Original Article

The identification of hub biomarkers and pathways in lung cancer and prognostic evaluation

Yi Yin[1#], Dong Li[2#], Muqun He[1], Jianfeng Wang[1]

[1]Department of Medical Oncology, Fujian Medical University Cancer Hospital, Fujian Cancer Hospital, Fuzhou, China; [2]Cancer Institute, Fujian Medical University Cancer Hospital, Fujian Cancer Hospital, Fuzhou, China

Contributions: (I) Conception and design: Y Yin, D Li; (II) Administrative support: Y Yin, D Li; (III) Provision of study materials or patients: All authors; (IV) Collection and assembly of data: All authors; (V) Data analysis and interpretation: All authors; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

[#]These authors equally contributed to this work.

Correspondence to: Dong Li. Cancer Institute, Fujian Medical University Cancer Hospital, Fujian Cancer Hospital, Fuzhou 350014, China. Email: jerryldong@163.com.

Background: Lung cancer is the most frequently diagnosed malignant tumor and the highest mortality worldwide, and can be divided into two differential histologic subtypes, non-small cell lung cancer (NSCLC) and small cell lung cancer (SCLC). However, there are significant differences in diagnosis and prognosis between NSCLC and SCLC. We aimed to identify hub differentially expressed genes (DEGs) and pathways for diagnostic and prognostic prediction in NSCLC and SCLC.

Methods: Three expression profiles (GSE43346, GSE40275 and GSE18842) were obtained through GEO2R tools from Gene Expression Omnibus (GEO) database. The Database for Annotation, Visualization, and Integrated Discovery (DAVID) was used to investigate functional enrichment of the DEGs. The protein–protein interaction network was constructed by the Search Tool for the Retrieval of Interacting Genes (STRING) and Cytoscape. Kaplan-Meier analysis was performed using Kaplan-Meier plotter and Gene Expression Profiling Interactive Analysis (GEPIA).

Results: We have identified 84 overlap DEGs that may play an important role in SCLC & NSCLC. However, we also found some genes were only significantly differential expressed in SCLC or NSCLC. There were 87 DEGs unique to SCLC tissues and 28 DEGs unique to NSCLC ones. Functional analysis results indicated that these DEGs had different biological functions and were significantly enriched in different pathways. Hub DEGs were identified via protein-protein interaction network and cross-validated using Kaplan-Meier plotter and GEPIA. The 14 hub DEGs were highly correlated with the overall survival of NSCLC. Kyoto Encyclopedia of Genes and Genome (KEGG) re-analysis of 14 hub DEGs showed that *RRM2*, *CHEK1* and *SERPINB5* enriched in the p53 signaling pathway, *RRM2* and *TYMS* enriched in pyrimidine metabolism pathway maybe play a key role in SCLC&NSCLC and were significantly related to overall survival in patients with NSCLC.

Conclusions: *RRM2*, *CHEK1*, *TYMS* and *SERPINB5*, which are mainly enriched in the p53 signaling pathway and pyrimidine metabolism pathway, were significantly associated with the overall survival of NSCLC patients. These genes could serve as potential prognostic markers in NSCLC and therapeutic target in lung cancer for personalized oncology.

Keywords: Non-small cell lung cancer (NSCLC); small cell lung cancer (SCLC); differentially expressed genes (DEGs); prognosis; biomarkers

Submitted Feb 03, 2022. Accepted for publication Jun 09, 2022.
doi: 10.21037/tcr-22-245
View this article at: https://dx.doi.org/10.21037/tcr-22-245

Introduction

Lung cancer is the most commonly diagnosed cancer and the leading cause of cancer death in 2020, which account for approximately one-tenth cancers diagnosed and one in 5 deaths. There are estimated 2.2 million new cancer cases and 1.8 million deaths in the world (1). Lung cancer can be divided into two histologic subtypes, including non-small cell lung cancer (NSCLC, about 85% of all lung cancer) and small cell lung cancer (SCLC, about 15%). NSCLC is made up of three major histologic subtypes: large-cell lung, two major pathological types adenocarcinoma (LUAD) and lung squamous-cell carcinoma (LUSC) (2). The treatments of lung cancer mainly included surgery, chemotherapy, radiotherapy and immunotherapy. As we know, there are different treatment options and prognosis for these subtypes. Therapeutic advances contributed to survival gains. With the progress on targeted therapies and immunotherapies, the 2-years survival rate for NSCLC increased from 34% during 2009–2010 to 42% during 2015–2016 in the United States, but SCLC survival remained low and steady at 14% to 15% (3-5). There are significant differences in survival between NSCLC and SCLC. Therefore, it is important to assess the difference of lung cancer to detect prognostic markers which are likely to affect future treatments and prognosis.

Gene expression profile array and bioinformatics analysis have been applied to study potential clinical biomarkers and molecular mechanisms. Some key differentially expressed genes (DEGs) have been identified by integrated bioinformatics analysis and are significantly associated with the treatments and prognosis in some cancers (6-11). Relevant biomarkers have been used as valuable tools in the prognosis and prediction of therapy response, significantly influencing the clinical course and outcome of the disease (12). In our study, we selected gene expression profile from the Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/) and analyzed the differential expression genes between differential lung cancer tissues and normal tissues to explore the hub pathways and key genes. We applied integrated bioinformatics methods to further investigate potential gene biomarkers and molecular mechanisms in lung cancer. Novel prognostic biomarkers will further inform clinical therapeutic decision-making. We present the following article in accordance with the REMARK reporting checklist (available at https://tcr.amegroups.com/article/view/10.21037/tcr-22-245/rc).

Methods

Microarray data and identification of DEGs

The gene expression profiles of lung cancer (GSE43346, GSE40275 and GSE18842) were downloaded from Gene Expression Omnibus public database (13-16). The above three gene expression profiles were performed by the Affymetrix Human Gene Expression Array. GSE43346 contains 42 normal tissue samples and 23 SCLC samples. GSE40275 includes 43 normal tissue samples, 19 SCLC samples and 16 NSCLC samples. GSE18842 includes 45 normal tissue samples and 46 NSCLC samples.

GEO2R online tool and Venn diagram software were applied to screen overlap DEGs from above three gene expression profiles. $|logFC| > 2$ and adjusted P value <0.05 were used as cutoff criteria by GEO2R online tools. The $logFC < -2$ or $logFC > 2$ were considered down-regulated or up-regulated genes, respectively.

Gene Ontology (GO) and Kyoto Encyclopedia of Gene and Genome (KEGG) functional enrichment analysis

The GO and KEGG were performed using Database for Annotation, Visualization and Integrated Discovery v6.8 (DAVID, https://david.ncifcrf.gov/) (17,18). The DAVID v6.8, an online set of functional annotation tools, was used to analyze biological process, cellular component, molecular function and pathways for DEGs. GO terms and KEGG pathways with P value <0.05 were considered statistical significant.

Protein-protein interaction network construction and module analysis

Protein-protein interaction network was obtained through the Search Tool for the Retrieval of Interacting Genes database (STRING, http://string-db.org) (19). The plugin MCODE of Cytoscape was applied to detect significant modules in the protein-protein interaction network (https://cytoscape.org/). The cutoff criteria were set with degree cutoff =2, node score cutoff =0.2, maximum depth =100, and k-core =2 (20). The interactions of module DEGs in the PPI networks were analysed using plugin cytoHubba (21). The specific connectivity genes that overlapped in the

Table 1 All overlap differentially expressed genes in SCLC & NSCLC, unique to SCLC or NSCLC tissues compared with normal tissues

Cancer	DEGs	No.	Genes name
SCLC & NSCLC	Up-regulated	54	*TPX2, CCNB1, HMGB3, DSP, GINS1, ANLN, UCHL1, EZH2, CHEK1, KIF11, CDC6, AURKA, KIF14, KIF4A, TYMS, CDCA7, MELK, NDC80, RFC4, CCNA2, BUB1, PBK, NUF2, PRR11, PTTG1, MMP12, UBE2T, ECT2, KIF23, DEPDC1, GGH, ASPM, ATAD2, BRIP1, UBE2C, CCNB2, PRC1, CEP55, RRM2, TOP2A, HELLS, CCNE2, BUB1B, RAD51AP1, MKI67, DTL, EXO1, KIF20A, KIAA0101, TTK, CDKN3, NCAPG, CENPF, NUSAP1*
	Down-regulated	30	*CHRDL1, EGR1, FHL1, AOC3, ZFP36, ZBTB16, EDNRB, ARRB1, SH2D3C, TNS1, C7, GPM6A, GADD45B, MFAP4, PTGDS, SDPR, NR4A1, FOS, TPSB2, AQP1, ADH1B, FABP4, FAM107A, PGM5, GPX3, FXYD1, FOSB, VIPR1, CFD, FBLN5*
Only SCLC	Up-regulated	70	*DONSON, DEPDC1B, RAB3IP, PCSK1, RFC3, BRCA1, ACTL6A, ISL1, PLK1, NUP62CL, MYEF2, ASF1B, CDH2, KIFC1, CBFA2T2, ZNF711, TMPO, DCX, RNF182, FANCA, GDAP1, PCNA, TPH1, SLC36A4, LOC643201, GNG4, H2AFY2, CCDC14, MIAT, STMN1, BEST3, TPD52, INTS7, NOL4, STXBP5L, TUBB2B, CEP78, GPR137C, PGAP1, HOXD10, NELL1, RAB3B, PMAIP1, DDC, MSH2, RRM1, SCG3, ESCO2, KIF1A, CBX3, MEST, MPHOSPH9, NRCAM, CDKN2C, GRP, RIMS2, MCM6, AGPAT5, CENPI, SCN8A, FZD3, GMNN, SSX2IP, SLCO5A1, ASCL1, CDKN2A, RFC5, USP1, LOC81691, ST18*
	Down-regulated	17	*LAMA2, HLA-DRB4, DPP4, MAOA, C3, RBPMS, ADAMTS1, AQP3, CD74, SYNE1, ANPEP, CCL21, RNASE1, MYL9, SNTB1, KIAA1462, DCN*
Only NSCLC	Up-regulated	20	*SERPINB5, AKR1B10, GJB6, ARNTL2, SLC7A11, SPRR1A, PLOD2, KRT6A, GCLC, FAP, FAM83B, NQO1, PSAT1, S100A2, SULF1, CLDN1, GPR87, CP, STEAP1, RPL39L*
	Down-regulated	8	*GRK5, KLF2, ADARB1, CLDN5, DENND3, CCBE1, MFNG, SELENBP1*

SCLC, small cell lung cancer; NSCLC, non-small cell lung cancer; DEGs, differentially expressed genes.

NSCLC, SCLC or NSCLC & SCLC PPI networks, are defined as hub DEGs.

Validation of hub DEGs and survival analysis

The online Kaplan-Meier plotter database (https://kmplot.com/analysis/) and The GEPIA server (http://gepia.cancer-pku.cn/) were applied to assess the survival rate of patients with LUAD or LUSC (22,23). The Kaplan-Meier plotter database, which the tool of a meta-analysis-based discovery and validation of survival biomarkers, includes 54,000 genes on survival in 21 cancer types including 3,452 lung cancer patients. The criteria we selected were HR with 95% CI and log-rank P<0.05 as a threshold.

