

## RESEARCH ARTICLE

# Identification of Specific Gene Modules in Mouse Lung Tissue Exposed to Cigarette Smoke

Yong-Hua Xing<sup>1,2</sup>, Jun-Ling Zhang<sup>1</sup>, Lu Lu<sup>1</sup>, De-Guan Li<sup>1</sup>, Yue-Ying Wang<sup>1</sup>, Song Huang<sup>1</sup>, Cheng-Cheng Li<sup>1</sup>, Zhu-Bo Zhang<sup>1</sup>, Jian-Guo Li<sup>1</sup>, Guo-Shun Xu<sup>1</sup>, Ai-Min Meng<sup>1,3\*</sup>

### Abstract

**Background:** Exposure to cigarette may affect human health and increase risk of a wide range of diseases including pulmonary diseases, such as chronic obstructive pulmonary disease (COPD), asthma, lung fibrosis and lung cancer. However, the molecular mechanisms of pathogenesis induced by cigarettes still remain obscure even with extensive studies. With systemic view, we attempted to identify the specific gene modules that might relate to injury caused by cigarette smoke and identify hub genes for potential therapeutic targets or biomarkers from specific gene modules. **Materials and Methods:** The dataset GSE18344 was downloaded from the Gene Expression Omnibus (GEO) and divided into mouse cigarette smoke exposure and control groups. Subsequently, weighted gene co-expression network analysis (WGCNA) was used to construct a gene co-expression network for each group and detected specific gene modules of cigarette smoke exposure by comparison. **Results:** A total of ten specific gene modules were identified only in the cigarette smoke exposure group but not in the control group. Seven hub genes were identified as well, including *Fip1l1*, *Anp32a*, *Acsl4*, *Evl*, *Sdc1*, *Arap3* and *Cd52*. **Conclusions:** Specific gene modules may provide better understanding of molecular mechanisms, and hub genes are potential candidates of therapeutic targets that may possible improve development of novel treatment approaches.

**Keywords:** Cigarette exposure - pulmonary - WGCNA - gene modules - hub gene

*Asian Pac J Cancer Prev*, 16 (10), 4251-4256

### Introduction

Cigarette smoke has multiple, highly diverse effects on human health, and leads more than 5 million deaths each year (Pirini et al., 2015). Exposure to cigarette smoke affects almost every organ and increases the risk of a wide range of diseases including cancers of esophagus, larynx, oral cavity, bladder, leukemia and pancreas in male and female smokers, and is also a major contributing cause to coronary heart disease, stroke and atherosclerosis (Alsanosy, 2014).

Particularly, cigarette smoke specifically causes respiratory tract remodeling, dysfunction and a range of respiratory diseases. Cigarette smoke is a complex mixture of over 4000 chemicals including carcinogens and toxin agents that first connect the bronchus and alveolus and cause several fatal pulmonary diseases. Polyaromatic hydrocarbons (PAHs), tobacco-specific nitrosamines (TSNAs), catechol, phenols, benzene and formaldehyde are well acknowledged as the carcinogens

that are able to induce lung cancers (de Groot and Munden, 2012). Cigarette smoking is also the main cause of chronic obstructive pulmonary disease (COPD) of which injuries include airway thickening and narrowing, irreversible airflow obstruction during expiration, mucus hypersecretion and emphysema (Moretto et al., 2012). Acrolein, one of chemical components of cigarette smoking acts as a trigger of Nucleotide Oligomerization Domain (NOD)-Like Receptors (NLRs) which can improve the allergic airway disease asthma via up-regulating the pro-inflammatory IL-1 family of cytokines, IL-18 or IL-1 $\beta$  (Kang et al., 2007). Cigarette smoking induces the severe progressive noncancerous lung disease, such as pulmonary fibrosis and emphysema by elevating ROS levels and impairing the DNA stability (Morse and Rosas, 2014). Most of these studies rely on the epidemiological studies and animal models or individual cell types *in vitro* that based on reductionist approach that resort to one pathogenic factor for one disease. It is acknowledged that reductionist study approach is over simple to clarify

<sup>1</sup>Tianjin Key Lab of Molecular Nuclear Medicine, Institute of Radiation Medicine of Chinese Academy of Medical Science and Peking Union Medical College, Tianjin, <sup>2</sup>Department of Public Hygiene, Medical College of Qing Hai University, Xining, Qinghai Province, <sup>3</sup>The institute of Laboratory Animal Science of Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China \*For correspondence: Ai\_min\_meng@126.com

the complex problem of cigarette smoke exposure on respiratory tract. Therefore, a network approach for research is necessarily required to elucidate the origin of lung disease induced by cigarette smoke.

