Iran Red Crescent Med J. 2018 March; 20(3):e65035. doi: 10.5812/ircmj.65035.

Published online 2018 March 31. **Research Article**

Cross-Sectional Study of Gene Expression Analysis Identifies Critical Biological Pathways and Key Genes Implicated in Non-Small Cell Lung Cancer

Tonglian Wang,[1] Jing Hu,[2] Lutong Xu,[1] Hongbo Zhao,[3] Yuanyue Li,[1] Tao Shou,[2] Xueshan Xia,[1] and Qiang Chen[1,*]

[1]Research Center of Molecular Medicine of Yunnan Province, Faculty of Life Science and Technology, Kunming University of Science and Technology, Kunming, P.R. China
[2]Medical Oncology, The First People's Hospital of Yunnan Province, Kunming, P.R. China
[3]Institute of Molecular and Clinical Medicine, Kunming Medical University, Kunming, P.R. China

[*]*Corresponding author*: Qiang Chen, Research Center of Molecular Medicine of Yunnan Province, Faculty of Life Science and Technology, Kunming University of Science and Technology, No. 727 of South Jingming Road, Chenggong New Town, Kunming, 650500, P.R. China. Tel: +86-87165939528, Fax: +86-87165939528, E-mail: chq@sjtu.edu.cn

Received 2017 December 18; **Revised** 2018 February 04; **Accepted** 2018 February 24.

Abstract

Background: Non-small cell lung cancer (NSCLC) is the most common type of lung Neoplasms, which accounts for about 85% of all lung cancer types. However, critical biological pathways and key genes implicated in NSCLC remain ambiguous.
Objectives: The present study aimed at identifying the critical biological pathways and key genes implicated in NSCLC, and providing insight into the molecular mechanism underlying NSCLC.
Methods: In this case-control bioinformatics study, the researchers used four microarray data of NSCLC from public gene expression omnibus (GEO) database at the national center for biotechnology information (NCBI) website. The microarray data came from studies of American, Spanish, and Taiwanese NSCLC patients, and in total contained 190 NSCLC tissue and 180 normal lung tissue. A standardized- microarray preprocessing and gene set enrichment analysis (GSEA) were used to analyze each microarray data and obtained significantly regulated pathways. Venn analysis was used to identify the common significantly regulated biological pathways. Protein and protein interaction (PPI) network analysis was used to identify the key genes within common significantly regulated pathways. The PPI information was retrieved from the STRING database, and Cytoscape software was used to construct and visualize the PPI network.
Results: Through integrating GSEA results of four microarray data, finally, the researchers identified 22 common up-regulated and 85 common down-regulated pathways. Many genes within 107 common significantly regulated pathways were significantly enriched within cell cycle pathway (P value of 2.58e-79) and focal adhesion pathway (P value of 2.44e-81). The PPI network showed that up-regulated CDK1 (P value = 1.33e-18 and logFC = 1.41) and down-regulated PIK3R1 (P value = 5.09e-22 and logFC = -1.13) genes shared the most abundant edges, and were associated with NSCLC.
Conclusions: This cross-sectional study showed increased concordance between gene expression profiling data. These identified pathways and genes provide some insight into the molecular mechanisms of NSCLC, and the genes may serve as candidate diagnostic and therapeutic targets of NSCLC.

Keywords: Cancer, Carcinogenesis, Critical Pathways, Gene Expression Profiling, Lung, Neoplasms Cancer, Profiling

1. Background

Lung cancer, as one of the most common malignancies, is the leading cause of cancer-related deaths all over the world (1). In the last decades, the incidence and mortality rates of lung cancer have been increasing rapidly, especially in regions where tobacco consumption is more common (2). Although many studies have shown that smoking-tobacco accelerated lung carcinogenesis, genetic factors still play a key role (3). In all lung cancer types, non-small cell lung cancer (NSCLC) is the most common type and accounts for 85% of all lung cancer types. However, despite extensive researches, the molecular mechanisms implicated in NSCLC are yet to be uncovered.

In the last decades, gene expression analysis-based microarray was widely used to study the NSCLC, and hundreds of differentially expressed genes (DEGs) were found by differentially expressed gene analysis (DEGA) (4-6). Furthermore, several key genes, including the well-known gene epidermal growth factor receptor (EGFR) and tumor

protein p53 (TP53) were identified (5). However, for most DEGs identified, their roles in NSCLC were obscure and needed to be discussed further. However, it is difficult to interpret the role of individual genes (7). Performing gene set analysis for gene expression profiling data is a more powerful method to reveal biological mechanisms implicated in NSCLC than conventional single-gene analysis methods, especially in identifying genes with subtle contributions (7, 8). Among some frequently used gene set analysis methods, gene set enrichment analysis (GSEA) is the most well-known and widely-used approach (7, 8), through which the significant difference in expression of pre-defined gene set between two groups of data can be identified. The pre-defined gene set can be a set of genes in a gene ontology category, in a biological pathway, or can be user-defined (9). Recently, using GSEA method, some biological pathways, such as Ras signaling pathway and Wnt signaling pathway were identified to be significantly regulated in NSCLC (9, 10), and explained the biological mechanisms of NSCLC. However, these studies mainly aimed at lung squamous cell carcinoma (LUSC), which is one of the major subtypes of NSCLC, or immune gene sets in females with NSCLC (9, 10). The identified biological pathways represented a fraction of the pathways implicated in NSCLC, and the biological pathways implicated in NSCLC need to be identified systematically.

In this study, the researchers collected four gene expression profiling data about NSCLC studies from Taiwanese, American and Spanish patients, and applied a standardized microarray preprocessing and GSEA to each gene expression profiling data to identify significantly regulated pathways. Furthermore, the researchers performed Venn analysis to identify common significantly regulated pathways and constructed the PPI network between genes within common significantly regulated pathways to identify key genes. This cross-study improved the concordance between gene expression profiling data and highlighted the genes weakly connected with NSCLC, which would provide some insight into the biological pathways implicated in NSCLC.

2. Methods

2.1. Microarray Data Collection

In this case-control bioinformatics study, the researchers used four microarray data of NSCLS for reanalysis. The microarray data were searched and downloaded from public gene expression omnibus (GEO) database at the national center for biotechnology information (NCBI) website (http://www.ncbi.nlm.nih.gov/geo/). These data following the criteria were used in this study: (1) the data were genome-wide, (2) the data included NSCLC and control data, (3) the raw or normalized data were complete and available, (4) the data were generated using the same chip platform.

