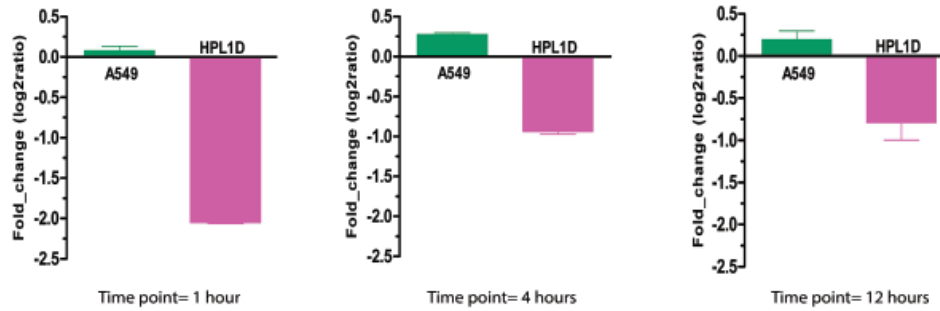
### 3.1. TGF- $\beta$ Regulated Gene Profiling in the Normal Lung Epithelial Cells and Lung Carcinoma Cells

#### 3.1.1. Data Description

To explore the network change when genes regulated by TGF- $\beta$  are knocked-out, we selected the gene expression profiling regulated by TGF- $\beta$  in the normal lung epithelial cells (HPL1D) and lung carcinoma cells (A549). HPL1D and A549 cells were treated with TGF- $\beta_1$  for 1, 4, and 12 hours, the total RNAs were extracted and were used for microarray using human 19 k arrays (4). This dataset was downloaded from Gene Expression Omnibus database (<http://www.ncbi.nlm.nih.gov/geo/>) (accession No. GSE7436). Genes which show log 2 ratios greater than 0.37 (1.3 fold induced) or less than -1.5 (0.3 fold induced) at any one of the time points are considered as up-regulated by TGF- $\beta$  or down-regulated by TGF- $\beta$ , respectively. To help observe the data distribution, we used the supra-hexagonal map (13) to display the samples' characteristics. The supra-hexagonal map provides a choice for calculating the covariance matrix based on a variety of different distance metrics. It converts the gene-sample matrix into the codebook matrix and genes with similar data patterns were taken as the same or nearby nodes in the map by applying a self-organizing learning algorithm for the symmetric topology of the supra-hexagonal map. We observed the obvious gene expression difference between normal lung epithelial cells (HPL1D) and lung carcinoma cells (A549) at different time points (supplementary Fig. 1).

#### 3.1.2. Differentially Expressed Genes Filtration under Different Time Points

In this analysis, we used R-bioconductor limma package (<http://www.bioconductor.org>) to select differentially expressed genes under different time points between normal lung epithelial cells (HPL1D) and lung carcinoma cells (A549). To select those genes that significantly distinguish the normal lung epithelial cells from the lung carcinoma cells as well as avoiding the constructed network complication which may cause the network can not be well explained; therefore, we kept the top 20 most significant differentially expressed genes distinguishing HPL1D from A549 at each time point. In addition, we also selected the top 20 ranked genes with the global differentially expression across all time points (supplementary Table 1). Finally, all of the reserved differentially expressed genes were used for further network construction.



**Figure 1.** The Fold\_change (log2ratio) of the HAND2 at three-time points. HAND2 displayed an obvious up-regulation trend (log2ratio>0) in the lung carcinoma cell (A549) and down-regulation trend (log2ratio<0) in the normal lung epithelial cells (HPL1D).

### 3.1.3. Construction of the Gene Regulation Network

In the present study, we used the time-dependent reverse-engineering method to construct gene regulatory network. This method relies on a Lasso penalized estimation of the linear regression model (14). Suppose the linear regression model is:

$$y_i = \sum_{j=1}^p \beta_j x_j + \eta_i (i=1,2,\dots,N) \quad (1)$$

Where  $y_i$  is the response, while  $x_{i.} = (x_{i1}, \dots, x_{ip})^T$  are the predictors, and  $\eta_i$  is a noise following some probabilistic distributions. Assume that the predictors are standardized and that the response is centered, the Lasso penalized estimation is then given by:

$$\hat{\beta}(\lambda) = \arg \min_{\beta \in R^p} \left[ \sum_{i=1}^N (y_i - \sum_{j=1}^p \beta_j x_j)^2 + \lambda \|\beta\|_1 \right] \quad (2)$$

Where  $\lambda$  is a non-negative scalar that determines the level of the constraints. Based on this method, the time dependent reverse-engineering method for constructing gene regulation network was described as followings simply (14):

Suppose that we have selected  $N$  genes across  $T$  time points and for  $P$  subjects; we define  $x_{npt}$  is the expression of gene  $n$  for individual  $p$  at the time-point  $t$ . Since each gene is considered as a response variable, therefore the model is composed of  $N$  linear regression models. It is defined:

$$\tilde{x}_{n.} = \begin{pmatrix} \tilde{x}_{n1.} \\ \dots \\ \tilde{x}_{nP.} \end{pmatrix}, \quad \hat{x}_{n.} = \begin{pmatrix} \hat{x}_{n1.} \\ \dots \\ \hat{x}_{nP.} \end{pmatrix} \quad (3)$$

