MEDICAL SCIENCE MONITOR

DATABASE ANALYSIS

e-ISSN 1643-3750
© Med Sci Monit, 2020; 26: e922680
DOI: 10.12659/MSM.922680

Received: 2020.01.07
Accepted: 2020.04.27
Available online: 2020.05.20
Published: 2020.07.16

Estimation of Hub Genes and Infiltrating Immune Cells in Non-Smoking Females with Lung Adenocarcinoma by Integrated Bioinformatic Analysis

Authors' Contribution:
Study Design A
Data Collection B
Statistical Analysis C
Data Interpretation D
Manuscript Preparation E
Literature Search F
Funds Collection G

ABCDEF 1 Jie Li
BCDF 2 Ben Wang
BF 3 Xin Li
AE 1 Yuxi Zhu

1 Department of Oncology, The First Affiliated Hospital of Chongqing Medical University, Chongqing, P.R. China
2 Department of Orthopedics, The First Affiliated Hospital of Chongqing Medical University, Chongqing, P.R. China
3 Department of Respiratory and Critical Care Medicine, The First Affiliated Hospital of Chongqing Medical University, Chongqing, P.R. China

Corresponding Author: Yuxi Zhu, e-mail: zhuyuxi@hospital.cqmu.edu.cn
Source of support: The Natural Science Foundation Project of Chongqing Science and Technology Commission (CSTC), China (grant no. cstc2018jcyjAX0012) supported this study

Background: In recent years, the morbidity and mortality rates of lung adenocarcinoma in non-smoking females have been increasing dramatically. Although much research has been done with some progress, the molecular mechanism remains unclear. In this study we aimed to estimate hub genes and infiltrating immune cells in non-smoking females with lung adenocarcinoma.

Material/Methods: Firstly, we obtained differentially expressed genes (DEGs) by GEO2R analysis based on 3 independent mRNA microarray datasets of GSE10072, GSE31547, and GSE32863. The DAVID database was utilized for functional enrichment analysis of DEGs. Moreover, we identified hub genes with prognostic value by STRING, Cytoscape, and Kaplan Meier plotter. Subsequently, these genes were further analyzed by Gene Expression Profiling Interactive Analysis, Oncomine, Tumor Immune Estimation Resource, and Human Protein Atlas. Finally, the immune infiltration analysis was performed by CIBERSORT and The Cancer Genome Atlas with R packages.

Results: We found 315 DEGs enriching in the extracellular matrix organization, cell adhesion, integrin binding, angiogenesis, and hypoxic response. And among these DEGs, we identified 10 hub genes (*SPP1, ENG, ATF3, TOP2A, COL1A1, PAICS, CAV1, CAT, TGFBR2*, and *ANGPT1*) of significant prognostic value. Simultaneously, we illustrated the distribution and differential expressions of 22 immune cell subtypes. and dendritic cells resting and macrophages M1 were identified with prognostic significance.

Conclusions: The results indicated that 10 hub genes and 2 immune cell subtypes might be promising biomarkers for lung adenocarcinoma in non-smoking females. This finding needs to be further evaluated.

MeSH Keywords: Female • Lung Neoplasms • Tobacco, Smokeless • Tumor Markers, Biological

Full-text PDF: https://www.medscimonit.com/abstract/index/idArt/922680

2929 2 7 65

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

Background

Lung cancer has become the chief cause of malignancy deaths worldwide, and adenocarcinoma is the most common histologic type of lung cancer [1]. Previously, smoking was thought to be the major cause of lung adenocarcinoma (LUAD). However, the morbidity of LUAD has increased in never-smokers, especially in females [2]. Studies have shown that non-smoking lung cancer should be considered as a separate subtype [3]. Epidemiological, pathological, and molecular evidence suggested that estrogen appears to participate in the carcinogenic effect of lung cancer besides smoking [4,5]. A study in South Korea found morbidity differences in gender and histological subtypes in smoking-related lung cancer. Compared with males, females were more likely to develop non-smoking related LUAD, thus, gender was also an independent prognostic factor [6,7]. One study reported that females benefited significantly more from immunotherapy for lung cancer than males [8]. Additionally, some studies have indicated that anti-estrogen could reduce non-small cell lung cancer (NSCLC) cell proliferation [9]. Therefore, more attention should be paid to the treatment and prognostic evaluation of LUAD in non-smoking females [10].

Although its pathogenesis remains unclear, the application of bioinformatics analysis in precision medicine might contribute to finding the key biomarkers in the big data era [11]. Data mining in cancers has played a vital part in cancer diagnosis and management [12]. Consequently, we explored the promising molecular mechanism of LUAD in non-smoking females by bioinformatics. We identified the differentially expressed genes (DEGs) between LUAD and normal samples of non-smoking females by data mining. Simultaneously, CIBERSORT and The Cancer Genome Atlas (TCGA) were utilized for immune infiltration analysis. Finally, we found 10 hub genes and 2 immune cell subtypes as promising biomarkers for LUAD in non-smoking females, which provided useful information for further exploration.

Material and Methods

Microarray data

We downloaded qualified datasets from the Gene Expression Omnibus (GEO) database (*http://www.ncbi.nlm.nih.gov/geo/*). In this study, datasets that meet the following criteria were included: a) it contained LUAD tissue samples and normal lung tissue samples of non-smoking females; b) at least 10 samples were included. Finally, GSE10072, GSE31547, and GSE32863 were qualified for further analysis. GSE10072 contained 13 LUAD tissue samples and 11 normal lung tissue samples of non-smoking females. GSE31547 contained 6 LUAD tissue samples and 5 normal lung tissue samples of non-smoking females. GSE32863 contained 23 LUAD tissue samples and 23 normal lung tissue samples of non-smoking females.

Identification of DEGs

GEO2R (*http://www.ncbi.nlm.nih.gov/geo/geo2r*), a web application using BioConductor R packages [13], could compare DEGs from 2 or more datasets in the GEO series. It was universally applied in various bioinformatics analyses [14–16], and it provided the native R script for researchers to replicate their analyses. We utilized GEO2R to screen DEGs between LUAD tissue samples and normal tissue samples of non-smoking females. $|\log FC| > 1$ and $P<0.01$ was set as the cutoff criterion. Moreover, we replicated this analysis by the native R script to ensure the reliability of the present study.

Functional enrichment analyses of DEGs

Gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis of DEGs were performed by the DAVID database (*http://david.ncifcrf.gov*) (version 6.8) [17]. $P<0.05$ was set as statistically significant.

PPI network and module analysis

STRING (*http://string-db.org*) (version 10.0) provides the prediction of quality-controlled protein-protein association networks. We performed the STRING database to construct a protein-protein interaction (PPI) network for DEGs, and combined score >0.4 was set as statistically significant. Cytoscape (version 3.6.1) [18] was performed to visualize molecular interaction networks. The plugin Molecular Complexity Detection (MCODE) (version 1.5.1) in Cytoscape was used to identify the most important module from the PPI network. And the condition was set as follows: Degree cutoff=2, k-core=5, max. Depth=100, and node score cutoff=0.2. Subsequently, functional enrichment analysis was performed for genes in this module by the online bioinformatics database Metascape (*http://metascape.org/*) [19].

Hub genes selection and analysis

The plugin cytoHubba of Cytoscape was performed to calculate the degree of genes in the PPI network. DEGs with degrees >10 were selected as hub genes. The Kaplan Meier plotter (*http://kmplot.com/analysis/*) [20] is an online platform to estimate the prognostic value of thousands of genes in several cancer types based on the data from GEO, Genomic Expression Archive, and TCGA database. And we performed overall survival (OS) analysis of hub genes in LUAD with non-smoking females by Kaplan Meier plotter. Gene Expression Profiling Interactive Analysis (GEPIA; *http://gepia.cancer-pku.cn*) [21] provides the

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

differential analysis based on the Genotype-Tissue Expression and the TCGA database. Moreover, we visualized the differential expression of the most significant hub genes in LUAD by GEPIA. Finally, further analyses were performed on *SPP1*, the hub gene with the highest degree found by cytoHubba. Tumor Immune Estimation Resource (TIMER; *https://cistrome.shinyapps.io/timer/*) [22] was used to assess the expression profile of *SPP1* in various human tumors based on TCGA database. And a meta-analysis of expression of *SPP1* in LUAD compared with normal tissues in different datasets was estimated based on the Oncomine database (*http://www.oncomine.com*) [23]. SPP1 protein expression analysis in LUAD tissues and normal tissues was performed by the Human Protein Atlas (*https://www.proteinatlas.org*) [24].

