European Review for Medical and Pharmacological Sciences 2015; 19: 3375-3384

Identification of potential therapeutic target genes and mechanisms in non-small-cell lung carcinoma in non-smoking women based on bioinformatics analysis

W. ZHOU, M. YIN, H. CUI, N. WANG, L.-L. ZHAO, L.-Z. YUAN, X.-P. YANG, X.-M. DING, F.-Z. MEN, X. MA, J.-R. NA

Department of Respiration, General Hospital of Ningxia Medical University, Yinchuan, China

Wei Zhou and Mei Yin are co-first authors

Abstract. – OBJECTIVE: The study was aimed to explore the underlying mechanisms and identify the potential target genes by bioinformatics analysis for non-small-cell lung carcinoma (NSCLC) treatment in non-smoking women.

MATERIALS AND METHODS: The microarray data of GSE19804 was downloaded from Gene Expression Omnibus (GEO) database. Paired samples (from the same patient) of tumor and normal lung tissues from 60 non-smoking female NSCLC patients were used to identify differentially expressed genes (DEGs). The functional enrichment analysis was performed. Furthermore, the protein-protein interaction (PPI) network of the DEGs was constructed by Cytoscape software. The module analysis was performed.

RESULTS: Totally, 817 DEGs including 273 up- and 544 down-regulated genes were identified. The up-regulated genes were mainly enriched in extracellular matrix (ECM)-receptor interaction, focal adhesion and cell cycle functions, while down-regulated genes were mainly enriched in the cytokine-cytokine receptor interaction pathway. DEGs including hyaluronan-mediated motility receptor (HMMR), collagen, type I alpha 2 (COL1A2), cyclin A2 (CCNA2), MAD2 mitotic arrest deficient-like 1 (MAD2L1), interleukin 6 (IL6) and interleukin 1, beta (IL1B) were identified in these functions. These genes were hub nodes in PPI networks. Besides, there were 3 up-regulated modules and 1 down-regulated module. The significant pathways were ECM-receptor interaction and focal adhesion in up-regulated modules, while in down-regulated module, the significant pathway was mitogen-activated protein kinase (MAPK) signaling pathway.

CONCLUSIONS: The ECM-receptor interaction, focal adhesion, cell cycle and cytokine-cytokine receptor interaction functions may be associated with NSCLC development. Genes such as HMMR, COL1A2, CCNA2, MAD2L1, IL6 and IL1B may be potential therapeutic target genes for NSCLC.

Key Words:
Non-small-cell lung carcinoma, Differentially expressed genes, Bioinformatics analysis.

Introduction

Non-small-cell lung carcinoma (NSCLC) is a type of lung cancer and it is the leading cause of cancer-related death in the world. The most common types of NSCLC are large cell carcinoma, squamous cell carcinoma and adenocarcinoma. Approximately 1.35 million new cases of lung cancers are diagnosed annually[1]. Patients with NSCLC, which account for about 80% of all lung cancer cases, are diagnosed with advanced stages of the disease[2]. The 5-year survival rate of NSCLC after surgery is only 17%[3]. Therefore, an improved understanding mechanism on the pathogenesis of NSCLC would supply new insights for the diagnosis and treatment of NSCLC.

The carcinogenesis of NSCLC is a multistep process of accumulation of genes mutations in cell growth, differentiation, migration, invasion and apoptosis[4]. Previous study suggested that multiple genetic alterations are associated with the process of tumorgenesis. For example, EPH receptor B3 is overexpressed in NSCLC and promotes tumor metastasis by enhancing cell survival and migration[5]. MiR-21 represses tumor suppressor phosphatase and tensin homolog (*PTEN*) and promotes cell growth and invasion in

Corresponding Author: Wei Zhou MM; e-mail: zhouweieifhsdhf@hotmail.com

NSCLC[6]. In addition, down-regulation of notch-1, via inhibition of NF-κB signaling pathways by delta-tocotrienol, could inhibit cell growth and induction of apoptosis in NSCLC[7]. The insulin-like growth factor (IGF) signaling pathway has a driving role in the development of NSCLC, and *IGF* has become a potential target gene for cancer therapy[8]. Moreover, it has been reported that tobacco smoking is a risk factor for NSCLC[9]. However, many patients with lung cancer are associated with non-smoking[10]. It has been indicated that non-smoking-associated lung cancer has an increased prevalence of adenocarcinoma[11]. Gene epidermal growth factor receptor mutations are enriched in non-smokers patients with NSCLC[12]. Besides, NSCLC in never smokers is clinically characterized by an increased incidence in females[13]. Although some factors have been reported to be associated with the development of NSCLC, the pathogenic mechanisms of NSCLC in non-smoking women are still not clearly demonstrated. There is also lack of effective target genes for NSCLC treatment despite some combined treatment was discussed for them[14,15].