Where

$$\tilde{x}_{\#.} = \begin{pmatrix} \tilde{x}_{\#it_2} \\ \dots \\ \tilde{x}_{\#it_T} \end{pmatrix}, \quad \hat{x}_{\#.} = \begin{pmatrix} \hat{x}_{\#it_1} \\ \dots \\ \hat{x}_{\#it_{T-1}} \end{pmatrix}, i=1,2,\dots,P \quad (4)$$

$x_{nit}$  is the expression of gene  $n$  for individual  $i$  at time-point  $t$ .  $\tilde{x}_{n.}$  is the regulated gene and  $\hat{x}_{n'..} (n'=1,2,\dots,N)$  are the regulators. The linear regression model is described as the following equation:

$$\tilde{x}_{n.} = \sum_{n'=1}^N F_{m(n')m(n)} w_{n'n} \hat{x}_{n'..} + \varepsilon_n \quad (5)$$

Where  $m(n)$  is the function that maps gene  $n$  to its time-cluster and  $F_{m(n')m(n)}$  is a  $T-1$  square matrix that describes the action of the genes.  $w_{n'n}$  is the strength of the connection from gene  $i$  toward gene  $j$ .  $\varepsilon$  is a noise vector of length  $T-1$  with  $E(\varepsilon) = 0$  and  $Var(\varepsilon) = \sigma^2$ . The Lasso estimate for linear regression is obtained with the following formula:

$$(\hat{w}, \hat{F}) = \arg \min_{\substack{w_{n'n} \in R, 1 \leq n', n \leq N \\ F_b \in M_{T-1}(R) 1 \leq a, b \leq T}} \left[ \sum_{n=1}^N (\tilde{x}_{n.} - \sum_{n'=1}^N F_{m(n')m(n)} w_{n'n} \hat{x}_{n'..})^2 \right] \quad (6)$$

Where  $n=1,2,\dots,N, \sum_{n'=1}^N w_{n'n} \leq \lambda_n$ .  $\lambda_n$  is a non-negative scalar that determines the level of the constraints. At each step, the estimation of matrices  $F_{m(n')m(n)}$  is done several times throughout the cross-validation. In this analysis, the reserved top ranked differentially expressed genes were used to construct gene regulation network. The Cascade package (15) of R software (<http://www.r-project.org>) was used to implement this program.

## 3.2. TGF- $\beta$ Regulated Gene Expression Profiling in the Lung Carcinoma Cells

### 3.2.1. Data Description

To further explore the gene regulation relationships associated with TGF- $\beta$  stimulation in the lung carcinoma cells, we performed the additional analysis for another time course microarray data. We downloaded this

dataset from Gene Expression Omnibus (accession No. GSE17708 (16)). This dataset was obtained from human A549 lung adenocarcinoma cells undergoing TGF- $\beta$ -induced epithelial-mesenchymal transition (EMT). A549 lung adenocarcinoma cell line was treated with 5 ng.mL<sup>-1</sup> TGF- $\beta$  for 0, 0.5, 1, 2, 4, 8, 16, 24, and 72 hours to induce EMT. Samples were assayed using Affymetrix HG\_U133\_plus\_2 arrays with 54,675 probe-sets.

### 3.2.2. Filtration of Differentially Expressed Genes at Different Time Points

For GSE17708 dataset, at each time point, excluding 0 hours, the differentially expressed genes were selected for their ability that distinguishes the expression at this time point from the expression at 0 hour. In the previous analysis, the genes with p-value <0.001 and >2-fold change at each time point were considered as differentially expressed genes (17). We kept these genes for further analysis, as well.

### 3.3. TGF- $\beta$ Associated Gene Clusters Identification in Lung Adenocarcinoma Cells; an Integration of the Two Datasets

To understand the functional gene clusters (i.e., networks) associated with TGF- $\beta$  in lung adenocarcinoma cells, we mapped the differentially expressed genes extracted from the two datasets; A)- the filtered differentially expressed genes for GSE7436 and B)- the filtered differentially expressed genes for GSE17708, to protein-protein interaction (PPI) networks using the STRING database (<http://string-db.org>) which is a database of known and predicted protein interactions. To avoid those pairs with a lower association to the lung adenocarcinoma into the analysis, we filtered the gene pairs which only include those genes related to the lung cancer using Online Mendelian Inheritance in Man (OMIM) database (18). In the current study, we consider a gene network to be a set of closely interrelated links. By calculating the similarity measures between links, we can determine the expected amount of overlap clusters around a gene. A gene belongs to multiple clusters means that this gene is important in the gene regulation network. The program was implemented with Linkcomm package (19) of R software (<http://www.r-project.org>).

