Mathematical Biosciences and Engineering

MBE, 16 (6): 7671–7687.
DOI: 10.3934/mbe.2019384
Received: 20 February 2019
Accepted: 10 July 2019
Published: 21 August 2019

http://www.aimspress.com/journal/MBE

Research article

The identification of key biomarkers in patients with lung adenocarcinoma based on bioinformatics

Kewei Ni[1,2] and Gaozhong Sun[1,2,*]

[1] Department of Cardio-Thoracic Surgery, Zhejiang Provincial People's Hospital
[2] People's Hospital of Hangzhou Medical College

*** Correspondence:** Tel: +86-13588806185; Email: sungz2002@163.com.

Abstract: Lung adenocarcinoma (LUAD) is one of the leading causes of cancer death globally. This study aims to investigate the underlying mechanisms implicated with LUAD and identify the key biomarkers. LUAD-associated gene expression dataset (GSE10072) was obtained from GEO database. Based on the GEO2R tool, we screened the differentially expressed genes (DEGs) between the patients with LUAD and normal individuals. Subsequently, Gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analyses were employed to find out the enriched pathways of these DEGs. Meanwhile, a protein-protein interaction (PPI) network was also employed to construct to visualize the interactions of these DEGs. Finally, the survival analysis of the top5 up-regulated and top5 down-regulated genes were conducted via GEPIA, aiming to figure out their potential effects on LUAD. In our study, a total of 856 DEGs were captured, including 559 up-regulated genes and 297 down-regulated genes. Among these DEGs, the top5 up-regulated genes were AGER, SFTPC, FABP4, CYP4B1 and WIF1 while the top5 down-regulated genes were GREM1, SPINK1, MMP1, COL11A1 and SPP1. GO analysis disclosed that these DEGs were mainly enriched in DNA synthesis, cell adhesion, signal transduction and cell apoptosis. KEGG analysis unveiled that the enriched pathway included pathways in cancer, PI3K/Akt signaling pathway, MAPK signaling pathway and cell cycle. Survival analysis showed that the expression level of ZG16 may correlate with the prognosis of LUAD patients. Furthermore, according to the connectivity degree of these DEGs, we selected the top15 hub genes, namely IL6, MMP9, EDN1, FOS, CDK1, CDH1, BIRC5, VWF, UBE2C, CDKN3, CDKN2A, CD34, AURKA, CCNB2 and EGR1, which were expected to be promising therapeutic target in LUAD. In conclusion, our study disclosed potential biomarkers and candidate targets in LUAD, which could be helpful to the diagnosis and treatment of LUAD.

Keywords: Lung adenocarcinoma, biomarkers, bioinformatics analysis

1. Introduction

Lung cancer, the most common cause of cancer-related mortality around the world, giving rise to over a million deaths each year [1]. Lung adenocarcinoma (LUAD) is one of the most common histological types of lung cancer [2]. It is well acknowledged that smoking is the major risk factor and cause of LUAD. Unexpectedly, more LUAD cases occur proportionally in people without smoking history (defined as less than 100 cigarettes in a lifetime) following the decreased smoking rates [3]. The pathogenesis of LUAD is a complicated process implicated with the progressive accumulation of gene alterations that pinpoint various molecular and cellular events involving autophagy, endoplasmic reticulum stress, oxidative stress, and abnormal cell cycle [4–8].

In recent years, apart from traditional surgery, chemotherapy and radiotherapy, targeted therapy has also greatly improved treatment for patients whose tumours harbour somatically activated oncogenes including mutant EGFR1, ERBB2 and BRAF or translocated RET, ALK or ROS1, however, the likelihood of a complete cure for the patients with LUAD is very slim [9,10]. Hence, the detection of early-stage biomarkers and identification of core therapeutic target appears significant to decrease LUAD-related deaths.

The recent high-throughput gene microarray has been widely employed to screen the differentially expressed genes (DEGs) between normal samples and tumor samples in human beings and animal models, which makes it accessible for us to further explore the entire molecular alterations of tumors at multiple levels involving DNA, RNA, proteins, epigenetic alterations, and metabolism [11]. However, there still exist obstacles to put these microarrays in application in clinic for the reason that the number of DEGs identified by gene profiling were huge and the statistical analyses were also too complicated [12,13]. Therefore, it is urgent to verify a proper number of genes and develop a suitable approach which can be operated by routine assay in clinic.

In this study, we downloaded the GSE10072 from Gene Expression Omnibus (GEO) and applied bioinformatics analysis to screen the DEGs between LUAD and normal control. Moreover, we carried out the functional analysis of these DEGs, including biological process (BP), molecular function (MF), cellular component (CC) and KEGG pathways. We chose top5 up-regulated and top5 down-regulated DEGs to make the overall survival (OS) analysis and stage plot. Finally, we used STRING to construct the protein-protein interaction (PPI) network to identify the hub genes with top15 degree of connectivity in LUAD. These genes will assist us to screen and identify significant biomarkers and therapeutic target of LUAD in the near future.

2. Materials and method

2.1. Screening database

The gene expression profile of GSE10072 was downloaded from the GEO database, which was a free and publicly available database. 58 LUAD tumor tissues and 49 non-tumor tissues from 20 never smokers, 26 former smokers, and 28 current smokers in this dataset were detected by GPL96 [HG-U133A] Affymetrix Human Genome U133A Array by Landi MT. We also downloaded the

Series Matrix File of GSE10072 from the GEO database.

2.2. DEGs analysis

In our study, the online software GEO2R was employed to analyze the tissue samples from GSE10072 dataset. GEO2R is an online software by which users can divide the samples into two and more groups and select out the DEGs. We used the Benjamini and Hochberg methods by default to discover false rate and used the adjust P value to reduce the errors of false positive. The choice criterion contains the adjust P value < 0.05 and $|\log FC| \geq 1$.

2.3. Gene ontology and KEGG pathway analysis of DEGs

Gene ontology (GO) analysis is a common framework annotating genes and gene products including functions of cellular components, biological pathways and molecular function [14]. Kyoto Encyclopedia of Genes and Genomes (KEGG) contains a set of genomes and biological pathways related with disease and drugs online database, which essentially is a resource for systematic understanding of biological system and certain high-level genome functional information [15]. The Database for Annotation, Visualization and Integrated Discovery (DAVID, http://david.ncifcrf.gov) is an online bioinformatics database [16]. It has widely covered a great many biological data and relevant analysis tools, then provided tools for the biological function annotation information for plenty of genes or proteins. $P < 0.05$ was considered as the cut-off criterion with significant difference. We could visualize the key biological processes, molecular functions, cellular components and pathways of DEGs by using DAVID online database. And further the scatter plot was performed by ImageGP according to the results of GO and KEGG pathway.

