Identification of a Cell Cycle Gene Signature Predicting Survival in Patients with Lung Squamous Cell Carcinoma

Lei Zhang
China Medical College Hospital: China Medical University Hospital
Shize Yang
China Medical College Hospital: China Medical University Hospital
Zhenglun Yu (alan5915@126.com)
China Medical University Hospital

Research

Keywords: lung squamous cell carcinoma, cell cycle, prognosis, mRNAs, meta-analysis survival

Posted Date: September 28th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-79401/v1

Abstract

Purpose: Lung cancer (LC) is one of the most important and common malignant tumours, and its incidence and mortality are increasing annually. Lung squamous cell carcinoma (LUSC) is the most common pathological type of LC. A small number of biomarkers have been certified to be consistent with its survival and prognosis by data excavation. However, the moderate forecast effect of a single gene biomarker is not accurate. Thus, we planned to find new gene signatures to preferably predict LUSC.

Methods: Using the mRNA mining method, we enforced mRNA expression analyzing in big LUSC cohorts (n=504) from The Cancer Genome Atlas (TCGA) database. Gene set enrichment analysis (GSEA) and gene set variation analysis (GSVA) were enforced, and relations between genes and the cell cycle were got with the Cox proportional hazards regression model.

Results: We confirmed a set of four genes (CDKN1A, CHEK2, E2F4 and RAD21) that was importantly related to overall survival (OS) in the test succession. Based on the four-gene signature, the patients were separated into high-risk and low-risk teams. Moreover ,multivariate Cox regression analysis showed that the prognostic value of the four-gene signature and clinical factors were mutual independent.

Conclusion: Our research demonstrated connections between the four-gene signature and LUSC. Novel insights into mechanisms of the cell cycle were also revealed after determining that the biomarkers were related to a poor prognosis in LUSC patients.

Background

In recent years, the incidence and mortality rates of lung cancer (LC) have been the highest worldwide, and LC is the dominant cause of malignant tumor-related death in China[1, 2]. There are two types of pathological categories, non-small cell lung cancer (NSCLC) and small cell lung cancer (SCLC), and lung squamous cell carcinoma (LUSC) is one type of NSCLC that accounts for approximately 25–30% of all LC patients[3]. The incidence of LUSC can be controlled by restricting tobacco smoking[4]. The identification of targeted treatment strategies for LUSC has remained a complicated goal[5]. A large amount of evidence shows that the detection and inhibition of molecular biomarkers will promote the prognostic valuation and identification of possible high-risk patients with LUSC. Therefore, we should identify additional biomarkers that affect the mechanisms of tumorigenesis and progression in LUSC. Therefore, the development of novel biomarkers that can predict the prognostic value for LUSC is of great significance. Recent research have provided insights into the mechanisms that induce DNA repair in particular cell cycle stages and have highlighted the mechanisms that ensure cell cycle arrest or progression between normal and cancerous cells. Some genes bind to and inhibit cyclin-dependent kinase activity, preventing the phosphorylation of final cyclin-dependent kinase substrates, stemming cell cycle advancement and promoting the development of LUSC.

In the past few years, many studies have confirmed that microRNAs (miRNAs) can control gene expression at the posttranscriptional level[6]; thus, miRNAs have been considered new biomarkers of

cancer with infinite clinical value[7]. Scientists have researched and detected many related genes and developed a new approach for further research and effective treatment methods. Some biomarkers have been confirmed as prognostic factors of LUSC. miR-324-3p and miR-1285 are diagnostic and prognostic biomarkers for early-stage LUSC[8]. Minichromosome maintenance protein (MCM) has been found to be a relevant marker for prognosis in many human cancers and may act as a potential therapeutic target for LUSC patients[9]. With the development of high-throughput sorting and magnificent data analysis techniques, numerous databases have enabled a strict realizing of genomic changes. Cancer-related genes and gather models can identified by bioinformatics and data integration[10]. Whereas, a single gene biomarker cannot stand for good prediction effects, and some research have found that a gene signature can be a better alternative to predict survival and prognosis and [11]. Kaplan-Meier (KM) plots and Cox proportional hazards regression methods have been used diffusely in biomarker screening[12, 13]. Multigene prognostic signatures can be used in clinical treatment because of the original cancer biopsies. Whereas, not all pathways have been detected to confirm new biomarkers of LUSC. Therefore, additional efforts are needed to identify more sensitive biomarkers of LUSC.

The primary public project The Cancer Genome Atlas (TCGA)[14] can select,collect and analyse human tissue for genome alterations effectively. The TCGA database provides an analysis of high-throughput data on a variety of genomic alterations, including those involving the cell cycle. Information on patient diagnostics, treatment, and tumour pathology can be combined with bioinformatics analysis. We applied gene set enrichment analysis (GSEA) and gene set variation analysis (GSVA) to search for genes and perform further analyses. GSEA does not require a obvious variant gene openning. The algorithm is also unique because several genes whose expression is based on the whole tendency of effective data and overall level can be identified, even without any experience before. GSVA was performed to explore the biological functions and functional annotations of the different genes. We conducted various bioinformatics and survival analyses, screened for one pathway associated with the cell cycle, and built a prognostic risk model to forecast the prognosis of LUSC patients. The results of this analysis could give us new insight into the molecular mechanisms on the base of the cell cycle in LUSC.

In our research, we obtained hallmark gene sets from LUSC patients with total mRNA expression data from the TCGA database. We identified the important mRNAs related to the cell cycle and structured a four-gene risk signature that can exactly forecast patient prognosis. Startingly, the cell cycle-related risk signature, which comprises fifteen pathways, can predict patients who are probably poor prognosis.

Methods

Patient Clinical Data and mRNA Expression Profiles

The mRNA expression profiles and clinical data concerning LUSC patients were collected from the TCGA database (https://cancergenome.nih.gov/) and divided into two groups: one with LUSC tissue and one with adjacent noncancerous tissue.In addition, the following clinical information was recorded: sex, age, tumour size, pathological stage, new tumour events after the initial treatment, status of distant organ

metastasis, neoplasm cancer status, and residual tumour. Finally, patients were classified. The ordinary clinical information are shown in Table 1.

Table 1.
Clinical pathological parameters of patients with Lung squamous cell carcinoma in this research

Clinical Characteristic	N	%
Age（year）		
<68	231	46.7
≥68	264	53.3
Gender		
Male	373	74.0
Female	131	26.0
T classification		
T1	114	22.8
T2	293	58.7
T3	70	14.0
T4	22	4.4
N classification		
N0	319	62.4
N1	133	25.6
N2	40	9.9
N3	5	1.0
N_X	6	1.1
M classification		
M0	411	82.7
M1	7	1.4
Mx	79	15.9
UICC stage		
Ⅰ stage	244	49.1
Ⅱ stage	165	33.2
Ⅲ stage	81	16.3
Ⅳ stage	7	1.4
New Event		

Yes	150	29.9
No	351	70.1
Cancer Status		
Tumor free	312	79.8
With tumor	79	20.2
Radiation therapy		
No	139	90.3
Yes	15	9.7
Residual tumor		
R0	398	90.7
R1	41	9.3

GSEA and GSVA

GSEA (http://www.broadinstitute.org/gsea/index.jsp) was performed to detect whether the identified gene sets showed significant differences between the groups[15]. The expression levels of mRNAs in LUSC tissue and adjacent noncancerous tissue were analysed. Normalized p values ($p < 0.05$) were used to determine which functions could be used for further investigation. GSVA is a GSE method that evaluates the variation in pathway activity over a specimen population in an unsupervised method. We showed the robustness of GSVA in a comparison with the state-of-the-art sample enrichment technique.