## 4. Results

### 4.1. TGF- $\beta$ Regulated Gene Expression Profiling in Normal Lung Epithelial Cells and Lung Carcinoma Cells (for GSE7436 Dataset)

#### 4.1.1. Extracting the Significant Differentially

#### Expressed Gene

For each time point, we selected the top 20 most significant differentially expressed genes between normal lung epithelial cells and lung carcinoma cells. We also kept the top 20 ranked genes with the global differentially expression. Finally, the total of 80 differentially genes was identified. Among the global differentially expressed genes, HAND2 displayed the obvious down-regulation in the normal lung epithelial cells after treatment with TGF- $\beta_1$  for 1 hour (Fig. 1). Although HAND2 did not display an obvious up- or down-regulation in the lung carcinoma cell at three-time points, the trend of up-regulation seems obvious ( $\log_2\text{ratio}>0$ ). There are some studies that have demonstrated HAND2 over-expression in the lung squamous cell carcinomas and de-regulated in the histological subtypes (4). Recent evidence has suggested that HAND2 methylation is a common and crucial molecular alteration in several types of cancers, and could be employed as a potential biomarker for the early detection as well as a predictor of the treatment response (20).

#### 4.1.2. Construction of the Gene Regulation Network Associated with TGF- $\beta$ Stimulation

We used the identified 80 differentially expressed genes to construct gene regulation network associated with the TGF- $\beta$  stimulation. Considering that a larger number of edges in the network makes it difficult to interpret the relationships between genes, therefore, we selected a cutoff to simplify the network. In order to choose the best cutoff, we used the evolution method which allows us to see the evolution of the network when the cut off is growing up. At each step, the positions of the genes are re-calculated. The current choice of cutoff relies on a p-value which corresponds to the adequacy of the data to a power law distribution. The p-value should ensure the scale-freeness of the network is reliable (14, 21). Finally, a cutoff of 0.10 was selected to filter the network (supplementary Fig. 2). The filtered network is shown in the supplementary Figure 3.

#### 4.1.3. Prediction of Gene Expression Modulations after a Knocked-out Experiment

Next, we wanted to predict the changed genes' regulations if expressions of some genes are knocked out. We found when HAND2 is knocked-out following to the 4 hours treatment with TGF- $\beta_1$ , NEFL, and LUC7A show the opposite change trend with C21orf90 which is up-regulated (Fig. 2A). When HAND2 is knocked out after treatment with TGF- $\beta_1$  for 12 hours, NEFL, LUC7A, C21orf90, and HSPC009 displayed



the same change trend (Fig. 2B). Meanwhile, we found C21orf90 to show a different expression trend after knocking out of HAND2 at different time points. Although few studies have confirmed the direct association between C21orf90 and lung carcinoma, the expression of C21orf90 was approved to be related to the lung parenchyma which is affected by the abnormal inflammatory immune response (22). When CSMD1 was knocked out following to 12 hours of TGF- $\beta$ , FGG and HSPC009 displayed the opposite change compared to the NEFL. FGG and HSPC009 were all up-regulated noticeably (Fig. 2C). We know that CSMD1 is a tumor suppressor, and a previous array-based comparative genomic hybridization (aCGH) analysis detected the loss of CSMD1 in lung squamous cell carcinomas (23). From the analysis, we can observe that knocking-down of CSMD1 causes the change in the FGG expression. Although the current evidence can not support the regulation relationships between CSMD1 and FGG, it has been reported that the expression of FGG changed during EMT of lung cancer by several genes such as FOXA1 knockdown in A549 cells (24). Therefore, these potential findings need to be validated by more molecular biology experiments in future studies.

#### 4.2. Identification of Gene Clusters Associated with TGF- $\beta$ in Lung Adenocarcinoma Cells (Integration of Two Datasets)

After 2,714 differentially expressed genes which include 80 differentially expressed genes extracted from GSE7436 and 2,634 differentially expressed genes extracted from GSE17708 were mapped to the protein-protein interaction network, 273 gene pairs which also include genes related to the lung cancer were filtered. Then we used Jaccard coefficient to cluster links between these genes. When cutting the dendrogram at a point (height = 0.5972) where the partition density is maximized, 17 gene clusters were identified (supplementary Fig. 4). Meanwhile, the number of nodes (genes) in the largest gene cluster is 19.

We observed 6 genes that belong to more than five gene clusters (Fig. 3A and Fig. 3B), including JUN (belongs to 9 clusters), VEGFA (belongs to 8 clusters), IL6 (belongs to 8 clusters), EGFR (belongs to 7 clusters), TGFB1 (belongs to 6 clusters), and EGR1 (belongs to 5 clusters). It is noted that these genes were all associated with the TGF- $\beta$  induction and activity in the lung cancer. For example, it has been reported that TGF- $\beta$  is the major inducer of the interleukin-6 (IL-6) and vascular endothelial growth factor (VEGF), and the increased production of TGF- $\beta$  is followed by the

increased IL-6 and VEGF secretion related to tumor cell proliferation (25). In addition, Finocchiaro *et al.* have suggested the role of TGFB as a mediator of the intrinsic resistance to EGFR tyrosine kinase inhibitors in non-small cell lung cancer (NSCLC) patients (26). As well, genes that belong to multiple gene clusters are possible to be discovered in the sets of genes that belong to clusters that are entirely nested within a larger cluster of the genes. For example, gene A, B C involved in a cluster (green color) are entirely nested within the larger cluster including gene A, B, C and D (brown color) (Fig. 3C). Therefore, genes involved in those nested clusters should be noted for their potential interaction effect with other genes. In the present study, we acquired some nested gene clusters. For example, Figure 3D shows the genes involved in the two nested gene clusters. We found that JUNB, EGR2, ZFP36, IRF1, IRF2, and MYD88 all are involved in the nested gene clusters.