The GEPIA is an interactive web serve of analyzing tumor/normal differential expression analysis, correlation analysis and patient survival analysis. The GEPIA was used to get stage plots and further validate the expression of hub DEGs between LUAD, LUSC and normal lung tissues (P<0.05).

Ethical consideration

The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

Results

Differentially expressed genes in SCLC and NSCLC

We extracted 1,945 and 1,137 DEGs using the online GEO2R tool between SCLC and normal lung tissues from GSE43346 and GSE40275. The 1,014 and 1,367 DEGs were extracted from GSE18842 and GSE40275 in NSCLC tissues compared with normal tissues. We identified 84 overlap DEGs in SCLC & NSCLC tissues by online Venn diagram software, including 54 up-regulated and 30 down-regulated DEGs. The 70 up-regulated and 17 down-regulated DEGs were identified unique to SCLC. The 20 up-regulated and 8 down-regulated DEGs were identified unique to NSCLC (*Table 1* and *Figure 1*).

Functional and pathway enrichment analysis

The online DAVID was utilized for GO and KEGG enrichment analysis of overlap DEGs in lung cancer. The GO analysis includes biological processes, cellular components, and molecular functions. For biological

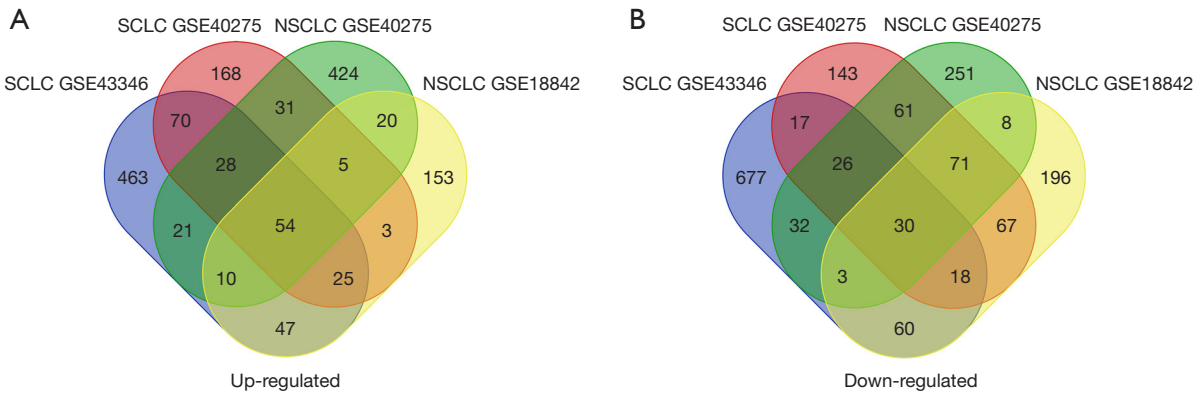

Figure 1 Venn diagrams of all screened differentially expressed genes identified from three gene expression profiles in SCLC & NSCLC, unique to SCLC or NSCLC. There are 54 up-regulated and 30 down-regulated DEGs in SCLC & NSCLC, 70 up-regulated and 17 down-regulated DEGs unique to SCLC, 20 up-regulated and 8 down-regulated DEGs unique to NSCLC were identified via online Venn diagram software, respectively. SCLC, small cell lung cancer; NSCLC, non-small cell lung cancer; DEGs, differentially expressed genes.

process analysis, 54 up-regulated and 30 down-regulated DEGs in SCLC & NSCLC were mainly related to mitotic nuclear division, cell division, mitotic cytokinesis, regulation of cell cycle, DNA replication and mitotic spindle organization, while the DEGs unique to SCLC were mainly involved in DNA damage response, detection of DNA damage, DNA replication, neuron migration, error-prone translesion synthesis, error-free translesion synthesis and negative regulation of cyclin-dependent protein serine/threonine kinase activity. The biological process analysis showed DEGs unique to NSCLC were associated with oxidation-reduction process, morphogenesis of an epithelium, aging, calcium-independent cell-cell adhesion via plasma membrane cell-adhesion molecules, positive regulation of vascular endothelial growth factor production and endothelial cell migration. Furthermore, cellular component analysis indicated that overlap DEGs in SCLC & NSCLC were located in the nucleus, nucleoplasm, midbody, spindle, cytoplasm and cytosol, while DEGs unique to SCLC in DNA replication factor C complex, chromatin, nuclear envelope, nuclear chromosome, telomeric region, transport vesicle membrane and nucleoplasm, and DEGs unique to NSCLC in extracellular space and extracellular exosome. Additionally, the results of molecular function analysis indicated that overlap DEGs in lung cancer were particularly enriched in protein binding, ATP binding, protein kinase binding, microtubule binding, chromatin binding and microtubule motor activity, while DEGs unique to SCLC in dinucleotide insertion or deletion binding, enzyme binding, mutLalpha complex binding, damaged DNA binding, DNA clamp loader activity and single-stranded DNA-dependent ATPase activity, and DEGs unique to NSCLC in structural molecule activity (*Table 2*).

The top 6 KEGG analysis results using DAVID software are shown in *Table 3*. The overlap DEGs in SCLC&NSCLC are mainly associated with cell cycle, the p53 signaling pathway, oocyte meiosis, progesterone-mediated oocyte maturation and HTLV-I infection pathways, and DEGs unique to SCLC enriched in mismatch repair, DNA replication, cell cycle, tryptophan metabolism, serotonergic synapse and nucleotide excision repair pathways. The DEGs unique to NSCLC were not enriched in any signaling pathway (all pathways P value >0.05).

Protein-protein interaction network and module analysis

The overlap DEGs in SCLC&NSCLC, unique to SCLC or NSCLC were used to construct the protein-protein interaction network using STRING and Cytoscape, respectively. A total of 84 DEGs in SCLC&NSCLC were imported into online STRING, which contained 84 nodes and 1,122 edges. The 2 important modules were identified using Cytoscape MCODE, which contains 47 hub genes and 5 hub genes, respectively (*Figure 2*). The 7 of 12 topological analysis methods using cytoHubba have identified 100% (47/47) hub DEGs that we have screened using plugin MCODE in the SCLC & NSCLC PPI networks. The remaining 5 methods have identified at least 25 of the 47 hub DEGs (Table S1).

Table 2 Gene ontology analysis of differentially expressed genes in SCLC & NSCLC, unique to SCLC or NSCLC

Cancer	Category	Term	Count	%	P value	FDR
SCLC & NSCLC	GOTERM_BP_DIRECT	GO:0007067~mitotic nuclear division	17	20.24	5.53E-14	2.09E-11
	GOTERM_BP_DIRECT	GO:0051301~cell division	19	22.62	6.21E-14	2.09E-11
	GOTERM_BP_DIRECT	GO:0000281~mitotic cytokinesis	6	7.14	2.66E-07	5.97E-05
	GOTERM_BP_DIRECT	GO:0051726~regulation of cell cycle	8	9.52	2.43E-06	4.09E-04
	GOTERM_BP_DIRECT	GO:0006260~DNA replication	8	9.52	1.06E-05	0.001430782
	GOTERM_BP_DIRECT	GO:0007052~mitotic spindle organization	5	5.95	1.31E-05	0.001476073
	GOTERM_CC_DIRECT	GO0005634~nucleus	54	64.29	1.21E-10	1.78E-08
	GOTERM_CC_DIRECT	GO0005654~nucleoplasm	36	42.86	3.51E-09	2.58E-07
	GOTERM_CC_DIRECT	GO0030496~midbody	10	11.90	7.12E-09	3.49E-07
	GOTERM_CC_DIRECT	GO0005819~spindle	9	10.71	7.78E-08	2.86E-06
	GOTERM_CC_DIRECT	GO0005737~cytoplasm	45	53.57	2.61E-06	7.67E-05
	GOTERM_CC_DIRECT	GO0005829~cytosol	34	40.48	3.40E-06	8.33E-05
	GOTERM_MF_DIRECT	GO:0005515~protein binding	64	76.19	1.20E-07	2.35E-05
	GOTERM_MF_DIRECT	GO:0005524~ATP binding	22	26.19	2.67E-06	2.61E-04
	GOTERM_MF_DIRECT	GO:0019901~protein kinase binding	10	11.90	5.77E-05	0.003748425
	GOTERM_MF_DIRECT	GO:0008017~microtubule binding	7	8.33	4.00E-04	0.016613263
	GOTERM_MF_DIRECT	GO:0003682~chromatin binding	9	10.71	4.39E-04	0.016613263
	GOTERM_MF_DIRECT	GO:0003777~microtubule motor activity	5	5.95	5.11E-04	0.016613263
Only SCLC	GOTERM_BP_DIRECT	GO:0042769~DNA damage response, detection of DNA damage	4	4.60	6.88E-04	0.34328918
	GOTERM_BP_DIRECT	GO:0006260~DNA replication	6	6.90	9.15E-04	0.34328918
	GOTERM_BP_DIRECT	GO:0001764~neuron migration	5	5.75	0.001660073	0.415018297
	GOTERM_BP_DIRECT	GO:0042276~error-prone translesion synthesis	3	3.45	0.003726164	0.558924545
	GOTERM_BP_DIRECT	GO:0070987~error-free translesion synthesis	3	3.45	0.003726164	0.558924545
	GOTERM_BP_DIRECT	GO:0045736~negative regulation of cyclin-dependent protein serine/threonine kinase activity	3	3.45	0.004986668	0.621663134
	GOTERM_CC_DIRECT	GO:0005663~DNA replication factor C complex	3	3.45	3.11E-04	0.046840782
	GOTERM_CC_DIRECT	GO:0000785~chromatin	5	5.75	7.61E-04	0.046840782
	GOTERM_CC_DIRECT	GO:0005635~nuclear envelope	6	6.90	8.41E-04	0.046840782
	GOTERM_CC_DIRECT	GO:0000784~nuclear chromosome, telomeric region	5	5.75	0.003071132	0.12821977
	GOTERM_CC_DIRECT	GO:0030658~transport vesicle membrane	3	3.45	0.013254179	0.376099937
	GOTERM_CC_DIRECT	GO:0005654~nucleoplasm	22	25.29	0.013512573	0.376099937
	GOTERM_MF_DIRECT	GO:0019899~enzyme binding	6	6.90	0.019926207	1

Table 2 (*continued*)

Table 2 (*continued*)

Cancer	Category	Term	Count	%	P value	FDR
	GOTERM_MF_DIRECT	GO:0032405~MutLalpha complex binding	2	2.30	0.0277565	1
	GOTERM_MF_DIRECT	GO:0003684~damaged DNA binding	3	3.45	0.035139311	1
	GOTERM_MF_DIRECT	GO:0003689~DNA clamp loader activity	2	2.30	0.036838541	1
	GOTERM_MF_DIRECT	GO:0043142~single-stranded DNA-dependent ATPase activity	2	2.30	0.045836807	1
Only NSCLC	GOTERM_BP_DIRECT	GO:0055114~oxidation-reduction process	5	17.86	0.01408762	1
	GOTERM_BP_DIRECT	GO:0002009~morphogenesis of an epithelium	2	7.14	0.022285495	1
	GOTERM_BP_DIRECT	GO:0007568~aging	3	10.71	0.028682612	1
	GOTERM_BP_DIRECT	GO:0016338~calcium-independent cell-cell adhesion via plasma membrane cell-adhesion molecules	2	7.14	0.033248128	1
	GOTERM_BP_DIRECT	GO:0010575~positive regulation of vascular endothelial growth factor production	2	7.14	0.042550374	1
	GOTERM_BP_DIRECT	GO:0043542~endothelial cell migration	2	7.14	0.045631924	1
	GOTERM_CC_DIRECT	GO:0005615~extracellular space	7	25.00	0.010282349	0.575811533
	GOTERM_CC_DIRECT	GO:0070062~extracellular exosome	9	32.14	0.037100315	1
	GOTERM_MF_DIRECT	GO:0005198~structural molecule activity	4	14.29	0.006274785	0.489433259

SCLC, small cell lung cancer; NSCLC, non-small cell lung cancer; FDR, false discovery rate.

