MEDICAL SCIENCE MONITOR

CLINICAL RESEARCH

e-ISSN 1643-3750
© Med Sci Monit, 2020; 26: e922070
DOI: 10.12659/MSM.922070

Received: 2019.12.10
Accepted: 2020.03.24
Available online: 2020.04.24
Published: 2020.06.24

Comprehensive Analysis of Candidate Diagnostic and Prognostic Biomarkers Associated with Lung Adenocarcinoma

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

AE 1 Jingyuan Li
CF 2,3 Xingyuan Liu
B 1 Zan Cui
DG 1 Guanying Han

1 Faculty of Pharmaceutical Sciences, The First Affiliated Hospital of Jinzhou Medical University, Jinzhou, Liaoning, P.R. China
2 Pathology Department, The First Affiliated Hospital of Jinzhou Medical University, Jinzhou, Liaoning, P.R. China
3 Pathology Department, Jinzhou Medical University, Jinzhou, Liaoning, P.R. China

Corresponding Author: Guanying Han, e-mail: hgy19800223@163.com
Source of support: The present study was supported by Natural Science Foundation Funding Scheme of Liaoning Province (No. 2019-MS-145)

Background: We aimed to screen and identify central genetic and molecular targets involved in advancement of lung adenocarcinoma (LUAD) and to perform an integrated analysis and clinical validation.

Material/Methods: The GEO2R technique was utilized to assess differentially expressed genes (DEGs) among the gene sets GSE75037, GSE85716, and GSE118370. Subsequently, gene Ontology (GO) analyses and Kyoto Encyclopedia of Genes and Genomes (KEGG) analytical methods were executed to determine related biofunctions and signaling pathways, which were annotated with tools from the Database for Annotation, Visualization and Integrated Discovery (DAVID) resource. Then, a protein-protein interaction (PPI) network complex consisting of all detected DEGs was built with the STRING web interface. Cytohubba and MCODE plug-ins for Cytoscape software and Gene Expression Profiling Interactive Analysis (GEPIA) were employed to identify the hub genes. Finally, the mRNA expression of the identified hub genes was quantitatively validated by The Cancer Genome Atlas (TCGA) database analysis and real-time quantitative polymerase chain reaction (RT-qPCR).

Results: We screened 146 upregulated DEGs and 431 downregulated DEGs with the criteria of |logFC| >1 and $P<0.05$, and the GO analysis indicated that DEGs were implicated in mitotic nuclear division (biological process, BP), the nucleus (cellular component, CC), and protein binding (molecular function, MF) and were associated with multiple KEGG pathways, such as the p53 signaling pathway in cancer. Then, the top 8 genes that predicted significantly different outcomes in LUAD patients were filtered from the DEGs and selected as hub genes. The TCGA database analysis and RT-qPCR results demonstrated that these genes were differentially expressed with the same trends in LUAD tissues compared with normal tissues.

Conclusions: Overall, we propose that 8 genes (PECAM1, CDK1, MKI67, SPP1, TOP2A, CHEK1, CCNB1, and RRM2) might be novel hub genes strongly associated with the progression and prognosis of LUAD.

MeSH Keywords: Biological Markers • Carcinoma, Non-Small-Cell Lung • Genes, vif

Full-text PDF: https://www.medscimonit.com/abstract/index/idArt/922070

3291 7 4 48

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 remains the principal culprit of cancer-related morbidity and mortality, accounting for 18.4% of cancer deaths based in a worldwide survey, corresponding to approximately 1.8 million deaths [1]. Lung adenocarcinoma (LUAD), which habitually occurs in peripheral portions of the lungs, is overtaking lung squamous cell carcinoma (LUSC) as the top prevailing pathological type of lung cancer and is characterized by glandular differentiation and/or mucin production [2]. IASLC/ATS/ERS corporately released a current multidisciplinary categorization of LUAD in 2011 that includes concordant expert terminology and comprehensive diagnostic criteria [3]. It reveals the correlation between the histological subtypes of LUAD and the risk of poor prognosis through comparison and analysis of big data from patients [4]. At present, the 5-year survival rate of MIA (minimally invasive adenocarcinoma) and AIS (adenocarcinoma *in situ*) after complete removal is approximately 100% and thus is associated with favorable prognosis [5]; however, the outcomes of other LUAD subtypes are unfavorable, and the 5-year survival rate is barely 21% [6]. Dissimilar clinical factors, including clinicopathological T classification and OSR (overall survival rate), exist between LUAD and LUSC owing to differences in genomic alternations, including conversion rate, mutation characteristics and frequently mutated genes [7]. The occurrence and progression of LUAD are closely related to driver mutations [8].

High-throughput sequencing, with extended read length and appropriate stability, has become more frequently applied for the exploration of candidate genes in cancer, diabetes, autism, and other genetic diseases and further increased the probability of identifying genes in non-model species. In the field of non-small-cell lung cancer (NSCLC)-targeted therapeutic strategies, studies concentrating on MET, RAT and BRAF as targets are being carried out in addition to studies on targeting ALK, EGFR, and ROS1 [9]. LUAD, which only shows symptoms similar to general respiratory diseases in the early stage, is probably overlooked, so uncovering biomarkers will substantially augment the potential to identify molecular targets in clinical practice. In addition, screening predictive biomarkers has been confirmed to be a necessary condition for discovering targeted anticancer drugs. In addition, patients with specific subtypes could potentially be recognized by applying biomarkers that can predict targeted drug response, invasive or malignant behavior, and drug resistance mechanisms. Accordingly, the effort of unrevealing beneficial and effective biomarkers for the preclinical diagnosis and prognostic prediction of LUAD is extremely urgent.

Comprehensive analysis of the gene expression profile array using bioinformatics techniques was carried out to pinpoint various differentially expressed genes (DEGs). In the current study, upregulated DEGs (uDEGs) and downregulated DEGs (dDEGs) were filtered among thousands of DEGs in 3 GSE datasets of patients with LUAD. Subsequently, we performed Gene Ontology (GO) analyses, including results in the BP, CC and MF categories, to explore biofunctions and singling pathway enrichment; Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis and protein-protein interaction (PPI) network construction was also performed with the indicated uDEGs and dDEGs. After the candidate genes were selected under certain conditions, the hub genes were obtained by survival analysis. Then, the differential messenger RNA (mRNA) expression levels of selected hub genes in LUAD tissues were confirmed by The Cancer Genome Atlas (TCGA) database analysis and real-time quantitative polymerase chain reaction (RT-qPCR). We intend to further clarify the intricate molecular biology of the pathogenesis of LUAD and substantiate potential pivotal genes that may be encouraging candidate biomarkers of diagnosis and prognosis, targets for the development of novel anticancer agents, or markers for drug resistance and precise treatment.

Material and Methods

Data source

GSE75037 data were detected with an Illumina HumanWG-6 v3.0 expression beadchip [10], GSE85716 analysis performed on an Agilent-062918 OE Human lncRNA Microarray V4.0 02800 [11], and GSE118370 analysis was performed on an Affymetrix Human Genome U133 Plus 2.0 Array [12]; these data were extracted from the GEO server, a free open genome database based on microarrays and sequences [13]. The characteristics of the 3 gene expression datasets are described in Table 1.