#### 4.3. Gene Set Enrichment Analysis (Integration of Two Datasets)

To explore the potential biological knowledge of the identified gene sets or gene networks, the total of 2,714 differentially expressed genes extracted from the two datasets were uploaded to ConceptGen software (16) which offers over 20,000 concepts comprising 14 different types of biological knowledge for performing the enrichment analysis. The significance of the over-representation is measured by a modified Fisher's exact test (p-value) and q-values which take into account the estimated proportion of false positives incurred based on p-values. By default, only concepts with q-values < 0.05 are displayed in this analysis. The results can partly reflect TGF- $\beta$  induction and the complexity of EMT process. For example, the protein interactions among the differentially expressed genes link the TGFB1, TGFB2, and TGFB3 interactions directly (Fig. 4A). Moreover, we found these differentially expressed genes set also links JUN interactions. It has been reported that dysregulated c-jun expression may be involved in the acquisition of anchorage independence in the process of human lung carcinogenesis (27). Transcription factors interactions among the differentially expressed genes link the transcription factor Egr-1 and p53 (Fig. 4B). Previous studies have found that Egr-1 mediates the stimulation of collagen transcription elicited by TGF- $\beta$ . Also, it was found that TGF- $\beta$  is necessary for the development of the pulmonary fibrosis (28). It has been reported that TGF- $\beta$  causes a time- and dose-dependent increase in Egr-1 protein and mRNA levels as well as enhancement of the Egr-1 gene transcription