Table 3 KEGG analysis of differentially expressed genes in SCLC & NSCLC, unique to SCLC or NSCLC

Cancer	Category	Term	Count	%	P value	FDR
SCLC & NSCLC	KEGG_PATHWAY	hsa04110: cell cycle	11	13.10	1.76E-09	1.46E-07
	KEGG_PATHWAY	hsa04115: p53 signaling pathway	6	7.14	4.31E-05	0.001786741
	KEGG_PATHWAY	hsa04114: oocyte meiosis	6	7.14	4.71E-04	0.013041937
	KEGG_PATHWAY	hsa04914: progesterone-mediated oocyte maturation	4	4.76	0.014735684	0.277141512
	KEGG_PATHWAY	hsa05166: HTLV-I infection	6	7.14	0.016695272	0.277141512
Only SCLC	KEGG_PATHWAY	hsa03430: mismatch repair	4	4.60	2.76E-04	0.029521615
	KEGG_PATHWAY	hsa03030: DNA replication	4	4.60	0.001057086	0.056554103
	KEGG_PATHWAY	hsa04110: cell cycle	5	5.75	0.00508098	0.181221637
	KEGG_PATHWAY	hsa00380: tryptophan metabolism	3	3.45	0.021332245	0.514320928
	KEGG_PATHWAY	hsa04726: serotonergic synapse	4	4.60	0.024529657	0.514320928
	KEGG_PATHWAY	hsa03420: nucleotide excision repair	3	3.45	0.028840426	0.514320928
Only NSCLC		None				

KEGG, Kyoto Encyclopedia of Genes and Genome; SCLC, small cell lung cancer; NSCLC, non-small cell lung cancer; FDR, false discovery rate.

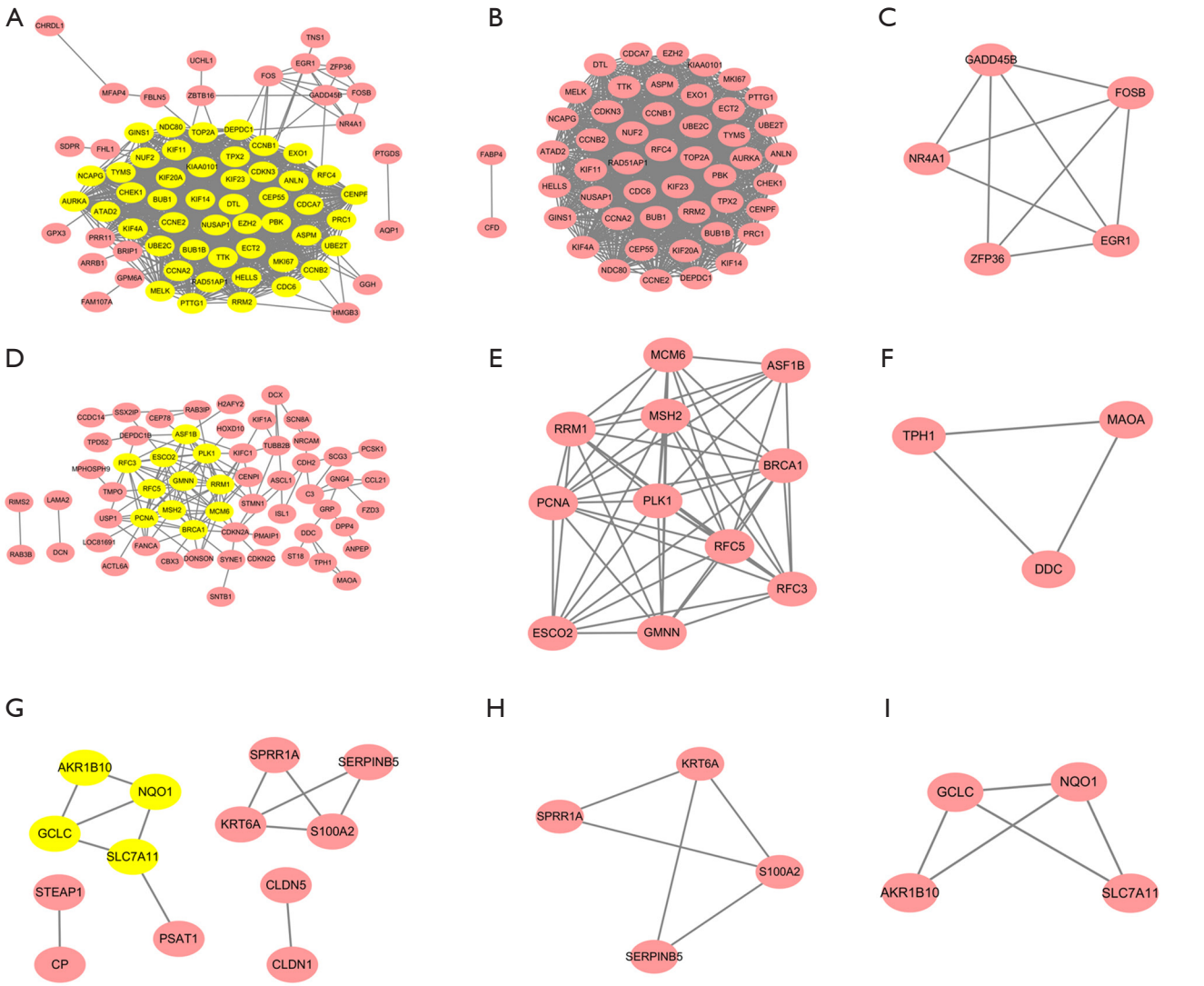

Figure 2 DEGs Protein-protein interaction network analysis. (A) DEGs protein-protein interaction network complex in SCLC & NSCLC, which contained 84 nodes and 1,122 edges. (B) Module 1 and (C) Module 2 identified by Cytoscape MCODE plugin in SCLC & NSCLC. (D) DEGs protein-protein interaction network complex unique to SCLC, which contained 84 nodes and 123 edges. (E) Module 1 and (F) Module 2 unique to SCLC. (G) DEGs protein-protein interaction network complex including 28 nodes and 13 edges unique to NSCLC. Two modules unique to NSCLC (H,I). The nodes represent proteins, and the edges represent protein interactions. DEGs, differentially expressed genes; SCLC, small cell lung cancer; NSCLC, non-small cell lung cancer.

The 87 DEGs unique to SCLC were analyzed using STRING. The results of protein-protein interaction network showed that a total of 84 nodes and 123 edges were acquired. We applied Cytotype MCODE for further analysis to get two hub modules including 11 hub nodes and 3 hub nodes. The 7 of 12 methods using cytoHubba have identified at least 8 of the 11 hub DEGs we have screened using plugin MCODE unique to SCLC PPI networks (Table S2). We also imported 28 DEGs unique to NSCLC into online STRING for protein-protein interaction network. The 28 nodes and 13 edges were included in protein-protein interaction network. Two modules were obtained using Cytotype MCODE. Every module contained 4 hub nodes (*Figure 2*). The 11 of 12 methods using cytoHubba also have identified at least 7 of the 8 hub DEGs unique to NSCLC PPI networks (Table S3).

Table 4 Survival analysis of hub DEGs using Kaplan-Meier plotter database and GEPIA

Cancer	Genes
SCLC & NSCLC	*ANLN, CHEK1, DTL, ECT2, KIF11, MKI67, NCAPG, PRC1, PTTG1, RRM2, TYMS, KIF14*
Only NSCLC	*KRT6A, SERPINB5*

The overlap 12 hub DEGs in SCLC & NSCLC and 2 hub DEGs unique to in NSCLC were significantly associated with overall survival of NSCLC patients (P<0.05). DEGs, differentially expressed genes; GEPIA, Gene Expression Profiling Interactive Analysis; SCLC, small cell lung cancer; NSCLC, non-small cell lung cancer.

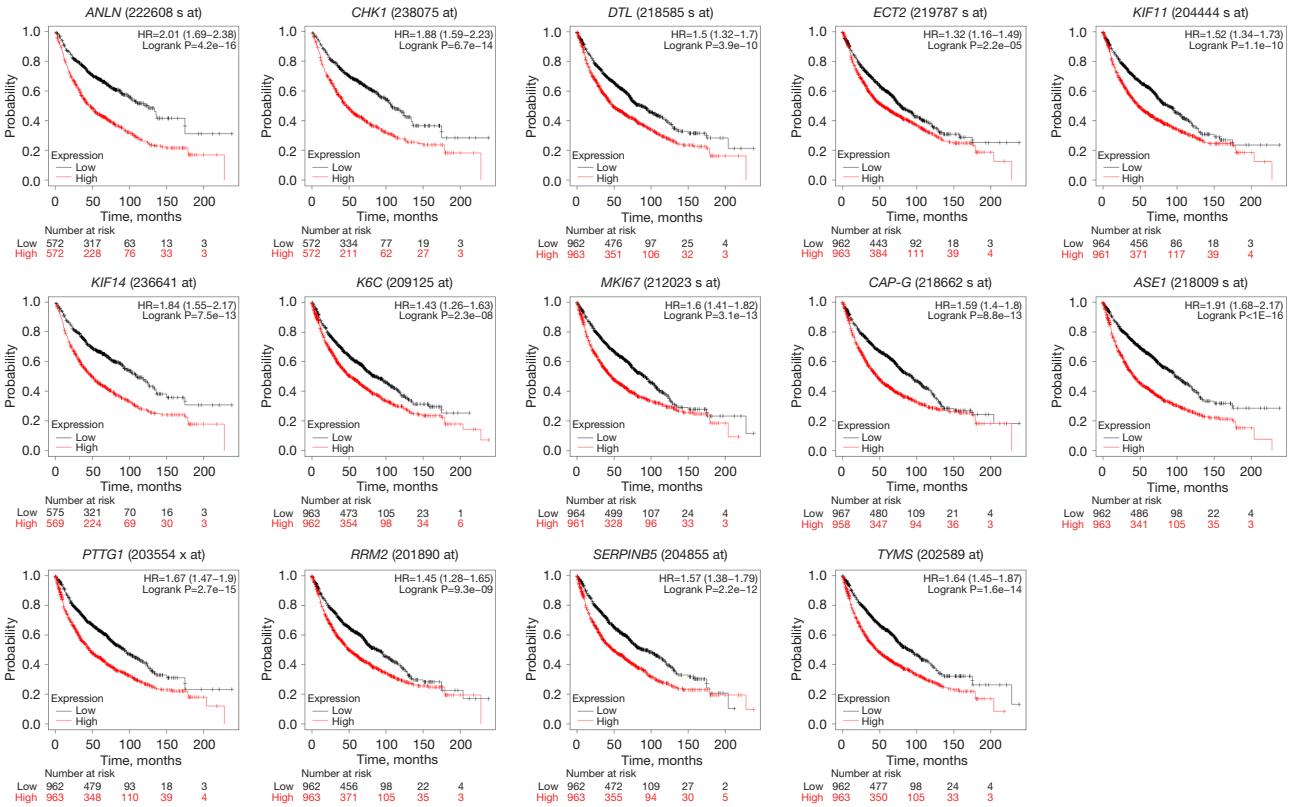

Figure 3 Overall survival analysis of 14 hub DEGs in NSCLC using Kaplan-Meier plotter database. K6C meant KRT6A; CAP-G meant NCAPG; ASE1 meant PRC1. DEGs, differentially expressed genes; NSCLC, non-small cell lung cancer.