Distribution and prognostic analysis of infiltrating immune cells in non-smoking female LUAD

Firstly, we downloaded the Transcriptome Profiling data and Clinical data of female LUAD from TCGA database. Among them, 34 normal female lung tissue samples and 47 non-smoking female LUAD tissue samples were included in this study (as for only 5 normal non-smoking female lung tissue samples were available in TCGA, we included all of 34 normal female lung tissue samples as the control group). And the raw data was converted to which could be matched with CIBERSORT [25] by Practical Extraction and Report Language (Perl). Moreover, we randomized the converted data by limma packages (version 3.8). After deleting samples with $P>0.05$, 32 normal samples and 42 tumor samples were left. Then we predicted the distribution of 22 infiltrating immune cells in these samples by CIBERSORT. Finally, vioplot packages and survival packages were performed to illustrate the distribution and prognostic analysis of 22 infiltrating immune cells of non-smoking female LUAD.

Results

Identification of DEGs

The detailed sample information of the included datasets was presented in Supplementary Table 1. We identified 315 overlapped DEGs among 3 datasets (Figure 1), consisting of 254 downregulated DEGs and 61 upregulated DEGs (Supplementary Table 2). Notably, the regulation of these 315 DEGs was consistent in all these 3 datasets.

Functional enrichment analyses of DEGs

The whole results of GO and KEGG enrichment analyses for 315 DEGs were presented in Supplementary Table 3, and the top 5 GO and KEGG terms were visualized in Table 1.

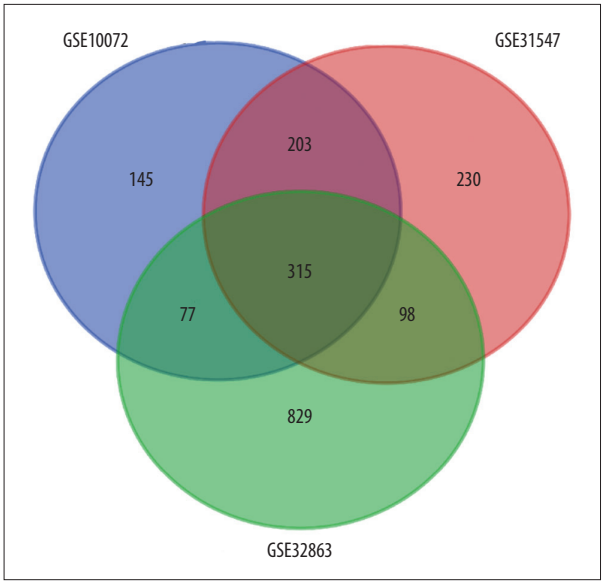


Figure 1. Venn diagram for overlapping DEGs in 3 microarray datasets. |log FC| >1 and $P<0.01$ was set as the cutoff criterion. There were 315 overlapped DEGs among 3 datasets (GSE10072, GSE31547, GSE32863) identified. DEGs – differentially expressed genes.

PPI network and module analysis

The PPI network of 315 DEGs was constructed (Figure 2), including 315 nodes and 708 edges. And the most significant module was illustrated in Figure 3A. Subsequently, our results suggested that DEGs in this module were mostly enriched in ERK1 and ERK2 cascade, vasculature development, cellular response to hormone stimulus and adrenomedullin receptor signaling pathway (Figure 3B).

Hub gene screening and analysis

In total, we identified 36 DEGs as hub genes with degrees >10 (Supplementary Table 4). Subsequently, 10 hub genes (Table 2) were screened out with prognostic value (Figure 4). Our results indicated that over-expression of *SPP1, ENG, ATF3, TOP2A, COL1A1,* and *PAICS* was related to worse OS for non-smoking females with LUAD ($P<0.05$). On the other hand, under-expression of *CAV1, CAT, TGFBR2,* and *ANGPT1* was associated with a poorer OS for non-smoking females with LUAD ($P<0.05$). As illustrated in Figure 5, compared with normal tissues, the expressing of *SPP1, TOP2A, COL1A1,* and *PAICS* increased in LUAD tissues, while *ENG, ATF3, CAV1, CAT, TGFBR2,* and *ANGPT1* decreased based on GEPIA. These results were coordinated with the results of differential expression analysis based on the GEO database, which validated the reliability of GEO analysis indirectly. Among these 10 most significant hub genes, *SPP1* accounts for the highest degree of 21, suggesting the potential significance. The result of TIMER indicated that *SPP1* was

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

Table 1. GO and KEGG pathway enrichment analysis of 315 DEGs.

Term	Description	Gene count	P-value
GO-CC: 0005615	Extracellular space	64	1.21E-13
GO-CC: 0070062	Extracellular exosome	96	8.15E-12
GO-CC: 0005578	Proteinaceous extracellular matrix	24	2.17E-10
GO-CC: 0005576	Extracellular region	60	1.26E-08
GO-CC: 0005887	Integral component of plasma membrane	55	1.52E-08
GO-BP: 0030198	Extracellular matrix organization	24	1.10E-12
GO-BP: 0007155	Cell adhesion	34	1.36E-11
GO-BP: 0016337	Single organismal cell-cell adhesion	13	2.73E-07
GO-BP: 0050900	Leukocyte migration	14	3.14E-07
GO-BP: 0001666	Response to hypoxia	16	5.41E-07
GO-MF: 0008201	Heparin binding	17	1.31E-08
GO-MF: 0005539	Glycosaminoglycan binding	7	3.44E-07
GO-MF: 0005515	Protein binding	189	1.11E-06
GO-MF: 0005178	Integrin binding	12	1.60E-06
GO-MF: 0050431	Transforming growth factor beta binding	6	4.97E-06
KEGG_pathway			
hsa05144	Malaria	8	1.37E-04
hsa04610	Complement and coagulation cascades	8	1.16E-03
hsa04514	Cell adhesion molecules (CAMs)	11	1.91E-03
hsa04530	Tight junction	8	4.40E-03
hsa04512	ECM-receptor interaction	8	4.40E-03

GO – Gene Ontology; KEGG – Kyoto Encyclopedia of Genes and Genomes; DEGs – differentially expressed genes; BP – biological processes; CC – cell component; MF – molecular function; ECM – extracellular matrix.

overexpressed in some cancers compared with normal tissues, including LUAD, breast invasive carcinoma (BRCA), colon adenocarcinoma (COAD), liver hepatocellular carcinoma (LIHC), stomach adenocarcinoma (STAD), uterine corpus endometrial carcinoma (UCEC), etc., (Figure 6A). A meta-analysis based on Oncomine datasets revealed that *SPP1* was over-expressed in LUAD compared with normal tissues (Figure 6B). As shown in Figure 6C, SPP1 protein was higher expressed in patients with LUAD compared with normal tissue.

Distribution and prognostic analysis of infiltrating immune cells in non-smoking female LUAD

The detailed clinical information of included samples was shown in Supplementary Table 5. The distribution of 22 kinds of infiltrating immune cells of non-smoking female LUAD indicated that T cells CD4 memory resting, macrophages M2 and macrophages M0 accounted for the largest proportion (Figure 7A). The results revealed that a series of cells are differentially expressed between tumor tissues and normal tissues with *P*-value <0.05. Some cells have higher expression in tumor tissue than that in normal tissue, including plasma cells, T cells regulatory, macrophages M1, and dendritic cells resting. In contrast, some cells have lower expression in tumor tissue than that in normal tissue, consisting of T cells CD4 memory resting, natural killer (NK) cells resting, monocytes, macrophages M0, mast cells resting, and neutrophils. In addition, among these 22 infiltrating immune cells, only dendritic cells resting and macrophages M1 were found to be statistically significant ($P<0.05$) for prognostic value. As shown in Figure 7B and 7C, lower expression of dendritic cells resting indicated a poor prognosis, while lower expression of macrophages M1 suggested a better prognosis.