Statistical Analysis

The expression profiles of 56,392 mRNAs are shown as raw data, and every mRNA was normalized by log2 for further analysis. We used univariate Cox regression to analyse and identify genes with visible relationships to overall survival (OS) (p values < 0.05). Next, we used multivariate Cox proportional hazards regression to analyse and further identify the prognostic genes from the preceding steps. The filtered mRNAs were classified into risk (hazard ratio (HR) > 1) and protective (0 < HR < 1) types.Subsequently,a prognostic risk score formula was established based on a linear combination of the expression levels weighted with the regression coefficients derived from the multivariate Cox regression analysis: Risk score = expression of gene 1 × β1 + expression of gene 2 × β2 + ⋯ + expression of gene n × βn. We separated patients into high-risk and low-risk subgroups using the median risk score as the cut-off. Next, we uselized KM curves and log-rank tests to verify the prognostic significance of the risk score. Subsequently, we inspected the diverse expression of ideal genes in LUSC tissue and adjacent noncancerous tissue and classified them into high-risk and low-risk groups by the median risk score. The KM method (multiplication of the positive limit) was used to predict the veracity of the survival state and survival time. The survival function was built by the KM approach, and the ROC curve was constructed.

The chosen gene alterations in specific cancer types can be found online (http://www.cbioportal.org/). All the statistical analyses were performed using SPSS 16.0 and GraphPad Prism 8 software.

Results

Initial Screening of Genes using GSEA and GSVA

We got the clinical data from patients with LUSC, along with an expression dataset for 56,392 mRNAs from the TCGA database. The Kyoto Encyclopedia of Genes and Genomes (KEGG) gene sets had expressed signatures derived by choosing multiple gene sets from the Molecular Signatures Database (MSigDB) to be on behalf of well-defined biological statuses or courses. GSEA was identified using the above detailed data to detect whether the identified gene sets showed statistically important differences between normal tissue and LUSC tissue. GSVA estimates the variation in gene set enrichment over the samples independently of any class label. Therefore, we obtained fifteen pathways (Fig. 1), including DNA replication, cell cycle, homologous recombination, mismatch repair, proteasome, base excision repair, spliceosome, aminoacyl tRNA biosynthesis, pyrimidine metabolism, nucleotide excision repair, P53 signalling pathway, basal transcription factors, RNA degradation, RNA polymerase and oocyte meiosis, and we selected the cell cycle pathway based on the number of genes and the normalized enrichment score (NES).

Identification of Cell Cycle-related mRNAs Associated with Patient Survival

First, we used univariate Cox regression analysis of the 125 genes for rudimentary screening and got 20 genes with p values < 0.1.In addition, multivariate Cox regression analysis was performed to further examine the correlation between the expression profiles of 20 mRNAs and the patient survival rate.Whereafter, 4 mRNAs (CDKN1A, CHEK2, E2F4 and RAD21) were identified as independent poor prognostic indicators. The filtered mRNAs were divided into a risk type (CDKN1A, E2F4 and RAD21), whose HR was > 1 with shorter survival, and a protective type (CHEK2), whose HR was < 1 with longer survival (Table 2). We calculated the Pearson correlation coefficients of the 4 mRNAs (as shown in Table 2) and found a correlation between E2F4 and CHEK2, between E2F4 and RAD21, and between CHEK2 and RAD21 (all R values were greater than 0.3) (Fig. 2).

Table 2
The information of four prognostic mRNAs importantly associated with overall survival in patients with LUSC

mRNA	Ensemble ID	Location	B(cox)	HR(95%CIs)	P
CDKN1A	ENSG00000124762	Chr6:36,676,460 – 36,687,339	0.14381	1.15466	0.00125
CHEK2	ENSG00000183765	Chr22:28,687,743 – 28,742,422	-0.30838	0.73464	0.000924
E2F4	ENSG00000205250	Chr16:67,192,155 – 67,198,918	0.25765	1.29388	0.023732
RAD21	ENSG00000164754	Chr8:116,845,935 – 116,874,866	0.23935	1.27042	0.035461

Construction of a Four-mRNA Signature to Predict Patient Prognosis

The prognostic risk score formula was fixed based on a linear combination of the expression levels weighted with the regression coefficients derived from the multivariate Cox regression analysis: risk score = 0.14381*expression of CDKN1A -0.30838*expression of CHEK2 + 0.25765*expression of E2F4 + 0.23935*expression of RAD21. Each patient with LUSC had single risk score. We calculated and arranged the scores and then divided the patients into high- and low-risk groups by the median value (Fig. 3A). The survival time (in days) of each patient is shown in Fig. 3B, and the patients with high risk scores showed higher mortality rates than those with low-risk scores (Fig. 3A and Fig. 3B). Additionally, a heatmap (Fig. 3C) was generated to display the expression profiles of the four mRNAs. Then, we compared the risk score to the prognosis of the 4-mRNA group, and the differential expression of the four genes between cancer tissue and adjacent normal tissue is shown in Fig. 4A. The differential expression of the four genes in each stage is displayed in Fig. 4B, and the graph of the survival curve between the risk score and the four genes is shown in Fig. 4C. Therefore, we can see that the P value of the risk score is dominant. With the increasing risk score of LUSC patients, the expression of risk-type mRNAs (CDKN1A, E2F4 and RAD21) was obviously upregulated. By contrast, the expression of the protective-type mRNAs (CHEK2) was downregulated.

Generation of the Risk Score from the Four-mRNA Signature as an Indicator of Prognosis

The prognostic value of the risk score was compared with the clinicopathological information by univariate and multivariate analyses. Samples with complete clinical data were used for analysis. The median age of the 504 patients with LUSC was 68 years and included 373 male patients and 131 female patients. Among 391 patients, 79 (20.2%) developed a tumour during the follow-up visit. Among 439 patients, 41 (9.3%) had residual tumours. Among 503 patients, 184 (36.6%) had lymph node metastasis, and 86 (17.3%) had distant metastases among 497 patients with LUSC. Among 154 patients, 15 (9.7%)

received radiation therapy. Additionally, we found that the risk score, new events, tobacco smoking history, and neoplasm cancer status were independent prognostic indicators because they showed significant differences in the univariate analysis, with p values < 0.05 (Table 3). In the subsequent multivariate analysis (Table 3), we found that the risk score, new events, neoplasm cancer status and tobacco smoking history showed statistical significance in the univariate and multivariate analyses (P < 0.05). Regardless of the analysis used (univariate or multivariate), the risk score showed prominent prognostic value, with p values < 0.05 (HR = 1.566, 95% CI (confidence interval) = 1.073–2.288). Additionally, the best clinical parameter to predict patient survival was neoplasm cancer status, and patients with tumours were 4.871 times more likely to experience death than those who were tumour free. Based on the P value, we can conclude that the risk score is more dominant than the TNM classification. The 4-mRNA expression-based survival risk score was applied to divide patients into a low-risk or high-risk group through the median risk score as the cut-off. The ROC curve analysis score was 0.661 (Fig. 5A),showing good sensitivity and specificity of the 4-mRNA signature in predicting survival in LUSC patients. We also generated an ROC curve of important clinical parameters (Fig. 5B-H) and found that the ROC curve of the risk score was obviously higher than and thus superior to the ROC curve of the other clinical parameters as an indicator of prognosis.