2.4. Comparison of the top5 upregulated and top5 downregulated DEGs

GEPIA (http://gepia.cancer-pku.cn/index.html), designed by Chenwei Li, Zefang Tang, and Boxi Kang of Zhang Lab, Peking University, is a newly developed interactive web server aiming at analyzing the RNA sequencing expression data of 9736 tumors and 8587 normal samples from the GTEx and TCGA projects in a standard processing pipeline [17]. In our study, we employed the boxplot to visualize the mRNA expression of top5 upregulated and top5 downregulated DEGs in LUAD tissues and and normal colorectum tissues.

2.5. The overall survival (OS) and stage plot of the top5 upregulated and top5 downregulated DEGs

Similarly, we used the GEPIA database to get the overall survival (OS) and stage plot information of these DEGs. The logrank P value and hazard ratio (HR) with 95% confidence intervals were showed on the plot. P < 0 05 was statistically significant.

2.6. PPI network and module analysis

Search Tool for the Retrieval of Interacting Genes (STRING) is an online tool for assessment and integration of the protein-protein interaction (PPI) information, containing physical and

functional associations. It covered 9,643,763 proteins from 2031 organisms in STRING version 10.0. We drew DEGs using STRING to evaluate the interactional associations among them, thereby utilized the Cytoscape sofrware to build a PPI network. We set maximum number of interactors = 0, confidence score ≥ 0.4 as the cut off criterion.

3. Results

3.1. Identification of DEGs

In our study, 58 tumor tissues from patients with LUAD and 49 non-tumor tissues from normal individuals were included and analyzed. We applied the GEO2R online analysis tool with default parameters to screen the DEGs, using adjusted P value < 0.05 and logFC ≤ −1 or logFC ≥ 1 as the cut-off criteria. We captured 856 DEGs were captured, including 559 up-regulated genes and 297 down-regulated genes. Whereafter, the DEGs were presented in the form of a volcano plot (Figure 1). Among the 856 DEGs, the top5 up-regulated genes involved AGER, SFTPC, FABP4, CYP4B1 and WIF1 while the top5 down-regulated genes were GREM1, SPINK1, MMP1, COL11A1 and SPP1. The gene tiles and biological functions of top5 upregulated and top5 down regulated genes were displayed in Table 1.

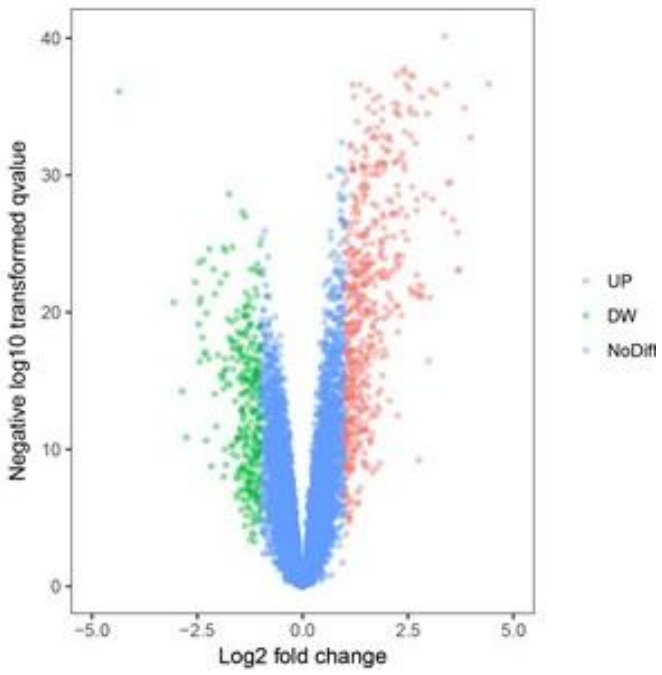

Figure 1. Heatmap of the differentially expressed genes (DEGs) between adjacent mucosa and carcinoma tissues from patients with lung adenocarcinoma (LUAD).

3.2. Validation of these DEGs by TCGA database

To ensure the credibility of the microarray of GSE10072 and proceed further credible analysis, we validated the top5 up-regulated genes and top5 down-regulated genes based on TCGA database

via GEPIA. The results based on GEPIA demonstrated that the mRNA expression levels of GREM1, SPINK1, MMP1, COL11A1 and SPP1 were significantly lower in carcinoma group compared to non-tumor group while the mRNA expression level of AGER, SFTPC, FABP4, CYP4B1 and WIF1 in carcinoma group were statistically higher than the non-tumor group ($P < 0.05$) (Figure 2A–J).