The microarray data (GSE19804) was analyzed with microarray analysis to identify the differentially expressed genes (DEGs)[16]. The result showed that sema domain, seven thrombospondin repeats (type 1 and type 1-like), transmembrane domain (TM) and short cytoplasmic domain, (semaphorin) 5A (*SEMA5A*) might be useful as prognostic biomarkers for NSCLC by Lu et al[16]. In this study, we downloaded this data and used different methods to identify DEGs between NSCLC and normal samples in non-smoking female patients. The functional enrichment analysis of DEGs was performed. Besides, protein-protein interaction (PPI) network construction, module analysis and functional analysis for module genes were performed to estimate the significant genes and their corresponding pathways. The purpose of this study was to identify the potential target genes and explore the underlying mechanisms in the pathogenesis of NSCLC in non-smoking women.

Materials and Methods

Affymetrix Microarray Data

The gene expression profile data of GSE19804 based on the platform of GPL570 (Affymetrix Human Genome U133 Plus 2.0 Array) (Affymetrix Inc., Santa Clara, CA, USA) was downloaded from Gene Expression Omnibus (GEO) database at the National Center for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov/geo/), it was deposited by Lu et al[16]. Paired samples (from the same patient) of tumor and normal lung tissues were collected from 60 non-smoking female NSCLC patients.

Data Preprocessing and Differential Expression Analysis

The probe-level data in .CEL files were converted into expression measures. The data were normalized by the robust multiarray average (RMA)[17] algorithm in R Affy package (http://www.bioconductor.org/packages/release/bioc/html/affy.html)[18]. The *t*-test method of the limma package (http://master.bioconductor.org/packages/release/bioc/html/limma.html)[19] in R was used to identify DEGs between NSCLC and normal samples. Then, $\log_2$-fold change ($\log_2$FC) was calculated to identify genes with expression-level differences. Only DEGs with $|\log_2 FC| > 1.0$ and p-value < 0.05 were selected.

Functional Enrichment Analysis

Gene Ontology (GO) database (http://geneontology.org/)[20] is a large set of gene annotation terms. Kyoto Encyclopedia of Genes and Genomes (KEGG) knowledge database (http://www.kegg.jp/)[21] was implemented to identify functional and metabolic pathways. Database for Annotation, Visualization and Integrated Discovery (DAVID) (http://david.abcc.ncifcrf.gov/)[22] is a tool which provides a comprehensive set of functional annotation for large list of genes. Both GO categories and KEGG pathway enrichment analyses for DEGs were conducted using DAVID. p-value < 0.05 and gene count > 5 were the cutoff criteria for the functional enrichment analysis.

Protein-Protein Interaction (PPI) Network Construction

Search Tool for the Retrieval of Interacting Genes (STRING) (http://www.string-db.org/)[23] is an online database which collects comprehensive information of proteins. The STRING online tool was applied to analyze the interactions of protein pairs and only the interaction with confidence score > 0.4 was selected as significant. Nodes with connectivity degree ≤ 5 were abandoned and the PPI network was constructed using Cytoscape software[24].

Module Analysis

The significant modules from the constructed PPI network were selected using the ClusterONE plug-in[25] of the Cytoscape software (http://cytoscape.org/plugins. html) with p-value < 0.01.

The InterPro database (http://www.ebi.ac.uk/interpro/)[26] is an integrated documentation resource for protein families, domains and functional sites. The protein domain analysis was conducted for module genes using InterPro database. Besides, the KEGG pathway enrichment analysis of the module genes was performed with p-value < 2.0E-3.

Results

Identification of DEGs

A total of 817 DEGs, including 273 up- and 544 down-regulated genes, were identified between NSCLC and normal samples.

Functional Enrichment Analysis

The top 5 GO terms enriched by up- and down-regulated genes were shown in Table I. The up-regulated genes were mainly enriched in GO biological process (BP) terms of nuclear division, cell cycle, multicellular organismal metabolic process, microtubule-based process, chromosome segregation and skin development. The GO cellular component (CC) terms contained proteinaceous extracellular matrix, condensed chromosome kinetochore and fibrillar collagen. The most enriched GO terms of down-regulated genes were relevant to BP terms of response to wounding, defense response, inflammatory response and vasculature development. Moreover, down-regulated genes were associated with GO CC terms of extracellular region, intrinsic to plasma membrane and plasma membrane part and molecular function (MF) terms of carbohydrate binding, polysaccharide binding and pattern binding.

The KEGG pathways of up- and down-regulated genes were shown in Table II. The up-regulated genes were mainly enriched in extracellular matrix (ECM)-receptor interaction, cell cycle, O-Glycan biosynthesis, focal adhesion and p53 signaling pathway. DEGs including hyaluronan-mediated motility receptor (*HMMR*), collagen, type I alpha 2 (*COL1A2*) and cartilage oligomeric matrix protein (*COMP*) were identified in ECM-receptor interaction pathway. Cell cycle related genes, such as cyclin A2 (*CCNA2*), MAD2 mitotic arrest deficient-like 1 (*MAD2L1*) and BUB1 mitotic checkpoint serine/threonine kinase B (*BUB1B*), were identified in this study. Additionally, down-regulated genes were related to cytokine-cytokine receptor interaction, complement and coagulation cascades and renin-angiotensin system pathways. The down-regulated genes including interleukin 6 (*IL6*), interleukin 1, beta (*IL1B*) and colony stimulating factor 3 (*CSF3*) were identified in the cytokine-cytokine receptor interaction pathway.