The genome-wide expression profiling is a useful technique to clarify the systemic influence of cigarette smoke on lung tissues. Sekine T and coworkers utilized the microarray approach to simultaneously demonstrate that total particulate matter (TPM) strongly activated Nrf2 pathway-mediated anti-oxidative stress reaction, whereas gas/vapor phase (GVP) caused notable DNA damage response. And they also identified that charcoal could not modify the effects of TPM but reduce the GVP biological effects (Sekine et al., 2015). In a prospective case-control study of gene expression in the oropharynx of children exposed to secondhand smoke, researchers found 65 genes were associated with lung cancers. Therein, 24 genes were involved in cell cycle and 18 genes were related to cell growth and proliferation (Ostrower et al., 2010). Meanwhile, microarray approach may provide sufficient information about gene expression pattern changes when exposure to cigarette. The gene ontology (GO) analysis and pathway enrichment analysis are meaningful methods that highlight the molecular mechanism of lung disease induced by cigarette smoke. Aside, it is well acknowledged that biological functions arise from complex interactions network of macromolecules, and diseases are the result of dysfunction of biological macromolecules network accordingly. Based on gene expression profiling of microarray, the gene co-expression network may be established to reflect the interaction patterns of genes and put insight into molecular mechanism of pulmonary diseases caused by the cigarette smoke exposure.

This study, based on the gene expression profiling, the weighted gene co-expression network analysis (WGCNA) was utilized to construct the gene co-expression network and detect gene modules of lung tissues that were exposed to cigarette. Comparison of networks between cigarette smoke exposure mice group and control group, allow investigation of individual gene level and determine the alteration of networks attributable to cigarette smoke exposure. We hypothesized that gene modules alterations in response to cigarette smoke exposure might provide insight into the pathogenesis of adverse effects of pulmonary caused by cigarette smoke. Furthermore, identification of hub genes may provide the biomarkers for evaluation of severity of pulmonary diseases and potential therapeutic targets for driving novel treatment in lung disease induced by cigarette smoke.

## Materials and Methods

### *Microarray datasets and data processing*

The transcription profiles of GSE18344 (Affymetrix Mouse Genome 430 2.0 Array) were downloaded from Gene Expression Omnibus (GEO). The RNA was isolated from left lobes of mice (n=20) exposed to cigarette for 5 months and sham-exposed mice (n=20), and were divided into cigarette smoke exposure group and control group accordingly. The raw data were preprocessed (correction, normalization) by rma function of affy package of R 3.03

software in Bioconductor (<http://www.bioconductor.org/>) (Gentleman et al., 2004). Based on  $p \leq 0.05$  t-test statistics standard, total 8146 genes were identified as differentially expressed genes (DEs) between cigarette smoke exposure and control group. Since the network construction is a high computationally cost, only 4000 most connected genes were considered for weighted gene co-expression network construction using WGCNA package of R software for each group in this study (Zhou et al., 2014).

### *Weighted co-expression network construction for cigarette smoke exposure and control group*

The weighted adjacency matrix assesses continuous connection strength ( $[0, 1]$ ) according to  $\beta$  parameter for each group. Adhering to the scale-free topology criterion,  $\beta=12$  was considered. The co-expression matrix and the topological overlap matrix (TOM) were constructed subsequently (Zhao et al., 2014). With average linkage hierarchical clustering, gene modules were identified for each group. The intramodular connectivity of each gene was also assessed via intramodular connectivity function (Langfelder et al., 2008). The module eigengene (ME) is the first principal component of a given module, and it is used to compute module membership (MM), which evaluate the importance of genes in the network (Langfelder and Horvath, 2007).

### *Identification of hub genes*

Identification of hub genes were mainly based on the gene significance (GS), intramodular connectivity and MM. The GS value can be obtained via formula  $GS_i = -\log(p_i)$ . Where  $p_i$  value was the value of t-test. GS refers differential expression of a gene between the exposure group and control group, and reflects the correlation between a gene expression and cigarette smoke exposure. The MM can be computed by the formula  $MM_i = |\text{cor}(x_i, ME)|$ . If  $MM_i$  closer to 1, the i-th gene is more important in a given gene module. The higher intramodular connectivity suggests that a gene is more connected with other genes in gene co-expression network. A gene with high GS, MM and high intramodular connectivity is an idea hub gene in a network (Zhang and Horvath, 2005).