Table 3. Univariable and multivariable analyses for each clinical feature

	Univariate analysis			Multivariate analysis		
Clinical feature	HR	95%CI	P	HR	95%CI	P
Risk score	1.702	1.292-2.241	0.000	1.566	1.073-2.288	0.020
Age	1.164	0.883-1.533	0.281			
Sex	1.217	0.881-1.681	0.234			
T classification	1.442	1.024-2.031	0.036	1.442	0.929-2.239	0.103
N classification	1.155	0.874-1.525	0.312			
M classification	1.658	1.142-2.409	0.008	1.153	0.695-1.913	0.582
Smoking history	0.637	0.479-0.847	0.002	0.578	0.398-0.840	0.004
UICC stage	1.304	0.991-1.715	0.058			
Neoplasm cancer status	3.485	2.493-4.871	0.000	2.199	1.364-3.547	0.001
New event	2.231	1.700-2.926	0.000	1.680	1.056-2.673	0.029
Radiation therapy	1.875	0.914-3.848	0.087			
Residual tumor	1.884	1.120-3.168	0.017	1.022	0.486-2.148	0.955

Validation of the Four mRNA Markers for Survival Prediction by KM Curve Analysis

KM curves and log-rank tests showed a poor prognosis in patients with high risk scores (p < 0.0010) (Fig. 6A). The univariate Cox regression analysis of OS showed that several clinicopathological factors, including age, sex, T classification, N classification, M classification, new events, neoplasm cancer status,

tobacco smoking history, radiation therapy and residual tumours, were valid in predicting the survival rate of LUSC patients. The KM method was then adopted to identify the above consequences. According to the curve, patients aged > sixty-eight, the presence of a tumour after treatment, a T classification > T1, distant organ metastasis, a tumour stage > stage I, a residual tumour, tobacco smoking history, or a positive tumour finding during the follow-up visit were correlated with a poor prognosis (Fig. 6B, 6C, 6E, 6G, 6K, 6M, 6N, 6O). These consequences provide further acknowledgement of the correctness of our analysis. Therefore, a further stratified analysis was achieved for data mining. The risk score is superior to other clinical indicators based on the results.

Validation of the Differentially Expressed Genes between the High-risk and Low-risk Groups

We classified LUSC patients into two groups by the risk score and determined the enriched pathways in the high-risk and low-risk groups by GSEA and GSVA. The differentially expressed genes were then subjected to mass survival analyses, and related genes were obtained. We calculated the Pearson correlation coefficients between related genes and the four genes and found a correlation between CDKN1A and KLK5 and between CDKN1A and KLK7, with R values greater than 0.3 (Fig. 7B).

Discussion

In recent years, some experts have confirmed that miRNAs play an significant role in the tumorigenesis and progression of LC. Accumulating evidence reveals the potential of miRNAs as biomarkers for estimating and predicting the prognosis for LC[16], showing their probable clinical importance in explore[17]. For example, Xu et al reported that *microRNA*-106b serves as a *prognostic* biomarker and is associated with cell proliferation, migration, and invasion in osteosarcoma[18]. Li et al concluded that miR-421 is overexpressed and promotes cell proliferation in NSCLC[19].Whereas, patient survival could not be predicted by these genes because varieties of factors can control a single gene, induce an wrong predictive effect. Therefore, a gene signature containing various genes which was built from the statistical model to predict malignant tumor results. The results show that the predictive outcomes of each gene involved can give a more precise prediction than a single biomarker[20].

As the development and popularity of gene signatures, statistical models can predict the prognosis of various of malignant tumor. The expression of miR-98-5p, miR-152-3p, miR-326 and miR-4289 expression could serve as a biomarker for prostate cancer diagnosis[21]. We also utilized GSEA through the mRNA expression data of the 503 LUSC patients and found that 4 exhibit important differences, with a P value < 0.05, and the minimum P value was used for further analysis. We found out peculiar functions to confirm genes by GSEA that could forecast the survival state of LUSC patients. More to the point, we performed univariate and multivariate Cox regression analyses and found a combination of 4 genes with prognostic value for LUSC patients take place of a unitary gene .After that, we found that our concerned risk signature may forcefully facilitate clinical results by comparison with known prognostic biomarkers. We

analysed cell cycle-related genes using the LUSC dataset in the TCGA and then compared one dataset with lung squamous cell carcinoma tissue to one dataset with adjacent noncancerous tissue. We draw conclusion that a high risk score was related to metastasis and a poor prognosis through KM analysis. Zhou et al showed that the expression of CDKN1A-interacting zinc finger protein 1 may contribute to the growth and angiogenesis of LSCC and may be a novel biomarker for LUSC[22]. Theasha Manicum et al demonstrated that *E2F4 is significantly associated with unfavourable OS in all gastric cancer patients*[23]. *RAD21* is also increased in bladder *cancer* tissues and cell lines, and it can be used as a potential therapeutic target in bladder *cancer*[24]. Traditional prognostic systems commonly make wrong predictions for risk stratification and evaluation of clinical results on account of heterogeneity between patients.As far as we known, compared with a unitary usual biomarker, the 4-mRNA signature can precisely predict survival time of LUSC patients.

The 4 genes were found to be correlated with the cell cycle. CDKN1A is a protein that plays numerous roles not only in the DNA lesion reaction but also in many cellular processes during cell growth. CDKN1A functions to arrest cell cycle progression by inhibiting the movement of cyclin-dependent kinases[25]. It has been found to peculiar in other important course, for example DNA replication/repair, cell death, gene transcription, and cell motility[26]. CDKN1A can control CDK activity by interacting with its N-terminal domain or by disturbing the phosphorylation of CDK1 and CDK2[27]. Moreover, it contributes to G1 arrest by inhibiting cyclin E and cyclin A/CDK2 activity[28]. CHEK2 is *a* serine/threonine protein kinase that is required for checkpoint-mediated cell cycle stop and the activation of DNA repair and apoptosis in response to the presence of DNA double-strand breaks. CHEK2 then reacts with the downstream phosphatase CDC25, serine/threonine protein kinase NEK6, transcription factor FOXM1, p53 protein and BRCA1 or BRCA2[29]. CHEK2 may also phosphorylate NEK6, which is involved in G2/M cell cycle arrest. CHEK2 controls apoptosis by phosphorylating p53/TP53, MDM4 and PML. E2F4 is a transcription activator that binds DNA cooperatively with DP proteins through the E2 recognition site, 5'-TTTC[CG]CGC-3', which is found in the originator region of a mass of genes whose manufactures are involved in cell cycle adjustment or DNA replication. RAD21 is contained in sister chromatid cohesion from the time of DNA duplication in S phase to isolation in mitosis, a function that is necessary for suitable chromosome segregation, post-replicative DNA repair, and the intervention of inadequate recombination between reduplicative places. It acts as a target gene for numerous genes or as a control gene for other genes to take part in extensively in the physiological and pathological processes of the human body[30–32]. RAD21 is highly expressed in breast cancer, squamous cell carcinoma of the oral cavity and hepatocellular carcinoma tissues and increased proliferation and metastasis[33–35]. Based on the Pearson correlation analysis, we also found the related genes KLK5 and KLK7, which are correlated with CDKN1A expression in lung cancer[36]. Therefore, the relationship between the 4 genes and the cell cycle and cell cycle arrest and apoptosis induce the progression of LC.

In a word, this is the first report of a four-gene risk signature associated with the cell cycle that can help predict survival and prognosis in LUSC patients. A poorer prognosis is showed by higher risk score. This discover will help future scientist find new treatments means for LUSC and offer added gene targets to cure LUSC in patients.

Abbreviations

LC: Lung cancer; LUSC: Lung squamous cell carcinoma; TCGA: The Cancer Genome Atlas; GSEA: Gene Set Enrichment Analysis; OS: overall survival; HR: hazard ratio; MSigDB: Molecular Signatures Database; GSVA: Gene Set Variation Analysis;EMT: epithelial-mesenchymal transition; NSCLC :Non-small cell lung cancer; SCLC small cell lung cancer;MCM: Minichromosome maintenance protein;

Declarations

Author contributions

ZY: conceived and designed the study. LZ and ZY analyzed the data and wrote the manuscript. SY: collected the data. ZY revised the manuscript. All authors read and approved the final manuscript.

Funding

Not applicable

Data availability statement

All data generated or analyzed during this study were included in this published article and its additional files.

Ethics statement

This research was approved by the Ethics Committee of the First Affiliated Hospital of China Medical University.

Competing interests

The authors declare that they have no competing interests to disclose.

Competing interests

The authors declare that they have no competing interests to disclose.