PPI Network Construction

The PPI networks of up- and down-regulated genes were shown in Figure 1. The up-regulated network was constructed with 133 nodes and 1453 edges (A). The proteins non-SMC condensin I complex, subunit G (NCAPG, degree = 53), PDZ binding kinase (PBK, degree = 52), MAD2L1 (degree = 52), cell division cycle associated 7 (CDCA7, degree = 52), CCNA2 (degree = 51) and HMMR (degree = 50) were hub nodes in this network. The down-regulated PPI network was constructed with 289 nodes and 1295 edges (B). Proteins IL6 (degree = 65), FBJ murine osteosarcoma viral oncogene homolog (FOS, degree = 46), IL1B (degree = 40) and chemokine (C-X-C motif) receptor 2 (CXCR2, degree = 38) were hub proteins in this network.

Module Analysis

A total of 3 significant modules were selected in the up-regulated PPI network. As shown in Figure 2, module 1 (p-value = 5.850E-7) was constructed with 20 nodes and 1216 edges (A); module 2 (p-value = 5.279E-6) was obtained with 17 nodes and 80 edges (B); module 3 (p-value = 0.002) contained with 11 nodes and 83 edges (C). One module (p-value = 8.481E-6) of down-regulated PPI network was obtained with 23 nodes and 170 edges (Figure 3).

As shown in Table III, the significant protein domains in up-regulated modules were fibrillar collagen, C-terminal, collagen triple helix repeat, von Willebrand factor, type C, LamG, thrombotype, N-terminal and con A-like lectin/glucanase, subgroup. In down-regulated module, the significant protein domains were basic-leucine zipper transcription, fos transforming protein, basic leucine zipper and so on (Table III).

In the up-regulated modules, the significant pathways were ECM-receptor interaction and focal adhesion, while in the down-regulated

Table I. Top 5 Gene Ontology (GO) terms enriched by up- and down-regulated genes.

Category	Term	Count	*p*-value
Up-regulated genes			
Cluster 1	Enrichment Score: 7.701339286862259		
BP	GO:0000280~nuclear division	25	2.19E-14
BP	GO:0007067~mitosis	25	2.19E-14
BP	GO:0000087~M phase of mitotic cell cycle	25	3.29E-14
BP	GO:0048285~organelle fission	25	5.39E-14
BP	GO:0000279~M phase	29	7.38E-14
Cluster 2	Enrichment Score: 4.3468704531868605		
BP	GO:0044259~multicellular orga macro metabolic process	9	1.26E-08
BP	GO:0044236~multicellular organismal metabolic process	9	5.73E-08
BP	GO:0032963~collagen metabolic process	8	1.35E-07
CC	GO:0005578~proteinaceous extracellular matrix	20	1.26E-06
CC	GO:0031012~extracellular matrix	20	3.83E-06
Cluster 3	Enrichment Score: 3.679263707199071		
BP	GO:0007017~microtubule-based process	16	6.01E-06
BP	GO:0007051~spindle organization	7	5.16E-05
BP	GO:0000226~microtubule cytoskeleton organization	11	7.10E-05
BP	GO:0007052~mitotic spindle organization	4	0.001307
BP	GO:0007010~cytoskeleton organization	14	0.013949
Cluster 4	Enrichment Score: 3.575395942162664		
BP	GO:0007059~chromosome segregation	12	3.05E-08
BP	GO:0000070~mitotic sister chromatid segregation	7	1.38E-05
BP	GO:0000819~sister chromatid segregation	7	1.63E-05
CC	GO:0000777~condensed chromosome kinetochore	8	4.19E-05
CC	GO:0000793~condensed chromosome	11	4.99E-05
Cluster 5	Enrichment Score: 2.8151554806198686		
CC	GO:0005583~fibrillar collagen	6	8.04E-07
CC	GO:0005581~collagen	7	1.94E-05
BP	GO:0043588~skin development	6	6.27E-05
BP	GO:0030199~collagen fibril organization	6	6.27E-05
CC	GO:0044420~extracellular matrix part	10	1.26E-04
Down-regulated genes			
Cluster 1	Enrichment Score: 17.086704678284505		
CC	GO:0005576~extracellular region	139	1.36E-20
CC	GO:0044421~extracellular region part	84	1.08E-17
CC	GO:0005615~extracellular space	65	3.72E-15
Cluster 2	Enrichment Score: 15.311658273172588		
BP	GO:0009611~response to wounding	58	3.81E-18
BP	GO:0006952~defense response	58	4.08E-15
BP	GO:0006954~inflammatory response	41	7.47E-15
Cluster 3	Enrichment Score: 11.047853620290875		
BP	GO:0001944~vasculature development	34	2.94E-13
BP	GO:0001568~blood vessel development	33	8.21E-13
BP	GO:0048514~blood vessel morphogenesis	29	1.58E-11
BP	GO:0001525~angiogenesis	22	1.69E-09
Cluster 4	Enrichment Score: 8.176809496800056		
CC	GO:0031226~intrinsic to plasma membrane	77	5.00E-09
CC	GO:0044459~plasma membrane part	117	5.98E-09
CC	GO:0005887~integral to plasma membrane	75	9.86E-09
Cluster 5	Enrichment Score: 6.857326135375957		
MF	GO:0030246~carbohydrate binding	38	5.99E-12
MF	GO:0030247~polysaccharide binding	19	3.61E-07
MF	GO:0001871~pattern binding	19	3.61E-07
MF	GO:0008201~heparin binding	14	6.58E-06
MF	GO:0005539~glycosaminoglycan binding	16	1.01E-05

BP: biological process; CC: cellular component; MF: molecular function. Count: enriched gene number in the category.