Discussion

In the present study, we found 10 prognostic hub genes and 2 kinds of significant infiltrating immune cells of LUAD in non-smoking females, which were verified with multiple databases. The biological functions and signaling pathways enriched in DEGs might participate in the tumorigenesis and development

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

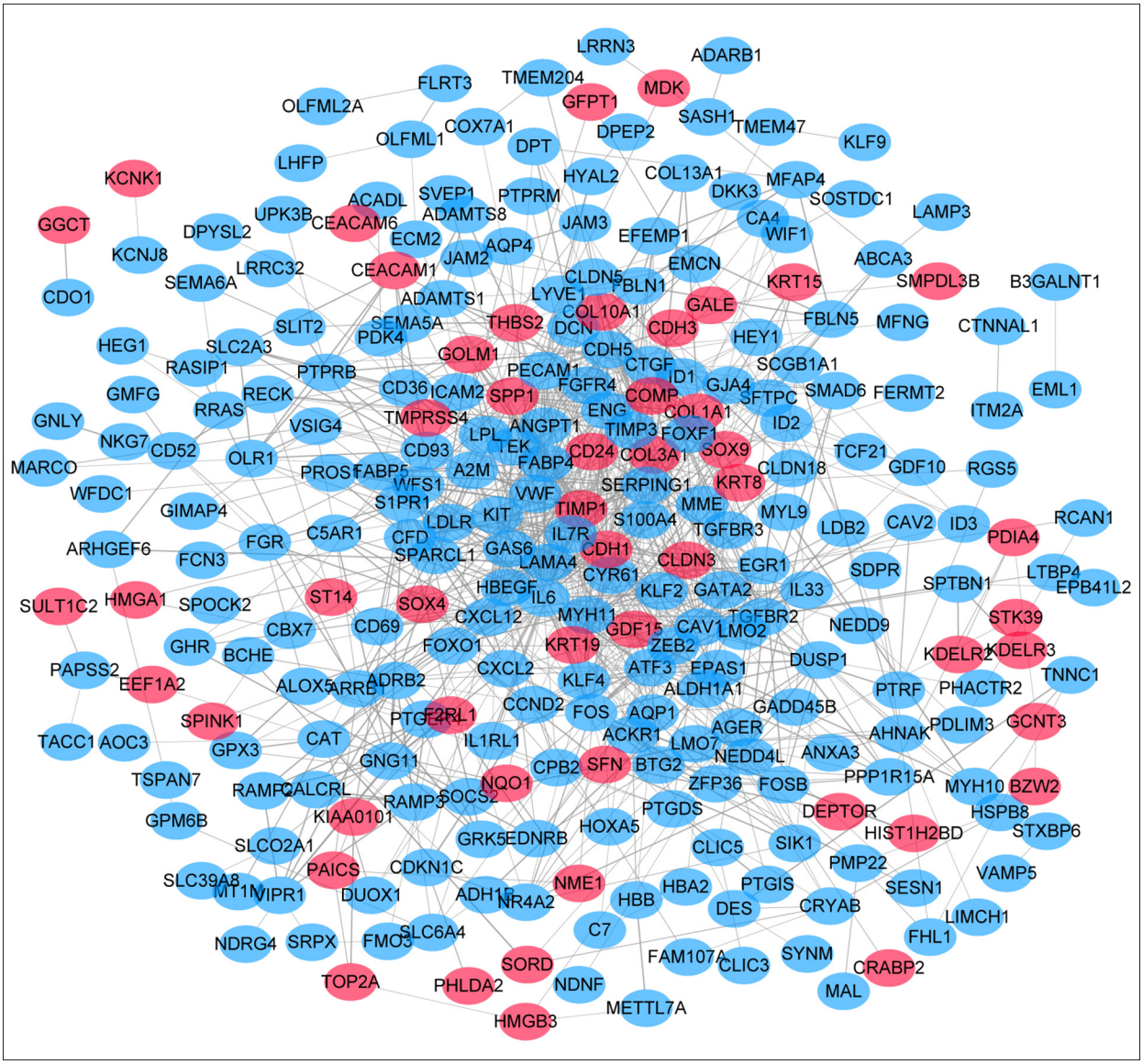


Figure 2. PPI network of 315 DEGs construction using STRING and Cytoscape. This network includes 315 nodes and 708 edges. Nodes stand for the DEGs and edges stand for the association of DEGs. Red nodes represent upregulated DEGs, while blue nodes represent downregulated DEGs. PPI – protein–protein interactions; DEGs – differentially expressed genes.

of LUAD in non-smoking females. Notably, this work was repeated 3 times by 3 individual researchers to ensure the reliability of the results.

Among these 10 most significant hub genes with prognostic value, *SPP1* accounted for the highest degree, suggesting its potential significance in non-smoking females with LUAD. *SPP1*, also known as osteopontin, has been reported to be upregulated in some tumors, such as colorectal cancer [26], cervical cancer [27], and breast cancer [28], which was consistent with our results (Figure 6). Both experimental and clinical analyses revealed that high expression of *SPP1* predicted a poor prognosis. Immunohistochemical analysis of 318 NSCLC tumor samples indicated that *SPP1* was significantly over-expressed in NSCLC tissues compared with normal tissues [29]. In clinical investigations, elevated plasma SPP1 level was found in early-stage and relapsed NSCLC patients, suggesting the potential diagnostic and prognostic value of *SPP1* [30]. However, the mechanism of *SPP1* in non-smoking female LUAD is poorly understood. As illustrated in Table 1 and Supplementary Table 3, DEGs in our study were enriched in integrin binding, extracellular matrix (ECM) organization, angiogenesis, phosphoinositide 3-kinase (PI3K)-Akt signaling pathway, and ECM-receptor interaction; this finding might help us understand the mechanism of LUAD in non-smoking females. It has been reported that *SPP1* interacts with various integrins and CD44 to

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

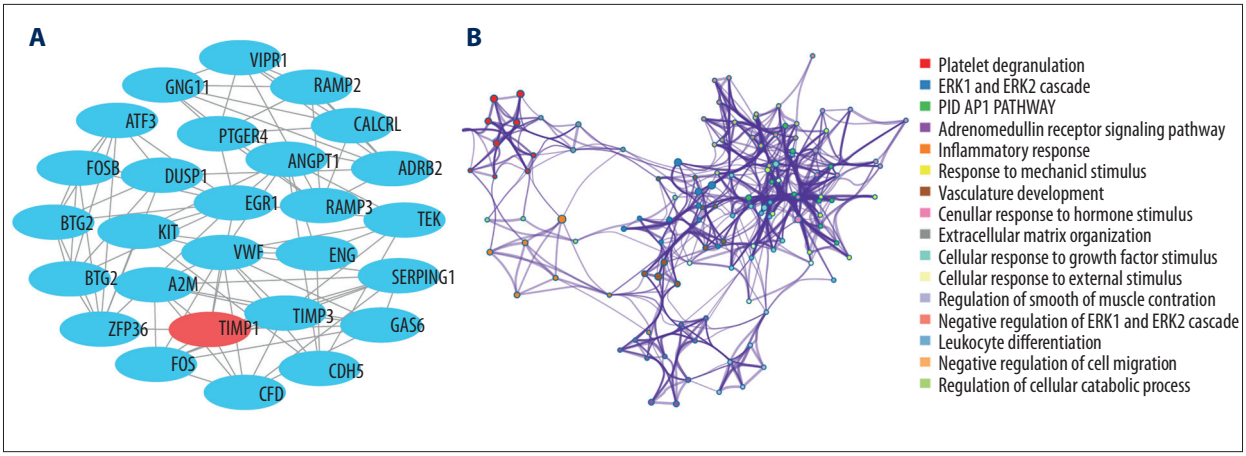
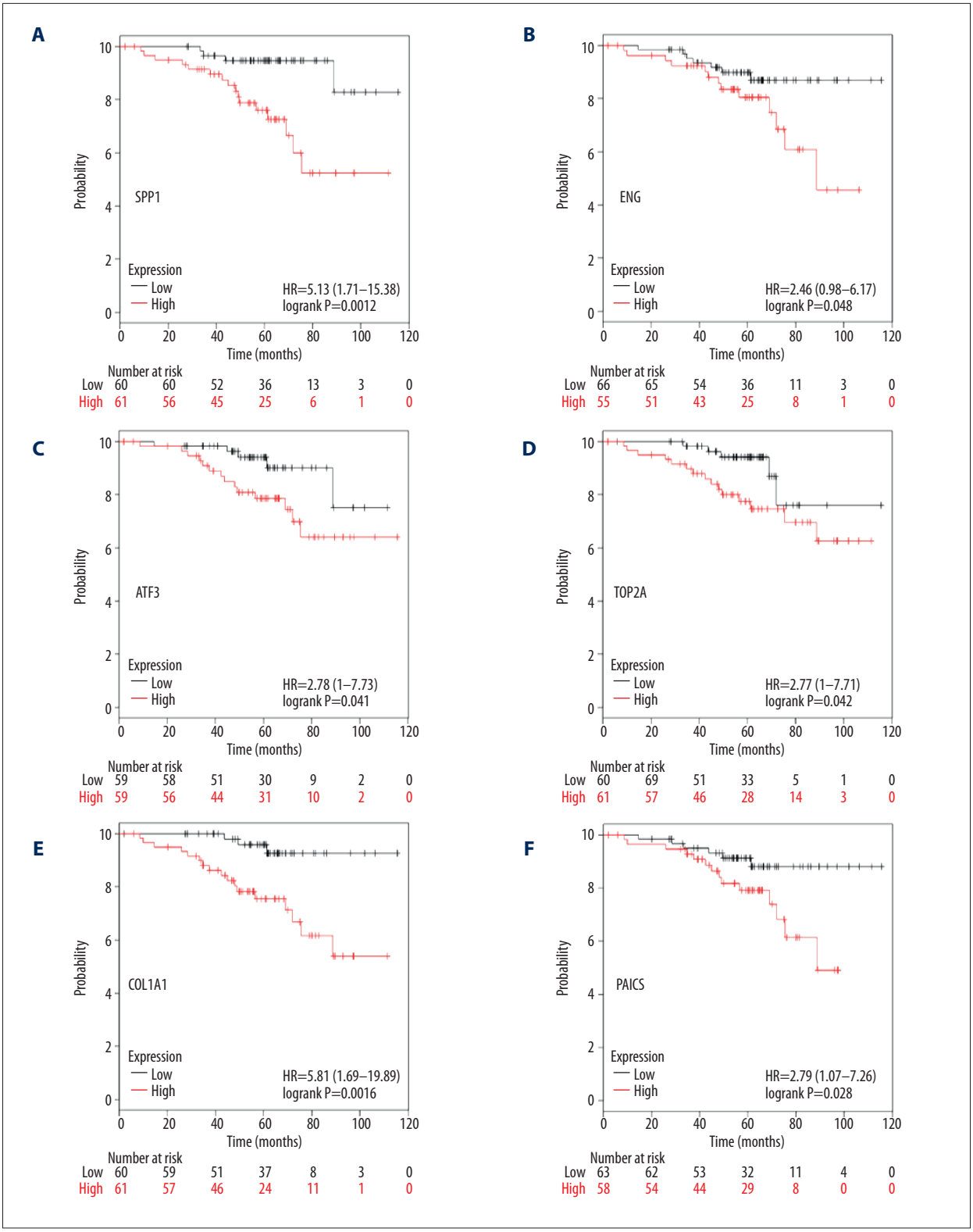