Finally, data sets with GEO accessions GSE7670, GSE10072, GSE18842, and GSE19804 were used in this study. Affymetrix microarray platform generated the microarray data. GSE7670 and GSE19804 data were from Taiwanese NSCLC studies, separately contained 52 and 120 pairwise samples. Furthermore, GSE10072 and GSE18842 data were from American and Spanish NSCLC studies, and 107 and 91 case-control samples, separately. The related information of microarray data listed in Table 1, such as author, sample source, GEO accession, chip platform, sample size, and sample type.

2.2. Microarray Data Preprocessing

To improve the efficiency of data reanalysis, all microarray data must be reprocessed. The researchers performed the data reprocessing using version 3.3.2 R language (http://www.r-project.org) and software packages version 3.4 Bioconductor project (http://www.bioconductor.org/). All data were subjected to background-adjust and normalized. Robust multichip averaging (RMA) algorithm was used to calculate the log2 probe-set intensities (11). Any gene failing to map any KEGG pathway was removed in the next analysis. The interquartile range (IQR) was used to measure the data variability. The cut-off value was set according to the resultant distribution of IQR values of all genes, and the genes with IQR values under 0.5 were removed. If one gene targeted multiple probe sets, the probe set with the most substantial variability was retained to be used in the next analysis.

2.3. Statistical Analysis of DEGs

Statistical analysis of DEGs was performed using the version 3.32.7 of Limma software package in Bioconductor project. Limma package employed the Voom method, Liner modeling, and empirical Bayes moderation to assess DEGs, and could acquire more robust results even in less of microarrays. The cut-off criteria of DEGs were fulfilled according to the following conditions: (1) a false discovery rate (FDR) was not more than 5%, and (2) a linear fold change (FC) was not less than 2 or not more than 0.5.

2.4. Statistical Analysis of Significant Pathways

Statistical analysis of significant pathway was accomplished using the GSEA method. The version 2.40.0 category package was used to perform GSEA of the pathway in the Bioconductor project. The purpose of performing

Table 1. Characteristics of Datasets Included in This Study

GEO Accession				GSE10072	GSE18842	GSE19804	GSE7670
Sample Source				America	Spain	Taiwan	Taiwan
First Author				Landi M	Sanchez-Palencia A	Lu TP	Su LJ
Submitted Year				2008	2009	2010	2007
Chip Platform				U133A [GPL96]	U133 Plus 2.0 [GPL570]	U133 Plus 2.0 [GPL570]	U133A [GPL96]
Probe Number				22k	55k	55k	22k
Histology				AC	AC (14, 30%) and SCC (32, 70%)	AC (56, 93%) and SCC	AC
Stages				I - IV (I, 38%. II, 36%. III, 21%. IV, 5%)	I - IV (I, 83%. II, 9%. III, 7%. IV, 2%)	I-IV (I + II, 78%. III + IV, 22%)	Unknown (early and late)
Experimental Design				Case-control	Paired except one tumor	Paired	Paired
Smoking	Never	Normal	Male	4	-	-	-
			Female	11	-	60	-
		Cancer	Male	3	-	-	-
			Female	13	-	60	-
	Former	Normal	Male	18	-	-	-
			Female	0	-	-	-
		Cancer	Male	16	-	-	-
			Female	2	-	-	-
	Current	Normal	Male	12	-	-	-
			Female	4	-	-	-
		Cancer	Male	16	-	-	-
			Female	0	-	-	-
	Unknown	Normal	Male	-	-	-	5
			Female	-	-	-	21
		Cancer	Male	-	-	-	5
			Female	-	-	-	21
Sample Number		Cancer tissue		58	46	60	26
		Normal tissue		49	45	60	26

Abbreviations: AC, adenocarcinoma, one major subtype of non-small cell lung cancer; NSCLC, non-small cell lung cancer, one of the most common types of lung cancer; SCC, squamous cell carcinomas, one major subtype of non-small cell lung cancer.

GSEA was to determine whether the members of a gene set S were randomly distributed throughout the entire reference gene list L or was just primarily found at the top or bottom. The most significant advantage of the GSEA method was the relative robustness to noise and outliers in the data. Gene sets including less than 10 genes were discarded. In each pathway, the t-statistic mean of the genes was computed. A permutation test of 1000 times was implemented, and the pathways with P value $\leq$ 0.05 were identified to significantly change (12).

2.5. Protein-Protein Interaction Network Construction

The interaction relationship between genes within common critical biological pathways was exhibited using the PPI network. The PPI information was predicted using the STRING database (https://string-db.org/), and the minimum required interaction score between gene and gene was set for 0.9. The PPI network was constructed and visualized using open-source version 3.5.1 Cytoscape software (http://www.cytoscape.org/).

3. Results

3.1. Identification of Significant Pathways

The researchers used GSEA to reanalyze four datasets to identify significantly regulated pathways implicated in NSCLC. According to the permutation 0.05 P value, the researchers found 28 (GSE7670), 48 (GSE10072), 63 (GSE18842), and 51 (GSE19804) up-regulated pathways, and 112 (GSE7670), 112 (GSE10072), 115 (GSE18842), and 118 (GSE19804) down-regulated pathways, separately (Table 2). The overlapping analysis showed that 22 common up-regulated pathways and 85 common down-regulated pathways were identified (Figure 1).

In common up-regulated pathways, the researchers observed that many pathways belonged to cell growth and death, carbohydrate metabolism, nucleotide metabolism, glycan biosynthesis and metabolism, replication and repair, translation and so on. In common down-regulated pathways, the researchers found that many pathways belonged to the immune system, cellular community, signal transduction, endocrine system, immune diseases, infectious diseases and so on (Table 3).

3.2. Identification of Key Genes

Overall, 412 genes were found within 22 common up-regulated pathways. Based on the minimum required interaction score of 0.9 for PPI information from STRING database, 370 of 412 genes were enriched in PPI networks (P value < 1.0e-16), and these genes were significantly enriched within cell cycle pathway (P value = 2.58e-79) and p53 signal pathway (P value = 1.08e-45). Besides, some pathways related to metabolism were also enriched at the top. These pathways included metabolic pathways (P value = 3.03e-64), purine metabolism (P value = 1.95e-44), pyrimidine metabolism (P value = 1.01e-29) and so on (Table 3). The PPI network showed that cyclin-dependent kinase 1 (CDK1) gene shared the most abundant edges (Figure 2A), and expression of CDK1 gene was significantly up-regulated in NSCLC samples (P value = 1.33e-18, logFC = 1.41, from GSE10072 data). In addition, the TP53 gene was also observed to share more abundant edges.