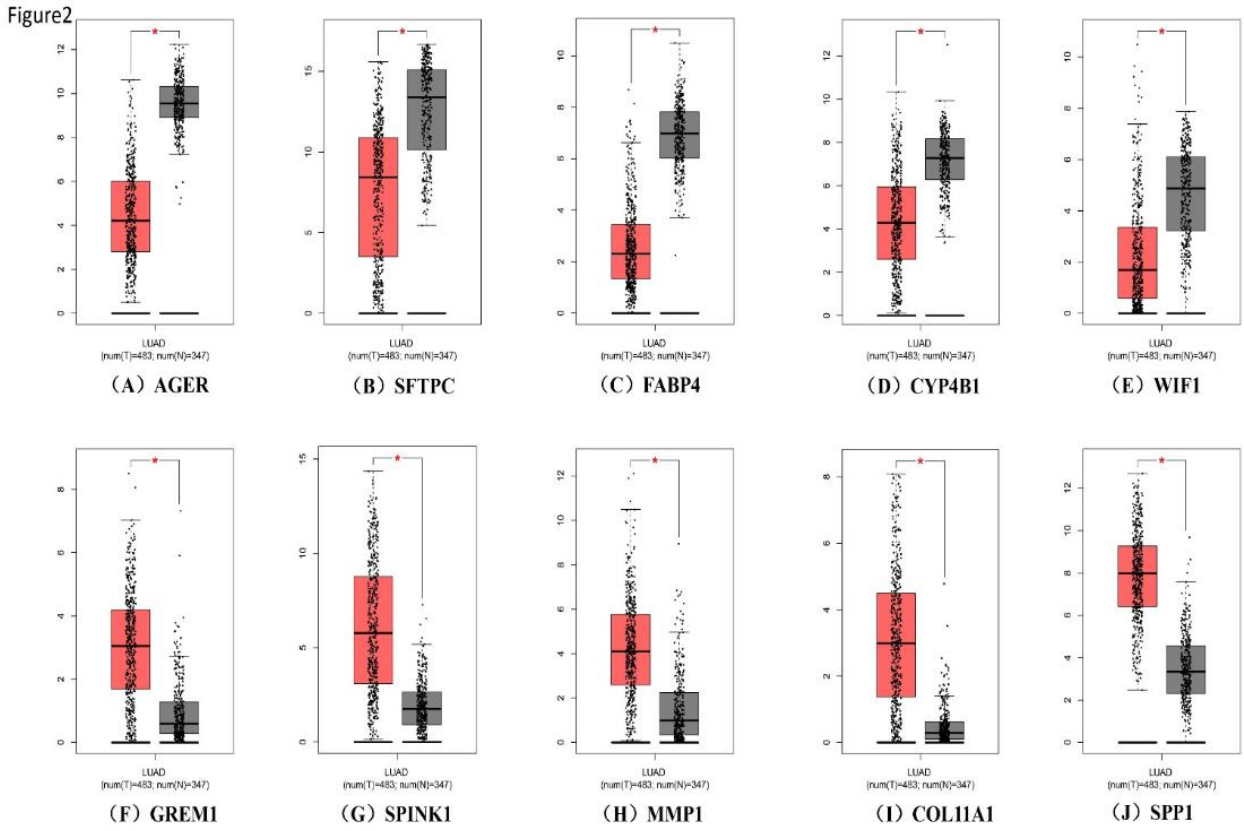

Figure 2. The mRNA expression of top5 up-regulated and top5 down-regulated genes based on TCGA database. (A-J) represents AGER, SFTPC, FABP4, CYP4B1, WIF1, GREM1, SPINK1, MMP1, COL11A1 and SPP1. T: tumor； N: Normal. * $P < 0.05$ versus normal group. $P < 0$ 05 was regarded statistically different.

3.3. Overall survival (OS)

Furthermore, we analyzed the potential association between the expression levels of top5 up-regulated genes as well as top5 down-regulated genes and the OS of patients with LUDA (Figure 3A–J). The Kaplan-Meier showed that AGER, SFTPC, CYP4B1, COL11A1 and SPP1 displayed significantly correlation with the OS of patients with LUAD. In detail, the high level of AGER and CYP4B1 may contribute to poorer prognosis of LUAD while the high level of SFTPC, COL11A1 and SPP1 may contribute to better prognosis ($P < 0.05$).

3.4. Correlation between DEGs expression and tumor stage in LUAD patients

Meanwhile, we analyzed the correlation between DEGs expression and tumor stage in LUAD patients. The results showed that the expression level of AGER, SFTPC, CYP4B1, WIF1, GREM1,

MMP1 and COL11A1 displayed strong correlation with the tumor stage in patients with LUAD while the FABP4, SPINK1 and SPINK1 groups did not significantly differ (Figure 4A–J).

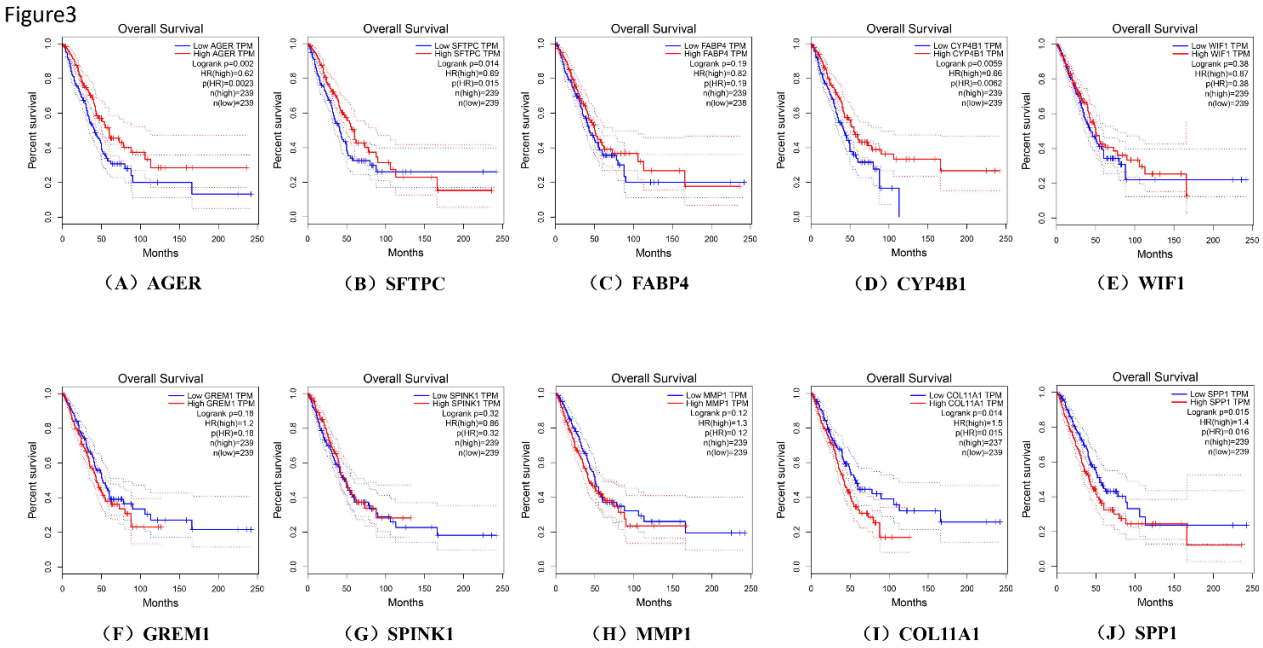

Figure 3. Prognostic value of top5 up-regulated and top5 down-regulated genes. (A–J) represents AGER, SFTPC, FABP4, CYP4B1, WIF1, GREM1, SPINK1, MMP1, COL11A1 and SPP1. P < 0 05 was regarded statistically different.