Data preprocessing

GEO2R (*www.ncbi.nlm.nih.gov/geo/geo2r/*) is based on R packages GEOquery and Limma, which is a GEO online analysis tool used to screen DEGs between normal lung tissue and LUAD samples in the current study. The genes that satisfied the conditions of |logFC| >1 and $P<0.05$ were determined to be statistically significant DEGs. The DEGs we obtained were grouped into 2 categories depending on logFC: uDEGs, logFC >1 and dDEGs, logFC <–1. The common intersection of identified uDEGs and dDEGs from GSE75037, GSE85716, and GSE118370 was separately acquired using Venny diagram plotter (*http://bioinfogp.cnb.csic.es/tools/venny/index.html*).

Functional enrichment of overlapping uDEGs and dDEGs

The Database for Annotation, Visualization and Integrated Discovery (DAVID), a bioinformatics database that provides integrated and detailed biofunctional annotation information

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. The main features of 3 studies of GEO gene expression microarray data.

GEO accession	Platform	Case/Control	Sample	Date
GSE75037	GPL6884	83/83	LUAD	2016
GSE85716	GPL19612	6/6	LUAD	2017
GSE118370	GPL570	6/6	LUAD	2019

LUAD – lung adenocarcinoma.

for sizable cohorts of genes or proteins [14], was applied to reveal the functional enrichment of uDEGs and dDEGs, including terms related to BP, CC, and MF. KEGG is a habitually used database for systematically uncovering the principal metabolic signaling pathways of gene expression products in cells. The DAVID web server was utilized to investigate the KEGG pathways of uDEGs and dDEGs in the current research.

PPI network complex construction and module analysis

STRING, a database to inspect interactions between recognized proteins [15], was implemented to assemble the PPI network complex of uDEGs and dDEGs. Cytoscape is a public serviceable program for improved visualization of interaction networks between molecules with integrated data [16]. The cytoHubba plug-in in Cytoscape software explores considerable nodes based on the score of all 11 methods in each node in the network. Accordingly, the PPI network module with a significant gene pair (combination score >0.2) was obtained with the MCODE plug-in. MCODE clustering score >5 and number of nodes >5 were used as the cutoff values for investigating meaningful modules. DAVID was applied to explore the functional enrichment of, and pathways involved with the genes in each module.

Determination of hub genes

GEPIA, which is implemented to assess the impact of hub genes on the prognostic outcome of the occurrence and progression of LUAD, is a well-performing web server that allows for interactive analysis of the specific functions of key genes, including profile analysis of differential expression in various tumors, pathological staging analysis, and survival analysis of patients according to DEGs [17]. In the course of survival analysis, patients were separated into high and low groups according to the median expression level of each selected gene. The significance between 2 groups was determined by the log-rank T test, and hazard ratio (HR) was estimated to evaluate the association between gene expression and survival rate. Kaplan-Meier charts with valid log-rank *P* values (<0.05 as the cutoff criterion) were applied to assess overall survival rate (OSR) and filter the hub genes between the 2 groups. GEPIA was used to observe the recognized differentially expressed hub genes in specific clinicopathological stages of LUAD. The expression of DEGs in different pathological stages was evaluated with one-way analysis of variance.

Table 2. Clinicopathologic features of patients with LUAD in JMU.

Variables	All patient
Patient number	36
Age, years	59.64±6.78
Body mass index, kg/m^2	23.3±2.99
Gender	
Male	17
Female	19
Clinicopathologic diagnosis (WHO)	
LUAD I	7
LUAD II	7
LUAD III	7
LUAD IV	7
LT	8
Lymph node involvement	
N	13
Y	15

JMU – Jinzhou Medical University; LUAD – lung adenocarcinoma; LT – lung tissue; WHO – World Health Organization.

TCGA database validation

TCGA was utilized, which has 2.5 petabytes (PB) of detailed data on 33 cancers from no less than 10 000 patients. The hub genes were verified using the GEPIA tool based on TCGA data of LUAD and corresponding paracancerous tissues.

Patient sample collection

To further validate the robustness of the hub genes identified, 28 LUAD specimens and 8 benign marginal normal lung tissues were collected in the First Affiliated Hospital of Jinzhou Medical University from January 2016 to December 2018. The characteristics of the patients participating in the study are shown in Table 2. Prior to the use of tissue samples, informed consent

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 3. Primer pair sequences for RT-qPCR of PECAM1, CDK1, MKI67, SPP1, TOP2A, CHEK1, CCNB1, RRM2, and GAPDH.

Gene	Forward (5'-3')	Reverse (5'-3')
PECAM1	atgccagtggaaatgtcc	tcagaagtggtactggtg
CDK1	ccgtcgtaacctgttgagtaactat	gtctacccttatacaccacaccgtaa
MKI67	cgcgaattcagagagcttttccagacaccatg	cgagcctcgaggaagattgttggggtacccac
SPP1	ttctgattgggacagccgtg	tctcatcattggctttccgct
TOP2A	ggtgagaaggactggcagaaat	cttgtcgatgaagtacagggcta
CHEK1	ccagatgctcagagattcttcca	tgttcaacaaacgctcacgatta
CCNB1	ttcgcctgagcctattttggta	agctccatcttctgcatccacat
RRM2	tttagtgagcttagcacagcggga	aaatctgcgttgaagcagtgaggc
GAPDH	tccaccaccctgttgctgta	gacttcaacagcaactcccac

was obtained from all patients. The study complied with the Helsinki Declaration and was approved by the Medical Ethics Committee of Jinzhou Medical University. All specimens were classified histologically by clinical pathologists according to the 2015 World Health Organization (WHO) classification of lung tumors. Each biopsy specimen was divided into 2 parts, one for routine histological diagnosis and the other for rapid freezing in RNAlater (Invitrogen) and storage at –20°C until total RNA separation and analysis.

RT-qPCR

Total RNA was obtained from 28 LUAD cases and 8 benign pulmonary tissue cases with TRIzol RNA separation reagent (Invitrogen) following the manufacturer's steps, and then the SuperScript IV System kit (Invitrogen) was used to synthesize cDNA. As described in a previous study [18], TB Green Premix Ex Taq II m (Takara) was used for RT-qPCR, and a 7500 Real-Time PCR Applied Bio-system (Thermo Fisher) was utilized to examine the expression difference with GAPDH as an internal reference. The primers used to amplify PECAM1, CDK1, MKI67, SPP1, TOP2A, CHEK1, CCNB1, RRM2, and GAPDH were synthesized and constructed by GenePharma Corporation (Shanghai, China). The primer sequences are shown in Table 3. The $2^{-\Delta\Delta Ct}$ method was utilized to compare the relative mRNA expression of the tumor samples and the control.

Results

DEGs screening

Criteria (|logFC| >1 and P<0.05) were adopted as thresholds for DEG filtration. Relevant data from GSE75037, GSE85716 and GSE118370 were used to estimate differentially expressed genes (DEGs) using GEO2R. A total of 577 DEGs were identified, among which 146 genes were upregulated (uDEGs) and 431 genes were downregulated (dDEGs), as demonstrated in the Venn diagram (Figure 1). The number of dDEGs exceeded the number of uDEGs. The top 5 most significant DEGs among the GSE75037, GSE85716 and GSE118370 datasets are presented in Table 4.

Pathway enrichment analysis

The outcomes of uDEGs GO analysis demonstrated that mitotic nuclear division, nucleus, and protein binding were the most crucial terms in the BP, CC, and MF categories, respectively. The dDEGs GO analysis results indicated that positive regulation of transcription from the RNA polymerase II promoter, integral component of the membrane, and protein binding were the most significant terms in the BP, CC, and MF categories (Table 5). KEGG analysis was conducted on the uDEGs and dDEGs to further explore the key pathways involved. The uDEGs were generally implicated in the p53 signaling pathway, cell cycle, and oocyte meiosis signaling pathway, and the dDEGs were substantially implicated in the pathways in cancer, cAMP signaling pathway, and neuroactive ligand-receptor interaction signaling pathway (Table 6). These significantly enriched GO terms and pathways could be revealed to further apprehend the function of these DEGs in the occasion and progress of LUAD.