Similarly, 1,972 genes were found within 85 common down-regulated pathways, and 905 genes were mainly enriched within focal adhesion pathway (P value = 2.44e-81), MAPK signaling pathway (P value = 2.36e-81), and chemokine signaling pathway (P value = 7.64e-55) (Table 3). The PPI network showed that phosphatidylinositol 3-kinase regulatory subunit alpha (PIK3R1) gene (P value = 5.09e-22, logFC = -1.13, from GSE10072 data) shared the most abundant edges, and was significantly down-regulated in NSCLC (Figure 2B). Besides, the researchers found that phosphatidylinositol 3-kinase 3 catalytic subunit alpha (PIK3CA) and EGFR genes also shared more abundant edges.

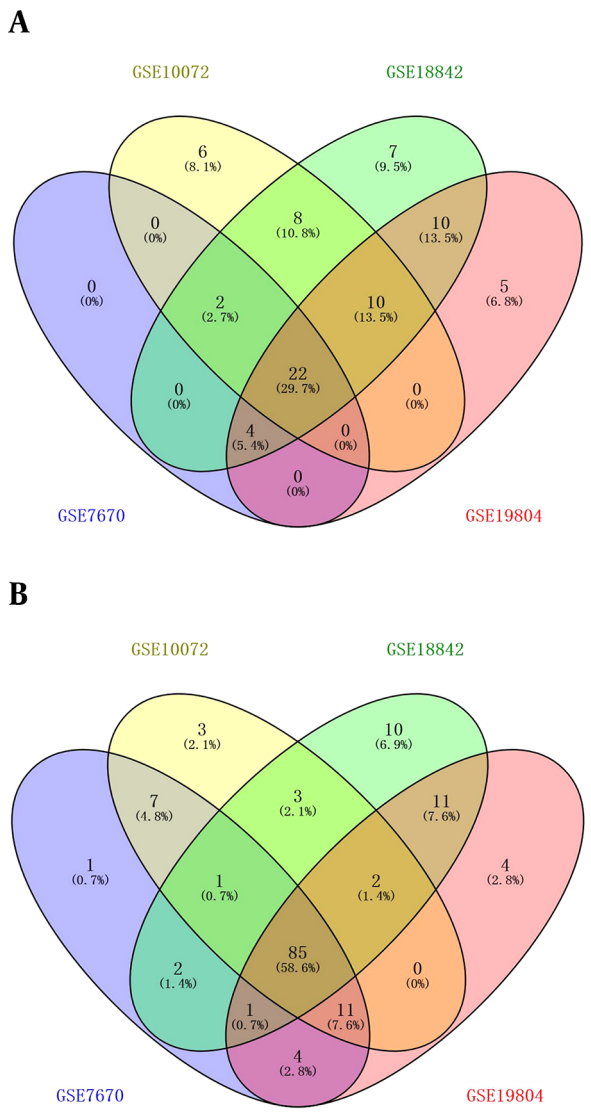

Figure 1. Significant pathways identified and overlapped. A and B respectively represented up-regulated and down-regulated pathways. "GSEXXXX" was GEO accession of microarray dataset. For each dataset, the researchers performed GSEA to generate P value for each pathway and used a permutation test with 1000 times, and obtained significant pathways with P values cut-off of $\leq$ 0.05. A, GSEA detected 28, 48, 63 and 51 up-regulated pathways and 22 common pathways were found; B, GSEA detected 112, 112, 115 and 118 down-regulated pathways and 85 common pathways were found.

4. Discussion

Finally, NSCLC mainly including adenocarcinoma and squamous cell carcinoma is the most common type of lung

Table 2. Reanalysis Results of Significantly Regulated Pathway Number

GEO Accession	No. of Genes After Preprocessed	No. of Pathways Have Genes $\geq$ 10	Up-Regulated Pathways	Down-Regulated Pathways
GSE7670	6268	184	28	112
GSE10072	5211	177	48	112
GSE18842	11588	203	63	115
GSE19804	11626	205	51	118
Common significantly regulated pathways	22	85		