References

1. Chen W, Zheng R, Baade PD, Zhang S, Zeng H, Bray F, et al.Cancer statistics in China, 2015. CA Cancer J Clin.2016;66:115-132.https://doi.org/10.3322/caac.21338.
2. org/Hu M, Hu Y, He J and Li B.Prognostic value of basic fibroblast growth factor (bFGF) in lung cancer: A systematic review with meta-analysis.PLoS One.2016;11(1):e0147374. https://doi.org/10.1371/journal.pone.0147374.

3. Lemjabbar-Alaoui H, Hassan OUI, Yang YW, Buchanan P. Lung cancer: biology and treatment options. Biochim Biophys Acta.2015;1856(2):189–210.https://doi.org/ 10.1016/j.bbcan.2015.08.002.
4. Meza R, Meernik C, Jeon J and Cote ML.Lung cancer incidence trends by gender, race and histology in the United States,1973-2010. PLoS One.2015;30:10(3):e0121323. https://doi.org/ 10.1371/journal.pone.0121323.
5. Perez-Moreno P, Brambilla E, Thomas R, Soria JC. Squamous cell carcinoma of the lung: molecular subtypes and therapeutic opportunities. Clin Cancer Res.2012;18(9):2443-51. https://doi.org/10.1158/1078-0432.CCR-11-2370.
6. Garzon R, Calin GA, Croce CM. MicroRNAs in Cancer.Annu Rev Med. 2009; 60:167-179. https://doi.org/10.1146/annurev.med.59.053006.104707.
7. Feng M, Zhao J, Wang L.Liu J.Upregulated Expression of Serum Exosomal microRNAs as Diagnostic Biomarkers of Lung Adenocarcinoma.Ann. Clin. Lab. Sci. 2018; 48(6): 712-18.
8. Gao X,Wang Y,Zhao H,Wei F,Zhang X,Su Y,etal. Plasma miR-324-3p and miR-1285 as diagnostic and prognostic biomarkers for early stage lung squamous cell carcinoma.Oncotarget. 2016;13:7(37): 59664- https://doi.org/10.18632/oncotarget.11198.
9. Wu W, Wang X,Shan C,Li Y,Li F. Minichromosome maintenance protein 2 correlates with the malignant status and regulates proliferation and cell cycle in lung squamous cell carcinoma. OncoTargets and Therapy.2018;11:5025-34.https://doi.org/10.2147/OTT.S169002.
10. Liu S, Miao C,Liu J,Wang CC,Lu XY. Four differentially methylated gene pairs to predict the prognosis for early stage hepatocellular carcinoma J Cell Physiol.2018;233(9):6583-90.https://doi.org/10.1002/jcp.26256.
11. Chen YL, Ge GJ,Qi C,Wang H,Wang HL,Li LY,et al. A five-gene signature may predict sunitinib sensitivity and serve as prognostic biomarkers for renal cell carcinoma.J Cell Physiol.2018; 233(10):6649-60. https://doi.org/10.1002/jcp.26441.
12. Sedgwick, P. Cox proportional hazards regression. British Medical Journal, 2013.347(aug09 1), f4919-f4919.
13. Dakhil, N. Analysis of breast cancer data using Kaplan-Meier survival analysis. Journal of Kufa for Mathematics and Computer.2012;6(1):7-14.
14. Tomczak K,Czerwińska P,Wiznerowicz M. The Cancer Genome Atlas (TCGA): An immeasurable source of knowledge. Contemp Oncol (Pozn).2015;19(1A):A68–A77. https://doi.org/10.5114/wo.2014.47136.
15. Subramanian A,Tamayo P,Mootha VK, Mukherjee S, Ebert BL, Gillette MA,et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles.ProcNatlAcadSciU S A.2005;102:15545-50.https://doi.org/1073/pnas.0506580102
16. Munson R, Jr., Bailey C, Grass S. Diversity of the outer membrane protein P2 gene from major clones of Haemophilus influenzae type b. Mol Microbiol.1989;3(12):1797-1803. https://doi.org/1111/j.1365-2958.1989.tb00165.x

17. Tian SS,Meng G,Zhang W, et al. A six-mRNA prognostic model to predict survival in head andneck squamous cell carcinoma.Cancer Manag Res.2018;11:131-42. https://doi.org/10.2147/CMAR.S185875.
18. Xu K, Xiong W, Zhao S, Wang B.MicroRNA-106b serves as a prognostic biomarker and is associated with cell proliferation, migration, and invasion in osteosarcoma. Oncol Lett.2019;18(3): 3342-48. https://doi.org/10.3892/ol.2019.10666.
19. Li X, Chen SH,Zeng JW. MiR-421 is Overexpressed and Promotes Cell Proliferation in Non-Small Cell Lung Cancer. Med Princ Pract.2019;Sep 2. https://doi.org/ 10.1159/000503020.
20. Niyazi M, Pitea A,Mittelbronn M, Steinbach J, Sticht C, Zehentmayr F.A 4-miRNA signature predicts the therapeutic outcome of glioblastoma. Oncotarget.2016;7(29):45764-775. https://doi.org/10.18632/oncotarget.9945.
21. Moya L, Meijer JSchubert S,Matin F,Batra J. Assessment of miR-98-5p, miR-152-3p, miR-326 and miR-4289 Expression as Biomarker for Prostate Cancer Diagnosis. J. Mol. Sci. 2019;20(5):1154. https://doi.org/10.3390/ijms20051154.
22. Zhou XJ,Liu Q, Wada Y, Liao L, Liu J.CDKN1A-interacting zinc finger protein 1 is a novel biomarker for lung squamous cell carcinoma. Oncol Lett.2018;15(1):183-88. https://doi.org/10.3892/ol.2017.7282.
23. Theasha Manicum et al. Prognostic values of E2F mRNA expression in human gastric cancer. Biosci. Rep. 2018;12:21:38(6).
24. Yu Z,Xu Q,Wang G, Rowe M, Driskell C, Xie Q,etal. DNA topoisomerase IIα and RAD21 cohesin complex component are predicted as potential therapeutic targets in bladder cancer. Oncol Lett.2019;18(1):518-28. https://doi.org/10.3892/ol.2019.10365.
25. El-Deiry WS, Tokino T, Velculescu VE, Levy DB, Parsons R, Trent JM,et al. WAF1,a potential mediator of p53 tumor suppression. Cell.1993;75(4):817–25.https://doi.org/ 10.1016/0092-8674(93)90500-p.
26. Stivala LA, Cazzalini O, Prosperi E.The cyclin-dependent kinase inhibitor p21CDKN1A as a target of anti-cancer drugs. Curr Cancer Drug Targets.2012;12(2):85–96.
27. Mandal M, Bandyopadhyay D, Goepfert TM, Kumar R.Interferon-induces expression of cyclin-dependent kinase-inhibitors p21WAF1 and p27Kip1 that prevent activation of cyclin-dependent kinase by CDK-activating kinase (CAK).Oncogene.1998;16(2):217–25. doi:10.1038/sj.onc.1201529.https://doi.org/10.1038/sj.onc.1201529.
28. Brugarolas J, Moberg K, Boyd SD, Taya Y, Jacks T, Lees JA .Inhibition of cyclin-dependent kinase 2 by p21 is necessary for retinoblastoma protein-mediated G1 arrest after gamma-irradiation. Proc Natl Acad Sci USA.1999;96(3):1002–7.
29. Magni M, Ruscica V, Buscemi G,Kim JE Nachimuthu BT, Fontanella E,et al. Chk2 and REGgamma-dependent DBC1 regulation in DNA damage induced apoptosis. Nucleic Acids Res.2014;42(21):13150–60.https://doi.org/10.1093/nar/gku1065.
30. Ahn JH, Kim TJ, Lee JH,Choi JH. Author Correction:Mutant p53 stimulates cell invasion through an interaction with Rad21 in human ovarian cancer cells. Sci Rep. 2018;8(1):11032.

https://doi.org/10.1038/s41598-018-29327-4.