Survival analysis and cross-validation of hub DEGs

We used the Kaplan-Meier plotter database and GEPIA to further analyze prognosis value of hub DEGs in lung cancer. As we know, 85% of lung cancer patients were NSCLC, which mainly contains LUAD and LUSC. There was survival information of unique to NSLSC patients in above two databases. There are 1,925 NSCLC patients for survival analysis using Kaplan-Meier plotter database. We conducted cross-validation survival analysis of 52 overlap hub DEGs in SCLC & NSCLC and 8 hub DEGs unique to NSCLC associated with NSCLC patients. The results demonstrated that 12 overlap hub DEGs in SCLC & NSCLC and 2 hub DEGs unique to NSCLC were significantly associated with the overall survival of NSCLC patients (P<0.05, *Table 4*, *Figures 3*,*4*).

The online GEPIA software was used to validate the expression of 14 hub DEGs in NSCLC tissues compared with normal lung tissues. A total of 11 of 14 hub DEGs were also overexpressed in LAUD and LUSC (P<0.05), but the other 3 hub DEGs *KIF14*, *KRT6A* and *SERPINB5* were only significantly different in LUSC not in LAUD (P>0.05) (*Figure 5*).

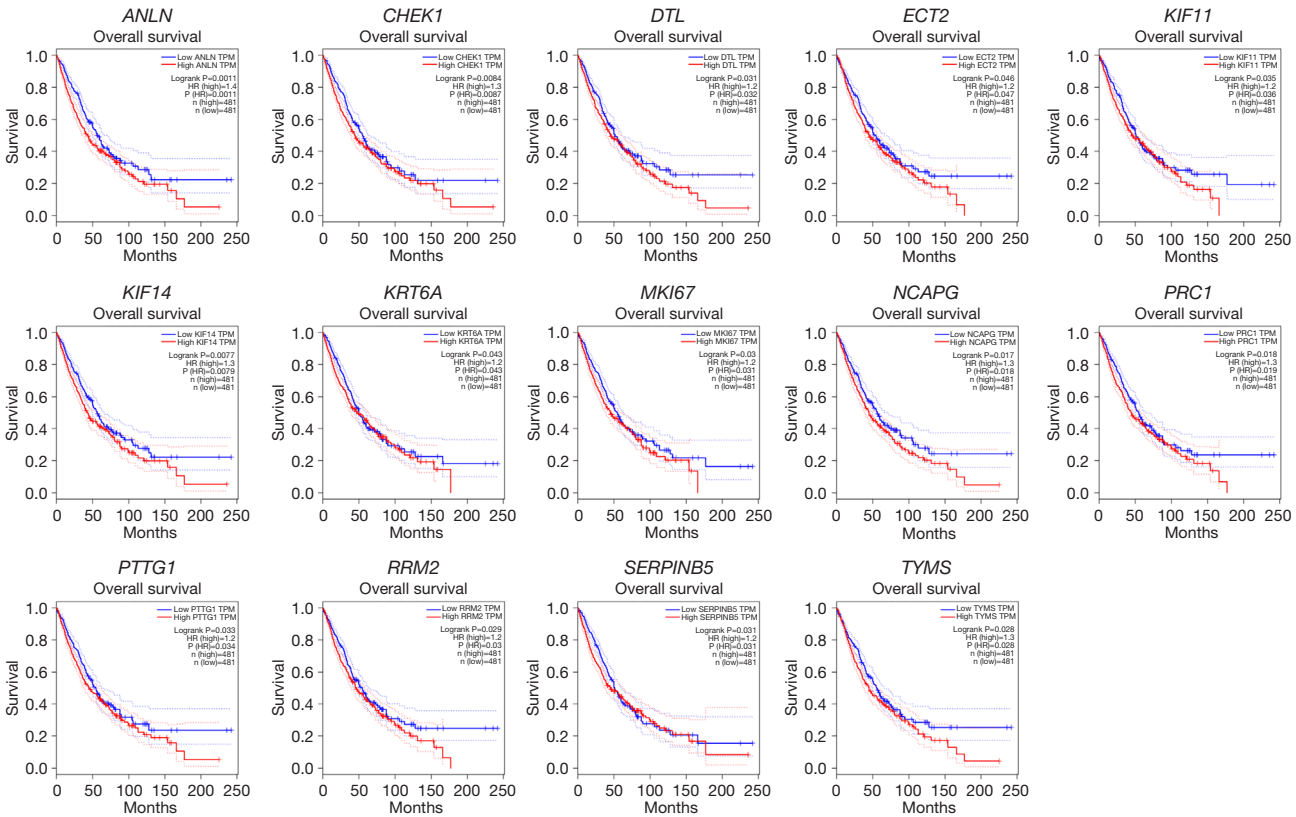

Figure 4 Overall survival analysis of 14 hub DEGs in NSCLC via GEPIA website. DEGs, differentially expressed genes; NSCLC, non-small cell lung cancer; GEPIA, Gene Expression Profiling Interactive Analysis.

Pathway enrichment re-analysis and stage analysis of hub DEGs

The 14 hub DEGs were re-analyzed to identify the more important pathways using DAVID software. The results showed that the p53 signaling pathway and pyrimidine metabolism pathway were significantly associated with the survival of NSCLC patients (P<0.05). The *RRM2*, *CHEK1* and *SERPINB5* enriched in the p53 signaling pathway, *RRM2* and *TYMS* enriched in pyrimidine metabolism pathway maybe play a key role in lung cancer. We used the GEPIA to validate the expression of these 4 hub DEGs in different stages of NSCLC. Statistical analysis identified the expression of *RRM2*, *CHEK1*, *TYMS* and *SERPINB5* were significant differential across different stages (*Table 5* and *Figure 6*).

Discussion

According to WHO criteria for lung tumors classification and diagnosis, lung cancer is generally divided into two histologic subtypes SCLC and NSCLC, and NSCLC is the main histological subtype of lung cancer. In order to identify the genetic differences between SCLC and NSCLC, we separately extracted DEGs in SCLC or NSCLC through the GEO database. Our study suggested hub DEGs play an important role not only in SCLC but also in NSCLC. We also found that some genes were only significantly differential expressed in SCLC or NSCLC. These results revealed that there are consistent differences and similarities between SCLC and NSCLC.

We further analyzed gene functional enrichment and interaction of DEGs using online DAVID, STRING database and Cytoscape MCODE. The overlap 84 DEGs found in both SCLC & NSCLC were mainly related to cell cycle, the p53 signaling pathway, oocyte meiosis, progesterone-mediated oocyte maturation and HTLV-I infection pathways. The DEGs unique to SCLC were enriched in mismatch repair, DNA replication, cell cycle, tryptophan metabolism, serotonergic synapse and

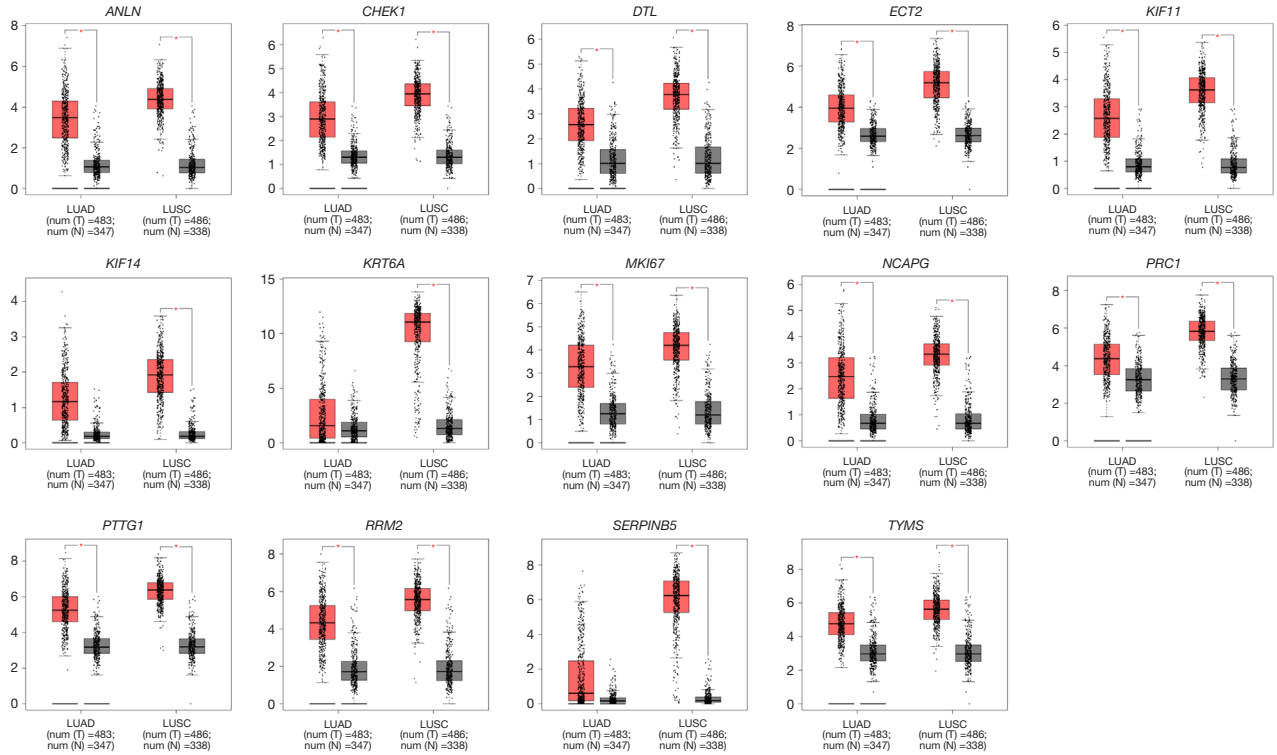

Figure 5 Validation of 14 hub DEGs expression by GEPIA website in LAUD and LUSC tissues compared with normal tissues. The red box indicates tumor samples, and the gray box indicates normal samples. *P<0.05. DEGs, differentially expressed genes; GEPIA, Gene Expression Profiling Interactive Analysis; LUAD, lung adenocarcinoma; LUSC, lung squamous-cell carcinoma.