PPI network and module analysis

The PPI network was constructed by using STRING to retrieve interactions between genes. The established PPI network consisted of 577 edges and 2166 nodes (Figure 2A), with a local clustering coefficient=0.4. The top 20 highest-level central nodes, including IL6, CDH1, VWF, PECAM1, CDK1, CD34, CXCL12, JUN, BMP2, MKI67, PPARG, CTGF, COL1A1, SPP1, CAV1, TOP2A, CHEK1, CCNB1, and BDNF, were screened out with the cytoHubba plug-in. To facilitate our understanding of the DEGs, we visualized the network complex in Cytoscape software and modularized it using the plug-in MCODE to obtain

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

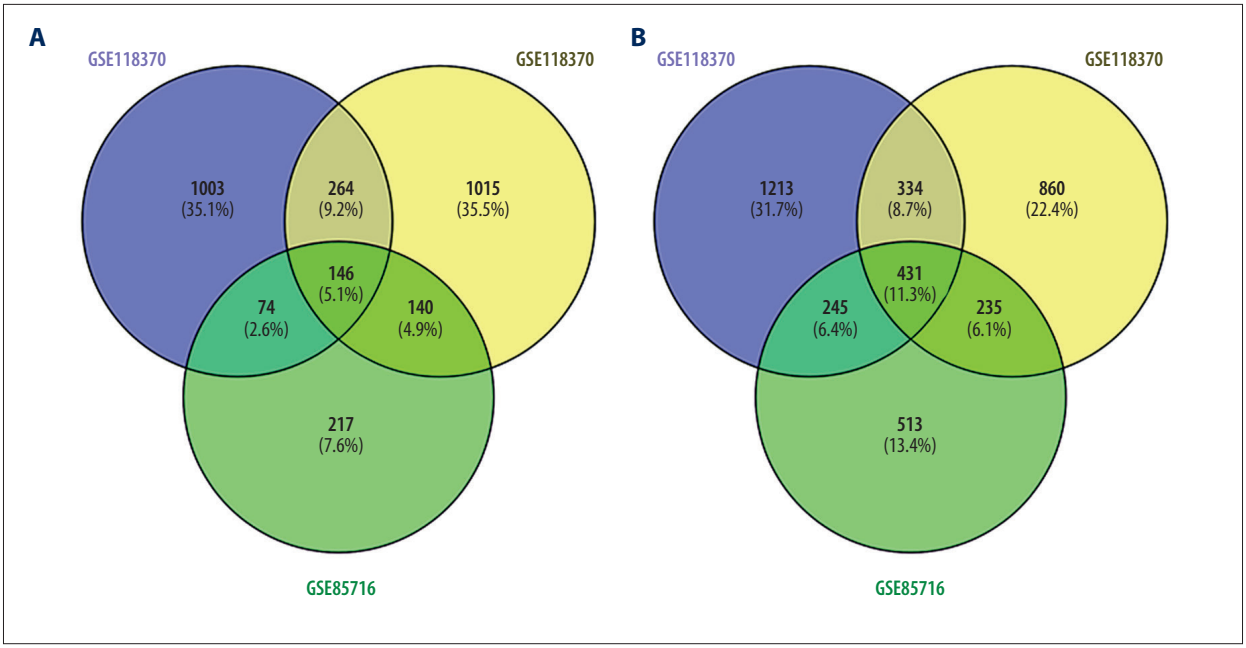

Figure 1. Venn diagram presenting the overlapping (**A**) uDEGs and (**B**) dDEGs in the 3 GEO datasets. In total, 146 uDEGs and 431 dDEGs were commonly identified between LUAD and corresponding normal adjacent lung samples. uDEGs – upregulated differentially expressed genes; dDEGs – downregulated differentially expressed genes; LUAD – lung adenocarcinoma.

Table 4. The most significant 5 uDEGs and 5 dDEGs in GSE75037, GSE85716, and GSE118370 datasets.

Gene symbol	GSE75037		Gene symbol	GSE85716		Gene symbol	GSE118370	
	P value	logFC		*P* value	logFC		*P* value	logFC
The most significant 5 uDEGs								
MMP11	1.90E-53	5.12	GLB1L3	9.61E-05	5.08254515	SPINK1	2.45E-05	4.7511875
EEF1A2	5.17E-29	5.05	CD1A	1.78E-03	3.63170602	HS6ST2	7.92E-03	4.3473223
GCNT3	2.88E-42	4.92	TMPRSS4	2.04E-03	3.87381578	TOX3	1.69E-04	4.1766185
CST1	4.27E-31	4.75	TUBB3	3.19E-03	4.16514507	CP	3.11E-04	4.1578082
FAM83A	4.68E-45	4.68	XAGE1A	3.46E-03	5.40792158	MSMB	2.86E-04	4.1265297
The most significant 5 dDEGs								
FABP4	1.45E-85	–6.03	AGER	5.93E-09	–5.94009052	CD300LG	4.19E-04	–4.9800372
CLDN18	1.16E-38	–6.03	SLC6A4	1.20E-06	–5.934542	FLJ34503	9.35E-05	–4.9718172
ITLN2	2.71E-71	–6.02	FAM107A	2.11E-06	–5.53509465	SLC19A3	5.58E-04	–4.9164017
AGER	6.39E-68	–5.72	SOSTDC1	6.81E-05	–6.41373342	FABP4	5.03E-05	–4.8213644
SFTPA1	1.85E-36	–5.54	CLDN18	3.02E-04	–5.67266395	SLC6A4	7.30E-06	–4.7189991

uDEGs – upregulated differentially expressed genes; dDEGs – downregulated differentially expressed genes.

3 consequential modules including module 1 (28 nodes and 343 edges, cluster score=25.407), module 2 (28 nodes and 118 edges, cluster score=8.741) and module 3 (5 nodes and 10 edges, cluster score=5) (Figure 2B–2D). The KEGG analysis laid out 4 related signaling pathways that were principally involved in module 1, including the cell cycle, p53 signaling pathway, and oocyte meiosis; 15 related signaling pathways were considerably involved in module 2, such as pathways in cancer; 9 related signaling pathways were significantly implicated in module 3, including basal cell carcinoma and others (Table 7).

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 5. TOP 5 GO function annotation of common uDEGs and dDEGs associated with LUAD.