Figure 3. Analysis of the most significant module. (**A**) Identification of the most significant module from PPI network using MCODE plugin of Cytoscape. Red nodes stand for upregulated DEGs, while blue nodes represent downregulated DEGs. (**B**) Functional enrichment analysis of the most significant module performed by Metascape. PPI – protein–protein interactions; MCODE – Molecular Complexity Detection; DEGs – differentially expressed genes.

Table 2. Functional roles of 10 hub genes with prognostic significance.

No.	Gene symbol	Regulation	Degree	Full name	Function
1	SPP1	Up	21	Secreted Phosphoprotein 1	SPP1 participates in a range of biological functions, including cell proliferation, adhesion, invasion, migration, and tumor angiogenesis
2	ENG	Down	18	Endoglin	ENG activated the TGF-β/ALK1 signaling pathway and promoted endothelial cell proliferation and migration in cancers
3	CAV1	Down	18	Caveolin 1	CAV1 transforms suppressor activity abnormal expressed in T cell leukemia in lung carcinoma and in breast carcinoma
4	ATF3	Down	17	Activating Transcription Factor 3	Over-expression of ATF3 upregulated p53 and inhibited the tumorigenesis of lung cancer
5	TOP2A	Up	17	DNA Topoisomerase II Alpha	TOP2A is involved in the process of chromosome condensation, chromatid separation, DNA transcription and replication. It is the target of several anti-cancer drugs
6	COL1A1	Up	16	Collagen Type I Alpha 1	COL1A1 encodes the pro-alpha1 chains of type I collagen. It is related to hypoxia and is significantly overexpressed in non-small cell lung cancer
7	CAT	Down	15	Catalase	CAT serves to protect cells from the toxic effects of hydrogen peroxide, and it changes the migration and invasion ability of lung cancer cell
8	TGFBR2	Down	15	Transforming Growth Factor Beta Receptor 2	TGFBR2 often alters during adenoma-carcinoma progression of some cancers
9	ANGPT1	Down	14	Angiopoietin 1	ANGPT1, a member of the angiopoietin family, plays an important role in vascular development and angiogenesis
10	PAICS	Up	11	Phosphoribosyl Aminoimidazole Carboxylase	PAICS is identified as an oncogene of various tumor types

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]



Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

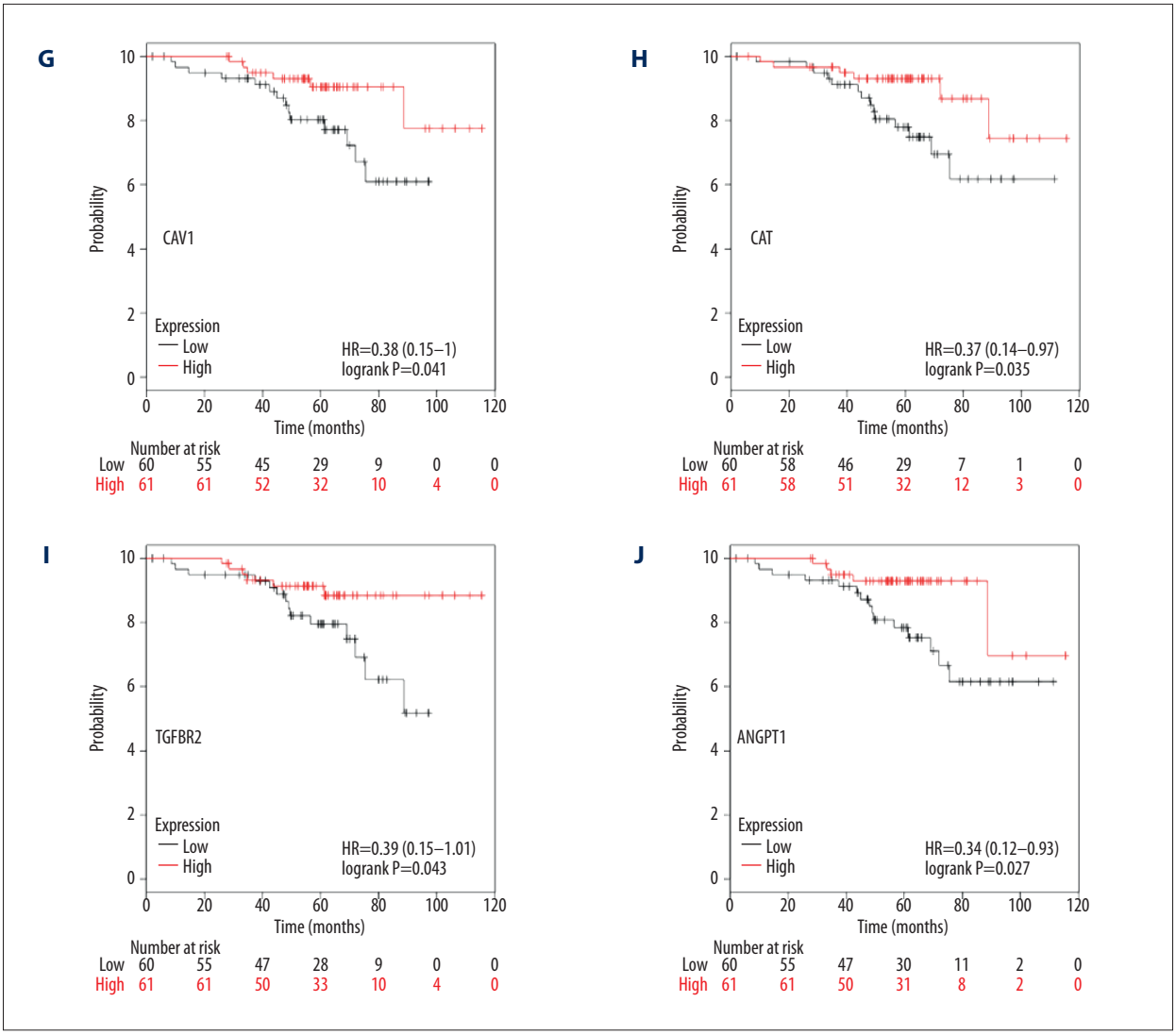


Figure 4. Overall survival analysis of 10 most significant hub genes in non-smoking females with LUAD based on Kaplan Meier plotter platform using the data of GEO, GEA, and TCGA databases, including *SPP1* (**A**), *ENG* (**B**), *ATF3* (**C**), *TOP2A* (**D**), *COL1A1* (**E**), *PAICS* (**F**), *CAV1* (**G**), *CAT* (**H**), *TGFBR2* (**I**), *ANGPT1* (**J**). LUAD – lung adenocarcinoma; GEO – Gene Expression Omnibus; GEA – Genomic Expression Archive; TCGA – The Cancer Genome Atlas.