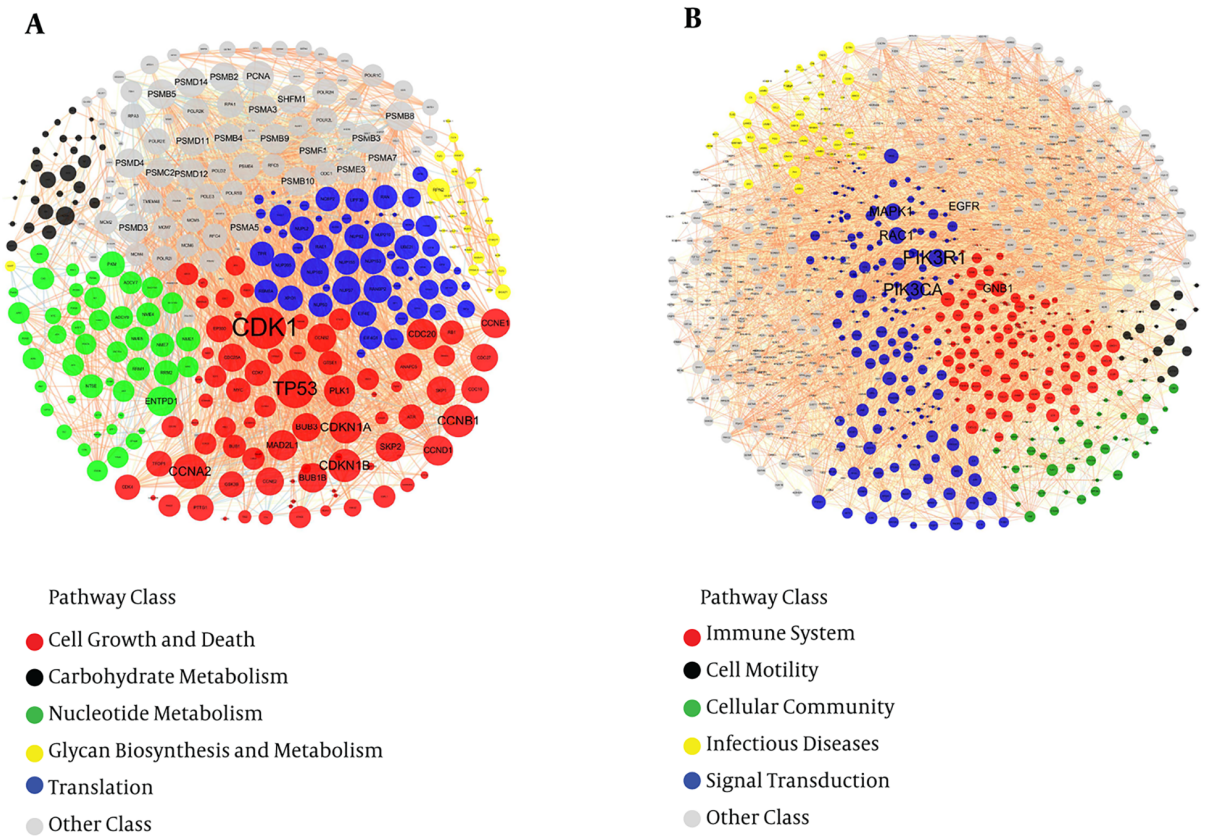

Figure 2. Protein and protein interaction(PPI) network of genes within significantly regulated pathways. A and B represented PPI network of the genes within up-regulated pathways and down-regulated pathways, respectively. Each node represented one gene. The node with color showed the gene belonging to the pathway class with the same color. Node size represented degree size of the node. The label of the node represented gene name. PPI, protein and protein interaction; A, genes of PPI network were mainly enriched within some pathways belonging to cell cycle and death, metabolism and so on. CDK1 and TP53 genes shared more abundant edges; B, Genes of PPI network were mainly enriched within some pathways belonging to the immune system, signal transduction and so on. PIK3R1 and PIK3CA genes shared more abundant edges.

cancer. However, early diagnosis and treatment of NSCLC are still difficult. One main reason is that the molecular mechanism implicated in NSCLC is vague. In this study, the researchers selected four microarray data of NSCLC to perform GSEA and PPI network analysis. Microarray data were from the same Affymetrix platform. The purpose was to minimize the error between chip platforms. In addition, these data included NSCLC patients from Asia, America, and Europe, and included two major subtypes of NSCLC and smoking status of NSCLC patients, which contributed to obtaining insight in the common molecular mechanism underlying NSCLC. Through GSEA and PPI network analysis, the results revealed that 107 pathways (22 up- and 85 down-regulated) were significantly dysreg-

ulated in NSCLC and the abnormal expression of CDK1 and PIK3R1 genes were associated with NSCLC.

Uncontrolled proliferation is one of the most prominent features of tumor cells. In the last decades, many studies have focused on the pathways related to cell growth and death in tumor formation. More and more results showed that cell cycle pathway and p53 signaling pathway played a key role in the formation of malignant tumors (9, 13-15). In this study, the researchers observed that cell cycle pathway (all P value < 0.001, from GSEA results of four independent microarray data) and p53 signaling pathway (all P value < 0.001, from GSEA results of four independent microarray data) were positively regulated in NSCLC. Furthermore, functional enrichment in PPI network showed that 65 and 37 genes were enriched within cell cycle pathway (P value = 2.58e-79, ranked first) and p53 signaling pathway (P value = 1.08e-45, ranked third), separately.

When the tumor was formed, sufficient energy, raw materials, and NADPH were required to provide for fast-growing cancer cells (16). Up-regulated pathways related to metabolism would provide the necessity to cancer cells. At present, many study results have shown that these pathways were related to cancers, such as prostate cancer and breast cancer and so on (16, 17). The current results showed that many metabolism pathways were significantly up-regulated in NSCLC, which mainly included purine metabolism (P value = 1.95e-44, ranked fourth), pyrimidine metabolism (P value = 1.01e-29, ranked fifth), glutathione metabolism (P value = 1.22e-27, ranked sixth), and amino sugar and nucleotide sugar metabolism (P value = 1.01e-21, ranked seventh).

Glycans, as important signaling molecules, attached to proteins or lipids and played an important role in malignant transformation (18). At present, glycans have been used as candidate diagnostic markers and therapeutic targets in clinics (19, 20). The modulation of N-Glycan biosynthesis (P value = 6.24e-14, ranked twelfth) changed the glycosylation of proteins and/or lipids, which made the functions and structures of glycoproteins and/or glycolipids change. The altered functions, such as cell signaling and cell adhesion, facilitated cancer invasion and metastasis (21).

An essential function of the immune system was immune surveillance, which played a key role in identifying and destroying tumors and defending against cancers (22). Once the immune system of the host was dysfunctional, tumors escaped the immune surveillance to transform cancers (23). Furthermore, tumor cells released some immunosuppressive cytokines, such as prostaglandins, vascular endothelial growth factor and transforming growth factor-beta to directly or indirectly inhibit the immune response (24). In the GSEA results, the researchers surprisingly found that 12 pathways related to the immune system were significantly down-regulated, which indicated that the immune system was strongly altered/inhibited in NSCLC (25).

In the pathways belonging to the cellular community, the researchers identified three pathways including tight junction pathway (P value = 7.24e-27, ranked nineteenth), gap junction pathway (P value = 2.65e-27, ranked eighteenth) and focal adhesion pathway (P value = 2.44e-81, ranked first), which were significantly down-regulated. Tight junctions played vital roles in creating an intercellular barrier, controlling para-cellular diffusion, and maintaining cell-cell junction and tissue integrity (26). The alterations in the expression or structures of tight junction proteins led to the loss of cohesion of tight junction structure, which resulted in the invasion and metastasis of cancer cells (26). Gap junction has been speculated to be essential in regular intercellular communication, and the loss of direct intercellular communication was found to be commonly associated with cancer onset and progression (27). A number of studies demonstrated that tumor promoters effectively inhibited the gap junctional between cells, while tumor suppressors effectively enhanced gap junction function (27-29). Focal adhesions, also called cell-matrix adhesions, similar to tight junction and gap junction, played crucial roles in mediating many processes, including migration and cell adhesion, tissue homeostasis, and tumorigenesis and so on (30). The loss or down-regulation of cell-cell and cell-matrix adhesion contributed to the invasion and metastasis of cancer cells (31). The current results showed that three pathways might play essential roles in cell migration of NSCLC.

At present, several studies have reported that some genes, such as EGFR, TP53, and PIK3CA, were associated with lung cancer (32-35). The current studies also showed that these genes shared more abundant edges in PPI networks (Figure 2), and further verified that the genes played an important role in NSCLC. However, the researchers observed that CDK1 and PIK3R1 genes shared the most abundant edges than the above genes in the sub-network, and were significantly up-regulated and down-regulated in NSCLC, separately. Currently, a few studies reported that the CDK1 gene were associated with carcinomas, including gastric, colorectal, breast, and lung cancers (36-39). Moreover, published results showed that some non-coding RNA inhibited cell proliferation of NSCLC by targeting CDK1 (39, 40). Despite these results, the role of CDK1 in NSCLC is still vague. The current results further proved the role of CDK1 in NSCLC, and CDK1 was the key gene in the PPI network. At present, PIK3R1 has been proved to be a double-sided factor in different cancers, and was a positive regulator in breast and endometrial cancers (41, 42) and was a negative regu-

lator in renal cancer (43). Few studies focused on the role of PIK3R1 in lung cancer. The current results showed that PIK3R1 was significantly down-regulated in NSCLC.