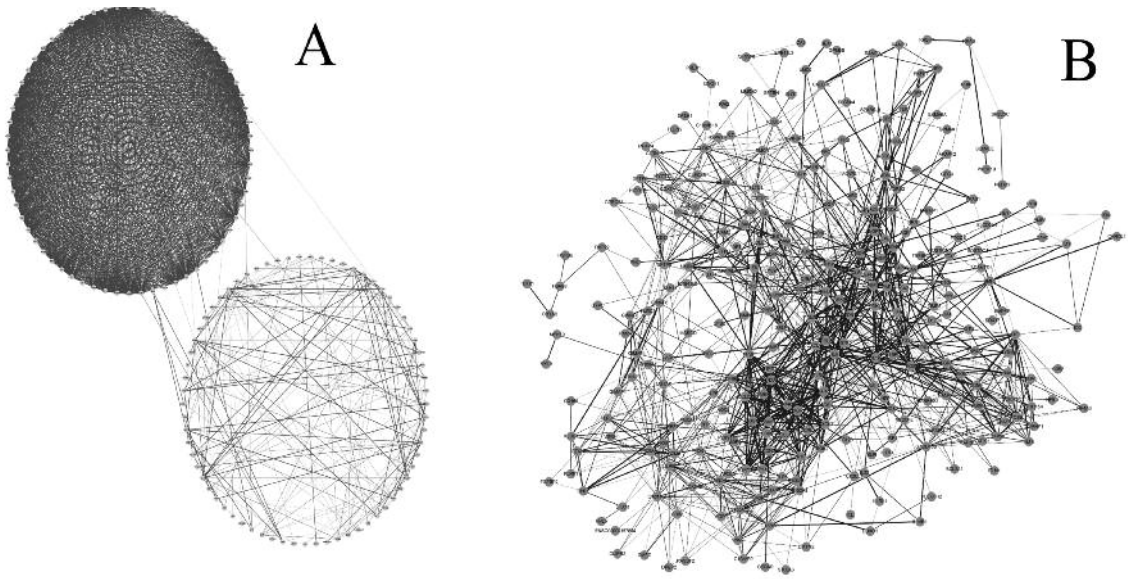

Figure 1. Protein-protein interaction (PPI) networks of differentially expressed genes (DEGs). ***A,*** PPI network of up-regulated genes. ***B,*** PPI network of down-regulated genes. Nodes stand for proteins and edges represent interactions between two proteins.

module, the significant pathway was mitogen-activated protein kinase (MAPK) signaling pathway (Table IV).

Discussion

In this study, the gene expression profile data of GSE19804 was downloaded from GEO database to identify DEGs between NSCLC and normal samples in non-smoking women using bioinformatics analysis. Totally, 817 DEGs including 273 up- and 544 down-regulated genes were selected. The functional enrichment analysis and module analysis results showed that up-regulated genes were related to ECM-receptor interaction, focal adhesion and cell cycle pathways, while down-regulated genes were mainly enriched in the cytokine-cytokine receptor interaction pathway. DEGs including *HMMR*, *COL1A2*, *CCNA2*, *MAD2L1*, *IL6* and *IL1B* were identified in these functions. Moreover, these genes were hub nodes in PPI networks. These DEGs and their related functions may be involved in NSCLC progress.

The up-regulated genes were significantly enriched in the pathways of ECM-receptor interaction and focal adhesion. Previous studies[27]

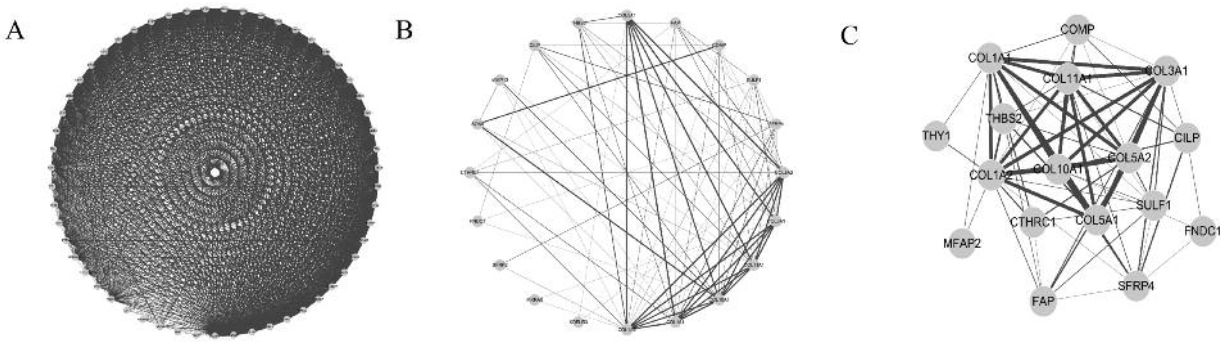

Figure 2. The significant modules in up-regulated protein-protein interaction (PPI) network. ***A***, Module 1 with *p*-value = 5.850E-7. ***B***, Module 2 with *p*-value = 5.279E-6. ***C***, Module 3 with *p*-value = 0.002. Nodes stand for proteins and edges represent interactions between two proteins.