Category	GO-ID	GO-term	Count	%	*P*-value
GO analysis of uDEGs (TOP5)					
BP	GO: 0007067	Mitotic nuclear division	12	8.22	5.6E-06
BP	GO: 0006260	DNA replication	9	6.16	3.8E-05
BP	GO: 0000086	G2/M transition of mitotic cell cycle	8	5.48	1.3E-04
BP	GO: 0051301	Cell division	12	8.22	1.3E-04
BP	GO: 0006977	DNA damage response, signal transduction by p53 class mediator resulting in cell cycle arrest	6	4.11	1.4E-04
CC	GO: 0005634	Nucleus	53	36.30	4.5E-02
CC	GO: 0005737	Cytoplasm	52	35.62	3.6E-02
CC	GO: 0016324	Apical plasma membrane	8	5.48	7.9E-03
CC	GO: 0005813	Centrosome	8	5.48	4.9E-02
CC	GO: 0030496	Midbody	7	4.79	5.2E-04
MF	GO: 0005515	Protein binding	79	54.11	3.3E-02
MF	GO: 0005524	ATP binding	22	15.07	4.6E-03
MF	GO: 0042802	Identical protein binding	14	9.59	4.8E-03
MF	GO: 0004674	Protein serine/threonine kinase activity	9	6.16	8.3E-03
MF	GO: 0019901	Protein kinase binding	9	6.16	8.3E-03
GO Analysis of dDEGs (TOP5)					
BP	GO: 0045944	Positive regulation of transcription from RNA polymerase II promoter	45	10.47	2.6E-05
BP	GO: 0007165	Signal transduction	44	10.23	1.9E-03
BP	GO: 0007155	Cell adhesion	33	7.67	4.0E-08
BP	GO: 0001525	Angiogenesis	32	7.44	1.2E-15
BP	GO: 0000122	Negative regulation of transcription from RNA polymerase II promoter	30	6.98	3.2E-03
CC	GO: 0016021	Integral component of membrane	157	36.51	4.7E-05
CC	GO: 0005886	Plasma membrane	156	36.28	1.5E-11
CC	GO: 0070062	Extracellular exosome	80	18.60	3.3E-02
CC	GO: 0005887	Integral component of plasma membrane	71	16.51	8.2E-10
CC	GO: 0005576	Extracellular region	70	16.28	3.3E-07
MF	GO: 0005515	Protein binding	222	51.63	7.6E-03
MF	GO: 0005509	Calcium ion binding	28	6.51	6.2E-03
MF	GO: 0042803	Protein homodimerization activity	27	6.28	1.4E-02
MF	GO: 0046982	Protein heterodimerization activity	19	4.42	1.8E-02
MF	GO: 0008134	Transcription factor binding	18	4.19	2.5E-04

GO – Gene Ontology; uDEGs – upregulated differentially expressed genes; dDEGs – downregulated differentially expressed genes; LUAD – lung adenocarcinoma, BP – biological process, CC – cellular component MF – molecular function.

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 6. KEGG pathway analysis of uDEGs and dDEGs.

ID	KEGG_PATHWAY description	Count	%	*P*-value
uDEGs				
hsa04115	p53 signaling pathway	8	5.48	1.6E-06
hsa04110	Cell cycle	8	5.48	9.1E-05
hsa04114	Oocyte meiosis	5	3.42	1.6E-02
hsa04914	Progesterone-mediated oocyte maturation	4	2.74	4.0E-02
hsa05219	Bladder cancer	3	2.05	4.9E-02
dDEGs				
hsa05200	Pathways in cancer	23	5.35	6.4E-04
hsa04024	cAMP signaling pathway	15	3.49	7.2E-04
hsa04080	Neuroactive ligand-receptor interaction	15	3.49	1.5E-02
hsa04550	Signaling pathways regulating pluripotency of stem cells	12	2.79	1.1E-03
hsa04514	Cell adhesion molecules (CAMs)	12	2.79	1.3E-03

KEGG – Kyoto Encyclopedia of Genes and Genomes; uDEGs – upregulated differentially expressed genes; dDEGs – downregulated differentially expressed genes.

Recognition of hub genes

To better define hub genes in LUAD, 3 conditions were set for hub genes: 1) the degree score of DEGs tested by cytohHubba was in the top 15; 2) DEGs were in the 3 modules obtained by MCODE; and 3) DEGs were associated with the survival rate of the patients in the survival curve (Figure 3A–3H). We used the top 15 genes identified by cytoHubba and the common genes in the 3 modules as candidate genes in the PPI network and then verified whether candidate genes influenced survival, as shown by the Kaplan-Meier chart. The obtained hub genes, including PECAM1 (F value=0.308), CDK1 (F value=5.91), MKI67 (F value=6.38), SPP1 (F value=1.31), TOP2A (F value=2.88), CHEK1 (F value=7.27), CCNB1 (F value=8.88), and RRM2 (F value=9.01), had altered expression, as shown by violin charts according to the pathological stage of the patient using GEPIA (Figure 3I–3P). Among these genes, PECAM1 was downregulated, while others were upregulated in LUAD compared to normal samples.

Verification of hub genes by TCGA database analysis and RT-qPCR

The mRNA expression level of hub genes was verified based on TCGA database analysis, and it was revealed that the mRNA expression levels of CDK1, MKI67, SPP1, TOP2A, CHEK1, CCNB1, and RRM2 in LUAD samples were statistically higher than those in non-lung cancer samples (Figure 4A–4H). This is consistent with previous bioinformatics research.

To determine whether hub genes identified in gene chip analysis can be non-selectively used to identify LUAD patients in clinical practice, the mRNA expression discrepancy of selected hub genes in LUAD samples and normal lung cancer tissues was verified by RT-qPCR examination (Figure 4I–4P). These results showed that CDK1, MKI67, SPP1, TOP2A, CHEK1, CCNB1, and RRM2 mRNA expression levels were significantly upregulated and PECAM1 was significantly downregulated, which was consistent with the gene chip analysis results. The results of RT-qPCR confirmed that the mRNA expression of selected hub genes in human LUAD tissues was higher than that in the control tissues, suggesting that these 8 hub genes might be new genetic markers for LUAD patients.

Discussion

Previous preclinical and clinical studies revealed a partial understanding of the underlying genetic mechanism of LUAD; nevertheless, the incidence and mortality rates of LUAD are continually increasing [5]. The rapid advancement of high-throughput technology has led to the identification of massive cohorts of applicable biomarkers, such as microRNAs (miRNAs) and long non-coding RNAs (lncRNAs), that can be used for early diagnosis, prognosis, and treatment decision making for many diseases, including LUAD [19]. In the current research, 3 gene expression profiling sets were subjected to integrated analysis, and 577 DEGs were identified, including 146 uDEGs and 431 dDEGs. Functional annotation of uDEGs and dDEGs was divided into 3 groups according to GO terminology (BP, CC, and MF), and KEGG pathway enrichment of uDEGs and dDEGs 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]