The strong point of the current study was to integrate gene profiling data of NSCLC from different subtypes, different people, and different smoking status to explore the molecular mechanism implicated in NSCLC using gene set analysis and PPI network analysis. Gene set analysis is more potent in revealing biological mechanisms than single-gene analysis, especially in identifying genes with subtle contributions. The PPI network analysis may confirm the interaction between genes, and contribute to the discovery of key genes in the biological process. Two methods are helpful for the in-depth understanding of the molecular mechanism of NSCLC. The primary limitation of the current study was that pure bioinformatics methods obtained the results. Experiments did not confirm the results. Next, greater attention to the results and verifying the genes by experiments to deepen the understanding of molecular mechanism of NSCLC is required.

5. Conclusions

A cross-sectional study of gene expression profiling data identified many pathways and genes implicated in NSCLC. Up-regulated cell cycle pathway and down-regulated focal adhesion pathway were significantly associated with NSCLC. The study increased the concordance between gene expression profiling data and provided insight into the molecular mechanisms of NSCLC. The CDK1 and PIK3R1 genes were identified as key genes of NSCLC and may serve as candidate diagnostic and therapeutic targets of NSCLC.

Acknowledgments

The authors thank Xiangqing Zhu and Xinghua Pan of national and regional integrated engineering laboratory of stem cell and immunocyte biological technology for helping discuss this manuscript.

Footnotes

Authors' Contribution: Qiang Chen designed and supervised the study. Tonglian Wang searched the microarrays from the NCBI GEO database with the help of Lutong Xu. Tonglian Wang and Hongbo Zhao performed the data processing and GSEA with the help of Qiang Chen. Tonglian Wang and Qiang Chen constructed the PPI network and wrote the manuscript with the help of Jing Hu and Yuanyue Li. Tao Shou and Xueshan Xia revised the manuscript. All authors have read this manuscript.

Conflict of Interests: The authors declare that they had no competing interests in this paper.

Financial Disclosure: The authors declare that they had no personal or professional financial conflicts.

Funding/Support: This work was supported by Yunnan Applied Basic Research Projects (No. 2016FB146), National Natural Science Foundation of China (No. 31660655), Funding for Fostering Talents of Kunming University of Science and Technology (No. KKSY201526053) and Funding of Health and Family Planning Commission of Yunnan Province (No. 2017NS261).

References

1. Edwards BK, Brown ML, Wingo PA, Howe HL, Ward E, Ries LA, et al. Annual report to the nation on the status of cancer, 1975-2002, featuring population-based trends in cancer treatment. *J Natl Cancer Inst*. 2005;**97**(19):1407–27. doi: 10.1093/jnci/dji289. [PubMed: 16204691].
2. Zhang H, Cai B. The impact of tobacco on lung health in China. *Respirology*. 2003;**8**(1):17–21. [PubMed: 12856737].
3. Hu Z, Wu C, Shi Y, Guo H, Zhao X, Yin Z, et al. A genome-wide association study identifies two new lung cancer susceptibility loci at 13q12.12 and 22q12.2 in Han Chinese. *Nat Genetics*. 2011;**43**(8):792–6. doi: 10.1038/ng.875.
4. Petty RD. Gene Expression Profiling in Non-Small Cell Lung Cancer: From Molecular Mechanisms to Clinical Application. *Clin Cancer Res*. 2004;**10**(10):3237–48. doi: 10.1158/1078-0432.ccr-03-0503.
5. Yu D, Li J, Han Y, Liu S, Xiao N, Li Y, et al. Gene expression profiles of ERCC1, TYMS, RRM1, TUBB3 and EGFR in tumor tissue from non-small cell lung cancer patients. *Chin Med J (Engl)*. 2014;**127**(8):1464–8. [PubMed: 24762590].
6. Wang J, Song J, Gao Z, Huo X, Zhang Y, Wang W, et al. Analysis of gene expression profiles of non-small cell lung cancer at different stages reveals significantly altered biological functions and candidate genes. *Oncol Rep*. 2017;**37**(3):1736–46. doi: 10.3892/or.2017.5380.
7. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*. 2005;**102**(43):15545–50. doi: 10.1073/pnas.0506580102. [PubMed: 16199517]. [PubMed Central: PMC1239896].
8. Mootha VK, Lindgren CM, Eriksson KF, Subramanian A, Sihag S, Lehar J, et al. PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet*. 2003;**34**(3):267–73. doi: 10.1038/ng1180. [PubMed: 12808457].
9. Cai B, Jiang X. Revealing Biological Pathways Implicated in Lung Cancer from TCGA Gene Expression Data Using Gene Set Enrichment Analysis. *Cancer Inform*. 2014;**13**(Suppl 1):113–21. doi: 10.4137/CIN.S13882. [PubMed: 25520551]. [PubMed Central: PMC4251186].
10. Araujo JM, Prado A, Cardenas NK, Zaharia M, Dyer R, Doimi F, et al. Repeated observation of immune gene sets enrichment in women with non-small cell lung cancer. *Oncotarget*. 2016;**7**(15):20282–92. doi: 10.18632/oncotarget.7943. [PubMed: 26958810]. [PubMed Central: PMC4991454].
11. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics*. 2003;**4**(2):249–64. doi: 10.1093/biostatistics/4.2.249. [PubMed: 12925520].