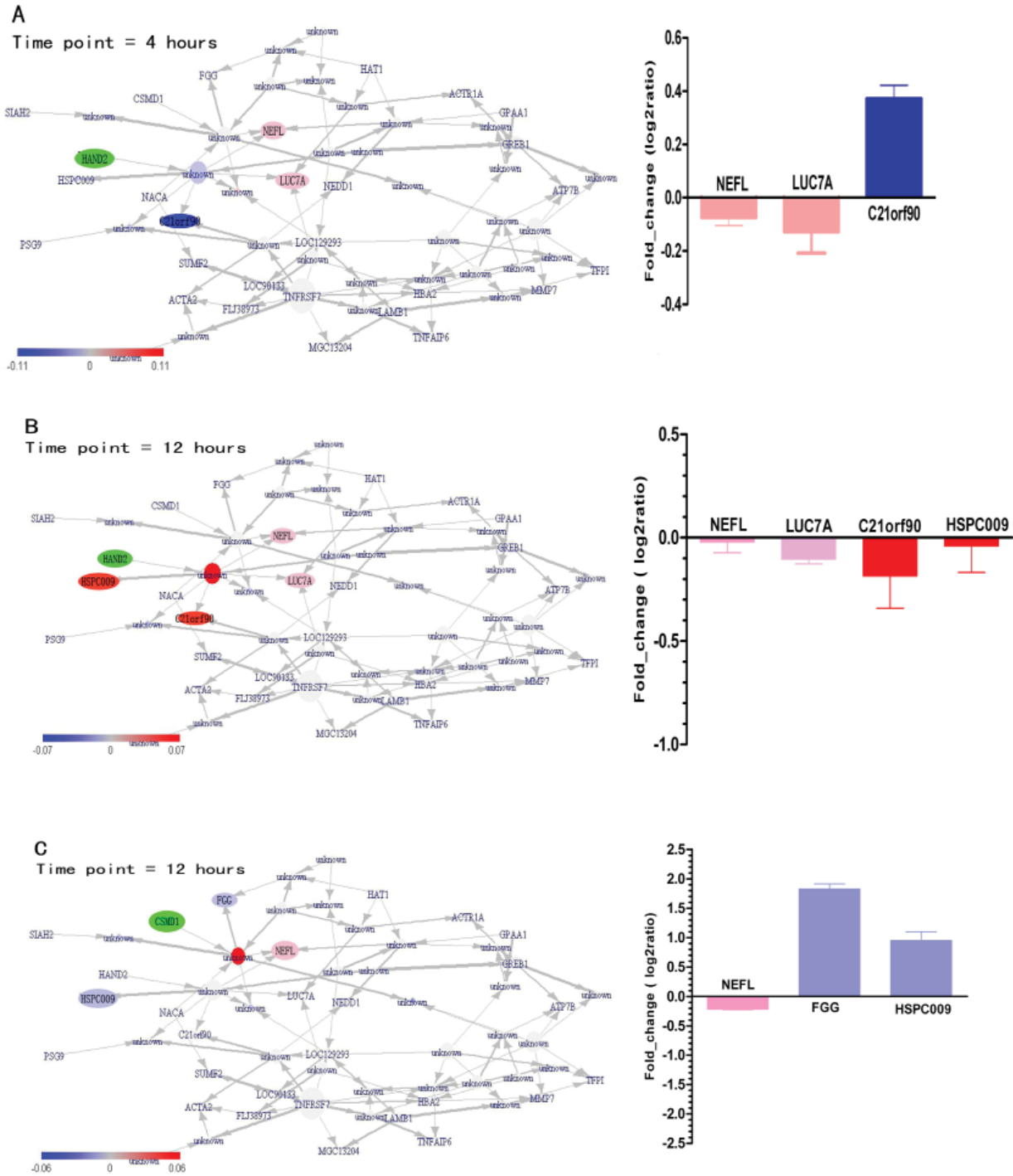
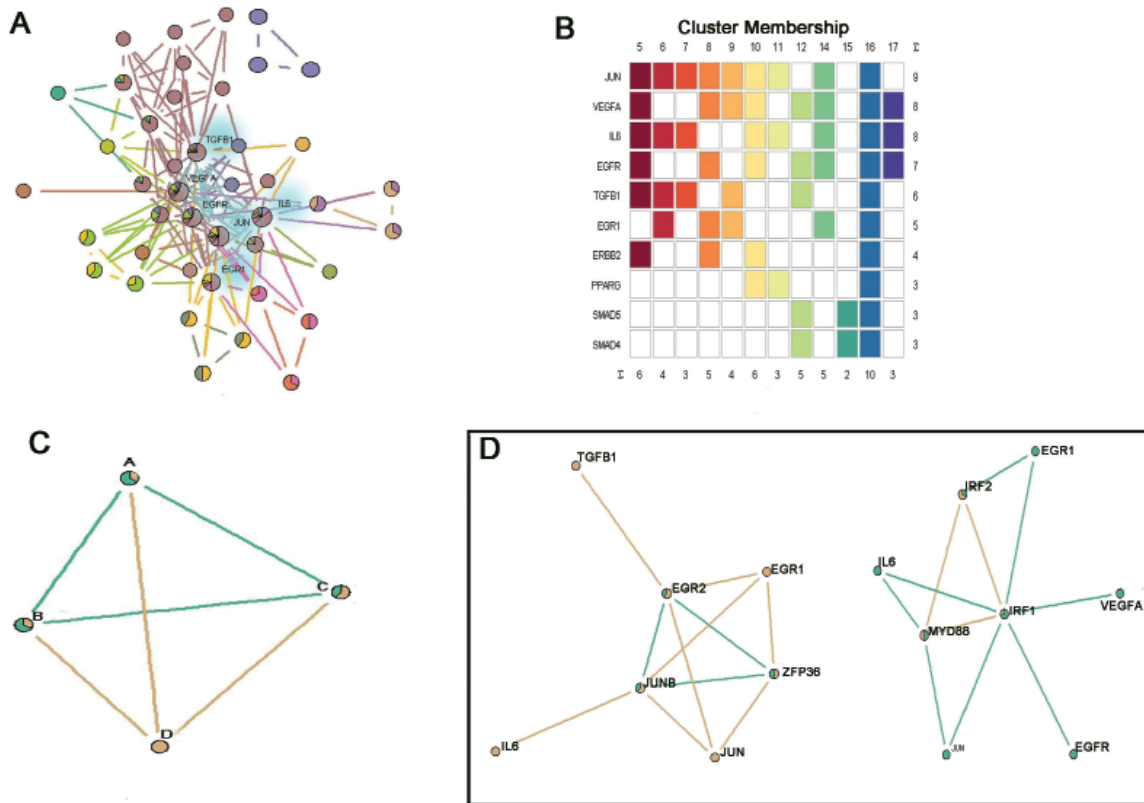


Figure 2. The changed gene expressions when special genes are knocked out after treatment with TGF-beta1. (A) The changed genes' regulations when HAND2 is knocked out after treatment with TGF- $\beta_1$  for 4 hours. When HAND2 (colored by green) is knocked out, NEFL and LUC7A (colored by pink) show the opposite change trend with C21orf90 (colored by deep blue). Meanwhile, C21orf90 is up-regulated ( $\log_2\text{ratio} > 0.37$ ). (B) The changed gene expressions when HAND2 is knocked out following to the treatment with TGF- $\beta_1$  for 12 hours. When HAND2 (colored by green) is knocked out, NEFL (colored by pink), LUC7A (colored by pink), C21orf90 (colored by red), and HSPC009 (colored by red) display the same changed trend. (C) The changed gene expression when CSMD1 is knocked out after treatment with TGF- $\beta_1$  for 12 hours. When CSMD1 (colored by green) is knocked out, FGG and HSPC009 (colored by purple) have displayed the opposite change with NEFL (colored by pink). Meanwhile, FGG and HSPC009 are all up-regulated markedly ( $\log_2\text{ratio} > 0.37$ ).