## Results

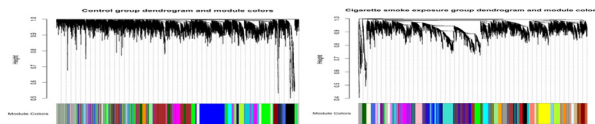
### *Weighted gene co-expression network construction and identification of specific cigarette related gene modules*

Total 8146 genes were identified as DEs by comparing the cigarette smoke exposure group with control group. Therein, 4000 DEs were considered for weighted gene co-expression network (WGCN) construction because of their highest connectivity. Finally, total 28 gene modules were identified in the control network and 38 gene modules were detected in cigarette smoke exposure network (Figure 1). Therein, the gene modules of darkmagenta, darkolivegreen, paleturquoise, plum1, saddlebrown, sienna3, skyblue3, steelblue, violet and yellowgreen were only detected in the exposure group. The ten gene modules of lung tissues were specific to cigarette stimuli. Those genes were not clustered into any modules were kept in the grey or gold modules in WGCNA package of R software,

and grey or gold module was discarded in present study.

### Identification of hub genes

These hub genes ought to be associated with the cigarette smoke exposure. Therefore, the hub genes were identified from the ten specific gene modules for cigarette smoke, and total ten hub genes were confirmed accordingly. Notably, the positive correlation between MM and GS were not observed in present study. Thus the hub genes were screened out mainly based on values of MM and intramodule connectivity (Table 1) (Figure 2).



**Figure 1. The Gene Co-Expression Modules of Control Group (left) and Cigarette Smoke Exposure Group (right).** Twenty eight modules were identified in figure 1 (left) and Thirty eight modules were detected in figure 1 (right). Colors in the bar indicate the modules

**Table 1. Hub Genes of Cigarette-Related Modules**

Module	Hub Gene	MM	Intramodular Connectivity
darkmagenta	Ythdc1	0.95	4.53
darkolivegreen	Lsm3	0.96	5.77
paleturquoise	Fip111	0.96	5.44
plum1	Anp32a	0.95	2.75
saddlebrown	Acs14	0.97	7.42
sienna3	Evl	0.94	3.77
skyblue3	Sdc1	0.95	5.01
steelblue	Sbf2	0.95	4.71
violet	Arap3	0.95	5.1
yellowgreen	Cd52	0.95	4.47

**Table 2. Top-Ranked Functional Annotations Enriched in Cigarette-Related Modules**

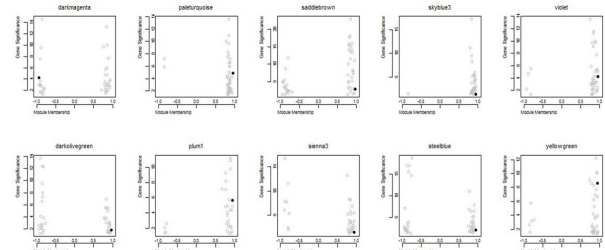
Module	GO ID	TermOntology	p-value	TermName
darkmagenta	GO:0043534	BP	1.36E-03	blood vessel endothelial cell migration
darkolivegreen	GO:0072350	BP	1.01E-03	tricarboxylic acid metabolic process
paleturquoise	GO:0010839	BP	1.36E-06	negative regulation of keratinocyte proliferation
plum1	GO:0044267	BP	1.25E-04	cellular protein metabolic process
saddlebrown	GO:0070265	BP	6.78E-05	necrotic cell death
sienna3	GO:0046631	BP	1.78E-03	alpha-beta T cell activation
skyblue3	GO:0018205	BP	7.81E-03	peptidyl-lysine modification
steelblue	GO:0044265	BP	3.53E-04	cellular macromolecule catabolic process
violet	GO:0060271	BP	1.07E-04	cilium morphogenesis
yellowgreen	GO:0050772	BP	1.25E-04	positive regulation of axonogenesis