12. Zhao H, Huang M, Chen Q, Wang Q, Pan Y. Comparative gene expression analysis in mouse models for identifying critical pathways in mammary gland development. *Breast Cancer Res Treat*. 2012;**132**(3):969–77. doi: 10.1007/s10549-011-1650-8. [PubMed: 21735046].
13. Shi I, Hashemi Sadraei N, Duan ZH, Shi T. Aberrant signaling pathways in squamous cell lung carcinoma. *Cancer Inform*. 2011;**10**:273–85. doi: 10.4137/CIN.S8283. [PubMed: 22174565]. [PubMed Central: PMC3236010].
14. Cancer Genome Atlas Research. Comprehensive genomic characterization of squamous cell lung cancers. *Nature*. 2012;**489**(7417):519–25. doi: 10.1038/nature11404.
15. Cooper WA, Lam DC, O'Toole SA, Minna JD. Molecular biology of lung cancer. *J Thorac Dis*. 2013;**5**(5):479–90. doi: 10.3978/j.issn.2072-1439.2013.08.03.
16. Schramm G, Surmann EM, Wiesberg S, Oswald M, Reinelt G, Eils R, et al. Analyzing the regulation of metabolic pathways in human breast cancer. *BMC Med Genomics*. 2010;**3**(1). doi: 10.1186/1755-8794-3-39.
17. Tsouko E, Khan AS, White MA, Han JJ, Shi Y, Merchant FA, et al. Regulation of the pentose phosphate pathway by an androgen receptor–mTOR-mediated mechanism and its role in prostate cancer cell growth. *Oncogenesis*. 2014;**3**(5):e103. doi: 10.1038/oncsis.2014.18.
18. Dennis JW, Granovsky M, Warren CE. Glycoprotein glycosylation and cancer progression. *Biochim Biophys Acta*. 1999;**1473**(1):21–34. eng. [PubMed: 10580127].
19. Kailemia MJ, Park D, Lebrilla CB. Glycans and glycoproteins as specific biomarkers for cancer. *Anal Bioanal Chem*. 2017;**409**(2):395–410. eng. doi: 10.1007/s00216-016-9880-6. [PubMed: 27590322]. [PubMed Central: PMCPmc5203967].
20. Lan Y, Hao C, Zeng X, He Y, Zeng P, Guo Z, et al. Serum glycoprotein-derived N- and O-linked glycans as cancer biomarkers. *Am J Cancer Res*. 2016;**6**(11):2390–415. [PubMed: 27904760]. [PubMed Central: PMC5126262].
21. Zhao YY, Takahashi M, Gu JG, Miyoshi E, Matsumoto A, Kitazume S, et al. Functional roles of N-glycans in cell signaling and cell adhesion in cancer. *Cancer Sci*. 2008;**99**(7):1304–10. doi: 10.1111/j.1349-7006.2008.00839.x. [PubMed: 18492092].
22. Swann JB, Smyth MJ. Immune surveillance of tumors. *J Clin Invest*. 2007;**117**(5):1137–46. doi: 10.1172/JCI31405. [PubMed: 17476343]. [PubMed Central: PMC1857231].
23. Seliger B. Strategies of tumor immune evasion. *BioDrugs*. 2005;**19**(6):347–54. [PubMed: 16392887].
24. Frumento G, Piazza T, Di Carlo E, Ferrini S. Targeting tumor-related immunosuppression for cancer immunotherapy. *Endocr Metab Immune Disord Drug Targets*. 2006;**6**(3):233–7. eng. [PubMed: 17017974].
25. Domagala-Kulawik J, Osinska I. [Immune alterations in lung cancer - the new therapeutic approach]. *Pneumonol Alergol Pol*. 2014;**82**(3):286–99. pol. doi: 10.5603/PiAP.2014.0034. [PubMed: 24793154].
26. Martin TA. The role of tight junctions in cancer metastasis. *Semin Cell Dev Biol*. 2014;**36**:224–31. eng. doi: 10.1016/j.semcdb.2014.09.008. [PubMed: 25239399].
27. Aasen T, Mesnil M, Naus CC, Lampe PD, Laird DW. Gap junctions and cancer: communicating for 50 years. *Nat Rev Cancer*. 2016;**16**(12):775–88. eng. doi: 10.1038/nrc.2016.105. [PubMed: 27782134]. [PubMed Central: PMCPmc5279857].
28. Trosko JE, Ruch RJ. Gap junctions as targets for cancer chemoprevention and chemotherapy. *Curr Drug Targets*. 2002;**3**(6):465–82. eng. [PubMed: 12448698].
29. Shi H, Shi D, Wu Y, Shen Q, Li J. Qigesan inhibits migration and invasion of esophageal cancer cells via inducing connexin expression and enhancing gap junction function. *Cancer Lett*. 2016;**380**(1):184–90. eng. doi: 10.1016/j.canlet.2016.06.015. [PubMed: 27345741].
30. Berrier AL, Yamada KM. Cell-matrix adhesion. *J Cell Physiol*. 2007;**213**(3):565–73. eng. doi: 10.1002/jcp.21237. [PubMed: 17680633].
31. Nigam AK, Savage FJ, Boulos PB, Stamp GW, Liu D, Pignatelli M. Loss of cell-cell and cell-matrix adhesion molecules in colorectal cancer. *Br J Cancer*. 1993;**68**(3):507–14. eng. [PubMed: 8353041]. [PubMed Central: PMCPmc1968382].
32. Toyooka S, Tsuda T, Gazdar AF. The TP53 gene, tobacco exposure, and lung cancer. *Hum Mutat*. 2003;**21**(3):229–39. eng. doi: 10.1002/humu.10177. [PubMed: 12619108].
33. Shigematsu H, Lin L, Takahashi T, Nomura M, Suzuki M, Wistuba ,I, et al. Clinical and biological features associated with epidermal growth factor receptor gene mutations in lung cancers. *J Natl Cancer Inst*. 2005;**97**(5):339–46. eng. doi: 10.1093/jnci/dji055. [PubMed: 15741570].
34. Martin P, Kelly CM, Carney D. Epidermal growth factor receptor-targeted agents for lung cancer. *Cancer Control*. 2006;**13**(2):129–40. eng. doi: 10.1177/107327480601300207. [PubMed: 16735987].
35. Yamamoto H, Shigematsu H, Nomura M, Lockwood WW, Sato M, Okumura N, et al. PIK3CA mutations and copy number gains in human lung cancers. *Cancer Res*. 2008;**68**(17):6913–21. eng. doi: 10.1158/0008-5472.can-07-5084. [PubMed: 18757405]. [PubMed Central: PMCPmc2874836].
36. Gao SY, Li J, Qu XY, Zhu N, Ji YB. Downregulation of Cdk1 and cyclinB1 expression contributes to oridonin-induced cell cycle arrest at G2/M phase and growth inhibition in SGC-7901 gastric cancer cells. *Asian Pac J Cancer Prev*. 2014;**15**(15):6437–41. eng. [PubMed: 25124639].
37. Sung WW, Lin YM, Wu PR, Yen HH, Lai HW, Su TC, et al. High nuclear/cytoplasmic ratio of Cdk1 expression predicts poor prognosis in colorectal cancer patients. *BMC Cancer*. 2014;**14**:951. eng. doi: 10.1186/1471-2407-14-951. [PubMed: 25511643]. [PubMed Central: PMCPmc4302138].
38. Nakayama S, Torikoshi Y, Takahashi T, Yoshida T, Sudo T, Matsushima T, et al. Prediction of paclitaxel sensitivity by CDK1 and CDK2 activity in human breast cancer cells. *Breast Cancer Res*. 2009;**11**(1):R12. eng. doi: 10.1186/bcr2231. [PubMed: 19239702]. [PubMed Central: PMCPmc2687717].
39. Shi Q, Zhou Z, Ye N, Chen Q, Zheng X, Fang M. MiR-181a inhibits non-small cell lung cancer cell proliferation by targeting CDK1. *Cancer Biomark*. 2017;**20**(4):539–46. eng. doi: 10.3233/cbm-170350. [PubMed: 28946554].
40. Pu S, Zhao Y, Zhou G, Zhu H, Gong L, Zhang W, et al. Effect of CDK1 shRNA on proliferation, migration, cell cycle and apoptosis in non-small cell lung cancer. *J Cell Physiol*. 2018;**233**(9):7514. eng. doi: 10.1002/jcp.26387. [PubMed: 29226963].
41. Yan LX, Liu YH, Xiang JW, Wu QN, Xu LB, Luo XL, et al. PIK3R1 targeting by miR-21 suppresses tumor cell migration and invasion by reducing PI3K/AKT signaling and reversing EMT, and predicts clinical outcome of breast cancer. *Int J Oncol*. 2016;**48**(2):471–84. eng. doi: 10.3892/ijo.2015.3287. [PubMed: 26676464]. [PubMed Central: PMCPmc4725461].
42. Cheung LW, Hennessy BT, Li J, Yu S, Myers AP, Djordjevic B, et al. High frequency of PIK3R1 and PIK3R2 mutations in endometrial cancer elucidates a novel mechanism for regulation of PTEN protein stability. *Cancer Discov*. 2011;**1**(2):170–85. eng. doi: 10.1158/2159-8290.cd-11-0039. [PubMed: 21984976]. [PubMed Central: PMCPmc3187555].
43. Lin Y, Yang Z, Xu A, Dong P, Huang Y, Liu H, et al. PIK3R1 negatively regulates the epithelial-mesenchymal transition and stem-like phenotype of renal cancer cells through the AKT/GSK3beta/CTNNB1 signaling pathway. *Sci Rep*. 2015;**5**:8997. eng. doi: 10.1038/srep08997. [PubMed: 25757764]. [PubMed Central: PMCPmc4355729].