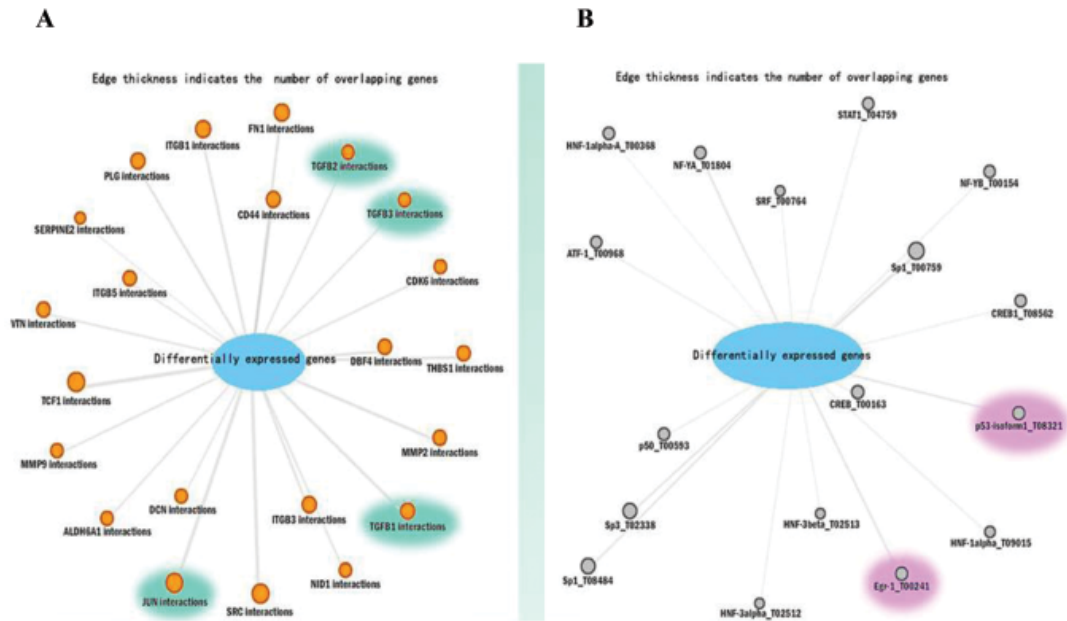
**Figure 3.** Visualization of the linked clusters. (A) Node pies graph. The fraction of the total number of edges that a gene has in each cluster is depicted using a pie chart. (B) The cluster membership for genes that belong to the most clusters. Colors indicate cluster-specific membership. Each color represents a gene cluster. For example, JUN, VEGFA, IL6, EGFR, TGFB1, and ERBB2 belong to the fifth gene cluster. (C) The graph definition for the nested clusters of genes. (D) Visualization of the two nested gene cluster structures.

via serum response elements in the normal fibroblasts (29). In addition, recent studies have shown that p53 affects TGF- $\beta$  /SMAD3-mediated signaling, cell migration, and tumorigenesis. Mutant p53 proteins also regulate Nox4-dependent signaling in TGF- $\beta$ -mediated cell motility (30). Other enrichment analyses such as those carried out based on Gene Ontology (GO) or Kyoto Encyclopedia of Genes and Genomes (KEGG) have also resulted in finding that these differentially expressed genes set are related to the TGF- $\beta$  induction and activities.

In summary, we observed NEFL and LUC7A show an opposite and the same change with C21orf90 if HAND2 is knocked-out following to TGF- $\beta_1$  treatment for 4 and 12 hours, respectively. In addition, FGG and HSPC009 display an opposite change with respect to NEFL if CSMD1 is knocked out after treatment with TGF- $\beta_1$  for 12 hours. Furthermore, by integrating the two datasets, we specifically have identified several nested clusters which include those genes regulated by the TGF- $\beta$  in the lung adenocarcinoma cells.

### 5. Discussion

It is known that TGF- $\beta$  plays important roles in cancer progression, affecting both tumor and stromal cells. Gene regulatory networks may have a potential influence on cell differentiation and cell metabolism. Most influence of the TGF- $\beta$  is brought about by regulation of gene expression. In the current study, our aim was to explore the intricate interplay between genes in response to the TGF- $\beta$  stimulation in the lung adenocarcinoma cells. Our analysis shows that when a number of genes involved in gene regulation network are knocked out, several other genes will show up-regulation or down-regulation trends. By integrating two time-course microarray data in human lung adenocarcinoma cells, we specifically have identified some nested gene clusters and found that TGFB1, EGR1, EGR2, EGFR, IL6, JUN, and JUNB all are involved in these clusters. The previous evidence has confirmed the potential regulation role of TGFB on these genes, such as EGR1, EGFR, and IL6 in the lung cancer. Taken together, our analysis can help to understand better how TGF- $\beta$  causes changes in the



**Figure 4.** Gene set enrichment analysis for 2,714 differentially expressed genes extracted from the two datasets using ConceptGen software. (A) Protein interactions among the differentially expressed genes treated with TGF- $\beta$  induction. (B) Transcription factors' interactions among the differentially expressed genes treated by TGF- $\beta$  induction.

expression of several other genes and gives an insight into TGF- $\beta$  effect on lung adenocarcinoma cells.