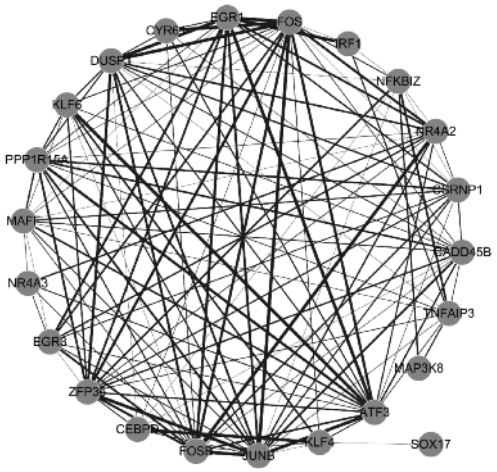

Figure 3. The significant module in down-regulated protein-protein interaction (PPI) network. *p*-value = 8.481E-6. Nodes stand for proteins and edges represent interactions between two proteins.

indicated that ECM-receptor interaction and focal adhesion were involved in the same function: cell adhesion. In our study, *HMMR* and *COL1A2* were identified in these pathways. *HMMR*, also known as *RHAMM* and *IHABP*, is a hyaluronan receptor. It has been reported that the ECM molecule hyaluronan and its receptor *RHAMM* induced focal adhesion to signal the cytoskeletal changes required for elevated cell motility seen in the processes of tumor cell progression, metastasis and invasion[28]. Wang et al[29] reported that the over-expressed *RHAMM* regulated ras signaling, and it was a significant parameter in breast cancer progression. However, the evidence concerning the impact of *HMMR* in NSCLC is rare. In this study, over-expressed *HMMR* was enriched in these functions and was hub node in PPI network, suggesting that *HMMR* may play an important role in NSCLC progression by regulating ECM-receptor interaction and focal adhesion. Additionally, *COL1A2* is a member of collagen family and is also associated with cell adhesion[30]. Adhesion of cells to collagen is mediated by fibronectin[31]. The study of Shintani et al[32] showed that collagen I promoted epithelial-to-mesenchymal transition in lung cancer cells through transforming growth factor-beta (TGF-β) signaling. In this study, *COL1A2* was up-regulated gene, which was consistent with a previous study that Iizasa et al[33] reported that collagen XVIII was over-expressed in NSCLC. Therefore, *HMMR* and *COL1A2* may be involved in NSCLC development. Their related pathways (ECM-receptor interaction and focal adhesion) may be potential pathogenic mechanisms of NSCLC.

Cell cycle is the series of events leading to cell duplication and division, which is closely involved in cell growth and proliferation[34]. Un-

Table II. The Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis of up- and down-regulated genes.

Term	Count	*p*-value
Up-regulated genes		
Up-regulated genes		
hsa04512:ECM-receptor interaction	10	7.22E-06
hsa04110:Cell cycle	10	1.74E-04
hsa00512:O-Glycan biosynthesis	5	0.001309
hsa04510:Focal adhesion	11	0.001455
hsa04115:p53 signaling pathway	5	0.024387
Down-regulated genes		
hsa04060:Cytokine-cytokine receptor interaction	22	8.59E-05
hsa04610:Complement and coagulation cascades	9	0.001545
hsa04614:Renin-angiotensin system	5	0.001787
hsa04080:Neu ligand-receptor interaction	18	0.003384
hsa04062:Chemokine signaling pathway	13	0.016985
hsa04360:Axon guidance	10	0.022776
hsa04514:Cell adhesion molecules (CAMs)	10	0.025994
hsa05332:Graft-versus-host disease	5	0.035698
hsa04670:Leukocyte transendothelial migration	9	0.035911

Count: enriched gene number in the KEGG term.

Table III. The significant protein domains in up- and down-regulated modules.