**Table 3. Top Pathway Enrichment in Cigarette-Related Modules**

Module	Pathway Name	Genes in Category	p-Value
darkmagenta	Protein processing in endoplasmic reticulum	3	1.33E-02
darkolivegreen	Glycolysis / Gluconeogenesis	2	7.22E-03
paleturquoise	Non-homologous end-joining	1	3.20E-02
plum1	Dorso-ventral axis formation	1	4.21E-02
saddlebrown	Primary bile acid biosynthesis	2	7.92E-04
sienna3	Arachidonic acid metabolism	2	1.69E-02
skyblue3	Aldosterone-regulated sodium reabsorption	2	5.45E-03
steelblue	Terpenoid backbone biosynthesis	1	4.53E-02
violet	Ribosome biogenesis in eukaryotes	2	2.44E-02
yellowgreen	Cell cycle	4	1.09E-04

### GO and KEGG pathway enrichment analysis of cigarette-related gene modules

To analyze the relevance of module genes with their biological roles in cigarette smoke exposure, functional annotations of the gene modules were performed by using DAVID software. Total 10 specific cigarette-related modules were enriched in total 60 GO terms. Paleturquoise gene modules were involved in regulation of keratinocyte proliferation, such as negative regulation of keratinocyte proliferation, regulation of keratinocyte proliferation and keratinocyte proliferation. These biological activities were strong related to lung fibrosis (Aguilar et al., 2009). Therefore, paleturquoise module was called as keratinocyte proliferation module. The sienna3 gene module was associated with inflammation and immune response, such as alpha-beta T cell activation, homeostasis of number of cells within a tissue, cell activation, positive T cell selection, actin nucleation, myeloid dendritic cell activation, interleukin-4 production, anatomical structure development and T cell selection. The sienna3 module was regarded as inflammation module in present study. The top functional enrichment of GO annotations of these 10



**Figure 2. Correlation Between the Module Membership (x-axis) and the Gene Significance (y-axis).** The grey circle indicates gene in the module. The black dot represents the hub gene of each module

gene modules were shown in the Table 2.

Kyoto Encyclopaedia of Genes and Genomes (KEGG) (<http://www.genome.jp/kegg/>) pathway enrichment analysis were performed by using the DAVID software. Total 37 biological pathways were successfully enriched in 10 specific cigarette-related modules. For example, darkmagenta module participated in the protein processing, leucine and isoleucine biosynthesis and glycolysis / gluconeogenesis. Therefore, darkmagenta module might be involved in protein metabolism in response to cigarette stimuli. Yellowgreen module was another important because it participated in the multiple biological pathways including cell cycle, drug metabolism-cytochrome P450, phenylalanine metabolism and mismatch repair. The top-ranked pathways enriched in cigarette-related modules were shown in Table 3.

## Discussion

With systemic view, present study was designed to construct mouse lung tissues gene co-expression network and identify gene modules by transcriptome data of lung tissue under cigarette smoke exposure. Many efforts have been made to study cigarette smoke exposure of lung tissue transcriptome by using differential expression analysis which prefers to isolate a list of DEs. However this study approach ignores the facts of correlation patterns between genes. In addition, the DEs are less reliable than the hub genes of a co-expression gene network. WGCNA analysis assesses the correlation between genes and considers degree of gene shared neighbors across the whole network. Moreover, WGCNA can provide connection strength between the genes, unlike general co-expression network. In this study, total 10 gene modules were specific related to cigarette smoke exposure. Consequently, total 10 hub genes were identified due to their importance and high connectivity in network.

The ten gene modules of darkmagenta, darkolivegreen, paleturquoise, plum1, saddlebrown, sienna3, skyblue3, steelblue, violet and yellowgreen were only identified in the cigarette smoke exposure network but not in control network. Based on KEGG enrichment analysis, total 37 pathways were observed in ten gene modules that appeared only in the cigarette smoke exposure network. Protein processing in endoplasmic reticulum (ER) pathway was top enriched in the darkmagenta module. When protein misfolding in the ER may cause ER stress and activate the unfolded protein response (UPR) to ER. And both of them are documented in development of many cancers including lung cancers. UPR was demonstrated that was involved in oncogenic transformation and interaction of oncogene and tumor suppressor gene networks to modulate the lung cancer development (Yadav et al., 2014). Glycolysis pathway has been known for providing the energy for cancer cell growth and as biomarker for pneumonocyte malignant transformation (Sasaki et al., 2012). Non-homologous end-joining (NHEJ) pathway was top enriched in the paleturquoise module. NHEJ is a prominent DNA double strand break (DSB) repair pathway in mammalian cells. And NHEJ is also an error-prone pathway that occurs throughout the cell cycle, which