It should point out the limitation of our study. The shortcoming of time-course microarray data is the low sample size. Although we integrated two datasets to implement the analysis in this paper, however, this integration is not enough. In the practice, utilizing more available time-course datasets and performing the synthesized analysis such as Meta-evaluation will help to improve the reliability of the results. Moreover, our prediction for the expression change of the NEFL, LUC7A, C21orf90, FGG, and HSPC009 if HAND2 and CSMD1 are knocked-out TGF- $\beta_1$  treatment at different time points need to be validated by further molecular biology experiments in the future studies.

### Acknowledgements

This work is supported by the Beijing Natural Science Foundation (Grant Nos. 7142015), and National Science Foundation of China (Grant Nos. 31100905). This study is also funded by the foundation-clinical cooperation project of the Capital Medical University (14JL43, 16JL58 and 17JL54).

### References

1. Massagué J. TGF- $\beta$  in Cancer. *Cell*. 2008;**134**(2):215-230. DOI: 10.1016/j.cell.2008.07.001
2. Ikushima H, Miyazono K. TGF- $\beta$  signalling: a complex web in cancer progression. *Nat Rev Cancer*. 2010;**10**(6):415-424. DOI: 10.1038/nrc2853
3. Yang Y-a, Dukhanina O, Tang B, Mamura M, Letterio JJ, MacGregor J, *et al.* Lifetime exposure to a soluble TGF- $\beta$  antagonist protects mice against metastasis without adverse side effects. *J Clin Invest*. 2002;**109**(12):1607-1615. DOI: 10.1172/JCI15333
4. Ranganathan P, Agrawal A, Bhushan R, Chavalmane AK, Kalathur RKR, Takahashi T, *et al.* Expression profiling of genes regulated by TGF-beta: Differential regulation in normal and tumour cells. *BMC Genomics*. 2007;**8**(1):98 DOI: 10.1186/1471-2164-8-98
5. Wang G, Yang X, Jin Y, Deng Y, Luo X, Hu J, Wang J. TGF- $\beta$  regulates the proliferation of lung adenocarcinoma cells by inhibiting PIK3R3 expression. *Mol Carcinog*. 2015;**suppl 1**:E162-E171. DOI: 10.1002/mc.22243
6. Yu JR, Tai Y, Jin Y, Hammell MC, Wilkinson JE, Roe JS, Vakoc CR, Van Aelst L. TGF- $\beta$ /Smad signaling through DOCK4 facilitates lung adenocarcinoma metastasis. *Genes Dev*. 2015;**29**(3):250-261. DOI: 10.1101/gad.248963.114
7. Risolino M, Mandia N, Iavarone F, Dardaei L, Longobardi E, Fernandez S, *et al.* Transcription factor PREP1 induces EMT and metastasis by controlling the TGF- $\beta$ -SMAD3 pathway in non-small cell lung adenocarcinoma. *Proc Natl Acad Sci U S A*. 2014;**111**(36):E3775-E3784. DOI: 10.1073/pnas.1407074111
8. Karlebach G, Shamir R. Modelling and analysis of gene regulatory networks. *Nat Rev Mol Cell Biol*. 2008;**9**:770-780. DOI: 10.1038/nrm2503
9. De Craene B and Berx G. Regulatory networks defining EMT during cancer initiation and progression. *Nat Rev Cancer*. 2013;**13**(2):97-110. DOI: 10.1038/nrc3447
10. Meng X, Lu P, Bai H, Xiao P, Fan Q. Transcriptional regulatory networks in human lung adenocarcinoma. *Mol Med Rep*.