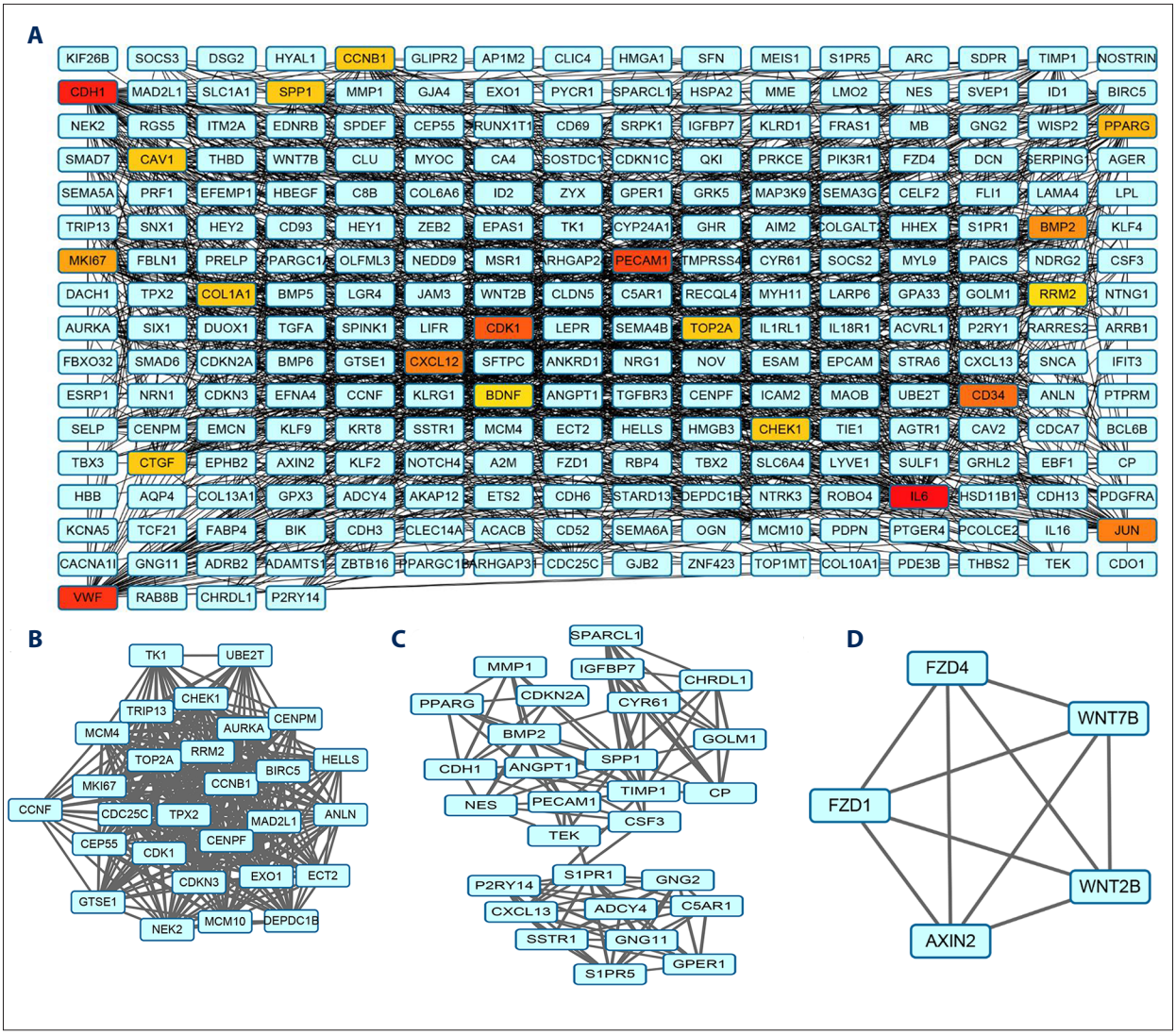

Figure 2. Construction of the protein-protein interaction (PPI) network complex with the STRING database and acquisition of 3 significant modules on the basis of the topological structure of the PPI network of the common uDEGs and dDEGs. (**A**) PPI network of uDEGs and dDEGs in LUAD tissues compared with corresponding normal adjacent tissues; (**B**) module 1, cluster score=25.407, 28 nodes and 343 edges; (**C**) module 2, cluster score=8.741, 28 nodes and 118 edges; (**D**) module 3, cluster score=5, 5 nodes and 10 edges. The blue squares indicate a node degree less than 30. uDEGs – upregulated differentially expressed genes; dDEGs – downregulated differentially expressed genes; LUAD – lung adenocarcinoma.

analyzed by the DAVID database. In line with KEGG pathway enrichment analysis, the uDEGs were mainly implicated in the pathways in cancer, cAMP signaling pathway, and neuroactive ligand-receptor interaction signaling pathway; the dDEGs were significantly involved in the p53 signaling pathway, cell cycle, and oocyte meiosis signaling pathway. According to previous studies [20–22], DEGs identified from LUAD are enriched in signaling pathways, such as the p53 signaling pathway, cell cycle, oocyte meiosis, pathways in cancer, cAMP signaling pathway, and neuroactive ligand-receptor interaction, which are related pathways involved in the advancement of malignant tumors.

A PPI network of uDEGs and dDEGs was constructed, and the top 15 candidate genes and the most important modules were screened based on the topological structures of the network. Through comprehensive bioinformatics analysis, 8 hub genes, including PECAM1, CDK1, MKI67, SPP1, TOP2A, CHEK1, CCNB1, and RRM2, were identified to have high connectivity and could distinguish the stage of LUAD from both benign lung disease and normal tissue. At present, the treatment options for LUAD include surgery, chemotherapy, radiotherapy and targeted drug therapy, which improve the survival of patients with mild effects [23].

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 7. KEGG pathway analysis of 3 key module in PPI network complex.

	ID	KEGG_PATHWAY	Count	%	P-value	Genes
Module 1	hsa04110	Cell cycle	6	21.43	1.3E-06	CCNB1, CDK1, MAD2L1, CHEK1, CDC25C, MCM4
	hsa04115	p53 signaling pathway	5	17.86	3.8E-06	CCNB1, CDK1, RRM2, CHEK1, GTSE1
	hsa04114	Oocyte meiosis	5	17.86	2.9E-05	CCNB1, CDK1, MAD2L1, AURKA, CDC25C
Module 2	hsa05200	Pathways in cancer	8	0.19	1.1E-04	ADCY4, BMP2, CDKN2A, PPARG, CDH1, GNG2, GNG11, MMP1
	hsa04151	PI3K-Akt signaling pathway	6	0.14	3.2E-03	CSF3, TEK, ANGPT1, GNG2, GNG11, SPP1
	hsa05219	Bladder cancer	3	0.07	6.8E-03	CDKN2A, CDH1, MMP1
Module 3	hsa05217	Basal cell carcinoma	5	100.00	3.4E-09	WNT7B, FZD1, AXIN2, FZD4, WNT2B
	hsa04310	Wnt signaling pathway	5	100.00	1.6E-07	WNT7B, FZD1, AXIN2, FZD4, WNT2B
	hsa04550	Signaling pathways regulating pluripotency of stem cells	5	100.00	1.6E-07	WNT7B, FZD1, AXIN2, FZD4, WNT2B

KEGG – Kyoto Encyclopedia of Genes and Genomes; PPI – protein–protein interaction

Platelet and endothelial cell adhesion molecule 1 (PECAM-1) is commonly detected in vascular endothelial cells and is involved in various physiological processes, including platelet aggregation, angiogenesis, and protection of the endothelium from endotoxin stress [24]. PECAM-1 is mediated by regulation of the tumor microenvironment (TME) and tumor cell proliferation, which is related to advanced tumor metastasis progression [25], and has a distinct differential expression in hepatocellular carcinoma (HCC) [26]. Cyclin-dependent kinase-1 (CDK1), a critical regulator of the G2/M checkpoint, can directly phosphorylate HIF-1α, upregulate HIF-1α transcription, and increase the invasion and migration of tumor cells [27]. Research has revealed that the nuclear/cytoplasmic expression ratio of CDK1 was determined as a self-sufficient prognostic reference point of colorectal cancer with histochemical staining [28]. CDK1, which is an effective prognostic indicator in patients with pancreatic ductal adenocarcinoma, could promote tumor progression [29]. Marker of proliferation Ki-67 (MKI67), a peripheral part of the mitotic chromosome, keeps the chromosome from collapsing into a single chromatin mass after the disintegration of the nuclear membrane; consequently, chromosomes can independently mobilize and effectively interact with mitotic spindles. The high expression of MKI67 mRNA is actually correlated with the advanced stage of non-muscle invasive bladder cancer [30]. Ki-67 functions in the proliferation of malignant cells at the front of infiltrating tumors, which is highly correlated with the invasiveness and poor clinical outcomes of triple-negative breast cancer [31], and has great potential as a biomarker of oral squamous cell carcinoma (OSCC) [32]. Decreased expression of secreted phosphoprotein 1 (SPP1), associated with EGFR mutation, is interconnected with the improvement of the overall survival rate and recurrence-free survival rate of LUAD [33]. SPP1 in hydrothorax can be applied in the detection of malignant pleural effusion and judgment of patients with NSCLC [34]. Additionally, the overexpression of SPP1 in OSCC and HCC is related to the occurrence and progression of tumors [35,36].