leads pneumonocyte genomic instability and lung cancer. Moreover, the research about anti-cancer therapy using the DSB repair inhibitor was reported already (Srivastava and Raghavan, 2015). Primary bile acid biosynthesis pathway was top enriched in saddlebrown module (Baptissart et al., 2013). Cells isolated from gastro-oesophageal reflux disease (GERD) that were exposed to bile acids over long time would induce interleukin-8 (IL-8), cyclooxygenase (COX-2), oxidative stress and DNA damage. Ultimately, esophagus adenocarcinoma would be developed (McQuaid et al., 2011). In addition, relative reports showed that bile acid biosynthesis pathway was implicated in hepato-carcinogenesis by affecting ROS levels (Yang et al., 2007). However, the relationship between primary bile acid biosynthesis pathway and lung cancer still was not reported yet. Arachidonic acid metabolism pathway was top enriched in sienna3 module. Arachidonic acid can be oxygenated by a variety of different enzymes and be converted to 5-oxoeicosatetraenoic acid (5-oxo-EETE) mediated by dehydrogenases during several steps. 5-oxo-EETE is involved in the asthma mediated by its selective OXE receptor. Moreover, 5-oxo-EETE also stimulates tumor cell proliferation and may be involved in lung cancer (Powell and Rokach, 2015). mTOR signaling pathway is significantly enriched in skyblue3 module. mTOR signaling pathway has been acknowledged that implicate in various human cancers. It can integrate the nutrient and growth factor to modulate the pneumonocyte proliferation in lung cancer (Han et al., 2013). Ribosome biogenesis in eukaryotes pathway was top enriched in violet module. Amount of ribosome affects the G1-S phase transition to regulate the cell cycle progression (Volarevic et al., 2000). Another report showed that stimulation or inhibition of rRNA synthesis led an acceleration or delay G1/S-phase progression. Therefore, ribosome biogenesis pathway can regulate pneumonocyte proliferation exposure to cigarette. Cell cycle pathway was top enriched in the yellowgreen module. Cell cycle is a tightly integrated process and is frequently aberrant in lung cancer. In brief, protein biosynthesis and processing, energy metabolism, inflammation, mTOR signaling regulation and cell cycle alteration simultaneously contribute to pulmonary diseases induced by cigarette, such as lung cancer, asthma and COPD.

There were total 10 hub genes were identified in the cigarette smoke exposure network based on their highest MM value and maximum intramodular connectivity. Based on GO analysis, paleturquoise module was top enriched in the negative regulation of keratinocyte proliferation. Fip111 were chosen as hub gene which was down regulated ( $p$ -value= 1.27E-05) in paleturquoise gene module. Tyrosine kinase - FIP1L1-PDGFRalpha is a known fusion gene that play role in chronic eosinophilic leukemia pathogenesis (Giacomini et al., 2013). A similarly study uncovered that FIP1L1/PDGFRalpha gene fusion caused hypereosinophilic syndrome (Cools et al., 2003). In common, gene fusions are ideal diagnostic markers and therapeutic targets. We hypothesized that FIP1L1/PDGFRalpha might have distinct potential in lung cancer because of its down regulation. According to GO analysis, plum1 gene module was top enriched in