Table 3. Common Significant Pathways Identified of Four Datasets by Gene Set Enrichment Analysis (GSEA)[a,b,c]

Entry	Pathway Name	Class	Number of Overlapping /Enriching Genes	Percentage of Common Genes, %	FDR
		Up-Regulated			
04110	Cell cycle	Cell growth and death	66/65	56.40	2.58E-79
03013	RNA transport	Translation	54/53	45.80	4.22E-52
04115	p53 signaling pathway	Cell growth and death	38/37	59.40	1.08E-45
00230	Purine metabolism	Nucleotide metabolism	52/49	39.10	1.95E-44
00240	Pyrimidine metabolism	Nucleotide metabolism	34/32	42.00	1.01E-29
00480	Glutathione metabolism	Metabolism of other amino acids	25/24	58.10	1.22E-27
00520	Amino sugar and nucleotide sugar metabolism	Carbohydrate metabolism	21/20	46.70	1.10E-21
03050	Proteasome	Folding, sorting and degradation	21/19	53.80	8.41E-21
00051	Fructose and mannose metabolism	Carbohydrate metabolism	18/16	58.10	1.26E-18
00250	Alanine, aspartate and glutamate metabolism	Amino acid metabolism	17/16	65.40	6.75E-18
03030	DNA replication	Replication and repair	15/15	44.10	2.62E-16
00510	N-Glycan biosynthesis	Glycan biosynthesis and metabolism	16/15	38.10	6.24E-14
03008	Ribosome biogenesis in eukaryotes	Translation	15/15	24.60	4.64E-11
00512	Mucin type O-Glycan biosynthesis	Glycan biosynthesis and metabolism	11/11	47.80	5.14E-11
00030	Pentose phosphate pathway	Carbohydrate metabolism	11/10	50.00	1.78E-10
03060	Protein export	Folding, sorting and degradation	9/9	45.00	9.48E-10
03410	Base excision repair	Replication and repair	10/10	38.50	2.34E-09
00983	Drug metabolism - other enzymes	Xenobiotics biodegradation and metabolism	12/10	52.20	3.46E-08
03430	Mismatch repair	Replication and repair	8/8	40.00	3.68E-08
00601	Glycosphingolipid biosynthesis - lacto and neolacto series	Glycan biosynthesis and metabolism	9/8	40.90	1.06E-07
03020	RNA polymerase	Transcription	9/8	34.60	3.26E-07
00860	Porphyrin and chlorophyll metabolism	Metabolism of cofactors and vitamins	6/6	26.10	0.000446
		Down-Regulated			
04510	Focal adhesion	Cellular community - eukaryotes	103/102	57.50	2.44E-81
04010	MAPK signaling pathway	Signal transduction	100/95	46.70	2.36E-61
04062	Chemokine signaling pathway	Immune system	84/77	54.90	7.64E-55
04144	Endocytosis	Transport and catabolism	84/79	47.20	2.60E-54
04145	Phagosome	Transport and catabolism	76/69	61.80	1.77E-53
04810	Regulation of actin cytoskeleton	Cell motility	82/79	46.30	3.19E-51