participate in a range of biological processes [31]. This interacting contributes to cell proliferation via PI3K/Akt signaling pathway [32]. Also, *SPP1*-induced cell motility and ECM-invasion are essential for tumor metastasis [33,34]. Moreover, vascular endothelial growth factor could induce tumor angiogenesis by promoting endothelial cell migration and capillary formation via *SPP1,* just like a hit falling dominoes [35]. Studies have shown that si*SPP1* increased the sensitivity of lung cancer cells to afatinib in afatinib-resistant lung cancer cells [36], suggesting that *SPP1* might also be involved in the tyrosine kinase inhibitor (TKI) resistance mechanisms. Additionally, *SPP1* also mediated tumor immunity, such as tumor-associated macrophages (TAMs) polarization, upregulation of PD-L1 and promotion of the immune escape of LUAD cells [37]. As a result, we speculated that *SPP1* might be involved in the progression of LUAD via these signaling pathways and biological functions. Interestingly, the correlations between *SPP1* and gender and smoking were also observed. *SPP1* polymorphism was found to be related to a higher risk of gastric precancerous lesions in males [38]. However, another study indicated that estrogen may upregulate the expression of *SPP1* [39]. Among lung cancer patients, non-smokers exhibited lower expression levels of *SPP1* [40]. Besides, tobacco extract could induce the expression of *SPP1 in vitro* [41]. Regrettably, with little information of *SPP1* on the evolvement of LUAD in non-smoking females, further investigations are required to confirm these results of the function of *SPP1*.

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

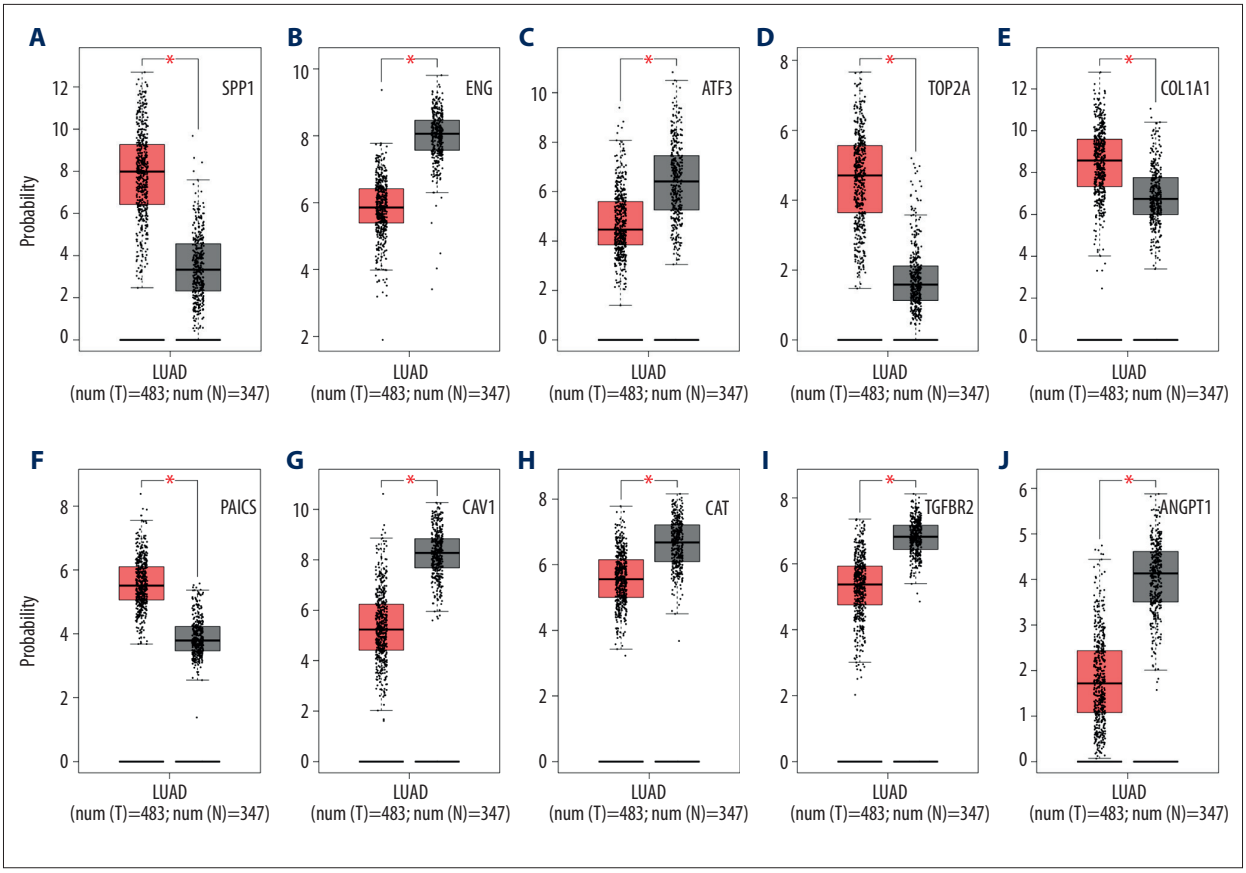


Figure 5. The differential expression analysis of 10 most significant hub genes in LUAD based on GEPIA using the data of TCGA and GTEx databases, including *SPP1* (**A**), *ENG* (**B**), *ATF3* (**C**), *TOP2A* (**D**), *COL1A1* (**E**), *PAICS* (**F**), *CAV1* (**G**), *CAT* (**H**), *TGFBR2* (**I**), *ANGPT1* (**J**). LUAD – lung adenocarcinoma; GEPIA – Gene Expression Profiling Interactive Analysis; TCGA – The Cancer Genome Atlas; GTEx – Genotype-Tissue Expression.

In order to further explore the pathogenesis of LUAD in non-smoking females, we screened the most important module (Figure 3A) from the PPI network. The DEGs in this module were mostly enriched in vasculature development, transforming growth factor (TGF)-beta signal pathway, and cellular response to hormone stimulus (Figure 3B), which could be involved in oncogenesis and progress of LUAD in non-smoking females. Among them, *ENG*, *ATF3*, and *ANGPT1* were down-regulated and were all included in the most important module. *ANGPT1*, a member of the angiopoietin family, plays a vital role in vascular development and angiogenesis [42]. *ANGPT1* was identified as a tumor suppressor gene related to female lung cancer in a sex-specific SNP-SNP interaction analysis based on the same dataset (GSE10072) [43], which is in line with our results. Moreover, the tumor metastasis of *ANGPT1* knockout mice increased significantly compared with the control group, which suggested that *ANGPT1* might be a prognostic marker [44]. However, the prognostic value of *ENG* and *ATF3* in lung cancer remains controversial. Over-expression of *ENG* was found to activate the TGF-β/ALK1 signaling pathway and promote endothelial cell proliferation and migration in one study [45], while another study showed that *ENG* haplo-insufficient mice with lung cancer could decrease tumor size and vascular density [46]. It was also reported that over-expression of *ATF3* significantly upregulated p53 and inhibited the tumorigenesis of lung cancer [47]. On the other hand, immunohistochemical staining results suggested that lung cancer cells proliferation was evidently inhibited through *ATF3* knockdown *in vitro* [48]. Therefore, the function of *ENG* and *ATF3* in the tumorigenesis is complicated, which might be related to specific regulating signals and variable tumor microenvironments. The specific molecular mechanisms deserve further exploration, and smoking and gender could be considered as regulators.

Our results indicated that over-expression of *COL1A1*, *PAICS*, and *TOP2A*, and under-expression of *CAT*, *CAV1*, and *TGFBR2* resulted in poor prognosis (Figure 4). Using the same dataset (GSE32863), Qiong Wu et al. identified *COL1A1* as a prognostic biomarker in NSCLC [49], which suggested the significance of *COL1A1* in lung cancer and the reliability of the present study. In addition, *COL1A1* was demonstrated to promote the migration of colorectal cancer cells by Transwell assays *in vitro* [50].