cellular protein metabolic process. Anp32a was chosen as hub gene which was down regulated ( $p$ -value= 2.29E-06) in the plum1 module by comparing control group. ANP32 proteins are implicated in cell differentiation, apoptotic cell death and cell proliferation (Reilly et al., 2014). ANP32A were recognized as tumor suppressor by inhibiting cell transformation (Reilly et al., 2014). Its expression level was deregulated in the breast, prostate cancer but up regulated in colorectal and liver cancer. Aside, ANP32A was a positive prognostic marker in non-small-cell lung cancer because it promote the cancer cells apoptosis (Hoffarth et al., 2008). Therefore, Anp32a may be a potential drug target and predictor of patient survival. As for saddlebrown module, necrotic cell death was top enriched according to GO analysis. Acs14 was considered as hub gene which was up regulated ( $p$ -value= 0.001) in saddlebrown module. The highly expressed ACSL4 tend to evaluate the cancer cell invasiveness by its ability of increasing the steroid and eicosanoid amount (Watkins and Ellis, 2012). ACSL4 also promote the breast cancer cell proliferation and reduce the cell apoptosis. And ACSL4 was regarded as biomarker and mediator of aggressive breast cancer (Wu et al., 2013). The role of ACSL4 in lung cancer development was not identified yet. However, it is deserved to explore performance of ACSL4 in lung cancer for possible candidate of drug target. In skyblue3 module, peptidyl-lysine modification was top enriched based on GO analysis. Sdc1 was determined as hub gene which up regulated ( $p$ -value= 0.04) in the skyblue3 module of cigarette exposure group. In an independent study, SDC1 highly expressed in well differentiated in squamous cell lung carcinoma. It suggested that level of SDC1 was predictor of outcome of lung cancer (Anttonen et al., 2001). Highly elevation of SDC1 was also considered as biomarker for identifying patients with early stage lung tumors (Linnerth et al., 2005). In sienna3 module, alpha-beta T cell activation was top enriched based on GO analysis. Evl was supposed to be hub gene which up regulated ( $p$ -value= 0.03) in the sienna3 module. A study showed that the EVL expression was higher in advanced stages of breast tumor tissues (Hu et al., 2008). Another report indicated that over expression of EVL might induce actin polymerization, promote actin bundling and suppress breast cancer cell invasion (Mouneimne et al., 2012). The roles of EVL in pulmonary cancer were not reported yet unfortunately, and it also provides us a new research area of cancer metastasis in our coming study. As for violet module, cilium morphogenesis was top enriched based on GO analysis. Arap3 was deemed to be hub gene which down regulated ( $p$ -value=5.70E-05) in violet module. In breast cancer, ARAP3 is a known regulator which participate in rearrangements to the cytoskeleton and cell shape, which is related to cell adhesion, cell invasion, tumor proliferation and metastasis (Blighe et al., 2014). In hypotrichosis-lymphedema-telangiectasia (HLT) demonstrated that ARAP3 was a key regulator which response to Vegfc signalling in lymphatic endothelial cells, and modulated the lymphatic vascular development and pathogenesis (Kartopawiro et al., 2014). In addition, ARAP3 guards neutrophils in their quiescent state, and participates neutrophils extravasate and chemotax to sites

of damage and/or infection by effecting the PI3K-integrin-dependent processes. ARAP3-deficient neutrophils are hyper responsive in adhesion, spreading and granule release in inflammation conditions (Gambardella et al., 2011). Therefore, Arap3 may contribute to the origin of COPD and pulmonary tumor proliferation and metastasis. Yellowgreen module was top enriched in positive regulation of axonogenesis based on GO analysis. Cd52 was identified as hub gene which up regulated ( $p$ -value=2.66E-09) in yellowgreen module. Cd52 is one of innate immunity related genes that expressed on the surface of microglia specific cell, and is significantly up regulated in infection conditions (Chatterjee et al., 2014). In addition, CD52 is also expressed on the surface of malignant leukaemia cells, and it becomes the target for acute lymphoblastic leukaemia and lymphocytic leukemia therapy (Pevna et al., 2014; Ai and Advani, 2015). Therefore, it suggested that Cd52 might be involved in the inflammation or tumor development. It is deserved to identify the mechanism of Cd52 in pulmonary diseases induced by cigarette exposure.

The roles in of rest hub genes Ythdc1, Sbf2 and Lsm3 from darkmagenta, steelblue and darkolivegreen module respectively, were not documented well in the literatures currently.

In summary, pulmonary pathogenesis caused by cigarette exposure is chronic complex pathologic process that cytokines, biomolecules, toxicants and particles are simultaneously involved in. It is necessary to put insight into the mechanism of pathogenesis by using gene co-expression network analysis for identifying the biomarker of risk exposure or therapeutic targets of these diseases. Identification of specific gene modules will be valuable for elucidating the mechanism and facilitating the treatment development of pulmonary diseases induced by cigarette. There were 10 gene modules were specific in the cigarette smoke exposure group, and they were involved in multiple biological behaviors including protein process, macromolecules metabolism, DNA repair, immune regulatory, inflammation response, cell proliferation and epigenetics alterations. It suggested that alterations of roles of 10 gene modules might be helpful to reverse above pathogenesis, as well as providing candidate targets of therapy. Accordingly, seven hub genes were identified from 10 specific gene modules, including Fip111, Anp32a, Acs14, Evl, Sdc1, Arap3 and Cd52. These seven hub genes can supply valuable information about both new research direction in our future study and potential therapeutic targets for pulmonary disease induced by cigarette smoke.