31. Zhu S, Zhao L, Li Y, Hou P, Yao R, Tan J ,et al. Suppression of RAD21 induces senescence of MDA-MB- 231 human breast cancer cells through RB1 pathway activation Via c- Myc downregulation. J Cell Biochem.2016;117(6):1359-69.https://doi.org/10.1002/jcb.25426.
32. Fisher JB, Peterson J, Reimer M, Stelloh C, Pulakanti K, Gerbec ZJ,et al.The cohesin subunit Rad21 is a negative regulator of hematopoietic self- renewal through epigenetic repression of Hoxa7 and Hoxa9. Leukemia. 2017;31(3):712-19.https://doi.org/10.1038/leu.2016.240.
33. Mahmood SF, Gruel N, Chapeaublanc E,Lescure A, Jones T, Reyal F,et al. A siRNA screen identifies RAD21, EIF3H, CHRAC1 and TANC2 as driver genes within the 8q23, 8q24.3 and 17q23 amplicons in breast cancer with effects on cell growth, survival and transformation. Carcinogenesis. 2014;35(3):670-82. https://doi.org/10.1093/carcin/bgt351.
34. Yamamoto G, Irie T, Aida T, Nagoshi Y, Tsuchiya R, Tachikawa T. Correlation of invasion and metastasis of cancer cells, and expression of the RAD21 gene in oral squamous cell carcinoma. Virchows Arch. 2006;448(4):435-41.https://doi.org/10.1007/s00428-005-0132-y.
35. Cai J, Li B, Zhu Y, Fang X, Zhu M, Wang M,et al. Prognostic biomarker identification through integrating the gene signatures of hepatocellular carcinoma properties. EBioMedicine. 2017;19:18-30. https://doi.org/10.1016/j.ebiom.2017.04.014.
36. Planque C, de Monte M, Guyetant S, Rollin J, Desmazes C, Panel V,et al. KLK5 and KLK7, two members of the human tissue kallikrein family, are differentially expressed in lung cancer. Biochem Biophys Res Commun.2005;329(4):1260-6.https://doi.org/ 10.1016/j.bbrc.2005.02.100.

Figures

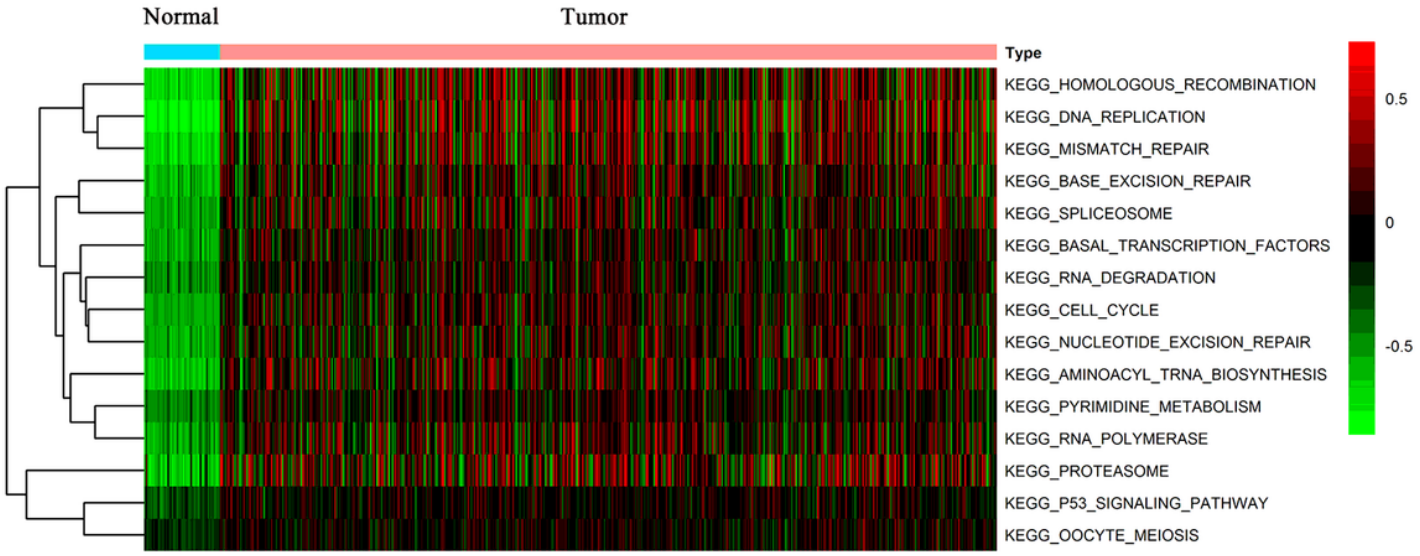

Figure 1

Variation analysis and enrichment plots of 4 gene sets which were importantly differentiated between normal tissue and lung squamous cell carcinoma tissues

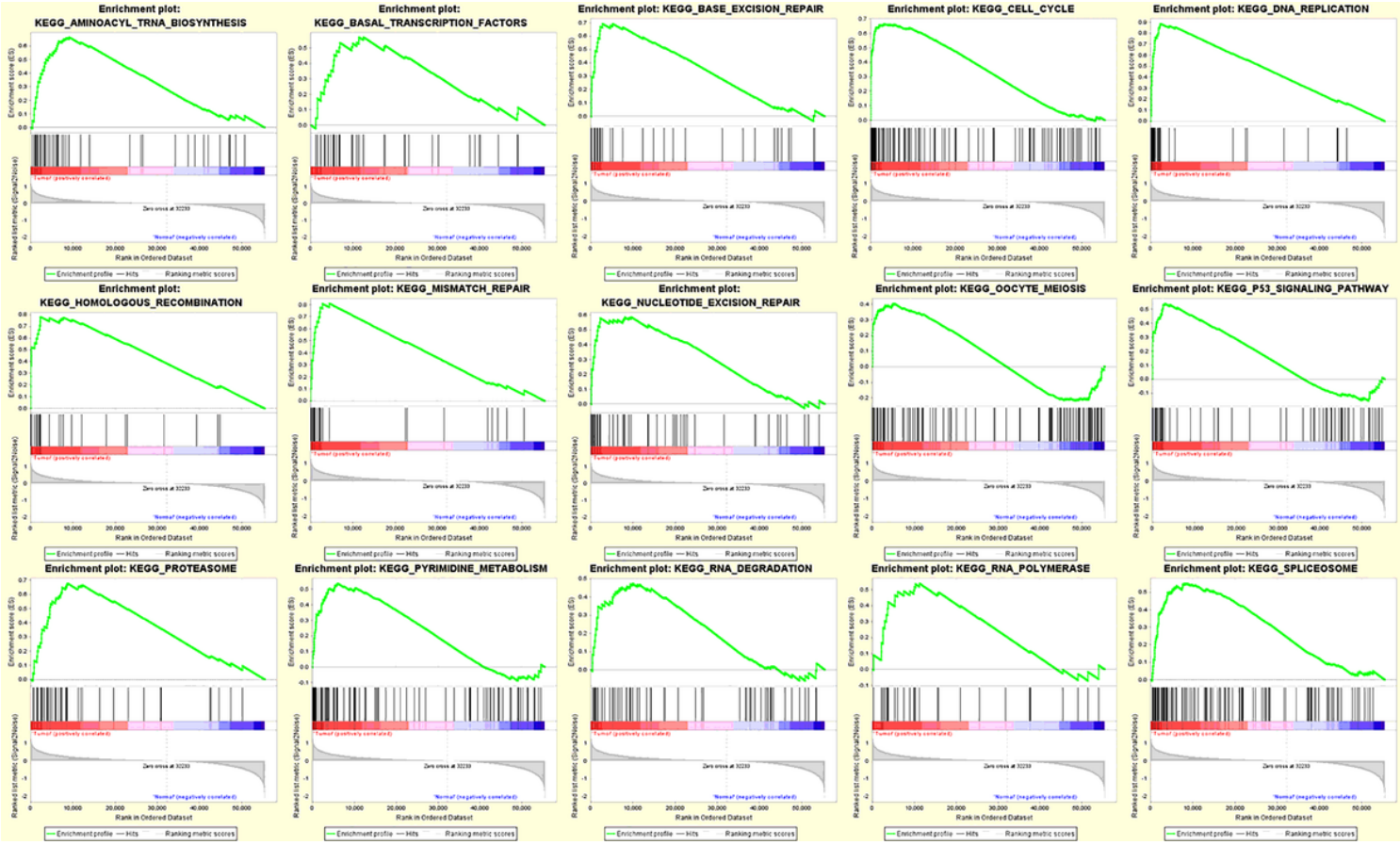

Figure 2

Correlations between the expression levels of 4 genes in lung squamous cell carcinoma were evaluated with the Pearson correlation coefficient.