The overexpression of DNA topoisomerase II alpha (TOP2A) has a negative effect on the prognosis of breast cancer patients; hence, TOP2A-targeted therapy might be beneficial to the treatment and prognosis prediction of breast cancer [37]. Patients with TOP2A-positive breast cancer are more sensitive to anthracyclines than TOP2A-negative patients [38]. TOP2A is highly expressed in more advanced uterine leiomyosarcoma with a high mitotic index but not in nonmalignant uterine diseases [39]. It has been previously reported that the protein encoded by the checkpoint kinase-1 (CHEK1) gene, which is essential for cell cycle arrest in the presence of DNA damage or unreplicated chromatid, belongs to a conserved serine/threonine kinase family. It is a meaningful oncogene associated with poor prognosis and is overexpressed in both esophageal squamous cell carcinoma and HCC [40,41]. The expression of CHEK1 can be applied not only as a prognostic indicator but also as a marker for the selection of CHEK1 inhibitors in patients with acute myeloid leukemia [42]. The high expression rate of CHEK1 in breast cancer was found to be 61%, and high expression was related to tumor size, triple-negative subtype, basal phenotype, epithelial-stromal transformation, dysfunction of the DNA homologous repair pathway and poor

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

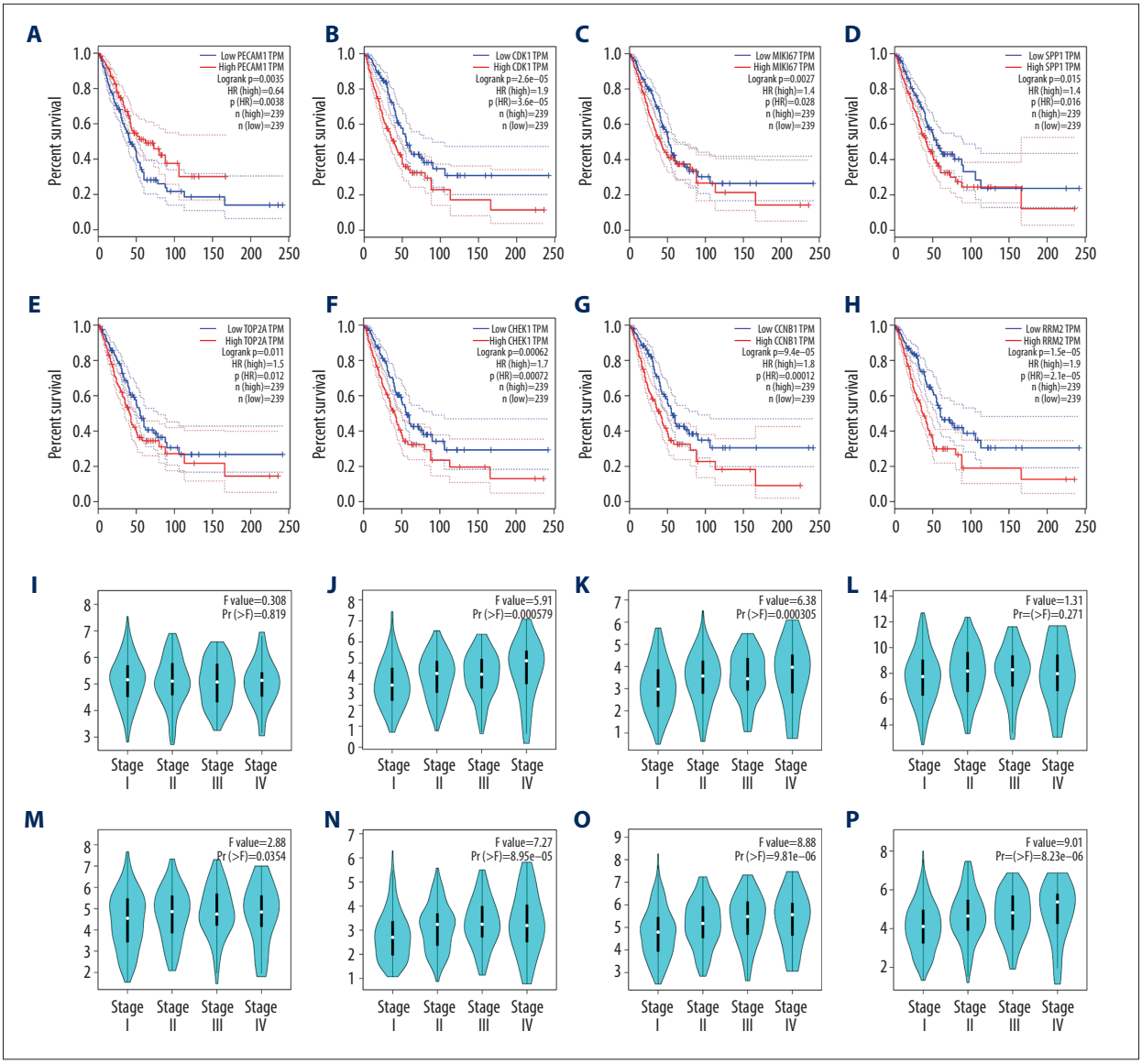

Figure 3. Kaplan-Meier survival curves indicated the prognostic value of hub genes in the form of overall survival (months) in human LUAD patients; (**A**) PECAM1, (**B**) CDK1, (**C**) MKI67, (**D**) SPP1, (**E**) TOP2A, (**F**) CHEK1, (**G**) CCNB1, (**H**) RRM2. Violin plots demonstrated an association between the pathological stage of human LUAD and the expression of 8 hub genes based on TCGA data analysis using GEPIA; (**I**) PECAM1, (**J**) CDK1, (**K**) MKI67, (**L**) SPP1, (**M**) TOP2A, (**N**) CHEK1, (**O**) CCNB1, (**P**) RRM2. LUAD – lung adenocarcinoma; TCGA – The Cancer Genome Atlas.

prognosis [43]. Abnormal expression of cyclin B1 (CCNB1) has been observed in a variety of tumors, including pituitary adenomas, and increases moderately with the escalation of invasiveness [18]. CCNB1 is a pivotal factor in the proliferation of hepatocellular carcinoma cells, and FOXM1 binds the CCNB1 promoter region to regulate the transcription of CCNB1 [44]. CCNB1 is an influential biomarker for the prognosis of estrogen receptor (ER)+ breast cancer; targeting CCNB1 can prevent or even reverse resistance to hormone therapy and facilitate personalized treatment [45]. The expression of ribonucleotide reductase regulatory subunit M2 (RRM2) correlates with clinical stage and is increased significantly in neuroblastoma compared to adjacent benign tissues [46]. HPVE7 facilitates to RRM2 upregulation and promotes the occurrence of cervical carcinoma through angiogenesis induced by Ros-ERK1/2-HIF-1, and α-VEGF. The overexpression of RRM2 is intimately related to the occurrence and progression of human ovarian carcinoma and breast cancer [47,48].