## Acknowledgements

This study was supported by the National Program on Key Basic Research Project (973 Program, 2011CB964800-G) and the National Natural Science Foundation of China (No 81129020 & 81372928).

## References

Aguilar S, Scotton CJ, McNulty K, et al (2009). Bone marrow stem cells expressing keratinocyte growth factor via an

- inducible lentivirus protects against bleomycin-induced pulmonary fibrosis. *PLoS One*, **4**, 8013.
- Ai J, Advani A (2015). Current status of antibody therapy in ALL. *Br J Haematol*, **168**, 471-80.
- Alsansoy RM (2014). Smokeless tobacco (shammah) in Saudi Arabia: a review of its pattern of use, prevalence, and potential role in oral cancer. *Asian Pac J Cancer Prev*, **15**, 6477-83.
- Anttonen A, Heikkilä P, Kajanti M, et al (2001). High syndecan-1 expression is associated with favourable outcome in squamous cell lung carcinoma treated with radical surgery. *Lung Cancer*, **32**, 297-305.
- Baptissart M, Vega A, Maqdasy S, et al (2013). Bile acids: from digestion to cancers. *Biochimie*, **95**, 504-17.
- Blighe K, Kenny L, Patel N, et al (2014). Whole genome sequence analysis suggests intratumoral heterogeneity in dissemination of breast cancer to lymph nodes. *PLoS One*, **9**, 115346.
- Chatterjee D, Addya S, Khan RS, et al (2014). Mouse hepatitis virus infection upregulates genes involved in innate immune responses. *PLoS One*, **9**, 111351.
- Cools J, DeAngelo DJ, Gotlib J, et al (2003). A tyrosine kinase created by fusion of the PDGFRA and FIP1L1 genes as a therapeutic target of imatinib in idiopathic hypereosinophilic syndrome. *N Engl J Med*, **348**, 1201-14.
- de Groot P, Munden RF (2012). Lung cancer epidemiology, risk factors, and prevention. *Radiol Clin North Am*, **50**, 863-76.
- Gambardella L, Anderson KE, Nussbaum C, et al (2011). The GTPase-activating protein ARAP3 regulates chemotaxis and adhesion-dependent processes in neutrophils. *Blood*, **118**, 1087-98.
- Gentleman RC, Carey VJ, Bates DM, et al (2004). Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol*, **5**, 80.
- Giacomini CP, Sun S, Varma S, et al (2013). Breakpoint analysis of transcriptional and genomic profiles uncovers novel gene fusions spanning multiple human cancer types. *PLoS Genet*, **9**, 1003464.
- Han D, Li SJ, Zhu YT, et al (2013). LKB1/AMPK/mTOR signaling pathway in non-small-cell lung cancer. *Asian Pac J Cancer Prev*, **14**, 4033-9.
- Hoffarth S, Zitzer A, Wiewrodt R, et al (2008). pp32/PHAPI determines the apoptosis response of non-small-cell lung cancer. *Cell Death Differ*, **15**, 161-70.
- Hu LD, Zou HF, Zhan SX, et al (2008). EVL (Ena/VASP-like) expression is up-regulated in human breast cancer and its relative expression level is correlated with clinical stages. *Oncol Rep*, **19**, 1015-20.
- Kang MJ, Homer RJ, Gallo A, et al (2007). IL-18 is induced and IL-18 receptor alpha plays a critical role in the pathogenesis of cigarette smoke-induced pulmonary emphysema and inflammation. *J Immunol*, **178**, 1948-59.
- Kartopawiro J, Bower NI, Karnezis T, et al (2014). Arap3 is dysregulated in a mouse model of hypotrichosis-lymphedema-telangiectasia and regulates lymphatic vascular development. *Hum Mol Genet*, **23**, 1286-97.
- Langfelder P, Horvath S (2007). Eigengene networks for studying the relationships between co-expression modules. *BMC Syst Biol*, **1**, 54.
- Langfelder P, Zhang B, Horvath S (2008). Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. *Bioinformatics*, **24**, 719-20.
- Linnerth NM, Sirbovan K, Moorehead RA (2005). Use of a transgenic mouse model to identify markers of human lung tumors. *Int J Cancer*, **114**, 977-82.