Table 5 Re-analysis of 14 hub DEGs via KEGG pathway enrichment

Pathway ID	Name	Count	%	P value	Genes
hsa04115	p53 signaling pathway	3	21.43	5.54E-04	*RRM2, CHEK1, SERPINB5*
hsa00240	Pyrimidine metabolism pathway	2	14.29	0.057460904	*RRM2, TYMS*
hsa04110	Cell cycle pathway	2	14.29	0.070192126	*PTTG1, CHEK1*

DEGs, differentially expressed genes; KEGG, Kyoto Encyclopedia of Genes and Genome.

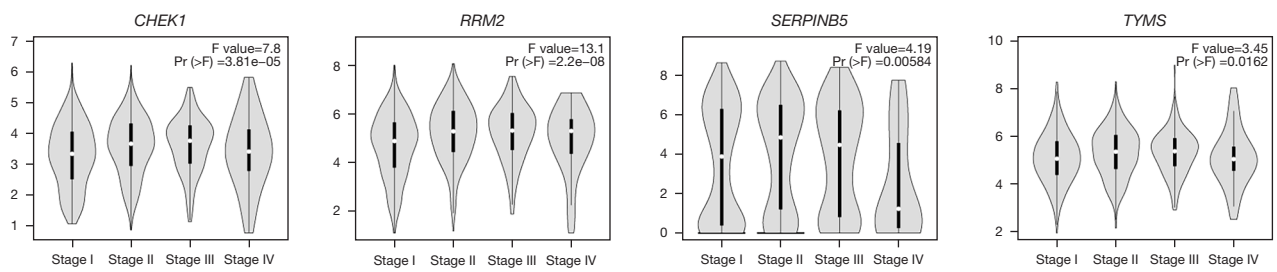

Figure 6 Pathological stage plot of hub DEGs in NSCLC. *RRM2*, *CHEK1*, *TYMS* and *SERPINB5* showed significant differential across different stages. DEGs, differentially expressed genes; NSCLC, non-small cell lung cancer.

nucleotide excision repair pathways. The DEGs unique to NSCLC were not associated with anyone pathway. Due to the lack of prognosis information of SCLC patients, we only carried out survival analysis of hub DEGs related to NSCLC patients. The results showed 14 hub DEGs were significantly associated with the overall survival of LAUD and LUSC patients. KEGG pathway enrichment re-analysis and stage analysis revealed that *RRM2*, *CHEK1*, *TYMS* and *SERPINB5* maybe new effective biomarkers in NSCLC prognosis, which were enriched in the p53 signaling pathway and pyrimidine metabolism pathway.

The ribonucleotide reductase regulatory subunit M2 (*RRM2*), one of two non-identical subunits for ribonucleotide reductase, catalyzes ribonucleotides to form deoxyribonucleotides. Transcription of *RRM2* results in two isoforms that differ in the lengths of their N-termini. RRM2 maintains to support DNA synthesis and repair and overexpressed in colorectal cancer, breast cancer, and cervical cancer (24-26). The expression of RRM2 in primary oral squamous cell carcinoma (OSCC) was significantly increased compared with normal tissues, and its overexpression was significantly associated with pathological grade, proliferation and migration, and recurrence in OSCC (27). High expression of RRM2 was associated with an immunosuppressive tumor-immune microenvironment and contributed to immune escape in prostate cancer (28). Studies have demonstrated that overexpressed miR-20a dramatically suppresses NSCLC cell proliferation and migration by inhibiting RRM2-mediated PI3K/Akt signaling, while the expression of RRM2 was upregulated in NSCLC (29). The RRM2 overexpression, as an independent predictive factor of poor prognosis in patients with lung adenocarcinoma, was significantly associated with tumor stage and TNM classification and reduced the activation of p53 signaling pathway (30).

The protein encoded by checkpoint kinase 1 (*CHEK1*), belongs to the Ser/Thr protein kinase family, and mediates cell cycle arrest in response to DNA damage or the presence of unreplicated DNA. This protein also integrates signals from ATM and ATR that are associated with chromatin in meiotic prophase I. CHEK1 promotes the phosphorylation of CDC25A protein phosphatase to delay cell cycle progression in response to double-stranded DNA breaks. High expression of CHEK1 was associated with poor clinical characteristics of multiple myeloma patients (31). CHEK1 inhibitors have been shown to potentiate in combination with chemotherapy and their single-agent antitumor, in particular gemcitabine (32-35). Studies indicated the therapeutic effects of CHEK1 inhibition are related to p53-deficiency (36).

Thymidylate synthase (*TYMS*) catalyzes the methylation of deoxyuridylate to deoxythymidylate. The function of TYMS is to maintain the dTMP (thymidine-5-prime monophosphate) pool critical for DNA replication and repair. TYMS has been a target for cancer chemotherapeutic agents, like 5-fluoro-2-prime-deoxyuridine, 5-fluorouracil, and some folate analogs. Some studies indicated that TYMS variants were associated with high-dose methotrexate in childhood acute lymphoblastic leukemia, severe hand-foot-syndrome, the risk of persistence of pre-neoplastic cervical lesions and the risk of head and neck cancer (37-40). TYMS levels associated with prognosis and chemotherapy response drives the phenotypes of epithelial-tomesenchymal transition in NSCLC. The results established TYMS as a theranostic NSCLC marker related with survival, chemo-resistance and epithelial-to-mesenchymal transition (41,42).

The serpin family B member 5 (*SERPINB5*), located on chromosome 18q21.33 as a tumor suppressor gene, plays a critical in cancer cell invasion and metastasis. SERPINB5 variants are significantly associated with gallbladder cancer risk (43). Upregulated Maspin inhibited the expression of IKKα to promote cell apoptosis and delayed the development of the precancerous lesions in precancerous rats (44). The high expression of SERPINB5 significantly increased the recurrence rate and shortened disease-free survival in patients with oral squamous cell carcinoma (45). Cytoplasmic immunoreactive scores results showed that the expression of SERPINB5 was significantly higher in cervical cancer patients compared to cervical intraepithelial neoplasia. These studies indicated SERPINB5 was related to the survival in cervical cancer, pancreatic ductal adenocarcinoma and oral squamous cell carcinoma (45-47). Wang *et al.* reported that SERPINB5 had a statistically negative correlation with NSCLC prognosis and might be a promising prognostic signature in NSCLC (48).

The above studies have reported that the 4 hub DEGs (*RRM2*, *CHEK1*, *TYMS* and *SERPINB5*) were closely related to progression and prognosis of different cancers. A small number of studies have demonstrated that above 4 hub DEGs enriched in the p53 signaling or pyrimidine metabolism pathways play a vital role in lung cancer and overall survival of NSCLC patients. The current study also has several limitations. Firstly, there is not survival information for SCLC patients currently available in the online Kaplan-Meier plotter and GEPIA database. So, survival analysis has been not carried out for patients with

SCLC. Secondly, we only assess the prognostic value and may miss some valuable information, which lacked of more clinical characteristics information from public databases, such as age, sex and treatment. Thirdly, our findings lacked molecular biological experimental validation of hub DEGs in NSCLC or SCLC.

Conclusions

Our study identified 4 hub DEGs (RRM2, CHEK1, TYMS and SERPINB5) in SCLC&NSCLC tissues compared with normal tissues. Functional analysis results indicated that these DEGs had different biological functions and were significantly enriched in different pathways. *RRM2*, *CHEK1*, *TYMS* and *SERPINB5*, which are mainly enriched in the p53 signaling and pyrimidine metabolism pathway, were significantly associated with the overall survival of NSCLC patients. These genes and pathways could serve as potential prognostic markers for personalized oncology in NSCLC or SCLC. However, more basic experiments and molecular mechanisms are needed to be confirmed for clinical applications.

Acknowledgments

Funding: This research was supported by startup funding for scientific research from Fujian Medical University (No. 2017XQ1216) and Research Project of Science and Technology Innovation Think Tank from Fujian Association for Science and Technology (No. FJKX-B2007).

Footnote

Reporting Checklist: The authors have completed the REMARK reporting checklist. Available at https://tcr.amegroups.com/article/view/10.21037/tcr-22-245/rc

Peer Review File: Available at https://tcr.amegroups.com/article/view/10.21037/tcr-22-245/prf

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at https://tcr.amegroups.com/article/view/10.21037/tcr-22-245/coif). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

References

1. Sung H, Ferlay J, Siegel RL, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin 2021;71:209-49.
2. Herbst RS, Heymach JV, Lippman SM. Lung cancer. N Engl J Med 2008;359:1367-80.
3. Siegel RL, Miller KD, Fuchs HE, et al. Cancer Statistics, 2021. CA Cancer J Clin 2021;71:7-33.
4. Howlader N, Forjaz G, Mooradian MJ, et al. The Effect of Advances in Lung-Cancer Treatment on Population Mortality. N Engl J Med 2020;383:640-9.
5. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. CA Cancer J Clin 2020;70:7-30.
6. Li D, Yin Y, He M, et al. Identification of Potential Biomarkers Associated with Prognosis in Gastric Cancer via Bioinformatics Analysis. Med Sci Monit 2021;27:e929104.
7. Giannos P, Kechagias KS, Gal A. Identification of Prognostic Gene Biomarkers in Non-Small Cell Lung Cancer Progression by Integrated Bioinformatics Analysis. Biology (Basel) 2021;10:1200.
8. Taniguchi H, Sen T, Rudin CM. Targeted Therapies and Biomarkers in Small Cell Lung Cancer. Front Oncol 2020;10:741.
9. Zengin T, Önal-Süzek T. Comprehensive Profiling of Genomic and Transcriptomic Differences between Risk Groups of Lung Adenocarcinoma and Lung Squamous Cell Carcinoma. J Pers Med 2021;11:154.
10. Wu J, Hao Z, Ma C, et al. Comparative proteogenomics profiling of non-small and small lung carcinoma cell lines using mass spectrometry. PeerJ 2020;8:e8779.
11. Yue C, Ma H, Zhou Y. Identification of prognostic gene