04060	Cytokine-cytokine receptor interaction	Signaling molecules and interaction	90/87	49.50	3.19E-51
04380	Osteoclast differentiation	Development	64/62	56.10	2.23E-49
04514	Cell adhesion molecules (CAMs)	Signaling molecules and interaction	65/59	59.10	3.00E-42
05146	Amoebiasis	Infectious diseases	2/56	1.50	3.13E-41
05145	Toxoplasmosis	Infectious diseases	60/51	56.10	7.16E-38
04670	Leukocyte transendothelial migration	Immune system	55/50	56.70	2.19E-36
05142	Chagas disease (American trypanosomiasis)	Infectious diseases	46/44	35.70	6.66E-33
04610	Complement and coagulation cascades	Immune system	38/38	69.10	1.07E-32
05323	Rheumatoid arthritis	Immune diseases	43/40	56.60	1.52E-30
05150	*Staphylococcus aureus* infection	Infectious diseases	35/32	83.30	6.28E-30
04630	Jak-STAT signaling pathway	Signal transduction	55/51	52.40	6.28E-30
04540	Gap junction	Cellular community - eukaryotes	40/37	56.30	2.65E-27
04530	Tight junction	Cellular community - eukaryotes	49/44	47.10	7.24E-27
04640	Hematopoietic cell lineage	Immune system	40/37	58.00	7.24E-27
04270	Vascular smooth muscle contraction	Circulatory system	46/42	52.30	9.53E-27
04660	T cell receptor signaling pathway	Immune system	45/39	46.40	1.57E-26
05140	Leishmaniasis	Infectious diseases	36/33	62.10	1.04E-25
04666	Fc gamma R-mediated phagocytosis	Immune system	40/35	47.10	4.16E-24
04650	Natural killer cell mediated cytotoxicity	Immune system	46/40	47.90	2.28E-23
04722	Neurotrophin signaling pathway	Nervous system	44/38	38.60	1.55E-22
04350	TGF-beta signaling pathway	Signal transduction	35/31	47.30	1.67E-21
04910	Insulin signaling pathway	Endocrine system	0/39	0	3.63E-21
04662	B cell receptor signaling pathway	Immune system	34/29	47.90	8.35E-21
04916	Melanogenesis	Endocrine system	35/33	45.50	2.52E-20
04210	Apoptosis	Cell growth and death	34/31	43.60	3.32E-20
05100	Bacterial invasion of epithelial cells	Infectious diseases	31/29	48.40	5.25E-20
05120	Epithelial cell signaling in *Helicobacter pylori* infection	Infectious diseases	27/27	45.80	2.70E-19
04972	Pancreatic secretion	Digestive system	33/31	49.30	3.32E-19
04020	Calcium signaling pathway	Signal transduction	46/42	38.70	3.92E-19
05416	Viral myocarditis	Cardiovascular diseases	2/25	1.50	4.14E-19
03320	PPAR signaling pathway	Endocrine system	27/27	54.00	9.82E-19
04150	mTOR signaling pathway	Signal transduction	23/25	48.90	3.13E-18
05144	Malaria	Infectious diseases	29/22	69.00	1.48E-17
04970	Salivary secretion	Digestive system	29/28	48.30	2.15E-17
05221	Acute myeloid leukemia	Cancers	25/23	47.20	1.01E-16

05210	Colorectal cancer	Cancers	29/23	48.30	2.53E-16
04920	Adipocytokine signaling pathway	Endocrine system	24/24	47.10	1.32E-15
04940	Type I diabetes mellitus	Endocrine and metabolic diseases	22/19	68.80	2.92E-15
04960	Aldosterone-regulated sodium reabsorption	Excretory system	22/18	68.80	4.92E-15
05020	Prion diseases	Neurodegenerative diseases	18/18	66.70	4.92E-15
05213	Endometrial cancer	Cancers	23/20	46.90	2.83E-14
04971	Gastric acid secretion	Digestive system	24/23	53.30	4.19E-14
04730	Long-term depression	Nervous system	27/21	50.90	6.33E-14
04912	GnRH signaling pathway	Endocrine system	30/25	39.50	9.73E-14
05143	African trypanosomiasis	Infectious diseases	16/16	66.70	3.15E-13
04370	VEGF signaling pathway	Signal transduction	30/20	46.90	1.03E-12
04664	Fc epsilon RI signaling pathway	Immune system	28/21	41.80	1.08E-12
05332	Graft-versus-host disease	Immune diseases	19/16	63.30	1.64E-12
04976	Bile secretion	Digestive system	22/21	45.80	2.05E-12
05414	Dilated cardiomyopathy	Cardiovascular diseases	30/22	44.80	6.07E-11
04962	Vasopressin-regulated water reabsorption	Excretory system	19/16	52.80	6.23E-11
05330	Allograft rejection	Immune diseases	17/14	65.40	1.51E-10
04964	Proximal tubule bicarbonate reclamation	Excretory system	12/11	75.00	1.62E-10
04720	Long-term potentiation	Nervous system	21/18	38.90	3.39E-10
00982	Drug metabolism - cytochrome P450	Xenobiotics biodegradation and metabolism	20/18	46.50	3.39E-10
00071	Fatty acid degradation	Lipid metabolism	16/15	47.10	4.92E-10
05412	Arrhythmogenic right ventricular cardiomyopathy (ARVC)	Cardiovascular diseases	27/19	45.80	6.03E-10
04930	Type II diabetes mellitus	Endocrine and metabolic diseases	18/15	50.00	9.89E-10
00380	Tryptophan metabolism	Amino acid metabolism	17/14	54.80	1.21E-09
04621	NOD-like receptor signaling pathway	Immune system	25/16	49.00	3.26E-09
00590	Arachidonic acid metabolism	Lipid metabolism	17/16	42.50	1.24E-08
05410	Hypertrophic cardiomyopathy (HCM)	Cardiovascular diseases	26/18	41.90	2.99E-08
05320	Autoimmune thyroid disease	Immune diseases	17/14	63.00	3.27E-08
00980	Metabolism of xenobiotics by cytochrome P450	Xenobiotics biodegradation and metabolism	19/16	45.20	5.21E-08
04672	Intestinal immune network for IgA production	Immune system	15/13	46.90	6.73E-08
00564	Glycerophospholipid metabolism	Lipid metabolism	19/18	31.10	1.58E-07
05310	Asthma	Immune diseases	12/10	66.70	3.38E-07
04710	Circadian rhythm	Environmental adaptation	9/10	45.00	4.86E-07
04973	Carbohydrate digestion and absorption	Digestive system	13/11	48.10	9.06E-07
02010	ABC transporters	Membrane transport	14/11	45.20	1.58E-06

04070	Phosphatidylinositol signaling system	Signal transduction	24/15	37.50	2.55E-06
00340	Histidine metabolism	Amino acid metabolism	11/8	45.80	2.67E-05
04260	Cardiac muscle contraction	Circulatory system	14/13	29.80	4.04E-05
04080	Neuroactive ligand-receptor interaction	Signaling molecules and interaction	27/26	24.10	0.000284
00562	Inositol phosphate metabolism	Carbohydrate metabolism	20/9	43.50	0.00224
04623	Cytosolic DNA-sensing pathway	Immune system	17/8	50.00	0.00856
04320	Dorso-ventral axis formation	Development	10/4	50.00	0.0232
04130	SNARE interactions in vesicular transport	Folding, sorting and degradation	16	48.50	Not be enriched
00830	Retinol metabolism	Metabolism of cofactors and vitamins	6	26.10	Not be enriched

Abbreviation: FDR, False discovery rate. FDR was obtained according to the results computed by STRING platform.
[a] Number of overlapping genes was obtained according to the overlap of genes within each common pathway of four datasets.
[b] Number of enriching genes was obtained according to the enriching results of all genes within all common pathways.
[c] Two pathways with 04130 and 00830 entry were not enriched in functional enrichment of Protein and Protein Interaction (PPI) network, but were significantly regulated by GSEA.