Modules	Term	Count	*p*-value
Up-regulated modules			
Module 1	IPR000885:Fibrillar collagen, C-terminal	6	1.88E-13
	IPR008160:Collagen triple helix repeat	8	7.08E-13
	IPR001007:von Willebrand factor, type C	4	7.03E-06
	IPR003129:LamG, thrombospondin-type, N-terminal	3	1.98E-04
	IPR013320:Con A-like lectin/glucanase, subgroup	3	0.002426
Module 2	IPR000885:Fibrillar collagen, C-terminal	6	5.01E-13
	IPR008160:Collagen triple helix repeat	8	3.08E-12
	IPR001007:von Willebrand factor, type C	4	1.21E-05
	IPR003129:LamG, throm-type, N-terminal	3	2.81E-04
	IPR013320:Con A-like lectin/glucanase, subgroup	3	0.003427
Module 3	IPR000885:Fibrillar collagen, C-terminal	6	1.88E-13
	IPR008160:Collagen triple helix repeat	8	7.08E-13
	IPR001007:von Willebrand factor, type C	4	7.03E-06
	IPR003129:Laminin G, throm-type, N-terminal	3	1.98E-04
	IPR013320:Con A-like lectin/glucanase	3	0.002426
Down-regulated module			
Module 1	IPR004827:Basic-leucine zipper transcription	6	5.26E-09
	IPR000837:Fos transforming protein	3	4.22E-05
	IPR011700:Basic leucine zipper	3	1.57E-04
	IPR011616:bZIP transcription factor, bZIP-1	3	6.88E-04
	IPR013087:Zinc finger, C2H2-type/integrase	4	0.041610044

Count: enriched gene number in the KEGG term.

Table IV. The significant Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis of up- and down-regulated modules.

Modules	Term	Count	*p*-value
Up-regulated modules			
Module 1	hsa04512:ECM-receptor interaction	8	9.14E-12
	hsa04510:Focal adhesion	8	4.58E-09
Module 2	hsa04512:ECM-receptor interaction	8	3.01E-11
	hsa04510:Focal adhesion	8	1.48E-08
Module 3	hsa04512:ECM-receptor interaction	8	9.14E-12
	hsa04510:Focal adhesion	8	4.58E-09
Down-regulated module			
Module 1	hsa04010:MAPK signaling pathway	4	0.002543

Count: enriched gene number in the KEGG term.

controlled cell proliferation is one of hallmarks of cancer[4]. Recent evidences suggested that cell cycle was mediated by several genes. For instance, *CCNA2* belongs to cyclin family, whose members regulate cell cycle[35]. Cyclin B1 (another member of the cyclin family) is a key molecule for G_2/M transition during cell cycle and is over-expressed in early-stage NSCLC[36]. Betticher et al[37] reported that genetic alteration of cyclin D1 was a key abnormality in NSCLC carcinogenesis. In this study, *CCNA2* was overexpressed and enriched in the function of cell cycle in NSCLC, suggesting that *CCNA2* may play a key role in NSCLC development by regulating cell cycle. Besides, *MAD2L1* was also identified in cell cycle. *MAD2L1* is the human orthologues of *MAD2* and is involved in maintaining the mitotic spindle checkpoint

function[38]. Reduced spindle checkpoint function of missense variation in *MAD2L1* gene affects susceptibility to lung cancer[39]. In present study, cell cycle checkpoint gene *MAD2L1* was up-regulated in NSCLC samples, which was consistent with previous study. Kato et al[40] reported that over-expression of *MAD2* was observed in NSCLC patients. *MAD2L1* was hub node in the up-regulated PPI network. Therefore, we inferred that cell cycle may play an important role in NSCLC through multiple genes including *CCNA2* and *MAD2L1*. These genes may be the potential therapeutic target genes for NSCLC.

Apart from up-regulated genes and their functions, down-regulated genes were mainly enriched in the cytokine-cytokine receptor interaction pathway in this study. *IL6* and *IL1B* were identified in this pathway. *IL6* and *IL1B* are secreted by lymphocytes to stimulate immune response. *IL6*, as an oncogene, was involved in the proliferation and differentiation of various malignant tumor cells[41]. Yi et al[42] showed that blockade of IL6/IL6 receptor (IL6R) signaling suppressed the proliferation of lung cancer stem cells. In this study, *IL6* was hub node in the down-regulated PPI network. It suggested that down-expression of *IL6* may inhibit NSCLC oncogenesis via regulating cytokine-cytokine receptor interaction pathway. Moreover, *IL1B* is a member of the IL1 family of cytokines. Zienoldiny et al[43] reported that polymorphisms of *IL1B* gene were associated with increased risk of NSCLC. A specific *IL1B* haplotype correlated with a high level of *IL1B* mRNA in the lung increases risk of NSCLC[44]. In this study, IL1B was down-regulated gene, suggesting that it may decrease the risk of NSCLC. These results showed that cytokine-cytokine receptor interaction pathway may be associated with the progression of NSCLC, and *IL6* and *IL1B* may be potential target genes for NSCLC treatment.

Conclusions

Our study shows that ECM-receptor interaction, focal adhesion, cell cycle and cytokine-cytokine receptor interaction functions may be closely associated with NSCLC development. Genes such as *HMMR*, *COL1A2*, *CCNA2*, *MAD2L1*, *IL6* and *IL1B* may be potential therapeutic target genes for NSCLC in non-smoking women. However, further studies are still required to confirm our results.

Conflict of Interest

The Authors declare that there are no conflicts of interest.