signature associated with microenvironment of lung adenocarcinoma. PeerJ 2019;7:e8128.
12. Šutić M, Vukić A, Baranašić J, et al. Diagnostic, Predictive, and Prognostic Biomarkers in Non-Small Cell Lung Cancer (NSCLC) Management. J Pers Med 2021;11:1102.
13. Tantai JC, Pan XF, Zhao H. Network analysis of differentially expressed genes reveals key genes in small cell lung cancer. Eur Rev Med Pharmacol Sci 2015;19:1364-72.
14. Liao Y, Yin G, Wang X, et al. Identification of candidate genes associated with the pathogenesis of small cell lung cancer via integrated bioinformatics analysis. Oncol Lett 2019;18:3723-33.
15. Li X, Ma C, Luo H, et al. Identification of the differential expression of genes and upstream microRNAs in small cell lung cancer compared with normal lung based on bioinformatics analysis. Medicine (Baltimore) 2020;99:e19086.
16. Ni M, Liu X, Wu J, et al. Identification of Candidate Biomarkers Correlated With the Pathogenesis and Prognosis of Non-small Cell Lung Cancer via Integrated Bioinformatics Analysis. Front Genet 2018;9:469.
17. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 2009;4:44-57.
18. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res 2000;28:27-30.
19. Szklarczyk D, Franceschini A, Wyder S, et al. STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res 2015;43:D447-52.
20. 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.
21. Chin CH, Chen SH, Wu HH, et al. cytoHubba: identifying hub objects and sub-networks from complex interactome. BMC Syst Biol 2014;8 Suppl 4:S11.
22. 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-W102.
23. Győrffy B. Survival analysis across the entire transcriptome identifies biomarkers with the highest prognostic power in breast cancer. Comput Struct Biotechnol J 2021;19:4101-9.
24. Kretschmer C, Sterner-Kock A, Siedentopf F, et al. Identification of early molecular markers for breast cancer. Mol Cancer 2011;10:15.
25. Grade M, Hummon AB, Camps J, et al. A genomic strategy for the functional validation of colorectal cancer genes identifies potential therapeutic targets. Int J Cancer 2011;128:1069-79.
26. Wang J, Yi Y, Chen Y, et al. Potential mechanism of RRM2 for promoting Cervical Cancer based on weighted gene co-expression network analysis. Int J Med Sci 2020;17:2362-72.
27. Wang S, Wang XL, Wu ZZ, et al. Overexpression of RRM2 is related to poor prognosis in oral squamous cell carcinoma. Oral Dis 2021;27:204-14.
28. Mazzu YZ, Armenia J, Nandakumar S, et al. Ribonucleotide reductase small subunit M2 is a master driver of aggressive prostate cancer. Mol Oncol 2020;14:1881-97.
29. Han J, Hu J, Sun F, et al. MicroRNA-20a-5p suppresses tumor angiogenesis of non-small cell lung cancer through RRM2-mediated PI3K/Akt signaling pathway. Mol Cell Biochem 2021;476:689-98.
30. Jin CY, Du L, Nuerlan AH, et al. High expression of RRM2 as an independent predictive factor of poor prognosis in patients with lung adenocarcinoma. Aging (Albany NY) 2020;13:3518-35.
31. Liu XP, Huang Q, Yin XH, et al. Strong Correlation between the Expression of CHEK1 and Clinicopathological Features of Patients with Multiple Myeloma. Crit Rev Eukaryot Gene Expr 2020;30:349-57.
32. Karp JE, Thomas BM, Greer JM, et al. Phase I and pharmacologic trial of cytosine arabinoside with the selective checkpoint 1 inhibitor Sch 900776 in refractory acute leukemias. Clin Cancer Res 2012;18:6723-31.
33. Daud AI, Ashworth MT, Strosberg J, et al. Phase I dose-escalation trial of checkpoint kinase 1 inhibitor MK-8776 as monotherapy and in combination with gemcitabine in patients with advanced solid tumors. J Clin Oncol 2015;33:1060-6.
34. Walton MI, Eve PD, Hayes A, et al. CCT244747 is a novel potent and selective CHK1 inhibitor with oral efficacy alone and in combination with genotoxic anticancer drugs. Clin Cancer Res 2012;18:5650-61.
35. Xiao Y, Ramiscal J, Kowanetz K, et al. Identification of preferred chemotherapeutics for combining with a CHK1 inhibitor. Mol Cancer Ther 2013;12:2285-95.
36. Ma CX, Cai S, Li S, et al. Targeting Chk1 in p53-deficient triple-negative breast cancer is therapeutically beneficial in human-in-mouse tumor models. J Clin Invest 2012;122:1541-52.
37. Al-Sheikh A, Yousef AM, Alshamaseen D, et al. Effects of thymidylate synthase polymorphisms on toxicities

associated with high-dose methotrexate in childhood acute lymphoblastic leukemia. Cancer Chemother Pharmacol 2021;87:379-85.
38. Silva NNT, Santos ACS, Nogueira VM, et al. 3'UTR polymorphism of Thymidylate Synthase gene increased the risk of persistence of pre-neoplastic cervical lesions. BMC Cancer 2020;20:323.
39. Hamzic S, Kummer D, Froehlich TK, et al. Evaluating the role of ENOSF1 and TYMS variants as predictors in fluoropyrimidine-related toxicities: An IPD meta-analysis. Pharmacol Res 2020;152:104594.
40. De Castro TB, Rodrigues-Fleming GH, Oliveira-Cucolo JG, et al. Gene Polymorphisms Involved in Folate Metabolism and DNA Methylation with the Risk of Head and Neck Cancer. Asian Pac J Cancer Prev 2020;21:3751-9.
41. Siddiqui MA, Gollavilli PN, Ramesh V, et al. Thymidylate synthase drives the phenotypes of epithelial-to-mesenchymal transition in non-small cell lung cancer. Br J Cancer 2021;124:281-9.
42. Agulló-Ortuño MT, García-Ruiz I, Díaz-García CV, et al. Blood mRNA expression of REV3L and TYMS as potential predictive biomarkers from platinum-based chemotherapy plus pemetrexed in non-small cell lung cancer patients. Cancer Chemother Pharmacol 2020;85:525-35.
43. Mahananda B, Vinay J, Palo A, et al. SERPINB5 Genetic Variants rs2289519 and rs2289521 are Significantly Associated with Gallbladder Cancer Risk. DNA Cell Biol 2021;40:706-12.
44. Wang N, Chang LL. The potential function of IKKα in gastric precancerous lesion via mediating Maspin. Tissue Cell 2020;65:101349.
45. Kawasaki M, Sakabe T, Kodani I, et al. Cytoplasmic-only Expression of Maspin Predicts Poor Prognosis in Patients With Oral Squamous Cell Carcinoma. Anticancer Res 2021;41:4563-70.
46. Isci Bostanci E, Guler I, Dikmen AU, et al. Prognostic role of maspin expression in patients with cervical dysplasia and cervical cancer. J Obstet Gynaecol Res 2020;46:759-64.
47. Uchinaka EI, Sakabe T, Hanaki T, et al. Cytoplasmic-only Expression of Maspin Predicts Unfavorable Prognosis in Patients With Pancreatic Ductal Adenocarcinoma. Anticancer Res 2021;41:2543-52.
48. Wang XF, Liang B, Zeng DX, et al. The roles of MASPIN expression and subcellular localization in non-small cell lung cancer. Biosci Rep 2020;40:BSR20200743.

Cite this article as: Yin Y, Li D, He M, Wang J. The identification of hub biomarkers and pathways in lung cancer and prognostic evaluation. Transl Cancer Res 2022;11(8): 2622-2635. doi: 10.21037/tcr-22-245

Table S1 The validation of top 47 hub DEGs using 12 topological analysis methods of Cytoscape plugin cytoHubba in the PPI networks of lung cancer (small cell lung cancer and non-small cell lung cancer)