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

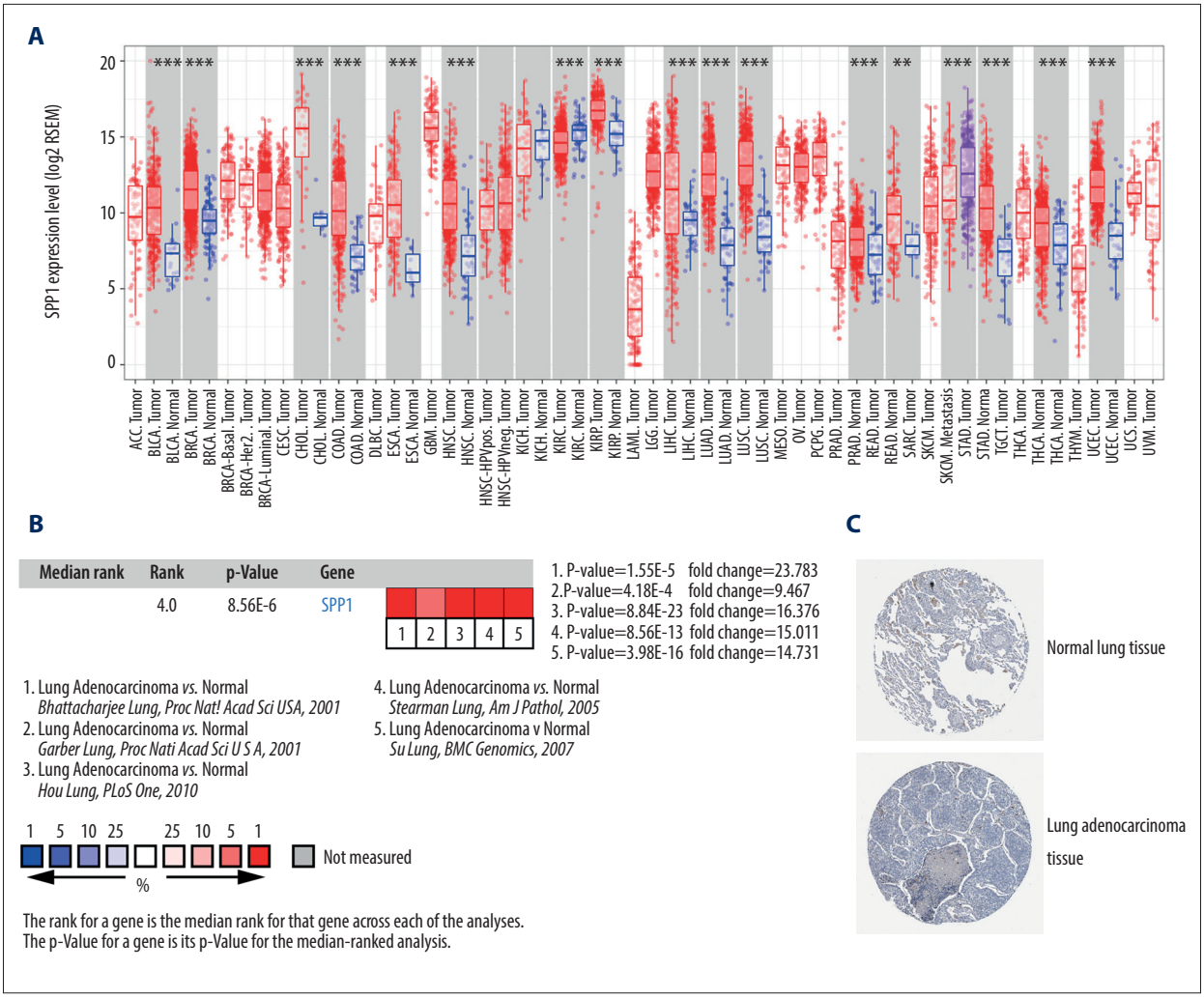


Figure 6. The upregulation of *SPP1* was validated in different databases. (**A**) The expression profiling of *SPP1* in various tumor types performed by TIMER according to the TCGA database. The red columns stand for tumor samples and the blue columns stand for normal samples. *** Represents that the *P*<0.001, and ** represents that the *P*<0.01. (**B**) A meta-analysis of *SPP1* across 5 analyses in the Oncomine database showed that *SPP1* was upregulated in LUAD. (**C**) Immunohistochemistry results of SPP1 protein expression in LUAD tissue and normal tissue from HPA database. TIMER – Tumor Immune Estimation Resource; TCGA – The Cancer Genome Atlas; HPA – Human Protein Atlas; LUAD – lung adenocarcinoma.

The proliferation of breast cancer cells and colon cancer cells was apparently inhibited through knockdown of *PAICS* and *TOP2A*, respectively [51,52]. It was reported that low expression of *CAT*, which could involve in oxidative stress defending, enhanced the invasion of lung cancer cells [53]. Moreover, *CAV1* was proven to inhibit LUAD cells proliferation [54]. Stephen et al. found that *TGFBR2* deletion in mouse airway epithelia increased migration and invasion, and led to poor survival of NSCLC patients [55]. The correlations between these genes and smoking have also been reported. Compared with smokers, the expression level of *CAV1* [56] and *TOP2A* [57] in non-smokers was reported to be lower, while the expression levels of *TGFBR2* [58] and *CAT* [59] were reported to be higher in non-smokers. Taken together, although these hub genes all played essential parts in the evolution and progression of LUAD, the specific mechanisms remain not clarified, and future experimental analyses are still demanded.

Furthermore, the immune microenvironment of LUAD in non-smoking females might also contribute to the tumorigenesis. A model of ovarian cancer indicated that increased immune infiltrates contributed to tumor progression, including dendritic cells and macrophages [60], which supported our findings (Figure 7A). However, Figure 7B illustrated that lower expression level of dendritic cells resting resulted in poor prognosis. As for relapsed colorectal cancer patients, fewer tumor-infiltrating dendritic cells were detected [61]. On the other hand, the increased tumor-infiltrating dendritic cells were observed