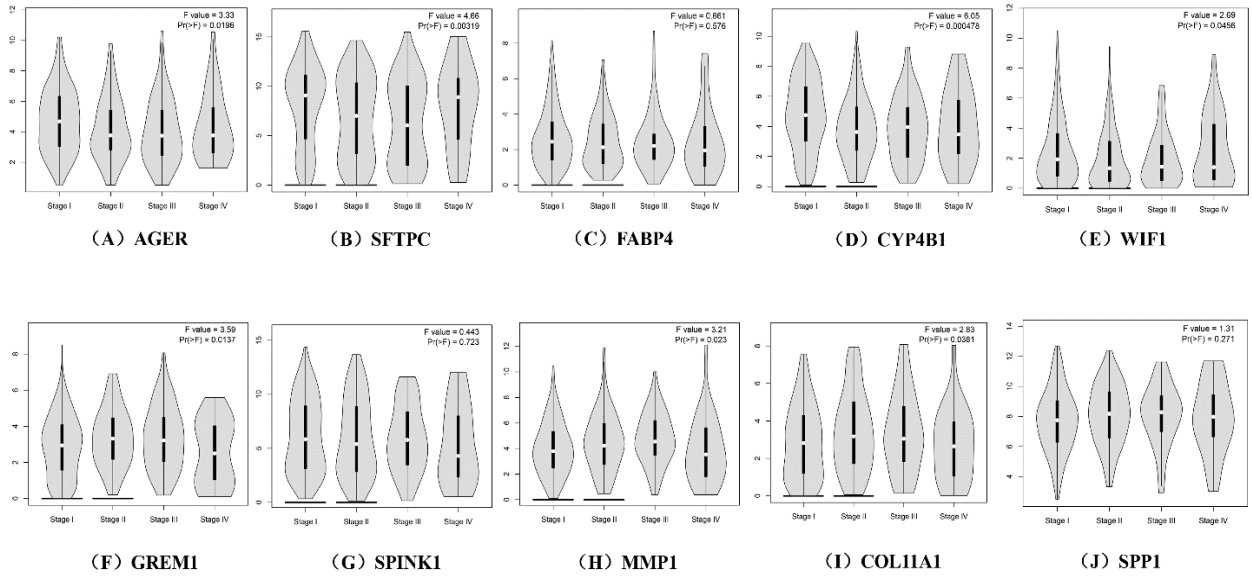

Figure 4. Correlation between DEGs expression and tumor stage in patients with LUAD (A-J) represents AGER, SFTPC, FABP4, CYP4B1, WIF1, GREM1, SPINK1, MMP1, COL11A1 and SPP1. P < 0 05 was regarded statistically different.

3.5. GO Enrichment Analysis

The results (Table 2 & Figure 5A–C) from GO term enrichment analysis varied from expression levels and GO classification of the DEGs. By analyzing GO enrichment of these up-regulated and down-regulated DEGs via DAVID, we found that the up-regulated DEGs in BP were mainly enriched in positive regulation of transcription from RNA polymerase II promoter, signal transduction, negative regulation of transcription from RNA polymerase II promoter, cell adhesion and positive regulation of GTPase activity while the up-regulated DEGs in BP mainly focused on cell division, apoptotic process, mitotic nuclear division, negative regulation of apoptotic process and cell adhesion. As for CC, the up-regulated DEGs were principally enriched in plasma membrane, integral component of membrane, extracellular exosome, extracellular region and extracellular space while the down-regulated DEGs were enriched in cytoplasm, nucleus, extracellular exosome, cytosol and extracellular space. MF analysis uncovered that the up-regulated DEGs were mainly enriched in protein binding, protein homodimerization activity, calcium ion binding, transcription factor activity, sequence-specific DNA binding and identical protein binding. By contrast, the down-regulated DEGs were enriched in extracellular matrix structural constituent, serine-type endopeptidase activity, identical protein binding, protein binding and platelet-derived growth factor binding.

3.6. KEGG pathway analysis

To acquire a more comprehensive information regarding to the critical pathways of those selected DEGs, KEGG pathways analysis were also carried out via DAVID. The results in Table 3 and Figure 5D unveiled the most vital KEGG pathways of the down-regulated and up-regulated DEGs. The up-regulated DEGs were mainly enriched in Pathways in cancer, PI3K-Akt signaling pathway, Endocytosis, MAPK signaling pathway and Complement and coagulation cascades while the down-regulated DEGs were mainly responsible for Cell cycle, PI3K-Akt signaling pathway, ECM-receptor interaction, Focal adhesion and p53 signaling pathway.

3.7. Establishing the PPI network

Applying the STRING online tool, 452 nodes with 311 PPI relationships were found, accounting for about 82.8% of these selected DEGs. According to the degree of connectivity of these DEGs, we constructed the PPI network and selected the top 15 hub genes (Table 4). The top 15 hub genes, possessing high degree of connectivity in LUAD are as follows, IL6, MMP9, EDN1, FOS, CDK1, CDH1, BIRC5, VWF, UBE2C, CDKN3, CDKN2A, CD34, AURKA, CCNB2 and EGR1. Among these 15 hub genes: IL6, EDN1, FOS, CDK1, VWF, CD34 and EGR1 were significantly up-regulated while MMP9, CDH1, BIRC5, UBE2C, CDKN3, CDKN2A, AURKA and CCNB2 significantly down-regulated ($P < 0.05$). The 15 hub genes could interact with 381 genes directly, and IL6 acted as the most intensive gene which could interact with 82 up-regulated genes and 44 down-regulated genes. Additionally, among these hub genes, there also displayed very strong interactions (Figure 6).