- McQuaid KR, Laine L, Fennerty MB, et al (2011). Systematic review: the role of bile acids in the pathogenesis of gastro-oesophageal reflux disease and related neoplasia. *Aliment Pharmacol Ther*, **34**, 146-65.
- Moretto N, Volpi G, Pastore F, et al (2012). Acrolein effects in pulmonary cells: relevance to chronic obstructive pulmonary disease. *Ann N Y Acad Sci*, **1259**, 39-46.
- Morse D, Rosas IO (2014). Tobacco smoke-induced lung fibrosis and emphysema. *Annu Rev Physiol*, **76**, 493-513.
- Mouneimne G, Hansen SD, Selfors LM, et al (2012). Differential remodeling of actin cytoskeleton architecture by profilin isoforms leads to distinct effects on cell migration and invasion. *Cancer Cell*, **22**, 615-30.
- Ostrower ST, Fischer TC, Smith RV, et al (2010). Gene expression in the oropharynx of children exposed to secondhand smoke. *Laryngoscope*, **120**, 2467-72.
- Pevna M, Doubek M, Coupek P, et al (2014). Residual cancer lymphocytes in patients with chronic lymphocytic leukemia after therapy show increased expression of surface antigen CD52 detected using quantitative fluorescence cytometry. *Clin Lymphoma Myeloma Leuk*, **14**, 411-8.
- Pirini F, Guida E, Lawson F, et al (2015). Nuclear and Mitochondrial DNA Alterations in Newborns with Prenatal Exposure to Cigarette Smoke. *Int J Environ Res Public Health*, **12**, 1135-55.
- Powell WS, Rokach J (2015). Biosynthesis, biological effects, and receptors of hydroxyeicosatetraenoic acids (HETEs) and oxoeicosatetraenoic acids (oxo-ETEs) derived from arachidonic acid. *Biochim Biophys Acta*, **1851**, 340-55.
- Reilly PT, Yu Y, Hamiche A, et al (2014). Cracking the ANP32 whips: important functions, unequal requirement, and hints at disease implications. *Bioessays*, **36**, 1062-71.
- Sasaki H, Shitara M, Yokota K, et al (2012). Overexpression of GLUT1 correlates with Kras mutations in lung carcinomas. *Mol Med Rep*, **5**, 599-602.
- Sekine T, Sakaguchi C, Fukano Y (2015). Investigation by microarray analysis of effects of cigarette design characteristics on gene expression in human lung mucoepidermoid cancer cells NCI-H292 exposed to cigarette smoke. *Exp Toxicol Pathol*, **67**, 143-51.
- Srivastava M, Raghavan SC (2015). DNA Double-Strand Break Repair Inhibitors as Cancer Therapeutics. *Chem Biol*, **22**, 17-29.
- Volarevic S, Stewart MJ, Ledermann B, et al (2000). Proliferation, but not growth, blocked by conditional deletion of 40S ribosomal protein S6. *Science*, **288**, 2045-7.
- Watkins PA, Ellis JM (2012). Peroxisomal acyl-CoA synthetases. *Biochim Biophys Acta*, **1822**, 1411-20.
- Wu X, Li Y, Wang J, et al (2013). Long chain fatty Acyl-CoA synthetase 4 is a biomarker for and mediator of hormone resistance in human breast cancer. *PLoS One*, **8**, 77060.
- Yadav RK, Chae SW, Kim HR, et al (2014). Endoplasmic reticulum stress and cancer. *J Cancer Prev*, **19**, 75-88.
- Yang F, Huang X, Yi T, et al (2007). Spontaneous development of liver tumors in the absence of the bile acid receptor farnesoid X receptor. *Cancer Res*, **67**, 863-7.
- Zhang B, Horvath S (2005). A general framework for weighted gene co-expression network analysis. *Stat Appl Genet Mol Biol*, **4**, 17.
- Zhao H, Cai W, Su S, et al (2014). Screening genes crucial for pediatric pilocytic astrocytoma using weighted gene coexpression network analysis combined with methylation data analysis. *Cancer Gene Ther*, **21**, 448-55.
- Zhou Y, Xu J, Liu Y, et al (2014). Rat hepatocytes weighted gene co-expression network analysis identifies specific modules and hub genes related to liver regeneration after partial hepatectomy. *PLoS One*, **9**, 94868.