- 2012;**6**(5):961-966. DOI: 10.3892/mmr.2012.1034
11. Genovese G, Ergun A, Shukla SA, Campos B, Hanna J, Ghosh P *et al.* microRNA regulatory network inference identifies miR-34a as a novel regulator of TGF- $\beta$  signaling in glioblastoma. *Cancer Discov.* 2012;**2**(8):736-749. DOI: 10.1158/2159-8290.CD-12-0111
  12. Vilar JMG, Jansen R, Sander C. Signal processing in the TGF- $\beta$  superfamily ligand-receptor network. *PLoS Comput Biol.* 2006;**2**(1):e3. DOI: 10.1371/journal.pcbi.0020003
  13. Fang H, Gough J. supraHex: An R/Bioconductor package for tabular omics data analysis using a supra-hexagonal map. *Biochem Biophys Res Commun.* 2014;**443** (1):285-289. DOI: 10.1016/j.bbrc.2013.11.103
  14. Tibshirani R. Regression shrinkage and selection via the lasso: a retrospective. *J Roy Statist Soc Ser B Metho.* 2011;**73** (3):273-282. DOI: 10.1111/j.1467-9868.2011.00771.x
  15. Jung N, Bertrand F, Bahram S, Vallat L, Maumy-Bertrand M. Cascade: a R package to study, predict and simulate the diffusion of a signal through a temporal gene network. *Bioinformatics.* 2014;**30**(4):571-573. DOI: 10.1093/bioinformatics/btt705
  16. Sartor MA, Mahavisno V, Keshamouni VG, Cavalcoli J, Wright Z, Karnovsky A, *et al.* ConceptGen: a gene set enrichment and gene set relation mapping tool. *Bioinformatics.* 2010;**26**(4):456-463. DOI: 10.1093/bioinformatics/btp683
  17. Keshamouni VG, Jagtap P, Michailidis G, Strahler JR, Kuick R, Papoulias AKRP, *et al.* Temporal Quantitative Proteomics by iTRAQ 2D-LC-MS/MS and Corresponding mRNA Expression Analysis Identify Post-Transcriptional Modulation of Actin-Cytoskeleton Regulators During TGF- $\beta$ -Induced Epithelial-Mesenchymal Transition. *J Proteome Res.* 2009;**8**(1):35-47. DOI: 10.1021/pr8006478
  18. Hamosh A, Scott AF, Amberger JS, Bocchini CA, McKusick VA. Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res.* 2005;**33**(Database issue):D514-D517. DOI: 10.1093/nar/gki033
  19. Kalinka AT, Tomancak P. linkcomm: an R package for the generation, visualization, and analysis of link communities in networks of arbitrary size and type. *Bioinformatics.* 2011;**27**(14):2011-2012. DOI: 10.1093/bioinformatics/btr311
  20. Jones A, Teschendorff AE, Li Q, Hayward JD, Kannan A, Mould T, *et al.* Role of DNA Methylation and Epigenetic Silencing of HAND2 in Endometrial Cancer Development. *PLoS Med.* 2013;**10**(11):e1001551. DOI: 10.1371/journal.pmed.1001551
  21. Clauset A, Shalizi CR, Newman ME. Power-law distributions in empirical data. *Siam Rev.* 2009;**51**(4):661-703. DOI: 10.1137/070710111
  22. Wang I-M, Stepaniants S, Boie Y, Mortimer JR, Kennedy B, Elliott M, *et al.* Gene Expression Profiling in Patients with Chronic Obstructive Pulmonary Disease and Lung Cancer. *Am J Respir Crit Care Med.* 2008;**177**(4):402-411. DOI: 10.1164/rccm.200703-390OC
  23. Ma C, Quesnelle KM, Sparano A, Rao S, Park MS, Cohen MA, *et al.* Characterization CSMD1 in a large set of primary lung, head and neck, breast and skin cancer tissues. *Cancer Biol Ther.* 2009;**8**(10):907-916. DOI: 10.4161/cbt.8.10.8132
  24. Wang H, Meyer CA, Fei T, Wang G, Zhang F, Liu XS. A systematic approach identifies FOXA1 as a key factor in the loss of epithelial traits during the epithelial-to-mesenchymal transition in lung cancer. *BMC Genomics.* 2013;**14**:680. DOI: 10.1186/1471-2164-14-680
  25. Kubiczakova L, Sedlarikova L, Hajek R, Sevcikova S. TGF- $\beta$ -an excellent servant but a bad master. *J Transl Med.* 2012;**10**:183. DOI: 10.1186/1479-5876-10-183
  26. Finocchiaro G, Toschi L, Gianoncelli L, Canal B, Rubino L, Russo CL, *et al.* TGF $\beta$  expression as predictor of outcome in EGFR wild type (WT) metastatic non-small cell lung cancer (NSCLC) patients (PTS) treated with EGFR tyrosine kinase inhibitors (TKIs). *J Clin Oncol.* 2012;**30**:2012 (suppl; abstr18104).
  27. Maeno K, Masuda A, Yanagisawa K, Konishi H, Osada H, Saito T, *et al.* Altered regulation of c-jun and its involvement in anchorage-independent growth of human lung cancers. *Oncogene.* 2006;**25**:271-277. DOI: 10.1038/sj.onc.1209018
  28. Lee CG, Cho SJ, Kang MJ, Chapoval SP, Lee PJ, Noble PW, *et al.* Early Growth Response Gene 1-mediated Apoptosis Is Essential for Transforming Growth Factor  $\beta_1$ -induced Pulmonary Fibrosis. *J Exp Med.* 2004;**200**(3):377-389. DOI: 10.1084/jem.20040104
  29. Bhattacharyya S, Chen S-J, Wu M, Warner-Blankenship M, Ning H, Lakos G, *et al.* Smad-Independent Transforming Growth Factor- $\beta$  Regulation of Early Growth Response-1 and Sustained Expression in Fibrosis: Implications for Scleroderma. *Am J Pathol.* 2008;**173**(4):1085-1099. DOI: 10.2353/ajpath.2008.080382
  30. Boudreau HE, Casterline BW, Burke DJ, Leto TL. Wild-type and mutant p53 differentially regulate NADPH oxidase 4 in TGF- $\beta$ -mediated migration of human lung and breast epithelial cells. *Brit J Cancer.* 2014;**110**:2569-2582. DOI: 10.1038/bjc.2014.165