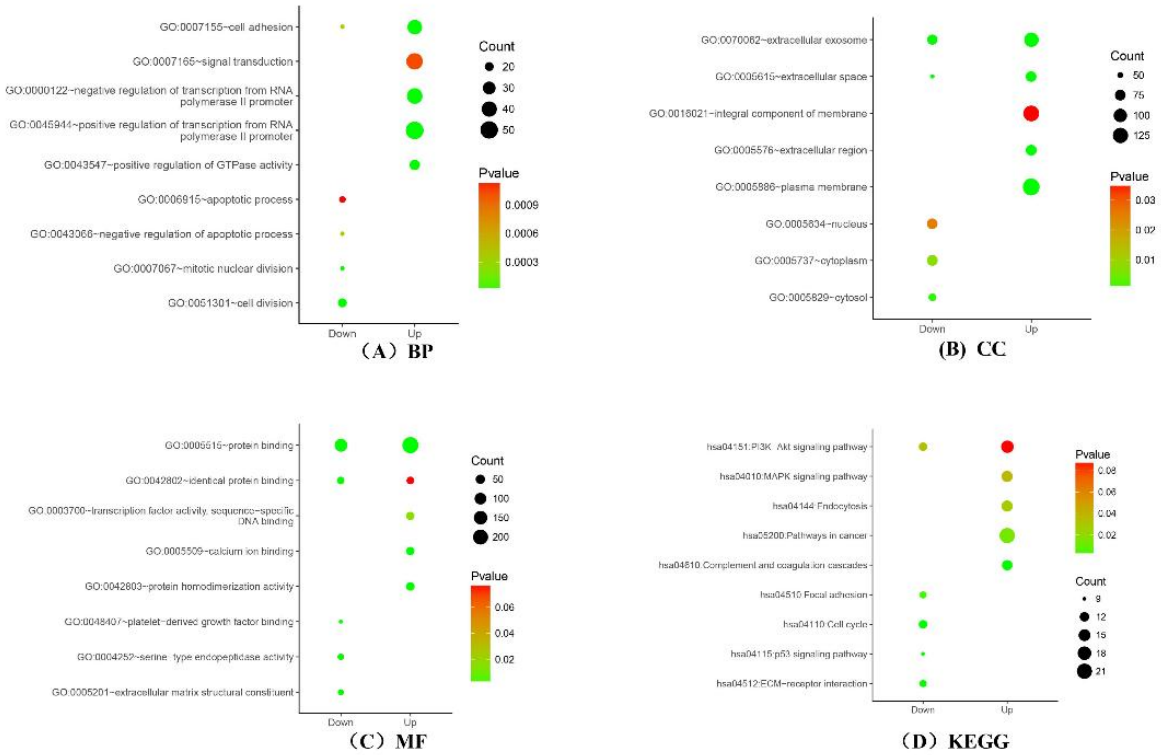

Figure 5. Gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathwayanalysis of LUAD. (A) The enriched GO terms in the biological process (BP); (B) The enriched GO terms in the cellular component (CC); (C) The enriched GO terms in the molecular function (MF); (D) The enriched KEGG pathway in LUAD.

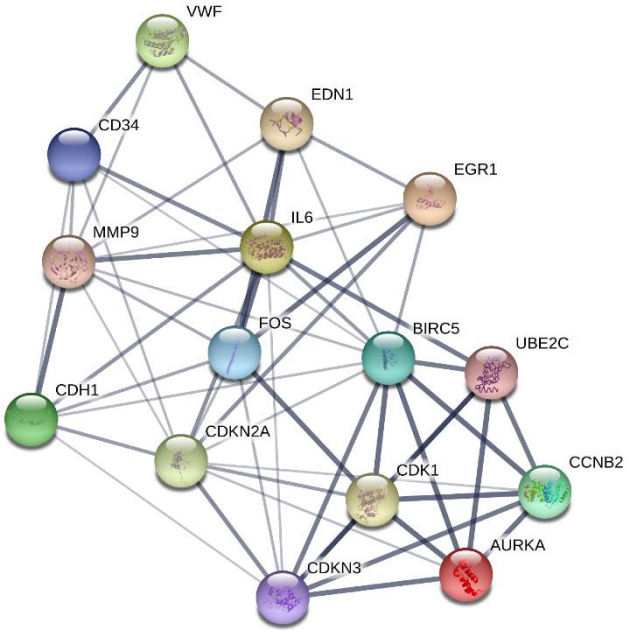

Figure 6. Protein-protein interaction (PPI) network. The PPI network of top15 hub genes with high connectivity degree.

Table 1. The top 5 up-regulated and down-regulated differentially expressed genes in patients with lung adenocarcinoma.

DEGs	Gene title	Gene symbol	LogFC	Biological function
Up-regulated	advanced glycosylation end-product specific receptor	AGER	4.4174695	A member of the immunoglobulin superfamily of cell surface receptors
	surfactant protein C	SFTPC	3.9898216	hydrophobic surfactant protein essential for lung function and homeostasis
	fatty acid binding protein 4	FABP4	3.8385413	fatty acid uptake, transport, and metabolism
	cytochrome P450 family 4 subfamily B member 1	CYP4B1	3.7097964	Metabolizing certain carcinogens
	WNT inhibitory factor 1	WIF1	3.6867095	inhibit WNT proteins
Down-regulated	gremlin 1, DAN family BMP antagonist	GREM1	−2.5483627	cell growth and differentiation factor
	serine peptidase inhibitor, Kazal type 1	SPINK1	−2.7583995	trypsin inhibitor
	matrix metallopeptidase 1	MMP1	−2.8620356	embryonic development, reproduction, and tissue remodeling
	collagen type XI alpha 1 chain	COL11A1	−3.061522	extracellular matrix
	secreted phosphoprotein 1	SPP1	−4.3644151	attachment of osteoclasts to the mineralized bone matrix

Table 2. Gene ontology analysis of differentially expressed genes associated with lung adenocarcinoma.

Expression	Category	Term	Count	%	P-Value	FDR
Up-regulated	GOTERM_BP_DIRECT	GO:0045944~positive regulation of transcription from RNA polymerase II promoter	51	0.080578904	1.60E-07	2.81E-04
	GOTERM_BP_DIRECT	GO:0007165~signal transduction	45	0.071099033	0.001027981	1.794107407
	GOTERM_BP_DIRECT	GO:0000122~negative regulation of transcription from RNA polymerase II promoter	41	0.064779119	3.51E-07	6.18E-04
	GOTERM_BP_DIRECT	GO:0007155~cell adhesion	37	0.058459205	2.09E-10	3.69E-07
	GOTERM_BP_DIRECT	GO:0043547~positive regulation of GTPase activity	23	0.036578904	1.60E-07	1.58217909
	GOTERM_CC_DIRECT	GO:0005886~plasma membrane	147	0.232256841	2.66E-10	0.006648866
	GOTERM_CC_DIRECT	GO:0016021~integral component of membrane	132	0.208557164	0.035243685	0.027678442
	GOTERM_CC_DIRECT	GO:0070062~extracellular exosome	114	0.18011755	4.84E-11	0.106773084
	GOTERM_CC_DIRECT	GO:0005576~extracellular region	75	0.118498388	9.67E-10	21.20178557
	GOTERM_CC_DIRECT	GO:0005615~extracellular space	74	0.11691841	6.08E-13	69.41801659
	GOTERM_MF_DIRECT	GO:0005515~protein binding	240	0.379194843	4.47E-06	0.006648866
	GOTERM_MF_DIRECT	GO:0042803~protein homodimerization activity	36	0.056879226	1.86E-05	0.027678442
	GOTERM_MF_DIRECT	GO:0005509~calcium ion binding	34	0.053719269	7.18E-05	0.106773084
	GOTERM_MF_DIRECT	GO:0003700~transcription factor activity, sequence-specific DNA binding	33	0.052139291	0.015885233	21.20178557
	GOTERM_MF_DIRECT	GO:0042802~identical protein binding	24	0.037919484	0.076530545	69.41801659