References

1) Askoxylakis V, Thieke C, Pleger ST, Most P, Tanner J, Lindel K, Katus HA, Debus J, Bischof M. Long-term survival of cancer patients compared to heart failure and stroke: a systematic review. BMC Cancer 2010; 10: 105.
2) Askoxylakis V, Tanner J, Kappes J, Hoffmann H, Nicolay NH, Rief H, Debus J, Thomas M, Bischof M. Trimodal therapy for stage III-N2 non-small-cell lung carcinoma: A single center retrospective analysis. BMC Cancer 2014; 14: 572.
3) National Cancer Institute: Bethesda, MD, USA. Seer cancer statistics review, 1975-2008. Available from http://seer.cancer.gov/csr/1975_2008/.
4) Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell 2011; 144: 646-674.
5) Ji XD, Li G, Feng YX, Zhao JS, Li JJ, Sun ZJ, Shi S, Deng YZ, Xu JF, Zhu YQ, Koeffler HP, Tong XJ, Xie D. EphB3 is overexpressed in non–small-cell lung cancer and promotes tumor metastasis by enhancing cell survival and migration. Cancer Res 2011; 71: 1156-1166.
6) Zhang JG, Wang JJ, Zhao F, Liu Q, Jiang K, Yang GH. MicroRNA-21 (miR-21) represses tumor suppressor PTEN and promotes growth and invasion in non-small cell lung cancer (NSCLC). Clin Chim Acta 2010; 411: 846-852.
7) Ji X, Wang Z, Geamanu A, Sarkar FH, Gupta SV. Inhibition of cell growth and induction of apoptosis in non-small cell lung cancer cells by delta-tocotrienol is associated with notch-1 down-regulation. J Cell Biochem 2011; 112: 2773-2783.
8) Scagliotti GV, Novello S. The role of the insulin-like growth factor signaling pathway in non-small cell lung cancer and other solid tumors. Cancer Treat Rev 2012; 38: 292-302.
9) Kim DH, Nelson HH, Wiencke JK, Zheng S, Christiani DC, Wain JC, Mark EJ, Kelsey KT. p16(INK4a) and histology-specific methylation of CpG islands by exposure to tobacco smoke in non-small cell lung cancer. Cancer Res 2001; 61: 3419-3424.
10) Wu PF, Lee CH, Wang MJ, Goggins WB, Chiang TA, Huang MS, Ko YC. Cancer aggregation and complex segregation analysis of families with female non-smoking lung cancer probands in Taiwan. Eur J Cancer 2004; 40: 260-266.
11) Sun S, Schiller JH, Gazdar AF. Lung cancer in never smokers—a different disease. Nat Rev Cancer 2007; 7: 778-790.
12) An SJ, Chen ZH, Su J, Zhang XC, Zhong WZ, Yang JJ, Zhou Q, Yang XN, Huang L, Guan JL, Nie Q, Yan HH, Mok TS, Wu YL. Identification of enriched driver gene alterations in subgroups of non-small cell lung cancer patients based on histology and smoking status. PLoS One 2012; 7: e40109.

13) Wakelee HA, Chang ET, Gomez SL, Keegan TH, Feskanich D, Clarke CA, Holmberg L, Yong LC, Kolonel LN, Gould MK. Lung cancer incidence in never smokers. J Clin Oncol 2007; 25: 472-478.

14) Mariotta S, Sposato B, Li Bianchi E, Fiorucci F, Ricci A, Carbone F, Crescenzi M, Schmid G. Combined treatment in advanced stages (IIIb-IV) of non-small cell lung cancer. Eur Rev Med Pharmacol Sci 2002; 6: 49-54.

15) Ma GF, Zhang RF, Ying KJ, Wang D. Effect evaluation of cisplatin-gemcitabine combination chemotherapy for advanced non-smallcell lung cancer patients using microarray data. Eur Rev Med Pharmacol Sci 2015; 19: 578-585.

16) Lu TP, Tsai MH, Lee JM, Hsu CP, Chen PC, Lin CW, Shih JY, Yang PC, Hsiao CK, Lai LC, Chuang EY. Identification of a novel biomarker, SEMA5A, for non-small cell lung carcinoma in nonsmoking women. Cancer Epidemiol Biomarkers Prev 2010; 19: 2590-2597.

17) Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, Speed TP. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics 2003; 4: 249-264.

18) Team RC. R: a language and environment for statistical computing. R foundation for statistical computing, Vienna, Austria, 2012. Available from http://www.R-project.org/

19) Smyth GK. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Stat Appl Genet Mol Biol 2004; 3: Article3.

20) Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 2000; 25: 25-29.

21) Altermann E, Klaenhammer TR. PathwayVoyager: pathway mapping using the Kyoto Encyclopedia of Genes and Genomes (KEGG) database. BMC Genomics 2005; 6: 60.

22) Huang DW, Sherman BT, Tan Q, Kir J, Liu D, Bryant D, Guo Y, Stephens R, Baseler MW, Lane HC, Lempicki RA. DAVID Bioinformatics Resources: expanded annotation database and novel algorithms to better extract biology from large gene lists. Nucleic Acids Res 2007; 35: W169-W175.

23) Szklarczyk D, Franceschini A, Kuhn M, Simonovic M, Roth A, Minguez P, Doerks T, Stark M, Muller J, Bork P, Jensen LJ, von Mering C. The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. Nucleic Acids Res 2011; 39: D561-D568.

24) Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 2003; 13: 2498-2504.