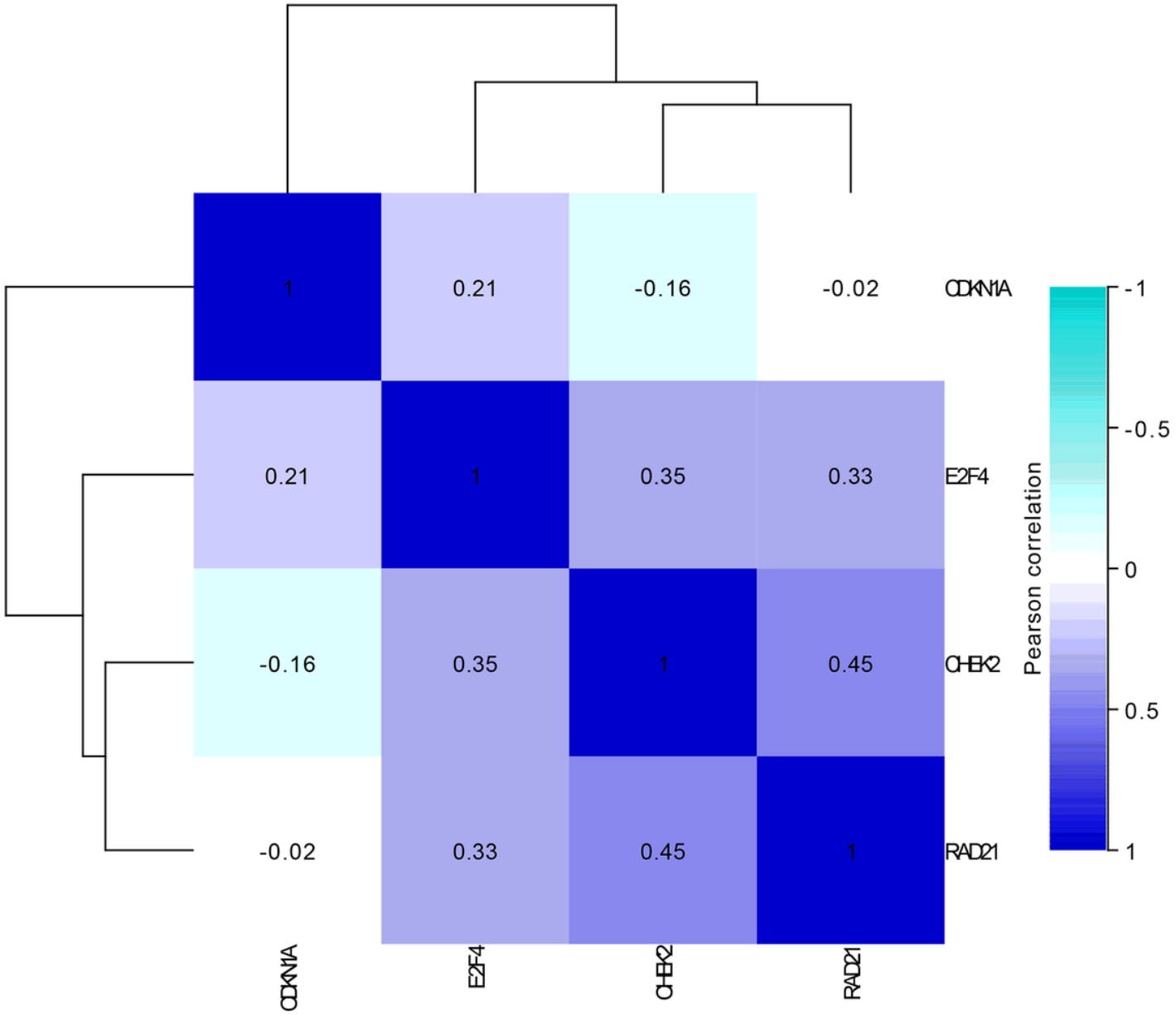

Figure 3

Construction of a Four-mRNA Signature to Predict Patient Prognosis

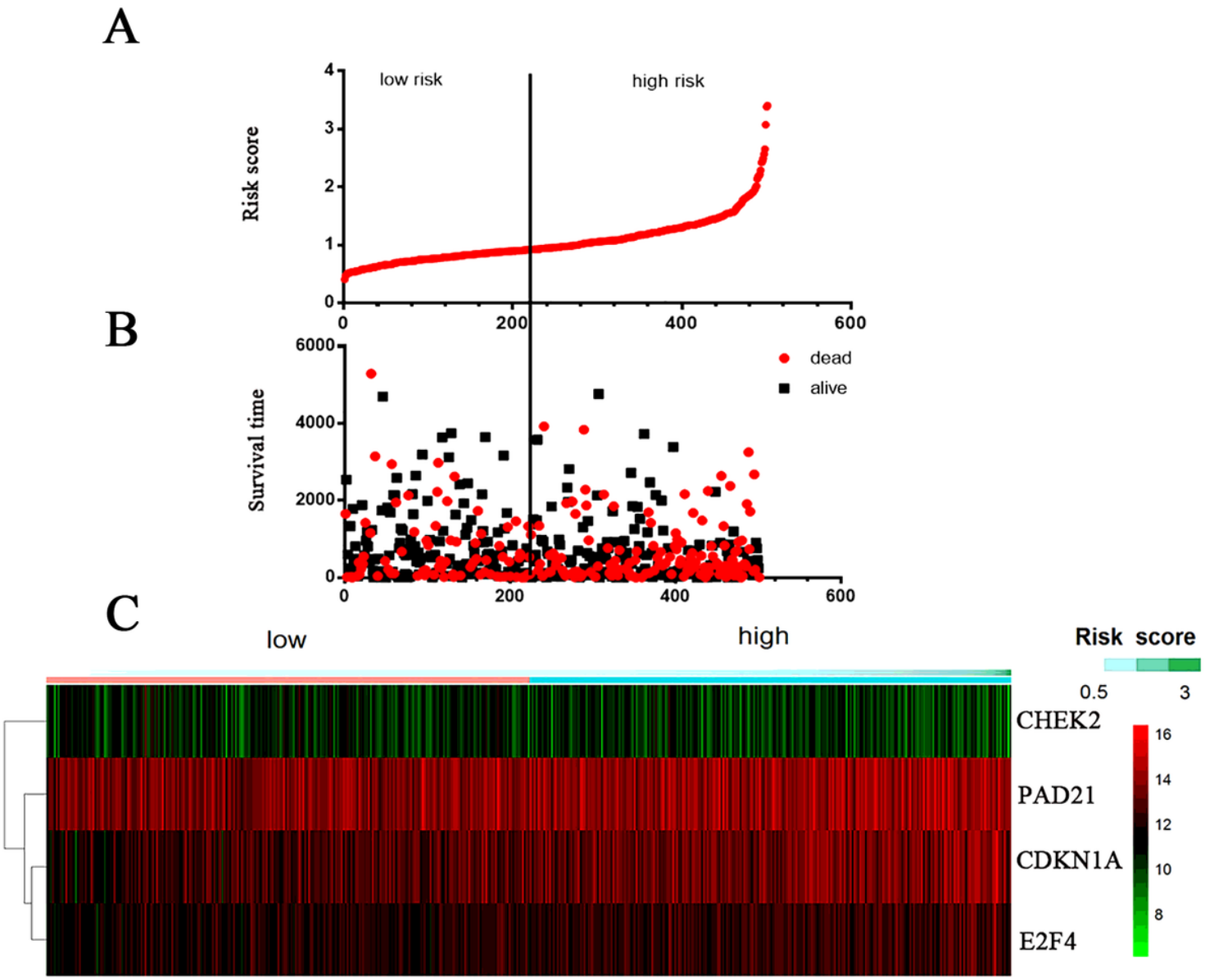

Figure 4

The four-mRNA signature related to risk score predicts overall survival in the patients with lung adenocarnicoma. (A)mRNA risk score distribution (B)Survival days of patients. (C) Heatmap of four genes' expression profile.