Continued on next page

Expression	Category	Term	Count	%	P-Value	FDR
Down-regulated	GOTERM_BP_DIRECT	GO:0051301~cell division	20	0.06355057	3.48E-08	5.71E-05
	GOTERM_BP_DIRECT	GO:0006915~apoptotic process	17	0.054017985	0.00114447	1.862881966
	GOTERM_BP_DIRECT	GO:0007067~mitotic nuclear division	16	0.050840456	2.59E-07	4.25E-04
	GOTERM_BP_DIRECT	GO:0043066~negative regulation of apoptotic process	16	0.050840456	3.35E-04	0.548772672
	GOTERM_BP_DIRECT	GO:0007155~cell adhesion	16	0.050840456	3.68E-04	0.603031926
	GOTERM_CC_DIRECT	GO:0005737~cytoplasm	75	0.238314639	0.007215425	8.951333631
	GOTERM_CC_DIRECT	GO:0005634~nucleus	74	0.23513711	0.025200114	28.14465235
	GOTERM_CC_DIRECT	GO:0070062~extracellular exosome	71	0.225604525	7.48E-12	9.69E-09
	GOTERM_CC_DIRECT	GO:0005829~cytosol	56	0.177941597	8.29E-04	1.068609771
	GOTERM_CC_DIRECT	GO:0005615~extracellular space	48	0.152521369	8.69E-13	1.13E-09
	GOTERM_MF_DIRECT	GO:0005201~extracellular matrix structural constituent	11	0.034952814	4.94E-09	6.87E-06
	GOTERM_MF_DIRECT	GO:0004252~serine-type endopeptidase activity	14	0.044485399	9.14E-06	0.012713971
	GOTERM_MF_DIRECT	GO:0042802~identical protein binding	23	0.073083156	6.07E-05	0.084464967
	GOTERM_MF_DIRECT	GO:0005515~protein binding	129	0.409901179	1.08E-04	0.150356421
	GOTERM_MF_DIRECT	GO:0048407~platelet-derived growth factor binding	4	0.012710114	2.37E-04	0.329852854

GO: Gene Ontology; FDR: False Discovery Rate.

Table 3. KEGG pathway analysis of differentially expressed genes associated with lung adenocarcinoma.

Category	Term	Count	%	P-Value	Genes	FDR
Up-regulated DEGs	hsa05200: Pathways in cancer	21	0.033179549	0.012311429	FGFR2, COL4A3, IL6, BMP2, EPAS1, PTGER4, TGFBR2, GNG11, ZBTB16, MECOM, CXCL12, COL4A5, EDNRA, AGTR1, FOS, EDNRB, LAMA4, ADCY9, PTCH1, AKT3, PIK3R1	14.56872647
	hsa04151: PI3K-Akt signaling pathway	16	0.025279656	0.087340443	FGFR2, COL4A3, FGFR4, IL6, NR4A1, GNG11, IL7R, COL4A5, VWF, LAMA4, ITGA8, TEK, ANGPT1, AKT3, PIK3R1, GHR	68.70335639
	hsa04144: Endocytosis	14	0.022119699	0.02683695	FGFR2, CAV2, FGFR4, CAV1, LDLR, TGFBR2, PIP5K1B, SNX1, HLA-E, ARRB1, FOLR1, NEDD4L, GRK5, RAB11FIP1	29.23295979
	hsa04010: MAPK signaling pathway	14	0.022119699	0.037175664	FGFR2, FGFR4, TGFBR2, NR4A1, MECOM, CACNA2D2, FOS, DUSP1, ARRB1, RPS6KA2, NTRK2, RRAS, GADD45B, AKT3	38.21659244
	hsa04610: Complement and coagulation cascades	13	0.020539721	6.68E-07	C7, A2M, C5AR1, F8, SERPING1, C4BPA, C1QB, VWF, CD55, THBD, CFD, CPB2, PROS1	8.49E-04
Down-regulated DEGs	hsa04110: Cell cycle	11	0.034952814	1.44E-05	CCNB1, CDK1, CDKN2A, MAD2L1, CCNB2, BUB1, TTK, BUB1B, CDC20, SFN, MCM4	0.017175633
	hsa04151: PI3K-Akt signaling pathway	11	0.034952814	0.033131219	COMP, TNC, COL3A1, COL1A2, EFNA4, COL1A1, COL11A1, THBS2, COL5A2, COL5A1, SPP1	33.06243008
	hsa04512: ECM-receptor interaction	10	0.031775285	5.18E-06	COMP, TNC, COL3A1, COL1A2, COL1A1, COL11A1, THBS2, COL5A2, COL5A1, SPP1	0.00617031
	hsa04510: Focal adhesion	10	0.031775285	0.003577108	COMP, TNC, COL3A1, COL1A2, COL1A1, COL11A1, THBS2, COL5A2, COL5A1, SPP1	4.179525016
	hsa04115: p53 signaling pathway	9	0.028597757	6.04E-06	CCNB1, CDK1, CDKN2A, CCNB2, RRM2, PMAIP1, SFN, PERP, IGFBP3	0.00719634

KEGG: Kyoto Encyclopedia of Genes and Genomes; FDR: False Discovery Rate.

Table 4. Top 15 hub genes with higher degree of connectivity.

Gene	Degree of connectivity	Adjusted P value
IL6	84	1.09E-05
MMP9	55	2.11E-13
EDN1	53	6.19E-12
FOS	47	9.42E-06
CDK1	44	1.99E-16
CDH1	44	4.41E-10
BIRC5	43	1.44E-05
VWF	42	7.41E-36
UBE2C	41	1.83E-14
CDKN3	36	1.10E-15
CDKN2A	36	1.03E-09
CD34	36	3.04E-27
AURKA	33	2.33E-14
CCNB2	33	3.01E-17
EGR1	33	2.58E-07

4. Discussion

Although cigarette smoking is one of dominating causes of lung cancer, surprisingly, among various major histological types of lung cancer, LUAD displayed the weakliest association with smoking, which often occurs in females and people without smoking history [18–20]. The somatic gene aberrations in LUAD have been most extensively explored. LUAD screening has been demonstrated to greatly decrease the morbidity and the mortality in a great many longstanding or newly economically developed countries [21,22]. However, at present, there is no an efficient and specific diagnostic methodology and treatment strategy for LUAD, which is mainly attributed to the intricate pathogenesis, and its symptoms that are difficult to diagnose in the first several years [23,24]. In the other one hand, the oncogenic pathway of LUAD is incompletely understood. In this study, 58 tumor tissues from patients with LUAD and 49 non-tumor tissues from normal individuals were analyzed. 856 DEGs including 559 up-regulated genes and 297 down-regulated genes were screened. To obtain a comprehensive understanding of these DEGs, we performed GO function and KEGG pathway analysis. Additionally, we analyzed the relationship between the 10 mosth significant DEGs and the overall survival as well as tumor stage.