25) Nepusz T, Yu H, Paccanaro A. Detecting overlapping protein complexes in protein-protein interaction networks. Nat Methods 2012; 9: 471-472.

26) Apweiler R, Attwood TK, Bairoch A, Bateman A, Birney E, Biswas M, Bucher P, Cerutti L, Corpet F, Croning MDR. The interpro database, an integrated documentation resource for protein families, domains and functional sites. Nucleic Acids Res 2001; 29: 37-40.

27) Albelda SM, Buck CA. Integrins and other cell adhesion molecules. FASEB J 1990; 4: 2868-2880.

28) Hall CL, Turley EA. Hyaluronan: RHAMM mediated cell locomotion and signaling in tumorigenesis. J Neurooncol 1995; 26: 221-229.

29) Wang C, Thor AD, Moore D, Zhao Y, Kerschmann R, Stern R, Watson PH, Turley EA. The overexpression of RHAMM, a hyaluronan-binding protein that regulates ras signaling, correlates with overexpression of mitogen-activated protein kinase and is a significant parameter in breast cancer progression. Clin Cancer Res 1998; 4: 567-576.

30) Soikkeli J, Podlasz P, Yin M, Nummela P, Jahkola T, Virolainen S, Krogerus L, Heikkilä P, von Smitten K, Saksela O, Hölttä E. Metastatic outgrowth encompasses COL-I, FN1, and POSTN up-regulation and assembly to fibrillar networks regulating cell adhesion, migration, and growth. Am J Pathol 2010; 177: 387-403.

31) Kleinman HK, Martin GR, Fishman PH. Ganglioside inhibition of fibronectin-mediated cell adhesion to collagen. Proc Natl Acad Sci U S A 1979; 76: 3367-3371.

32) Shintani Y, Maeda M, Chaika N, Johnson KR, Wheelock MJ. Collagen i promotes epithelial-to-mesenchymal transition in lung cancer cells via transforming growth factor-β signaling. Am J Respir Cell Mol Biol 2008; 38: 95-104.

33) Iizasa T, Chang H, Suzuki M, Otsuji M, Yokoi S, Chiyo M, Motohashi S, Yasufuku K, Sekine Y, Iyoda A. Overexpression of collagen XVIII is associated with poor outcome and elevated levels of circulating serum endostatin in non-small cell lung cancer. Clin Cancer Res 2004; 10: 5361-5366.

34) MacLachlan T K, Sang N, Giordano A. Cyclins, cyclin-dependent kinases and cdk inhibitors: implications in cell cycle control and cancer. Crit Rev Eukaryot Gene Expr 1995; 5: 127-156.

35) Pagano M, Pepperkok R, Verde F, Ansorge W, Draetta G. Cyclin A is required at two points in the human cell cycle. EMBO J 1992; 11: 961-971.

36) Soria J-C, Jang SJ, Khuri FR, Hassan K, Liu D, Hong WK, Mao L. Overexpression of cyclin B1 in early-stage non-small cell lung cancer and its clinical implication. Cancer Res 2000; 60: 4000-4004.

37) Betticher DC, Heighway J, Hasleton PS, Altermatt H, Ryder W, Cerny T, Thatcher N. Prognostic significance of CCND1 (cyclin D1) overexpression in primary resected non-small-cell lung cancer. Br J Cancer 1996; 73: 294-300.

38) Cahill DP, da Costa LT, Carson-Walter EB, Kinzler KW, Vogelstein B, Lengauer C. Characterization of MAD2B and other mitotic spindle checkpoint genes. Genomics 1999; 58: 181-187.

39) Guo Y, Zhang X, Yang M, Miao X, Shi Y, Yao J, Tan W, Sun T, Zhao D, Yu D. Functional evaluation of missense variations in the human MAD1L1 and MAD2L1 genes and their impact on susceptibility to lung cancer. J Med Genet 2010; 47: 616-622.

40) Kato T, Daigo Y, Aragaki M, Ishikawa K, Sato M, Kondo S, Kaji M. Overexpression of MAD2 predicts clinical outcome in primary lung cancer patients. Lung Cancer 2011; 74: 124-131.

41) Adachi Y, Yoshio-Hoshino N, Nishimoto N. The blockade of IL-6 signaling in rational drug design. Curr Pharm Des 2008; 14: 1217-1224.

42) Yi H, Cho H-J, Cho S, Jo K, Park J, Kim NH, Amidon GL, KIM J, Shin HC. Blockade of interleukin-6 receptor suppresses the proliferation of H460 lung cancer stem cells. Int J Oncol 2012; 41: 310-316.

43) Zienolddiny S, Ryberg D, Maggini V, Skaug V, Canzian F, Haugen A. Polymorphisms of the interleukin-1 beta gene are associated with increased risk of non-small cell lung cancer. Int J Cancer 2004; 109: 353-356.

44) Landvik NE, Hart K, Skaug V, Stangeland LB, Haugen A, Zienolddiny S. A specific interleukin-1B haplotype correlates with high levels of IL1B mRNA in the lung and increased risk of non-small cell lung cancer. Carcinogenesis 2009; 30: 1186-1192.