MCODE		MCC		Betweenness		BottleNeck		Closeness		ClusteringCoefficient		Degree		DMNC		EcCentricity		EPC		MNC		Radiality		Stress	
Rank	Gene	Gene	Score	Name	Score	Name	Score	Name	Score	Name	Score	Name	Score	Name	Score	Name	Score	Name	Score	Name	Score	Name	Score	Name	Score
1	*CHEK1*	*KIF11*	1.11E+44	*CEP55*	394.97	*EZH2*	5	*CCNB1*	57.58	*HMGB3*	1.00	*RFC4*	49	*ANLN*	1.49	*CEP55*	0.32	*CCNB1*	14.57	*CCNB1*	49	*CCNB1*	5.42	*EZH2*	1,796
2	*MKI67*	*KIF23*	1.11E+44	*EZH2*	352.05	*CCNB1*	5	*UBE2C*	57.42	*ZFP36*	1.00	*KIF11*	49	*KIF4A*	1.48	*CHEK1*	0.24	*KIF20A*	14.08	*UBE2C*	49	*RFC4*	5.41	*RFC4*	1,130
3	*UBE2T*	*CCNB1*	1.11E+44	*RFC4*	317.57	*CEP55*	4	*RFC4*	57.42	*GGH*	1.00	*CCNB1*	49	*KIF23*	1.48	*MKI67*	0.24	*CCNB2*	14.06	*RFC4*	48	*CCNA2*	5.41	*FOS*	1,092
4	*KIAA0101*	*NCAPG*	1.11E+44	*FBLN5*	264.00	*FBLN5*	3	*TOP2A*	57.42	*TNS1*	1.00	*TOP2A*	49	*CDC6*	1.48	*UBE2T*	0.24	*CHEK1*	14.02	*KIF11*	48	*TOP2A*	5.41	*CCNB1*	1,016
5	*ATAD2*	*RAD51AP1*	1.11E+44	*CDKN3*	166.49	*FOS*	3	*KIF11*	57.25	*BRIP1*	1.00	*UBE2C*	49	*DTL*	1.48	*KIAA0101*	0.24	*TOP2A*	14.01	*BUB1*	48	*CCNB2*	5.41	*CEP55*	962
6	*KIF20A*	*CDC6*	1.11E+44	*UBE2C*	165.22	*ECT2*	3	*CCNA2*	57.08	*PRR11*	1.00	*BUB1*	48	*KIAA0101*	1.48	*ATAD2*	0.24	*RFC4*	14.01	*TPX2*	48	*UBE2C*	5.41	*CDKN3*	950
7	*HELLS*	*KIF20A*	1.11E+44	*TOP2A*	164.05	*TPX2*	3	*CCNB2*	57.08	*GINS1*	0.99	*TPX2*	48	*PTTG1*	1.47	*HELLS*	0.24	*EXO1*	13.96	*CCNA2*	48	*KIF11*	5.39	*TOP2A*	628
8	*BUB1*	*BUB1*	1.11E+44	*KIF11*	157.05	*RFC4*	2	*TPX2*	56.92	*CCNE2*	0.99	*CCNA2*	48	*NUF2*	1.47	*EZH2*	0.24	*MELK*	13.94	*TOP2A*	48	*CEP55*	5.39	*UBE2C*	612
9	*EZH2*	*TPX2*	1.11E+44	*CCNB1*	152.87	*KIF11*	2	*BUB1*	56.75	*CDCA7*	0.99	*CCNB2*	48	*MKI67*	1.46	*CCNE2*	0.24	*NCAPG*	13.91	*CCNB2*	48	*TPX2*	5.39	*CCNB2*	608
10	*ECT2*	*PBK*	1.11E+44	*FOS*	142.43	*RAD51AP1*	2	*CEP55*	56.50	*ANLN*	0.98	*NCAPG*	47	*CENPF*	1.46	*RFC4*	0.24	*ASPM*	13.86	*NCAPG*	47	*BUB1*	5.38	*GADD45B*	602
11	*CCNE2*	*ASPM*	1.11E+44	*TPX2*	141.93	*AURKA*	2	*NCAPG*	56.25	*PTTG1*	0.98	*RAD51AP1*	47	*ATAD2*	1.46	*NDC80*	0.24	*CCNA2*	13.84	*RAD51AP1*	47	*CDKN3*	5.38	*CCNA2*	594
12	*CEP55*	*CCNA2*	1.11E+44	*ZBTB16*	136.05	*FHL1*	2	*RAD51AP1*	56.25	*ATAD2*	0.98	*KIF20A*	47	*CEP55*	1.46	*KIF11*	0.24	*KIF23*	13.83	*KIF20A*	47	*EZH2*	5.37	*TPX2*	578
13	*RFC4*	*TOP2A*	1.11E+44	*GPM6A*	134.00	*ZBTB16*	2	*KIF20A*	56.25	*NUF2*	0.97	*CEP55*	47	*NCAPG*	1.46	*KIF23*	0.24	*CDKN3*	13.82	*PBK*	47	*NCAPG*	5.37	*FBLN5*	544
14	*KIF4A*	*MELK*	1.11E+44	*FHL1*	134.00	*GPM6A*	2	*PBK*	56.25	*KIF4A*	0.97	*PBK*	47	*KIF20A*	1.46	*CCNB1*	0.24	*UBE2C*	13.79	*ASPM*	47	*RAD51AP1*	5.37	*KIF11*	538
15	*NDC80*	*CCNB2*	1.11E+44	*MFAP4*	134.00	*MFAP4*	2	*ASPM*	56.25	*HELLS*	0.97	*ASPM*	47	*ASPM*	1.46	*NCAPG*	0.24	*NUSAP1*	13.79	*MELK*	47	*KIF20A*	5.37	*ECT2*	480
16	*TPX2*	*TTK*	1.11E+44	*PRC1*	125.92	*TOP2A*	2	*MELK*	56.25	*KIAA0101*	0.96	*MELK*	47	*MELK*	1.46	*RAD51AP1*	0.24	*CEP55*	13.74	*TTK*	47	*PBK*	5.37	*PRC1*	452
17	*KIF11*	*NUSAP1*	1.11E+44	*ECT2*	121.38	*UBE2C*	2	*TTK*	56.25	*KIF14*	0.96	*TTK*	47	*TTK*	1.46	*AURKA*	0.24	*MKI67*	13.73	*NUSAP1*	47	*ASPM*	5.37	*GPM6A*	426
18	*PRC1*	*UBE2C*	1.11E+44	*CCNB2*	92.71	*CDKN3*	2	*NUSAP1*	56.25	*KIF23*	0.96	*NUSAP1*	47	*NUSAP1*	1.46	*CDC6*	0.24	*KIF11*	13.71	*BUB1B*	47	*MELK*	5.37	*FHL1*	390
19	*KIF23*	*BUB1B*	1.11E+44	*GADD45B*	90.79	*CHEK1*	1	*BUB1B*	56.25	*CDC6*	0.96	*BUB1B*	47	*BUB1B*	1.46	*ANLN*	0.24	*ANLN*	13.64	*RRM2*	47	*TTK*	5.37	*BUB1*	346
20	*PBK*	*RRM2*	1.11E+44	*CCNA2*	66.09	*MKI67*	1	*RRM2*	56.25	*DTL*	0.96	*RRM2*	47	*KIF14*	1.45	*FBLN5*	0.24	*TYMS*	13.64	*EXO1*	47	*NUSAP1*	5.37	*PBK*	310
21	*CCNB1*	*EXO1*	1.11E+44	*PBK*	45.21	*UBE2T*	1	*EXO1*	56.25	*DEPDC1*	0.95	*EXO1*	47	*AURKA*	1.45	*GINS1*	0.24	*KIAA0101*	13.64	*CHEK1*	46	*BUB1B*	5.37	*ZBTB16*	304
22	*CENPF*	*DTL*	1.11E+44	*TYMS*	42.04	*KIAA0101*	1	*CDKN3*	56.08	*MKI67*	0.95	*CHEK1*	46	*NDC80*	1.45	*PRR11*	0.24	*RRM2*	13.60	*MKI67*	46	*RRM2*	5.37	*TYMS*	288
23	*NCAPG*	*CHEK1*	1.11E+44	*RRM2*	30.40	*ATAD2*	1	*CHEK1*	55.75	*CENPF*	0.95	*MKI67*	46	*CCNE2*	1.45	*KIF20A*	0.24	*BUB1B*	13.59	*KIF23*	46	*EXO1*	5.37	*MFAP4*	274
24	*KIF14*	*MKI67*	1.11E+44	*EGR1*	30.34	*HELLS*	1	*MKI67*	55.75	*NDC80*	0.95	*KIF23*	46	*CHEK1*	1.44	*BUB1*	0.24	*CDC6*	13.59	*AURKA*	46	*CHEK1*	5.35	*RAD51AP1*	272
25	*RAD51AP1*	*CENPF*	1.11E+44	*BUB1*	23.05	*CCNE2*	1	*KIF23*	55.75	*UBE2T*	0.95	*AURKA*	46	*DEPDC1*	1.44	*ECT2*	0.24	*PBK*	13.56	*CDC6*	46	*MKI67*	5.35	*EXO1*	272
26	*ASPM*	*CEP55*	1.11E+44	*NR4A1*	21.92	*NDC80*	1	*AURKA*	55.75	*AURKA*	0.94	*CDC6*	46	*HELLS*	1.44	*KIF4A*	0.24	*AURKA*	13.55	*CEP55*	46	*KIF23*	5.35	*RRM2*	254
27	*AURKA*	*KIF4A*	1.11E+44	*RAD51AP1*	19.16	*KIF23*	1	*CDC6*	55.75	*NCAPG*	0.94	*CENPF*	46	*RAD51AP1*	1.43	*TPX2*	0.24	*PRC1*	13.55	*CENPF*	46	*AURKA*	5.35	*EGR1*	252
28	*CCNA2*	*ANLN*	1.11E+44	*EXO1*	19.16	*NCAPG*	1	*CENPF*	55.75	*KIF20A*	0.94	*CDKN3*	46	*EXO1*	1.43	*PRC1*	0.24	*BUB1*	13.50	*CDKN3*	46	*CDC6*	5.35	*CHEK1*	238
29	*CDC6*	*AURKA*	1.11E+44	*CHEK1*	17.81	*CDC6*	1	*DTL*	55.75	*ASPM*	0.94	*DTL*	46	*RRM2*	1.43	*PBK*	0.24	*NUF2*	13.50	*DTL*	46	*PRC1*	5.35	*UBE2T*	188
30	*ANLN*	*PRC1*	1.11E+44	*UBE2T*	15.11	*ANLN*	1	*PRC1*	55.42	*MELK*	0.94	*KIAA0101*	45	*PBK*	1.43	*CENPF*	0.24	*KIF4A*	13.48	*KIAA0101*	45	*CENPF*	5.35	*NCAPG*	186
31	*TOP2A*	*CDKN3*	1.11E+44	*NCAPG*	5.93	*UCHL1*	1	*KIAA0101*	55.25	*TTK*	0.94	*NDC80*	45	*PRC1*	1.42	*KIF14*	0.24	*TTK*	13.45	*NDC80*	45	*DTL*	5.35	*KIF20A*	186
32	*MELK*	*KIAA0101*	1.11E+44	*KIF20A*	5.93	*HMGB3*	1	*EZH2*	55.25	*NUSAP1*	0.94	*KIF4A*	45	*ECT2*	1.42	*ASPM*	0.24	*NDC80*	13.42	*KIF4A*	45	*KIAA0101*	5.34	*ASPM*	186
33	*CCNB2*	*NDC80*	1.10E+44	*ASPM*	5.93	*GPX3*	1	*KIF4A*	55.25	*BUB1B*	0.94	*PRC1*	45	*KIF11*	1.41	*CCNA2*	0.24	*CENPF*	13.36	*PRC1*	45	*ECT2*	5.34	*MELK*	186
34	*TTK*	*NUF2*	1.10E+44	*MELK*	5.93	*SDPR*	1	*TYMS*	55.25	*CHEK1*	0.93	*TYMS*	45	*BUB1*	1.41	*TOP2A*	0.24	*DEPDC1*	13.33	*TYMS*	45	*KIF4A*	5.34	*TTK*	186
35	*NUSAP1*	*PTTG1*	8.30E+43	*TTK*	5.93	*FAM107A*	1	*NDC80*	55.08	*ECT2*	0.93	*EZH2*	44	*TYMS*	1.41	*MELK*	0.24	*KIF14*	13.28	*EZH2*	44	*TYMS*	5.34	*NUSAP1*	186
36	*UBE2C*	*KIF14*	8.26E+43	*NUSAP1*	5.93	*FOSB*	1	*ECT2*	54.92	*PRC1*	0.93	*ANLN*	44	*CCNA2*	1.41	*CCNB2*	0.24	*ECT2*	13.17	*ANLN*	44	*NDC80*	5.32	*BUB1B*	186
37	*CDCA7*	*RFC4*	5.55E+43	*BUB1B*	5.93	*NR4A1*	1	*ANLN*	54.75	*RAD51AP1*	0.92	*ECT2*	44	*TOP2A*	1.41	*TTK*	0.24	*DTL*	13.15	*ECT2*	44	*ANLN*	5.32	*AURKA*	174
38	*BUB1B*	*TYMS*	5.54E+43	*AURKA*	5.58	*ZFP36*	1	*NUF2*	54.58	*EXO1*	0.92	*NUF2*	44	*CDCA7*	1.41	*NUSAP1*	0.24	*PTTG1*	13.14	*NUF2*	44	*NUF2*	5.31	*NR4A1*	170
39	*RRM2*	*ECT2*	5.51E+43	*MKI67*	4.89	*EGR1*	1	*PTTG1*	54.25	*TYMS*	0.92	*KIF14*	43	*TPX2*	1.41	*UBE2C*	0.24	*TPX2*	13.13	*KIF14*	43	*PTTG1*	5.31	*MKI67*	152
40	*NUF2*	*DEPDC1*	5.51E+43	*CENPF*	4.89	*CHRDL1*	1	*KIF14*	54.08	*RRM2*	0.92	*DEPDC1*	43	*CCNB2*	1.41	*BUB1B*	0.24	*RAD51AP1*	13.08	*DEPDC1*	43	*ATAD2*	5.30	*CENPF*	152
41	*EXO1*	*UBE2T*	2.79E+43	*NDC80*	4.52	*FABP4*	1	*DEPDC1*	53.92	*PBK*	0.92	*PTTG1*	43	*UBE2T*	1.40	*RRM2*	0.24	*EZH2*	12.97	*PTTG1*	43	*KIF14*	5.30	*NDC80*	136
42	*CDKN3*	*ATAD2*	2.76E+43	*DEPDC1*	3.37	*CFD*	1	*ATAD2*	53.75	*CEP55*	0.90	*ATAD2*	42	*GINS1*	1.38	*NUF2*	0.24	*HELLS*	12.91	*ATAD2*	42	*UBE2T*	5.28	*KIF23*	114
43	*DEPDC1*	*EZH2*	4.24E+41	*KIF23*	3.04	*GGH*	1	*UBE2T*	53.25	*BUB1*	0.90	*UBE2T*	41	*CCNB1*	1.36	*EXO1*	0.24	*ATAD2*	12.81	*UBE2T*	41	*HELLS*	5.28	*CDC6*	114
44	*GINS1*	*CCNE2*	6.20E+40	*CDC6*	3.04	*TNS1*	1	*HELLS*	53.25	*CCNA2*	0.90	*HELLS*	41	*UBE2C*	1.36	*CDKN3*	0.24	*UBE2T*	12.12	*HELLS*	41	*DEPDC1*	5.28	*DTL*	114
45	*PTTG1*	*HELLS*	2.13E+40	*DTL*	3.04	*GADD45B*	1	*CCNE2*	52.25	*FOSB*	0.90	*CCNE2*	39	*RFC4*	1.33	*DEPDC1*	0.24	*CDCA7*	11.83	*CCNE2*	39	*CCNE2*	5.25	*KIAA0101*	100
46	*DTL*	*CDCA7*	9.48E+36	*KIF14*	3.00	*BRIP1*	1	*CDCA7*	50.12	*TPX2*	0.90	*CDCA7*	36	*CDKN3*	1.33	*PTTG1*	0.24	*CCNE2*	11.79	*CDCA7*	36	*CDCA7*	5.14	*DEPDC1*	98
47	*TYMS*	*GINS1*	1.70E+34	*KIAA0101*	2.64	*CDCA7*	1	*GINS1*	48.92	*CCNB2*	0.90	*GINS1*	33	*EZH2*	1.24	*DTL*	0.24	*GINS1*	11.46	*GINS1*	33	*GINS1*	5.14	*KIF14*	86
	Reference (47 genes)	47/47		38/47		25/47		47/47		39/47		47/47		47/47		45/47		47/47		47/47		47/47		38/47	