Our analysis selected out 856 DEGs with 2-fold change between carcinoma tissues and normal tissues. According to the rank of the fold change of these DGEs, we picked up the top5 up-regulated DEGs and top5 down-regulated DEGs. From our perspective, these DEGs would be possible candidates for the diagnosis of LUAD in near future. At present, some of these DEGs, in fact, have been already disclosed to be novel indicators of LUAD. For instance, Tang Z et al. [25] found that the elevated expression of fatty acid binding proteins 4 (FABP4) in non-small cell lung cancer was not only significantly correlated with advanced tumor node metastasis (TNM) stage, but also exhibited a negative effect on the overall survival. WIF1, a vital component in the Wnt signaling pathway, was found to be down-regulated in multiple cancers, including breast, prostate, bladder, and

lung cancer [26]. WIF-1 promoter region hypermethylation contributes to aberrant activation of Wnt signaling pathway in NSCLC. Meanwhile, WIF-1 promoter region hypermethylation is also a novel diagnostic marker for LUAD-related malignant pleural effusions [27]. An interesting finding in our study was that WIF-1 is highly expressed in LUAD tissue while its level in patients with LUAD is not associated with overall survival, which hints that WIF-1 is a switch of LUAD but could not promote the progression of LUAD. However, the role of other DEGs has not been explored in LUAD. The receptor for advanced glycation end products (AGER) as an oncogenic transmembranous receptor, was up-regulated in many human cancers. Elevated AGER may promote cervical cancer cell migration, proliferation, and inhibited its apoptosis [28]. Hence, the role of AGER in LUAD needs to be further investigated. Our study screened the DEGs of LUAD from the angle of bioinformatics for the first time, which not only verified the reported genes, but also prompted new biomarkers in LUAD.

At the same time, we picked out 15 hub genes imlpicated with LUAD, all of which were located in the core nodes in PPI network, meaning that these genes could be critical therapeutic targets to protect against LUAD. It is reported that IL-6 may serve as a mediator of many reactions involving an inflammatory response in patients with lung cancer [29–31]. Autocrine IL-6-induced Stat3 activation could result in the occurrence of LUAD and the production of malignant pleural effusion [32–34]. In our study, we disclosed that IL6 is the hub genes with the highest connective degree, suggesting that IL6 plays a core role in the occurrence and development of LUAD. CDH1, encoded by the CDH1 gene in humans, was also a hub gene with high connective degree in our study. Study showed that CDH1 methylation is closely correlated with an elevated risk of lung cancer, the hypermethylation of which could inactive CDH1, thereby influencing proliferation, invasion, and metastasis of lung cancer cells [35]. Hence, these hub genes are potential therapeutic targets in LUAD.

Abnormal uncontrolled cell growth and cell cycle-mediated cell transformation are the basical biological features of LUAD. Our study unveiled that the DEGs are mainly enriched in the events and pathways associated with cell proliferation and apoptosis. Furthermore, the GO analysis in CC and KEGG pathway analysis also proved that extracellular matrix (ECM) exerts an essential role in LUAD. Indeed, ECM is the direct tumor microenvironment, consisting of proteoglycans, non-proteinaceous glycosaminoglycans, and collagens. It is recognized that ECM molecules could activate autocrine or paracrine cell signaling directly, form a biomechanical scaffold for adherent cells, eventually remodelling tissue architecture during inflammation. Laura E. Stevens et al. [36] found that up-regulation of the hyaluronan receptor HMMR in mice with LUAD was correlated with a significant inflammatory molecular signature as well as poor prognosis. Finally, our study proved the critical roles of PI3K/Akt signaling pathway and MAPK pathway from the angle of bioinformatics, which further solid the previous experimental researches.

5. Conclusion

In conclusion, we provided a comprehensive and novel analysis of gene expression profiles patients with LUAD. Particularly, the top5 up-regulated genes including AGER, SFTPC, FABP4, CYP4B1 and WIF1 and the top5 down-regulated genes including GREM1, SPINK1, MMP1, COL11A1 and SPP1, which are expected to sensitive biomarkers in diagnosis of LUAD. Meanwhile, we also screened the top 15 hub genes involving IL6, MMP9, EDN1, FOS, CDK1, CDH1, BIRC5,

VWF, UBE2C, CDKN3, CDKN2A, CD34, AURKA, CCNB2 and EGR1, which could be promising therapeutic targets of LUAD. Additionally, genes and pathways involved in ECM were also significantly altered in patients with LUAD. Anyway, this analysis may offer the powerful evidence and clues for the future genomic individualized treatment of LUAD.

Conflict of interest

The authors declare that they have no conflict of interest.