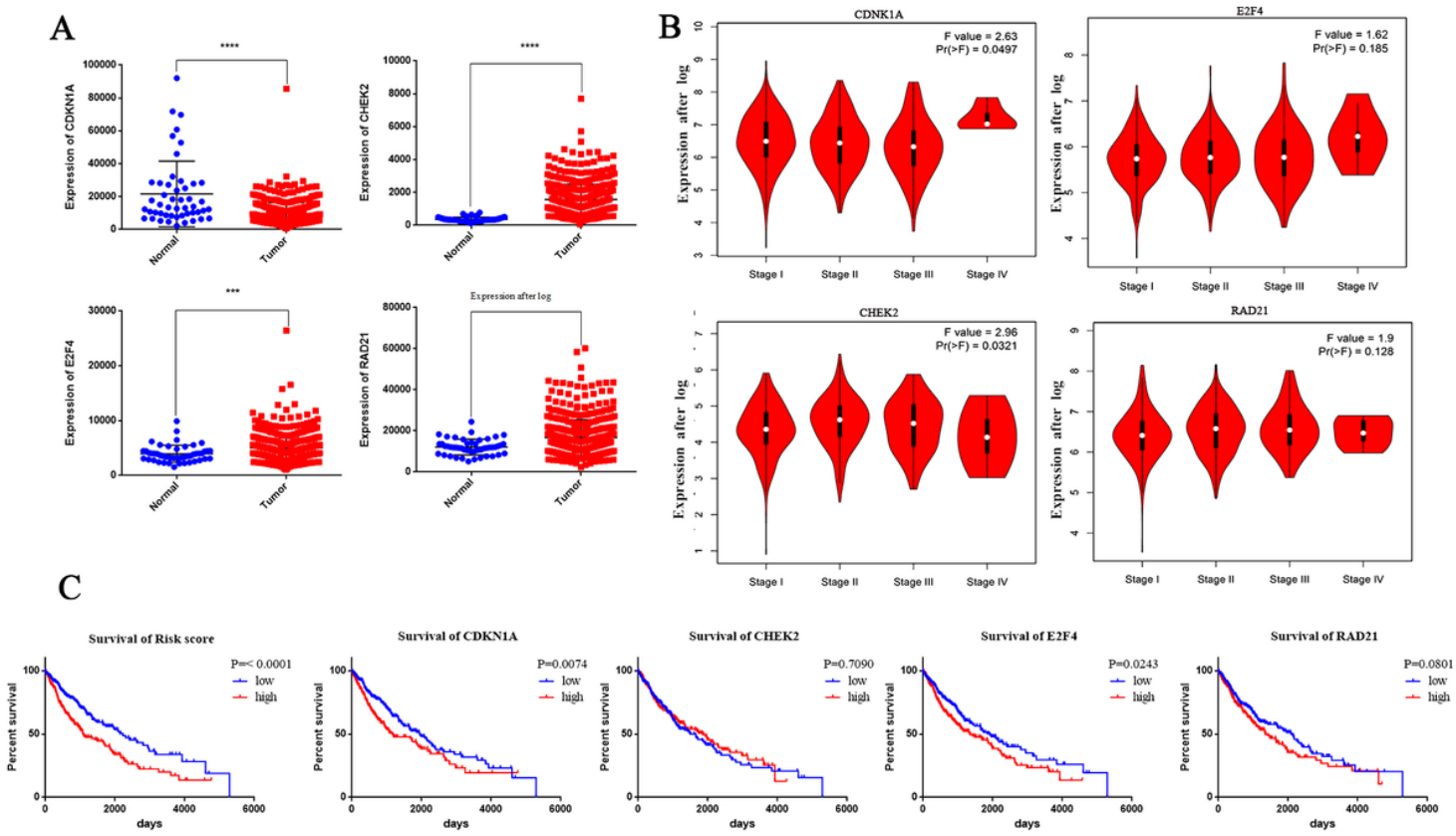

Figure 5

(A)The different expression of four gene between cancer tissue and adjacent normal tissue.(B) The different expression of four gene in each stage between cancer tissue and adjacent normal tissue.(C) The graph of survival curve of between risk score and four gene.