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

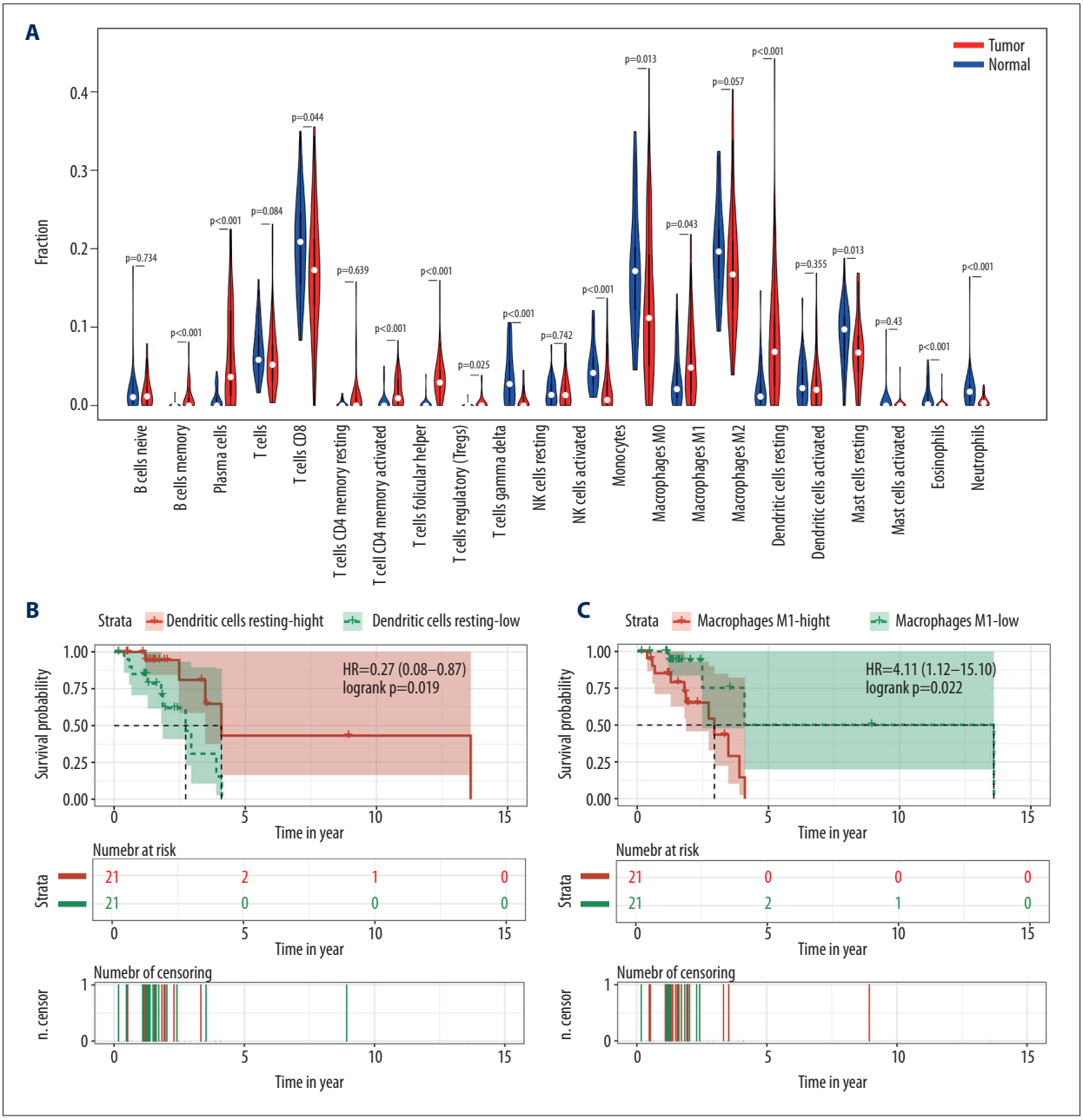


Figure 7. Distribution and prognostic analysis of infiltrating immune cells in non-smoking females with LUAD based on TCGA database and CIBERSORT. (**A**) Distribution landscape of infiltrating immune cell in non-smoking females with LUAD. The overall survival analysis of dendritic cells resting (**B**) and macrophages M1 (**C**). LUAD – lung adenocarcinoma; TCGA – The Cancer Genome Atlas.

in a mouse model of ovarian cancer as tumor progressed [62]. This suggested that during the process of tumorigenesis and development, the amount, subtypes, and functions of dendritic cells were changing [63], and gender and smoking might be included as impact factors to understand its complex functions. Accumulated evidence revealed the essential value of TAMs, M1 phenotype of TAMs, was identified as a tumor-suppressing factor in LUAD [64]. Based on TCGA and CIBERSORT, our results suggested that lower level of macrophages M1 was associated with better prognosis, which is contrary to previous reports, suggesting that female and smoking might be independent factors. It has been reported that the ratio of macrophages M1 increased from 26% to 84% with smoking severity [65]. Regrettably, the impact of gender and smoking on macrophages M1 in LUAD has not been assessed in previous studies, and further research is urgently demanded.

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

Conclusions

Importantly, in this study, the mechanisms of LUAD in non-smoking females were explored by bioinformatics methods, and promising biomarkers and possible signaling pathways were identified and validated based on multiple databases were combined to confirm these results. Ten hub genes and 2 immune cell subtypes were found with prognostic significance, including *SPP1, ENG, ATF3, TOP2A, COL1A1, PAICS*, *CAV1, CAT, TGFBR2, ANGPT1*, dendritic cells resting, and macrophages M1. However, as a result of the limitation of the relatively small sample size of online data in this field, the specific mechanism of these hub genes and infiltrating immune cells were still unrevealed. Therefore, further research on the mechanism of LUAD in non-smoking females is necessary.

Conflict of interest

None.

Supplementary Data

Supplementary Table 1. Information for samples in the included datasets (GEO database).

Supplementary Table 2. All 315 commonly differentially expressed genes (DEGs) were detected from three profile datasets, including 254 downregulated genes and 61 up-regulated genes.

Supplementary Table 3. GO enrichment analysis of 315 DEGs.

Supplementary Table 4. Functional roles of 36 hub genes with degree >10.

Supplementary Table 5. Information for included samples from TCGA database.

Supplementary/raw data available from the corresponding author on request.

References:

1. Siegel RL, Miller KD, Jemal A: Cancer statistics, 2019. Cancer J Clin, 2019; 69: 7–34
2. Cancer Genome Atlas Research Network: Comprehensive molecular profiling of lung adenocarcinoma. Nature, 2014; 511: 543–50
3. Donner I, Katainen R, Sipila LJ et al: Germline mutations in young non-smoking women with lung adenocarcinoma. Lung Cancer, 2018; 122: 76–82
4. Cheng TD, Darke AK, Redman MW et al: Smoking, sex, and non-small cell lung cancer: Steroid hormone receptors in tumor tissue (S0424). J Natl Cancer Inst, 2018; 110: 734–42
5. Hsu LH, Chu NM, Kao SH: Estrogen, estrogen receptor and lung cancer. Int J Mol Sci. 2017; 18(8): pii: E1713
6. O'Keeffe LM, Taylor G, Huxley RR et al: Smoking as a risk factor for lung cancer in women and men: A systematic review and meta-analysis. BMJ Open, 2018; 8: e021611
7. Kinoshita FL, Ito Y, Morishima T et al: Sex differences in lung cancer survival: Long-term trends using population-based cancer registry data in Osaka, Japan. Jpn J Clin Oncol, 2017; 47: 863–69
8. Conforti F, Pala L, Bagnardi V et al: Cancer immunotherapy efficacy and patients' sex: A systematic review and meta-analysis. Lancet Oncol, 2018; 19: 737–46
9. Skov BG, Fischer BM, Pappot H: Oestrogen receptor beta over expression in males with non-small cell lung cancer is associated with better survival. Lung Cancer, 2008; 59: 88–94
10. Sun S, Schiller JH, Gazdar AF: Lung cancer in never smokers – a different disease. Nat Rev Cancer, 2007; 7: 778–90
11. Nakagawa H, Fujita M: Whole genome sequencing analysis for cancer genomics and precision medicine. Cancer Sci, 2018; 109: 513–22
12. Wu D, Wang X: Application of clinical bioinformatics in lung cancer-specific biomarkers. Cancer Metastasis Rev, 2015; 34: 209–16
13. Barrett T, Wilhite SE, Ledoux P et al: NCBI GEO: Archive for functional genomics data sets – update. Nucleic Acids Res, 2013; 41: D991–95
14. Feng H, Gu ZY, Li Q et al: Identification of significant genes with poor prognosis in ovarian cancer via bioinformatical analysis. J Ovarian Res, 2019; 12: 35
15. Sun C, Yuan Q, Wu D et al: Identification of core genes and outcome in gastric cancer using bioinformatics analysis. Oncotarget, 2017; 8: 70271–80
16. Zhang L, Peng R, Sun Y et al: Identification of key genes in non-small cell lung cancer by bioinformatics analysis. Peer J, 2019; 7: e8215
17. Jiao X, Sherman BT, Huang da W et al: DAVID-WS: A stateful web service to facilitate gene/protein list analysis. Bioinformatics, 2012; 28: 1805–6
18. Doncheva NT, Morris JH, Gorodkin J, Jensen LJ: Cytoscape stringApp: Network analysis and visualization of proteomics data. J Proteome Res, 2019; 18: 623–32
19. Tripathi S, Pohl MO, Zhou Y et al: Meta- and orthogonal integration of influenza "OMICs" data defines a role for UBR4 in virus budding. Cell Host Microbe, 2015; 18: 723–35
20. Gyorffy B, Surowiak P, Budczies J, Lanczky A: Online survival analysis software to assess the prognostic value of biomarkers using transcriptomic data in non-small-cell lung cancer. PLoS One, 2013; 8: e82241
21. Tang Z, Li C, Kang B et al: GEPIA: A web server for cancer and normal gene expression profiling and interactive analyses. Nucleic Acids Res, 2017; 45: W98–102
22. Li T, Fan J, Wang B et al: TIMER: A web server for comprehensive analysis of tumor-infiltrating immune cells. Cancer Res, 2017; 77: e108–10
23. Rhodes DR, Kalyana-Sundaram S, Mahavisno V et al: Oncomine 3.0: Genes, pathways, and networks in a collection of 18,000 cancer gene expression profiles. Neoplasia, 2007; 9: 166–80
24. Uhlen M, Fagerberg L, Hallstrom BM et al: Proteomics. Tissue-based map of the human proteome. Science, 2015; 347: 1260419
25. Newman AM, Liu CL, Green MR et al: Robust enumeration of cell subsets from tissue expression profiles. Nat Methods, 2015; 12: 453–57
26. Choe EK, Yi JW, Chai YJ, Park KJ: Upregulation of the adipokine genes ADIPOR1 and SPP1 is related to poor survival outcomes in colorectal cancer. J Surg Oncol, 2018; 117: 1833–40
27. Chen X, Xiong D, Ye L et al: SPP1 inhibition improves the cisplatin chemosensitivity of cervical cancer cell lines. Cancer Chemother Pharmacol, 2019; 83: 603–13
28. Insua-Rodríguez J, Pein M, Hongu T et al: Stress signaling in breast cancer cells induces matrix components that promote chemoresistant metastasis. EMBO Mol Med, 2018; 10: pii: e9003
29. Hu Z, Lin D, Yuan J et al: Overexpression of osteopontin is associated with more aggressive phenotypes in human non-small cell lung cancer. Clin Cancer Res, 2005; 11: 4646–52
30. Blasberg JD, Pass HI, Goparaju CM et al: Reduction of elevated plasma osteopontin levels with resection of non-small-cell lung cancer. J Clin Oncol, 2010; 28: 936–41