References

1. T. Akhurst, Staging of non-small-cell lung cancer, *PET. Clin.*, **13** (2018), 1–10.
2. Y. Lavin, S. Kobayashi, A. Leader, et al., Innate immune landscape in early lung adenocarcinoma by paired single-cell analyses, *Cell*, **169** (2017), 750–765.e717.
3. L. Han, C. K. Lee, H. Pang, et al., Genetic predisposition to lung adenocarcinoma among never-smoking Chinese with different epidermal growth factor receptor mutation status, *Lung Cancer*, **114** (2017), 79–89.
4. E. Tsiambas, A. Y. Lefas, S. N. Georgiannos, et al., EGFR gene deregulation mechanisms in lung adenocarcinoma: A molecular review, *Pathol. Res. Pract.*, **212** (2016), 672–677.
5. J. F. Fahrmann, D. Grapov, B. S. Phinney, et al., Proteomic profiling of lung adenocarcinoma indicates heightened DNA repair, antioxidant mechanisms and identifies LASP1 as a potential negative predictor of survival, *Clin. Proteom.*, **13** (2016), 31.
6. T. Yamaoka, T. Ohmori, M. Ohba, et al., Acquired resistance mechanisms to combination Met-TKI/EGFR-TKI exposure in met-amplified EGFR-TKI-resistant lung adenocarcinoma harboring an activating EGFR mutation, *Mol. Cancer. Ther.*, **15** (2016), 3040–3054.
7. E. B. Choi, A. Y. Yang, S. C. Kim, et al., PARP1 enhances lung adenocarcinoma metastasis by novel mechanisms independent of DNA repair, *Oncogene*, **35** (2016), 4569–4579.
8. D. Yang, D. Cheng, Q. Tu, et al., HUWE1 controls the development of non-small cell lung cancer through down-regulation of p53, *Theranostics*, **8** (2018), 3517–3529.
9. I. Mender, R. LaRanger, K. Luitel, et al., Telomerase-Mediated Strategy for Overcoming Non-Small Cell Lung Cancer Targeted Therapy and Chemotherapy Resistance, *Neoplasia*, **20** (2018), 826–837.
10. Comprehensive molecular profiling of lung adenocarcinoma, *Nature*, **511** (2014), 543–550.
11. N. Li, L. Li and Y. Chen, The identification of core gene expression signature in hepatocellular carcinoma, *Oxid. Med. Cell. Longev.*, **2018** (2018), 3478305.
12. T. Braunschweig, J. Y. Chung and S. M. Hewitt, Tissue microarrays: bridging the gap between research and the clinic, *Expert Rev. Proteom.*, **2** (2005), 325–336.
13. L. Li, N. Li, C. He, et al., Proteomic analysis of differentially expressed proteins in kidneys of brain dead rabbits, *Mol. Med. Rep.*, **16** (2017), 215–223.
14. R. P. Huntley, T. Sawford, P. Mutowo-Meullenet, et al., The GOA database: Gene ontology annotation updates for 2015, *Nucleic Acids Res.*, **43** (2015), D1057–1063.
15. F. Fabris and A. A. Freitas, New KEGG pathway-based interpretable features for classifying ageing-related mouse proteins, *Bioinformatics*, **32** (2016), 2988–2995.
16. G. Dennis, Jr., B. T. Sherman, D. A. Hosack, et al., DAVID: Database for annotation,

visualization, and integrated discovery, *Genome Biol.*, **4** (2003), P3.
17. Z. Tang, C. Li, B. Kang, et al., GEPIA: A web server for cancer and normal gene expression profiling and interactive analyses, *Nucleic Acids Res.*, **45** (2017), W98–W102.
18. Z. Yin, Z. Cui, Y. Ren, et al., Association between polymorphisms in pre-miRNA genes and risk of lung cancer in a Chinese non-smoking female population, *Lung Cancer*, **94** (2016), 15–21.
19. T. Vu, L. Jin and P. K. Datta, Effect of cigarette smoking on epithelial to mesenchymal transition (EMT) in lung cancer, *J. Clin. Med.*, **5** (2016).
20. Y. Zhang, M. Elgizouli, B. Schottker, et al., Smoking-associated DNA methylation markers predict lung cancer incidence, *Clin. Epigenet.*, **8** (2016), 127.
21. F. Ciompi, K. Chung, S. J. van Riel, et al., Towards automatic pulmonary nodule management in lung cancer screening with deep learning, *Sci. Rep.*, **7** (2017), 46479.
22. U. Yousaf-Khan, C. van der Aalst, P. A. de Jong, et al., Risk stratification based on screening history: the NELSON lung cancer screening study, *Thorax*, **72** (2017), 819–824.
23. Y. Zhang, J. Sui, X. Shen, et al., Differential expression profiles of microRNAs as potential biomarkers for the early diagnosis of lung cancer, *Oncol. Rep.*, **37** (2017), 3543–3553.
24. M. Reck and K. F. Rabe, Precision diagnosis and treatment for advanced non-small-cell lung cancer, *N. Engl. J. Med.*, **377** (2017), 849–861.
25. Z. Tang, Q. Shen, H. Xie, et al., Elevated expression of FABP3 and FABP4 cooperatively correlates with poor prognosis in non-small cell lung cancer (NSCLC), *Oncotarget*, **7** (2016), 46253–46262.
26. T. M. Yang, S. W. Leu, J. M. Li, et al., WIF-1 promoter region hypermethylation as an adjuvant diagnostic marker for non-small cell lung cancer-related malignant pleural effusions, *J. Cancer Res. Clin. Oncol.*, **135** (2009), 919–924.
27. C. Wissmann, P. J. Wild, S. Kaiser, et al., WIF1, a component of the Wnt pathway, is down-regulated in prostate, breast, lung, and bladder cancer, *J. Pathol.*, **201** (2003), 204–212.
28. X. Zhu, L. Zhou, R. Li, et al., AGER promotes proliferation and migration in cervical cancer, *Biosci. Rep.*, **38** (2018).
29. H. Yanagawa, S. Sone, Y. Takahashi, et al., Serum levels of interleukin 6 in patients with lung cancer, *Br. J. Cancer*, **71** (1995), 1095–1098.
30. Y. P. Chen, Y. Zhang, J. W. Lv, et al., Genomic analysis of tumor microenvironment immune types across 14 solid cancer types: Immunotherapeutic implications, *Theranostics*, **7** (2017), 3585–3594.
31. C. Xie, J. Zhu, Y. Jiang, et al., Sulforaphane inhibits the acquisition of tobacco smoke-induced lung cancer stem cell-like properties via the IL-6/DeltaNp63alpha/Notch axis, *Theranostics*, **9** (2019), 4827–4840.
32. H. H. Yeh, W. T. Chang, K. C. Lu, et al., Upregulation of tissue factor by activated Stat3 contributes to malignant pleural effusion generation via enhancing tumor metastasis and vascular permeability in lung adenocarcinoma, *PLoS One*, **8** (2013), e75287.
33. L. Li, J. Xu, G. Qiu, et al., Epigenomic characterization of a p53-regulated 3p22.2 tumor suppressor that inhibits STAT3 phosphorylation via protein docking and is frequently methylated in esophageal and other carcinomas, *Theranostics*, **8** (2018), 61–77.
34. L. Jiang, R. Wang, L. Fang, et al., HCP5 is a SMAD3-responsive long non-coding RNA that promotes lung adenocarcinoma metastasis via miR-203/SNAI axis, *Theranostics*, **9** (2019), 2460–2474.

35. Q. Yu, Q. Guo, L. Chen, et al., Clinicopathological significance and potential drug targeting of CDH1 in lung cancer: A meta-analysis and literature review, *Drug Des. Devel. Ther.*, **9** (2015), 2171–2178.
36. L. E. Stevens, W. K. C. Cheung, S. J. Adua, et al., Extracellular matrix receptor expression in subtypes of lung adenocarcinoma potentiates outgrowth of micrometastases, *Cancer Res.*, **77** (2017), 1905–1917.

AIMS Press