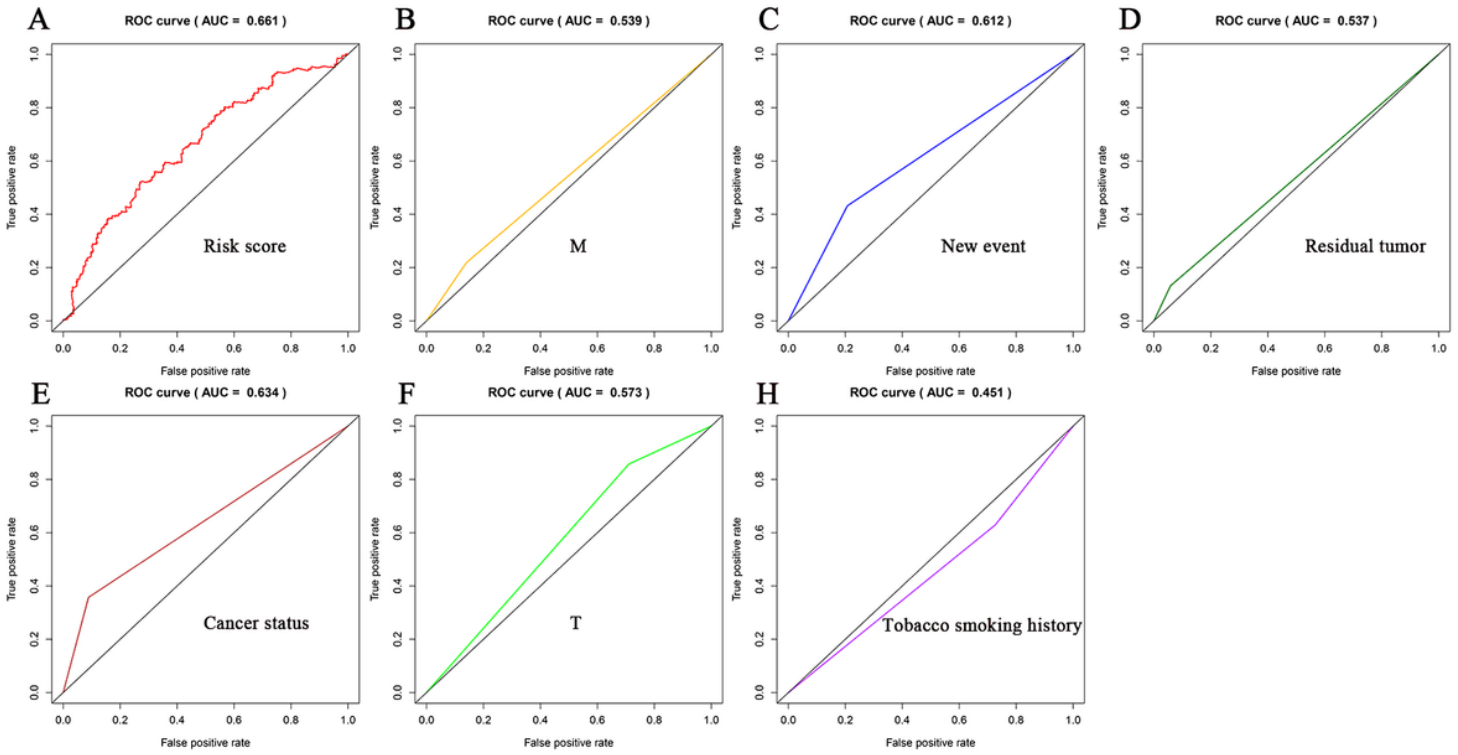

Figure 6

Receiver operating characteristic (ROC) analysis of the sensitivity and specificity of every factor.(A)Risk score(B)M classification (C)New event (D)Residual tumor (E) Cancer status (F) T classification (H)Tobacco smoking history

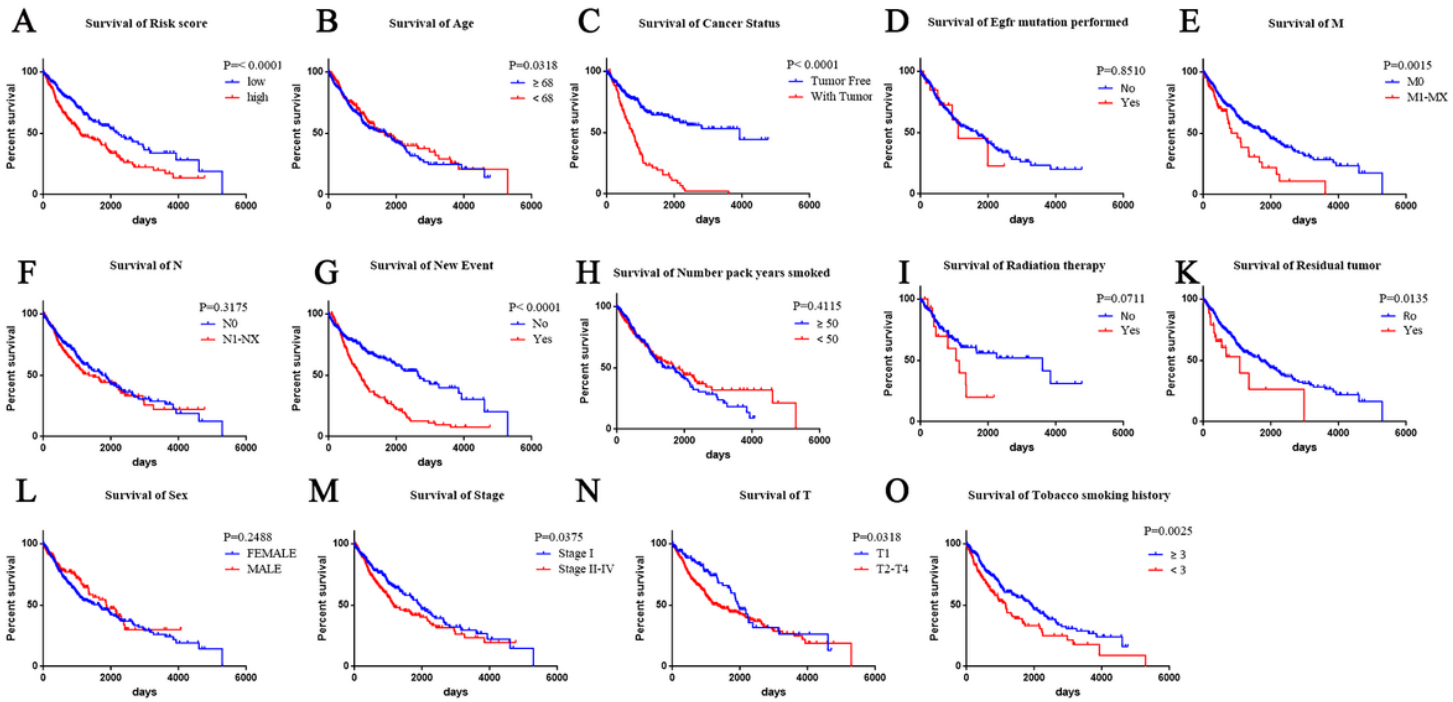

Figure 7

Kaplan-Meier survival analysis for the patients with lung adenocarcinoma in TCGA dataset.(A)The Kaplan-Meier curve for patients divided into high-risk and low-risk.Different clinical features predict patients' survival. (B)Age (C) Cancer status (D)EGFR mutation (E) M classification (F) N classification (G) New event (H)Number pack years smoked(I)Radiation therapy (K) Residual tumor (L)Sex(M)UICC stage (N) T classification (O)Tobacco smoking history

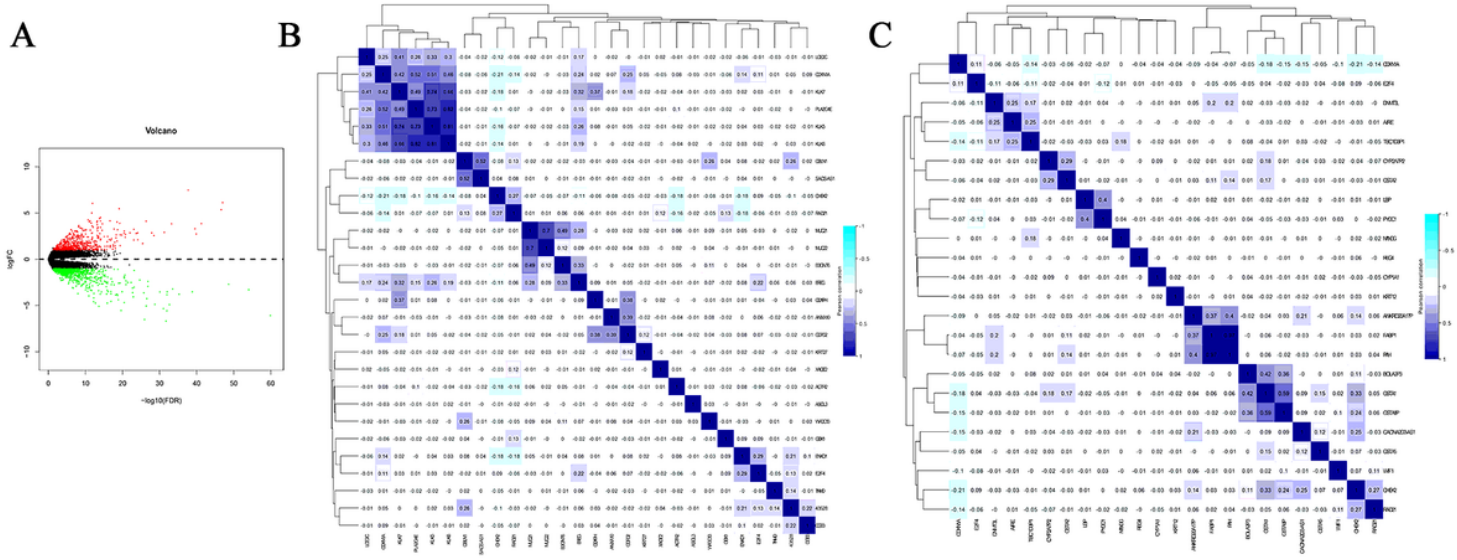

Figure 8

Correlations between the expression levels of 4 genes and genes relate to survival were evaluated with the Pearson correlation coefficient.