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

31. Hao C, Cui Y, Owen S et al: Human osteopontin: Potential clinical applications in cancer (review). Int J Mol Med, 2017; 39: 1327–37
32. Ogata T, Ueyama T, Nomura T et al: Osteopontin is a myosphere-derived secretory molecule that promotes angiogenic progenitor cell proliferation through the phosphoinositide 3-kinase/Akt pathway. Biochem Biophys Res Commun, 2007; 359: 341–47
33. Pang X, Xie R, Zhang Z et al: Identification of SPP1 as an extracellular matrix signature for metastatic castration-resistant prostate cancer. Front Oncol, 2019; 9: 924
34. Das R, Philip S, Mahabeleshwar GH et al: Osteopontin: it's role in regulation of cell motility and nuclear factor kappa B-mediated urokinase type plasminogen activator expression. IUBMB Life, 2005; 57: 441–47
35. Babarović E, Valković T, Budisavljević I et al: The expression of osteopontin and vascular endothelial growth factor in correlation with angiogenesis in monoclonal gammopathy of undetermined significance and multiple myeloma. Pathol Res Pract, 2016; 212: 509–16
36. Wang X, Zhang F, Yang X et al: Secreted phosphoprotein 1 (SPP1) contributes to second-generation EGFR tyrosine kinase inhibitor resistance in non-small cell lung cancer. Oncol Res, 2019; 27: 871–77
37. Zhang Y, Du W, Chen Z, Xiang C: Upregulation of PD-L1 by SPP1 mediates macrophage polarization and facilitates immune escape in lung adenocarcinoma. Exp Cell Res, 2017; 359: 449–57
38. Chang WL, Lin MY, Kuo HY et al: Osteopontin polymorphism increases gastric precancerous intestinal metaplasia susceptibility in *Helicobacter pylori* infected male. Future Oncol, 2017; 13: 1415–25
39. Banerjee A, Rose R, Johnson GA et al: The influence of estrogen on hepatobiliary osteopontin (SPP1) expression in a female rodent model of alcoholic steatohepatitis. Toxicol Pathol, 2009; 37: 492–501
40. Maneechotesuwan K, Kasetsinsombat K, Wongkajornsilp A, Barnes PJ: Simvastatin up-regulates adenosine deaminase and suppresses osteopontin expression in COPD patients through an IL-13-dependent mechanism. Respir Res, 2016; 17: 104
41. Bishop E, Theophilus EH, Fearon IM: *In vitro* and clinical studies examining the expression of osteopontin in cigarette smoke-exposed endothelial cells and cigarette smokers. BMC Cardiovasc Disord, 2012; 12: 75
42. Fisher TE, Molskness TA, Villeda A et al: Vascular endothelial growth factor and angiopoietin production by primate follicles during culture is a function of growth rate, gonadotrophin exposure and oxygen milieu. Hum Reprod, 2013; 28: 3263–70
43. Yao S, Dong SS, Ding JM et al: Sex-specific SNP-SNP interaction analyses within topologically associated domains reveals ANGPT1 as a novel tumor suppressor gene for lung cancer. Genes Chromosomes Cancer, 2019 [Epub ahead of print]
44. Michael IP, Orebrand M, Lima M et al: Angiopoietin-1 deficiency increases tumor metastasis in mice. BMC Cancer, 2017; 17: 539
45. Lebrin F, Goumans MJ, Jonker L et al: Endoglin promotes endothelial cell proliferation and TGF-beta/ALK1 signal transduction. EMBO J, 2004; 23: 4018–28
46. Dallas NA, Samuel S, Xia L et al: Endoglin (CD105): A marker of tumor vasculature and potential target for therapy. Clin Cancer Res, 2008; 14: 1931–37
47. Du A, Jiang Y, Fan C: NDRG1 Downregulates ATF3 and inhibits cisplatin-induced cytotoxicity in lung cancer A549 cells. Int J Med Sci, 2018; 15: 1502–7
48. Li X, Zhou X, Li Y et al: Activating transcription factor 3 promotes malignance of lung cancer cells *in vitro*. Thorac Cancer, 2017; 8: 181–91
49. Wu Q, Zhang B, Sun Y et al: Identification of novel biomarkers and candidate small molecule drugs in non-small-cell lung cancer by integrated microarray analysis. Onco Targets Ther, 2019; 12: 3545–63
50. Zhang Z, Wang Y, Zhang J et al: COL1A1 promotes metastasis in colorectal cancer by regulating the WNT/PCP pathway. Mol Med Rep, 2018; 17: 5037–42
51. Gallenne T, Ross KN, Visser NL et al: Systematic functional perturbations uncover a prognostic genetic network driving human breast cancer. Oncotarget, 2017; 8: 20572–87
52. Zhang R, Xu J, Zhao J, Bai JH: Proliferation and invasion of colon cancer cells are suppressed by knockdown of TOP2A. J Cell Biochem, 2018; 119: 7256–63
53. Tsai JY, Lee MJ, Dah-Tsyr Chang M, Huang H: The effect of catalase on migration and invasion of lung cancer cells by regulating the activities of cathepsin S, L, and K. Exp Cell Res, 2014; 323: 28–40
54. Yan Y, Xu Z, Qian L et al: Identification of CAV1 and DCN as potential predictive biomarkers for lung adenocarcinoma. Am J Physiol Lung Cell Mol Physiol, 2019; 316: L630–43
55. Malkoski SP, Haeger SM, Cleaver TG et al: Loss of transforming growth factor beta type II receptor increases aggressive tumor behavior and reduces survival in lung adenocarcinoma and squamous cell carcinoma. Clin Cancer Res, 2012; 18: 2173–83
56. Singh DP, Kaur G, Bagam P et al: Membrane microdomains regulate NLRP10- and NLRP12-dependent signalling in A549 cells challenged with cigarette smoke extract. Arch Toxicol, 2018; 92: 1767–83
57. Koomägi R, Stammler G, Manegold C et al: Expression of resistance-related proteins in tumoral and peritumoral tissues of patients with lung cancer. Cancer Lett, 1996; 110: 129–36
58. Szymanowska-Narloch A, Jassem E, Skrzypski M et al: Molecular profiles of non-small cell lung cancers in cigarette smoking and never-smoking patients. Adv Med Sci, 2013; 58: 196–206
59. Alzoubi KH, Halboup AM, Alomari MA, Khabour OF: The neuroprotective effect of vitamin E on waterpipe tobacco smoking-induced memory impairment: The antioxidative role. Life Sci, 2019; 222: 46–52
60. Scarlett UK, Rutkowski MR, Rauwerdink AM et al: Ovarian cancer progression is controlled by phenotypic changes in dendritic cells. J Exp Med, 2012; 209: 495–506
61. Krempski J, Karyampudi L, Behrens MD et al: Tumor-infiltrating programmed death receptor-1+ dendritic cells mediate immune suppression in ovarian cancer. J Immunol, 2011; 186: 6905–13
62. Kocián P, Šedivcová M, Drgáč J et al: Tumor-infiltrating lymphocytes and dendritic cells in human colorectal cancer: Their relationship to KRAS mutational status and disease recurrence. Hum Immunol, 2011; 72: 1022–28
63. Tran Janco JM, Lamichhane P, Karyampudi L, Knutson KL: Tumor-infiltrating dendritic cells in cancer pathogenesis. J Immunol, 2015; 194: 2985–91
64. Rakaee M, Busund LR, Jamaly S et al: Prognostic value of macrophage phenotypes in resectable non-small cell lung cancer assessed by multiplex immunohistochemistry. Neoplasia, 2019; 21: 282–93
65. Bazzan E, Turato G, Tine M et al: Dual polarization of human alveolar macrophages progressively increases with smoking and COPD severity. Respir Res, 2017; 18: 40

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]