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

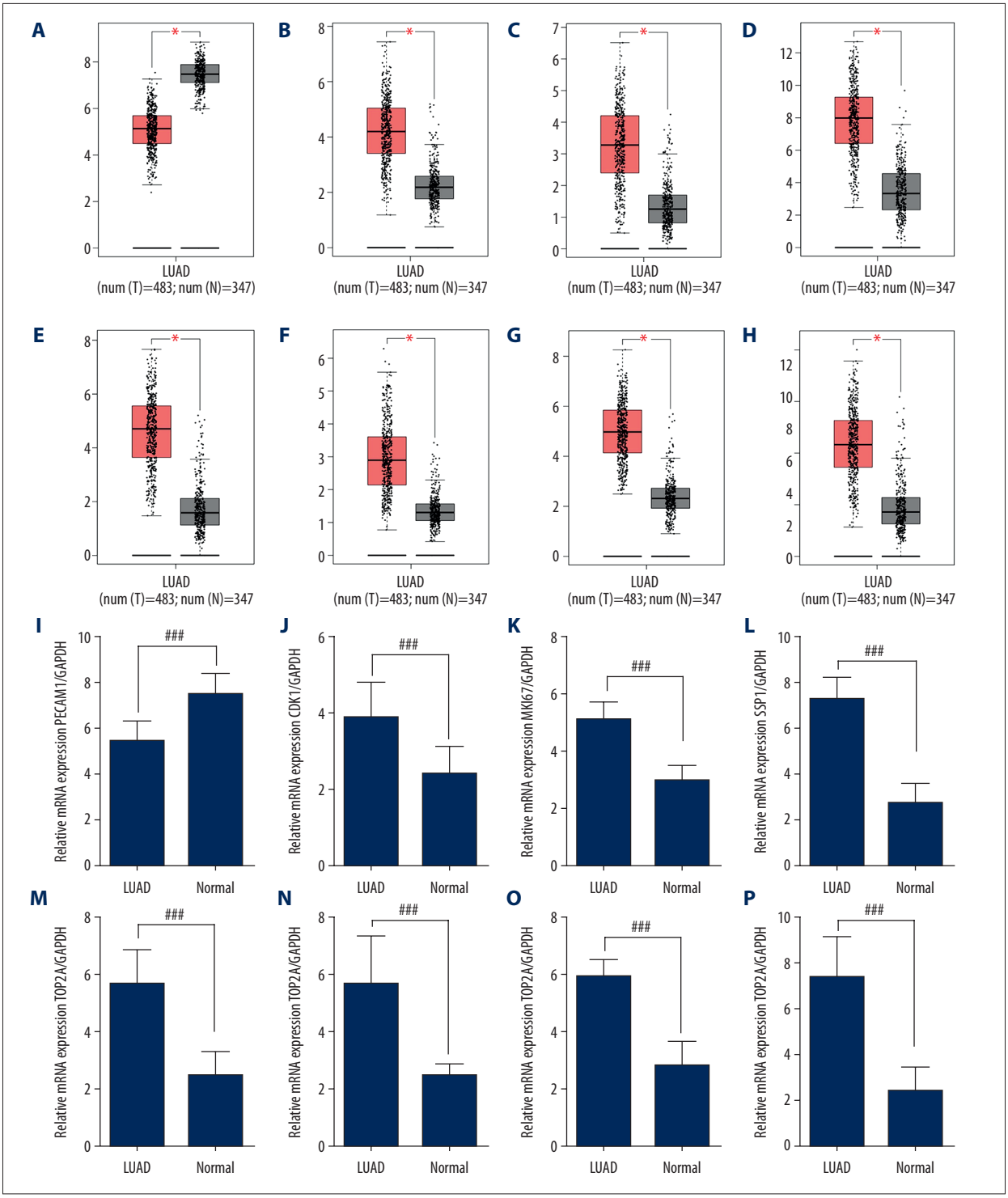

Figure 4. DEGs were validated in LUAD and normal lung tissue. The gene expression of (**A**) PECAM1, (**B**) CDK1, (**C**) MKI67, (**D**) SPP1, (**E**) TOP2A, (**F**) CHEK1, (**G**) CCNB1, (**H**) RRM2 in 483 LUAD patients and 347 normal tissues was evaluated by GEPIA. The relative mRNA expression levels of (**I**) PECAM1, (**J**) CDK1, (**K**) MKI67, (**L**) SPP1, (**M**) TOP2A, (**N**) CHEK1, (**O**) CCNB1, and (**P**) RRM2 between 28 LUAD patients and 8 normal lung tissues were determined by RT-qPCR. All data are shown as the mean±SD, ### $P<0.001$ versus the normal group. DEGs – differentially expressed genes; LUAD – lung adenocarcinoma; GEPIA – Gene Expression Profiling Interactive Analysis; mRNA – messenger RNA; RT-qPCR – real-time quantitative polymerase chain reaction; SD – standard deviation.

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

8 hub genes in LUAD were identified from hundreds of candidate DEGs. The hub genes were markedly correlated with the overall survival of LUAD patients, and their expression was histologically validated. Studies have shown that dysregulated gene expression can lead to the occurrence of tumors. This study improves our understanding of the molecular determinants of LUAD progression and the expression reliability of biomarkers and provides new biomarkers that may potentially support the diagnosis and precise identification of treatment targets for LUAD. However, *in vivo* and *in vitro* trials and multicenter randomized controlled clinical trials are still needed before these biomarkers can be accurately applied in clinical laboratory diagnostics.

References:

1. Bray F, Ferlay J, Soerjomataram I et al: Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. Cancer J Clin, 2018; 68: 394–424
2. Travis WD, Brambilla E, Nicholson AG et al: The 2015 World Health Organization classification of lung tumors: Impact of genetic, clinical and radiologic advances since the 2004 classification. J Thorac Oncol, 2015; 10: 1243–60
3. Travis WD, Brambilla E, Noguchi M et al: International Association for the Study of Lung Cancer/American Thoracic Society/European Respiratory Society international multidisciplinary classification of lung adenocarcinoma. J Thorac Oncol, 2011; 6: 244–85
4. Nakamura H, Takagi M: Clinical impact of the new IASLC/ATS/ERS lung adenocarcinoma classification for chest surgeons. Surg Today, 2015; 45: 1341–51
5. Behera M, Owonikoko TK, Gal AA et al: Lung adenocarcinoma staging using the 2011 IASLC/ATS/ERS classification: A pooled analysis of adenocarcinoma *in situ* and minimally invasive adenocarcinoma. Clin Lung Cancer, 2016; 17: e57–64
6. Macheleidt IF, Dalvi PS, Lim SY et al: Preclinical studies reveal that LSD1 inhibition results in tumor growth arrest in lung adenocarcinoma independently of driver mutations. Mol Oncol, 2018; 12: 1965–79
7. Meng F, Zhang L, Ren Y, Ma Q: The genomic alterations of lung adenocarcinoma and lung squamous cell carcinoma can explain the differences of their overall survival rates. J Cell Physiol, 2019; 234: 10918–25
8. Inamura K: Clinicopathological characteristics and mutations driving development of early lung adenocarcinoma: tumor initiation and progression. Int J Mol Sci, 2018; 19: pii: E1259
9. Kadota K, Yeh YC, D'Angelo SP et al: Associations between mutations and histologic patterns of mucin in lung adenocarcinoma: Invasive mucinous pattern and extracellular mucin are associated with KRAS mutation. Am J Surg Pathol, 2014; 38: 1118–27
10. Girard L, Rodriguez-Canales J, Behrens C et al: An expression signature as an aid to the histologic classification of non-small cell lung cancer. Clin Cancer Res, 2016; 22: 4880–89
11. Peng Z, Wang J, Shan B et al: Genome-wide analyses of long noncoding RNA expression profiles in lung adenocarcinoma. Sci Rep, 2017; 7: 15331
12. Xu L, Lu C, Huang Y et al: SPINK1 promotes cell growth and metastasis of lung adenocarcinoma and acts as a novel prognostic biomarker. BMB Rep, 2018; 51: 648–53
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. Huang da W, Sherman BT, Lempicki RA: Bioinformatics enrichment tools: Paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res, 2009; 37: 1–13
15. Szklarczyk D, Morris JH, Cook H et al: The STRING database in 2017: Quality-controlled protein-protein association networks, made broadly accessible. Nucleic Acids Res, 2017; 45: D362–68
16. Shannon P, Markiel A, Ozier O et al: Cytoscape: A software environment for integrated models of biomolecular interaction networks. Genome Res, 2003; 13: 2498–504
17. 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
18. Zhao P, Zhang P, Hu W et al: Upregulation of cyclin B1 plays potential roles in the invasiveness of pituitary adenomas. J Clin Neurosci, 2017; 43: 267–73
19. Liu XX, Yang YE, Liu X et al: A two-circular RNA signature as a noninvasive diagnostic biomarker for lung adenocarcinoma. J Transl Med, 2019; 17: 50
20. Ning Y, Wang C, Liu X et al: CK2-mediated CCDC106 phosphorylation is required for p53 degradation in cancer progression. J Exp Clin Cancer Res, 2019; 38: 131
21. Batra A, Winquist E: Emerging cell cycle inhibitors for treating metastatic castration-resistant prostate cancer. Expert Opin Emerg Drugs, 2018; 23: 271–82
22. Ma R, Zhai X, Zhu X, Zhang L: LINC01585 functions as a regulator of gene expression by the CAMP/CREB signaling pathway in breast cancer. Gene, 2019; 684: 139–48
23. Zappa C, Mousa SA: Non-small cell lung cancer: Current treatment and future advances. Transl Lung Cancer Res, 2016; 5: 288–300
24. Abraham V, Cao G, Parambath A et al: Involvement of TIMP-1 in PECAM-1-mediated tumor dissemination. Int J Oncol, 2018; 53: 488–502
25. DeLisser H, Liu Y, Desprez PY et al: Vascular endothelial platelet endothelial cell adhesion molecule 1 (PECAM-1) regulates advanced metastatic progression. Proc Natl Acad Sci USA, 2010; 107: 18616–21
26. Sarathi A, Palaniappan A: Novel significant stage-specific differentially expressed genes in hepatocellular carcinoma. BMC Cancer, 2019; 19: 663
27. Warfel NA, Dolloff NG, Dicker DT et al: CDK1 stabilizes HIF-1alpha via direct phosphorylation of Ser668 to promote tumor growth. Cell Cycle, 2013; 12: 3689–701
28. Sung WW, Lin YM, Wu PR et al: High nuclear/cytoplasmic ratio of Cdk1 expression predicts poor prognosis in colorectal cancer patients. BMC Cancer, 2014; 14: 951
29. Piao J, Zhu L, Sun J et al: High expression of CDK1 and BUB1 predicts poor prognosis of pancreatic ductal adenocarcinoma. Gene, 2019; 701: 15–22
30. Breyer J, Wirtz RM, Laible M et al: ESR1, ERBB2, and Ki67 mRNA expression predicts stage and grade of non-muscle-invasive bladder carcinoma (NMIBC). Virchows Arch, 2016; 469: 547–52
31. Borges US, Costa-Silva DR, da Silva-Sampaio JP et al: A comparative study of Ki-67 antigen expression between luminal A and triple-negative subtypes of breast cancer. Med Oncol, 2017; 34: 156
32. Verma R, Singh A, Jaiswal R et al: Association of Ki-67 antigen and p53 protein at invasive tumor front of oral squamous cell carcinoma. Indian J Pathol Microbiol, 2014; 57: 553–57
33. Shen XY, Liu XP, Song CK et al: Genome-wide analysis reveals alcohol dehydrogenase 1C and secreted phosphoprotein 1 for prognostic biomarkers in lung adenocarcinoma. J Cell Physiol, 2019; 234: 22311–20
34. Zhang H, Liu HB, Yuan DM et al: Prognostic value of secreted phosphoprotein-1 in pleural effusion associated with non-small cell lung cancer. BMC Cancer, 2014; 14: 280
35. Huang CF, Yu GT, Wang WM et al: Prognostic and predictive values of SPP1, PAI and caveolin-1 in patients with oral squamous cell carcinoma. Int J Clin Exp Pathol, 2014; 7: 6032–39
36. Zheng Y, Huang Q, Ding Z et al: Genome-wide DNA methylation analysis identifies candidate epigenetic markers and drivers of hepatocellular carcinoma. Brief Bioinform, 2018; 19: 101–8
37. Sahin S, Isik Gonul I, Cakir A et al: Clinicopathological significance of the proliferation markers Ki67, RacGAP1, and topoisomerase 2 alpha in breast cancer. Int J Surg Pathol, 2016; 24: 607–13
38. Nikolenyi A, Uhercsak G, Csenki M et al: Tumour topoisomerase II alpha protein expression and outcome after adjuvant dose-dense anthracycline-based chemotherapy. Pathol Oncol Res, 2012; 18: 61–68

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

39. Baiocchi G, Poliseli FL, De Brot L et al: TOP2A copy number and TOP2A expression in uterine benign smooth muscle tumours and leiomyosarcoma. J Clin Pathol, 2016; 69: 884–89
40. Xie Y, Wei RR, Huang GL et al: Checkpoint kinase 1 is negatively regulated by miR-497 in hepatocellular carcinoma. Med Oncol, 2014; 31: 844
41. Li J, Tang Y, Huang L et al: Genetic variants in CHEK1 gene are associated with the prognosis of thoracic esophageal squamous cell carcinoma patients treated with radical resection. J Huazhong Univ Sci Technolog Med Sci, 2016; 36: 828–33
42. David L, Fernandez-Vidal A, Bertoli S et al: CHK1 as a therapeutic target to bypass chemoresistance in AML. Sci Signal, 2016; 9: ra90
43. Ebili HO, Iyawe VO, Adeleke KR et al: Checkpoint kinase 1 expression predicts poor prognosis in Nigerian breast cancer patients. Mol Diagn Ther, 2018; 22: 79–90
44. Chai N, Xie HH, Yin JP et al: FOXM1 promotes proliferation in human hepatocellular carcinoma cells by transcriptional activation of CCNB1. Biochem Biophys Res Commun, 2018; 500: 924–29
45. Ding K, Li W, Zou Z et al: CCNB1 is a prognostic biomarker for ER+ breast cancer. Med Hypotheses, 2014; 83: 359–64
46. Li J, Pang J, Liu Y et al: Suppression of RRM2 inhibits cell proliferation, causes cell cycle arrest and promotes the apoptosis of human neuroblastoma cells and in human neuroblastoma RRM2 is suppressed following chemotherapy. Oncol Rep, 2018; 40: 355–60
47. Wang LM, Lu FF, Zhang SY et al: Overexpression of catalytic subunit M2 in patients with ovarian cancer. Chin Med J (Engl), 2012; 125: 2151–56
48. Bell R, Barraclough R, Vasieva: Gene expression meta-analysis of potential metastatic breast cancer markers. Curr Mol Med, 2017; 17: 200–10

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