PPI, protein-protein interaction; MCODE, Molecular Complex Detection; EPC, Percolated Component; MNC, Maximum Neighborhood Component; DMNC, Density of Maximum Neighborhood Component; MCC, Maximal Clique Centrality.

Table S2 The validation of top 11 hub DEGs using 12 topological analysis methods of Cytoscape plugin cytoHubba in the PPI networks of small cell lung cancer

MCODE		MCC		Betweenness		BottleNeck		Closeness		ClusteringCoefficient		Degree		DMNC		EcCentricity		EPC		MNC		Radiality		Stress	
Rank	Gene	Gene	Score	Gene	Score	Gene	Score	Gene	Score	Gene	Score	Gene	Score	Gene	Score	Gene	Score	Gene	Score	Gene	Score	Gene	Score	Gene	Score
1	*MSH2*	*MSH2*	86,430	*CDKN2A*	1,348.78	*CDKN2A*	53	*PLK1*	30.97	*KIF1A*	1.00	*PLK1*	17	*ESCO2*	0.72	*CDH2*	0.19	*BRCA1*	16.28	*BRCA1*	15	*CDKN2A*	8.06	*CDKN2A*	3,120
2	*PCNA*	*PCNA*	86,426	*CDH2*	1,276.61	*CDH2*	49	*BRCA1*	30.14	*DONSON*	1.00	*BRCA1*	16	*FANCA*	0.71	*NRCAM*	0.16	*PLK1*	16.27	*PLK1*	15	*PLK1*	8.03	*CDH2*	2,706
3	*RFC3*	*RFC3*	86,412	*C3*	900.00	*BRCA1*	26	*MSH2*	28.89	*MAOA*	1.00	*PCNA*	15	*USP1*	0.71	*SCG3*	0.16	*MSH2*	16.25	*RFC3*	14	*BRCA1*	7.98	*C3*	1,800
4	*RFC5*	*RFC5*	86,406	*GNG4*	766.00	*PLK1*	13	*PCNA*	28.00	*TPH1*	1.00	*RFC3*	14	*GMNN*	0.70	*C3*	0.16	*PCNA*	16.10	*MSH2*	14	*MSH2*	7.91	*PLK1*	1,718
5	*PLK1*	*PLK1*	85,780	*PLK1*	612.11	*C3*	11	*RFC3*	27.84	*CEP78*	1.00	*MSH2*	14	*PCNA*	0.68	*CDKN2A*	0.16	*RFC3*	16.07	*RRM1*	13	*GMNN*	7.84	*GNG4*	1,486
6	*RRM1*	*RRM1*	85,758	*GRP*	604.00	*MSH2*	10	*RRM1*	27.67	*CCL21*	1.00	*RRM1*	13	*RFC5*	0.66	*ISL1*	0.16	*RFC5*	15.98	*RFC5*	13	*CDH2*	7.74	*BRCA1*	1,236
7	*BRCA1*	*BRCA1*	81,423	*BRCA1*	430.66	*GNG4*	9	*GMNN*	27.55	*FANCA*	1.00	*RFC5*	13	*RRM1*	0.64	*PCSK1*	0.13	*RRM1*	15.73	*PCNA*	13	*STMN1*	7.67	*GRP*	1,144
8	*GMNN*	*GMNN*	80,689	*DDC*	316.00	*GRP*	7	*CDKN2A*	27.32	*USP1*	1.00	*GMNN*	12	*MCM6*	0.63	*SCN8A*	0.13	*MCM6*	15.54	*MCM6*	12	*RRM1*	7.64	*MSH2*	1,028
9	*MCM6*	*MCM6*	45,438	*MSH2*	308.22	*STMN1*	6	*RFC5*	27.17	*CENPI*	0.83	*MCM6*	12	*ASF1B*	0.62	*PMAIP1*	0.13	*GMNN*	15.28	*GMNN*	11	*RFC3*	7.60	*KIFC1*	780
10	*ESCO2*	*ESCO2*	40,327	*GMNN*	248.54	*DDC*	4	*MCM6*	26.67	*ESCO2*	0.67	*ESCO2*	10	*MSH2*	0.62	*CDKN2C*	0.13	*ESCO2*	14.56	*ESCO2*	9	*PCNA*	7.59	*STMN1*	780
11	*ASF1B*	*ASF1B*	5,071	*PCNA*	244.83	*PCNA*	4	*ESCO2*	25.50	*RFC5*	0.67	*ASF1B*	10	*RFC3*	0.61	*CCL21*	0.13	*ASF1B*	14.49	*ASF1B*	9	*RFC5*	7.57	*PCNA*	766
	Reference (11 genes)	11/11		5/11		4/11		10/11		2/11		11/11		9/11		0/11		11/11		11/11		8/11		4/11	

PPI, protein-protein interaction; MCODE, Molecular Complex Detection; EPC, Percolated Component; MNC, Maximum Neighborhood Component; DMNC, Density of Maximum Neighborhood Component; MCC, Maximal Clique Centrality.

Table S3 The validation of top 8 hub DEGs using 12 topological analysis methods of Cytoscape plugin cytoHubba in the PPI networks of non-small cell lung cancer

MCODE		MCC		Betweenness		BottleNeck		Closeness		ClusteringCoefficient		Degree		DMNC		EcCentricity		EPC		MNC		Radiality		Stress	
Rank	Gene	Gene	Score	Gene	Score	Gene	Score	Gene	Score	Gene	Score	Gene	Score	Gene	Score	Gene	Score	Gene	Score	Gene	Score	Gene	Score	Gene	Score
1	*S100A2*	*S100A2*	4	*SLC7A11*	6	*SLC7A11*	2	*SLC7A11*	3.50	*SERPINB5*	1.00	*S100A2*	3	*S100A2*	0.31	*S100A2*	0.31	*NQO1*	3.18	*S100A2*	3	*SLC7A11*	1.44	*SLC7A11*	8
2	*KRT6A*	*KRT6A*	4	*NQO1*	2	*GCLC*	2	*NQO1*	3.50	*SPRR1A*	1.00	*KRT6A*	3	*KRT6A*	0.31	*KRT6A*	0.31	*GCLC*	3.16	*KRT6A*	3	*NQO1*	1.44	*NQO1*	4
3	*NQO1*	*NQO1*	4	*GCLC*	2	*PSAT1*	1	*GCLC*	3.50	*AKR1B10*	1.00	*SLC7A11*	3	*NQO1*	0.31	*SLC7A11*	0.19	*SLC7A11*	3.12	*NQO1*	3	*GCLC*	1.44	*GCLC*	4
4	*GCLC*	*GCLC*	4	*S100A2*	1	*SERPINB5*	1	*S100A2*	3.00	*S100A2*	0.67	*NQO1*	3	*GCLC*	0.31	*NQO1*	0.19	*AKR1B10*	2.86	*GCLC*	3	*AKR1B10*	1.25	*S100A2*	2
5	*SLC7A11*	*SLC7A11*	3	*KRT6A*	1	*SPRR1A*	1	*KRT6A*	3.00	*KRT6A*	0.67	*GCLC*	3	*SERPINB5*	0.31	*GCLC*	0.19	*KRT6A*	2.86	*SERPINB5*	2	*PSAT1*	1.15	*KRT6A*	2
6	*SERPINB5*	*SERPINB5*	2	*PSAT1*	0	*S100A2*	1	*AKR1B10*	2.83	*NQO1*	0.67	*SERPINB5*	2	*SPRR1A*	0.31	*SERPINB5*	0.15	*S100A2*	2.85	*SPRR1A*	2	*S100A2*	0.92	*PSAT1*	0
7	*SPRR1A*	*SPRR1A*	2	*SERPINB5*	0	*KRT6A*	1	*SERPINB5*	2.50	*GCLC*	0.67	*SPRR1A*	2	*SLC7A11*	0.31	*SPRR1A*	0.15	*SERPINB5*	2.63	*SLC7A11*	2	*KRT6A*	0.92	*SERPINB5*	0
8	*AKR1B10*	*AKR1B10*	2	*SPRR1A*	0	*STEAP1*	1	*SPRR1A*	2.50	*SLC7A11*	0.33	*AKR1B10*	2	*AKR1B10*	0.31	*STEAP1*	0.15	*SPRR1A*	2.62	*AKR1B10*	2	*SERPINB5*	0.82	*SPRR1A*	0
	Reference (8 genes)	8/8		7/8		6/8		8/8		8/8		8/8		8/8		7/8		8/8		8/8		7/8		7/8	

PPI, protein-protein interaction; MCODE, Molecular Complex Detection; EPC, Percolated Component; MNC, Maximum Neighborhood Component; DMNC, Density of Maximum Neighborhood Component; MCC, Maximal Clique Centrality.